The Annals of Statistics
2000, Vol. 28, No. 1, 195–214

SIZE AND POWER OF PRETEST PROCEDURES

BY WILLEM ALBERS, PIETA C. BOON AND WILBERT C. M. KALLENBERG

University of Twente

A pretest procedure consists of a preliminary test on a nuisance parameter, investigating whether it equals a given value or not, followed by the main testing problem on the parameter of interest. In case of acceptance of the preliminary test, the main test is applied in the restricted family with the given value of the nuisance parameter, while otherwise the test is performed in the complete family, including the nuisance parameter. For an appropriate class of tests, containing all standard first-order optimal tests, an attractive expression for the difference in size and power between the pretest procedure and the test in the complete family is derived using second-order asymptotics. For a very great part the result is the same for all members of the class. From this expression considerable insight can be obtained in a qualitative and quantitative sense. The results can be applied easily as is illustrated by a number of practical examples, where also the accuracy of the approximations is seen from comparison with numerical results.

1. Introduction. In many textbooks on statistics [see, e.g., Hoel (1984), pages 296 and 300] it is correctly stated that when testing for equality of the means in two normal samples with the two sample t-test, the variances should be equal. Often, this is followed by a treatment of the F-test for testing equality of the variances in the two samples, thus suggesting that if the F-test does not reject, equality of the variances is acceptable. Most textbooks give no further comments. Sometimes there is some discussion on this combined procedure, warning that by repeated use of the same data the size of the whole testing procedure may differ from the presumed level. Usually in such a case, there is no indication of the magnitude of the involved error. In a still smaller minority it is advised to use separate data sets for the F-test and the t-test.

In fact, the combined procedure is a two-step procedure, consisting of a preliminary F-test, followed by the t-test in case of acceptance of the null hypothesis of equality of variances and, for example, by the Welch–Satterthwaite test in case of rejection. We call such a combined procedure a pretest procedure. Albers, Boon and Kallenberg (1997a) discuss the contrast between the implicitly optimistic view of most of the textbooks on the one hand and papers on the subject appearing in statistical literature (which do not recommend the pretest procedure) on the other hand. Application of second-order asymptotics in that paper provides simple and transparent approximations to the size and

Received December 1997; revised September 1999.
AMS 1991 *subject classifications*. 62F05, 62E20.
Key words and phrases. Pretest procedure, second-order asymptotics, locally most powerful test, likelihood ratio test, Rao's test, Wald's test, two-parameter exponential families, location-scale families.

power of the pretest procedure. Thus answers are given to practitioners, who prefer the more simple t-test if at all possible, to questions such as: How wrong can it be to follow the textbooks? What is the best level for the pretest? etc.

The same issue occurs in the one-sample testing problem, where the t-, Welch–Satterthwaite and F-test are replaced by the Gauss-, t- and χ^2-test, respectively. Since the main feature of the problem is present in the one-sample problem as well, we restrict ourselves in the extension to general densities, treated in this paper, to the one-sample situation.

Here we have two parameters, θ and τ, say. The main testing problem concerns a testing problem on θ, while τ is a nuisance parameter. The pretest procedure consists of a preliminary test on τ, to decide whether it equals a given value or not, followed by a suitable test in the restricted family with the given value of τ in case of acceptance, while otherwise a test is performed in the complete family, including the nuisance parameter. The idea is of course that people prefer the test in the more simple model as long as possible, either because of the convenience of greater simplicity itself and/or because of a possibly higher power in that case, due to "knowing the value of the nuisance parameter". It is the aim of the paper to reveal the differences in size and power between the pretest procedure and the one-stage test in the complete family.

In the special case of normal distributions, the χ^2-test and the Gauss- or t-test are (almost) independent. In general, independence between the preliminary test and the main tests does not necessarily hold. However, if the preliminary test already is almost a test for the main problem, a two-step procedure is not very appealing: the two steps should not be mixed up too much. More than that: for larger correlations of the test statistic applied for the preliminary test and the main tests, the size of the pretest procedure varies wildly and unacceptable violations of the prescribed level cannot be avoided. Hence, pretest procedures are only of interest if the correlations of the test statistic applied for the preliminary test and the main tests are small. This conclusion follows already from first-order asymptotics and is presented in Section 2.

So, in the analysis of pretest procedures a correlation parameter ρ plays an important role. In view of the above mentioned arguments, based on fixed ρ and first-order asymptotics w.r.t. n, the only interesting case occurs when ρ is small and this is assumed further on. However, if ρ is small (e.g., $\rho = 0$) first-order asymptotics w.r.t. n are of little help anymore; see Albers, Boon and Kallenberg (1997a). However, the use of second-order asymptotics makes clear what is going on. This more complicated analysis is tackled by taking into account that the two main tests are tests for essentially the same testing problem (although in different models) and therefore have some common part. This makes it possible to reduce substantially the rather many terms in the second-order asymptotic expansions. This argument is worked out in Section 2 and may be of independent interest.

The fact that the two main tests are not that much different is partly due to considering small ρ. In other words, second-order asymptotic analysis w.r.t.

n for fixed ρ is rather hopeless and will not lead to useful expressions. But fortunately, as indicated before, the only interesting case is that of small ρ. In principle one could imagine ρ to be tied to n in some way. For example, in many statistical papers when studying powers of tests, alternatives are linked to sample sizes in order to discuss nontrivial power. However, for the parameter ρ no such link is available: it is given in the problem, no matter how large n is. Since there is no natural relation between ρ and n, we apply asymptotics *both* w.r.t. the important parameter $\rho \to 0$ and w.r.t. $n \to \infty$. In our opinion, this is the proper way to arrive at a meaningful and attractive description of the behavior of pretest procedures, which can be easily applied in practice.

The accuracy of the approximations is $O(\rho^3+\rho n^{-1/2})+o(n^{-1/2})$. This means that there exist a constant C and a sequence $\{a_n\}$ with $\lim_{n\to\infty} a_n = 0$ such that the error terms due to the approximations are bounded by $C(\rho^3+\rho n^{-1/2})+a_n n^{-1/2}$ for all ρ and n. Hence, our results are *uniformly* valid in ρ and n, but of course they are only meaningful for $n \to \infty$ and $\rho \to 0$. If we should add the terms of order $\rho n^{-1/2}$ to the approximation, a large number of terms would come in. When we really would use these terms, it would make the expression less insightful due to too many complications. The same holds for the ρ^3-terms. Ignoring terms $O(\rho^3+\rho n^{-1/2})$- and $o(n^{-1/2})$-terms seems to be the right compromise between needed accuracy and transparency.

Section 3 contains the main results. The (bivariate) Edgeworth expansions of the test statistics are presented, followed by a transparent expression for the difference in size and power between the pretest procedure and the one-stage test in the complete family. This expression gives much insight in a qualitative and quantitative sense.

It turns out that for all members of the class of test statistics the same expression holds, except for one quantity coming from the test in the more simple model. In particular, it means that (up to the considered order) there is no difference between applying, for example, the locally most powerful test with the restricted or unrestricted maximum likelihood estimator of the nuisance parameter inserted, the likelihood ratio test, Rao's efficient score test or Wald's test in the pretest and/or in the testing problem on θ in the complete family.

One of the ideas behind the pretest procedure is that practitioners will prefer a more simple test if at all possible: it is more easily explained to clients or even familiar to them, directly available in a statistical package, etc. The main point then is control of the size. The results of Section 3 are very helpful for judging it, since the family of distributions and the class of tests is involved through only four parameters; compare (4.1). After a general discussion on the consequences for the actual size in the beginning of Section 4, two important classes of distributions are considered: two-parameter exponential families and symmetric location-scale families. In a general symmetric location-scale family, the approximation for the deviation from the nominal size of the pretest procedure is very simple: it is the same as that for the normal case, except for a multiplicative constant.

For a great part the hope for a higher power with the pretest procedure is not realized, since the power change is often mainly nothing else but a factor times the size change. This implies that essentially a higher power is only obtained if the size exceeds the nominal level and that a higher power at (θ, τ) goes hand in hand with a lower power at some other point $(\theta, \tilde{\tau})$, while we do not know the nuisance parameter. Further details are provided in Section 4, where also the accuracy of the approximations is seen from comparison with some numerical results.

2. Notation, assumptions and preliminaries Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with density $f(x; \theta, \tau)$ w.r.t. some measure μ on the measurable space $(\mathscr{X}, \mathscr{A})$. Our main testing problem concerns H_0: $\theta = \theta_0$ against H_1: $\theta > \theta_0$. If the nuisance parameter τ would be known and be equal to τ_0, say, we would like to test $\theta = \theta_0$ in the family $f(x; \theta, \tau_0)$. To see whether we may use this information, we perform a pretest of $\bar{H}_0$: $\tau = \tau_0$ against $\bar{H}_1$: $\tau \neq \tau_0$. This strategy leads to the following procedure: if $\bar{H}_0$ is accepted, test H_0 in the family $f(x; \theta, \tau_0)$; otherwise, test H_0 against H_1 in the family $f(x; \theta, \tau)$. We call this procedure the *pretest procedure*.

To define the tests of the separate testing problems we give some notation. Due to lack of space we do not present the regularity conditions. They are of the same type as those used in classical large sample theory; see, for example, Lehmann [(1983), page 429] and compare also Albers, Boon and Kallenberg (1997b). Without loss of generality, let $\theta_0 = 0$ and $\tau_0 = 0$. As usual (θ, τ) will denote the true value of the parameter as well as a variable in $\mathbb{R}^2$. Its meaning is plain from the context.

By $P_{\theta,\tau}$ we denote that X_i has density $f(x; \theta, \tau)$. Further, let

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

$$I_{11} = E(\psi^*_{10})^2, \qquad I_{12} = E\psi^*_{10}\psi^*_{01} \quad \text{and} \quad I_{22} = E(\psi^*_{01})^2.$$

Here and in the sequel we often write E for $E_{0,0}$, ψ^*_{10} or $\psi^*_{10}(X)$ for $\psi^*_{10}(X; 0, 0)$, etc., to avoid unnecessarily complicated notation. Furthermore, let

$$\psi_{ij}(x; \theta, \tau) = \psi^*_{ij}(x; \theta, \tau) I_{11}^{-i/2} I_{22}^{-j/2}.$$

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

$$E_{\theta,\tau}\psi_{ij}(X; \theta, \tau) = 0 \quad \text{for } i, j = 0, 1, 2, 3, \qquad (i, j) \neq (0, 0). \tag{2.1}$$

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

$$\rho = I_{12}(I_{11} I_{22})^{-1/2}. \tag{2.2}$$

Define

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

The statistic S corresponds to the locally most powerful (LMP) test for testing $\theta = 0$ in the family $f(x;\theta,0)$. Other candidates for testing H_0 are, for example, the likelihood ratio (LR) test, the test based on the maximum likelihood estimator (MLE) and Wald's test. The test statistics of all of these tests can be written (up to order $n^{-1/2}$) in the following form:

$$SK = S + n^{-1/2}S\left\{n^{-1/2}\sum_{i=1}^{n}k(X_i)\right\} \tag{2.4}$$

with $Ek = 0$. Therefore, we use this class of test statistics for testing H_0: $\theta = 0$ against H_1: $\theta > 0$, rejecting for large values of SK. (Here "K" refers to the known value of τ.) In particular, we have

$$\begin{aligned} \text{LMP} &: k = 0,\\ \text{LR} &: k = \tfrac{1}{2}\left\{\psi_{20} - (\psi_{10}^2 - 1)\right\} + (\tfrac{1}{3}E\psi_{10}^3 - \tfrac{1}{2}E\psi_{10}\psi_{20})\psi_{10},\\ \text{MLE} &: k = \psi_{20} - (\psi_{10}^2 - 1) + (E\psi_{10}^3 - \tfrac{3}{2}E\psi_{10}\psi_{20})\psi_{10},\\ \text{Wald} &: k = \psi_{20} - (\psi_{10}^2 - 1) + \tfrac{1}{2}(E\psi_{10}^3 - E\psi_{10}\psi_{20})\psi_{10}. \end{aligned} \tag{2.5}$$

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

These tests are more familiar in case of two-sided alternatives. The one-sided forms as needed here are presented, for example, as directed versions on page 82 of Barndorff-Nielsen and Cox (1994). A brief justification for LR is as follows. (We use $\doteq$ for approximately equal up to order $n^{-1/2}$.) Here we deal with testing in the family $f(x;\theta,0)$ and we write for the corresponding MLE of $\theta : \hat{\theta}_0$. The likelihood equations yield

$$\begin{aligned} 0 = n^{-1/2}\sum_{i=1}^{n}\psi_{10}(X_i;\hat{\theta}_0,0) &\doteq S - n^{1/2}\hat{\theta}_0 I_{11}^{1/2}\\ &+ n^{-1/2}\left[n^{1/2}\hat{\theta}_0 I_{11}^{1/2}\left\{n^{-1/2}\sum_{i=1}^{n}z(X_i)\right\} + an\hat{\theta}_0^2 I_{11}\right], \end{aligned} \tag{2.6}$$

where $z = \psi_{20} - (\psi_{10}^2 - 1)$ and $a = E\psi_{10}^3 - \frac{3}{2}E\psi_{10}\psi_{20}$. The directed LR test statistic is given by

$$\operatorname{sgn}(\hat{\theta}_0)\sqrt{2\{l(\hat{\theta}_0) - l(0)\}} \quad \text{with } l(\theta) = \sum_{i=1}^{n}\log f(X_i;\theta,0).$$

By the Taylor expansion we get

$$l(\hat{\theta}_0) - l(0) \doteq \hat{\theta}_0 n^{1/2} I_{11}^{1/2} S + \tfrac{1}{2}\hat{\theta}_0^2 I_{11}\left\{\sum_{i=1}^{n}z(X_i) - n\right\} + \tfrac{1}{6}n\hat{\theta}_0^3 2aI_{11}^{3/2}. \tag{2.7}$$

In view of (2.6) we obtain

$$n^{1/2}\hat{\theta}_0 I_{11}^{1/2} \doteq S + n^{-1/2} S \left\{ n^{-1/2} \sum_{i=1}^{n} z(X_i) + aS \right\}, \tag{2.8}$$

which by the way gives (2.4) and (2.5) for the MLE. Inserting (2.8) in (2.7) gives

$$l(\hat{\theta}_0) - l(0) = \tfrac{1}{2} S^2 \left\{ 1 + n^{-1/2} \left[n^{-1/2} \sum_{i=1}^{n} z(X_i) + \tfrac{2}{3} aS \right] \right\}$$

and (2.4) and (2.5) for the LR statistic now easily follow.

It is easily seen that under $f(x; 0, cI_{22}^{-1/2} n^{-1/2})$ the test statistic SK converges in distribution to a normal r.v. with expectation $c\rho$ and variance 1. Therefore, if ρ is not small, the actual size of SK under $f(x; 0, cI_{22}^{-1/2} n^{-1/2})$ with $c \neq 0$ differs drastically from the prescribed level. So, even small departures for τ from 0 lead to unacceptable deviations in size.

The statistics of the preliminary test, defined later on, equal to first order $n^{-1/2} \sum_{i=1}^{n} (\psi_{01} - \rho\psi_{10})(X_i)/\sqrt{1-\rho^2}$ and converge in distribution to a normal r.v. with expectation $c\sqrt{1-\rho^2}$ and variance 1 under $f(x; 0, cI_{22}^{-1/2} n^{-1/2})$. The preliminary tests are intended to prevent application of the test SK (in favor of SU) when unacceptable deviations in size occur. If r is not small, $c\sqrt{1-\rho^2}$ is of the same order as (or even smaller than) $c\rho$ and therefore the preliminary tests do not have sufficient power and the protection fails. Consequently, the pretest procedure will have the same problem with size, unless ρ is small. Hence, for ρ not small the situation is clear: the pretest procedure is unacceptable. For a practical application see Example 4.2.

As a consequence it is only of interest to consider small ρ, as will be assumed from now on. This implies that the preliminary tests, (mainly) based on $\psi_{01} - \rho\psi_{10}$, and the main tests, (essentially) based on ψ_{10} and $\psi_{10} - \rho\psi_{01}$, are at most weakly dependent and have, so to speak, different aims: the preliminary tests concentrates on the nuisance parameter and the main tests mainly deal with the parameter of interest, thus avoiding an unwanted mix-up.

With respect to n, first-order asymptotics are not sufficient. This is clearly seen in the special case that $\rho = 0$. The test statistics for testing $\theta = 0$ when τ is unknown, are at first order equal to [cf. also (2.9)] $n^{-1/2} \sum_{i=1}^{n} (\psi_{10} - \rho\psi_{01})(X_i)/\sqrt{1-\rho^2}$. And hence, for $\rho = 0$ they are equal at first-order to SK. This would imply that, based on first order asymptotics, for $\rho = 0$ there is no problem with the pretest procedure. However, in the normal case, where we do have $\rho = 0$, the actual size of the pretest procedure may differ substantially from its nominal level; compare Albers, Boon and Kallenberg (1997a, 1998), dealing with pretest procedures for tests on normal means as well in the one-sample as in the two-sample problem and compare Moser, Stevens and Watts (1989) and Markowski and Markowski (1990) for the latter testing problem. Therefore, to make clear what is going on, second-order asymptotics in n will be applied.

If τ is unknown, the class of tests of H_0: $\theta = 0$ is in principle based on

$$(2.9)\qquad \left[S - \rho T + n^{-1/2}\left\{Sn^{-1/2}\sum_{i=1}^{n} q(X_i) + Tn^{-1/2}\sum_{i=1}^{n} r^*(X_i)\right\}\right](1-\rho^2)^{-1/2}$$

with $Eq = Er^* = 0$. However, the (second-order) limiting distribution of these statistics under $(0, cI_{22}^{-1/2}n^{-1/2})$ depends on c, since their expectation and variance under $(0, cI_{22}^{-1/2}n^{-1/2})$ are, up to order $n^{-1/2}$ and ρ^2, but ignoring terms of order $\rho n^{-1/2}$,

$$n^{-1/2}\left\{Eq\psi_{10} + Er^*\psi_{01} + c^2\left(\tfrac{1}{2}E\psi_{10}\psi_{02} + Er^*\psi_{01}\right)\right\}$$

and

$$1 + cn^{-1/2}\left\{E\psi_{10}^2\psi_{01} + 2\left[Eq\psi_{01} + Er^*\psi_{10}\right]\right\},$$

respectively. Standardizing and plugging in $n^{1/2}\hat{\tau}I_{22}^{1/2}$, or, equivalently, T as estimator of c solves the problem. As a result we use as test statistic SU, which is (2.9) with r^* replaced by

$$(2.10)\qquad \begin{aligned} r = r^* &- \tfrac{1}{2}\left\{E\psi_{10}^2\psi_{01} + 2\left[Eq\psi_{01} + Er^*\psi_{10}\right]\right\}\psi_{10} \\ &- \left(\tfrac{1}{2}E\psi_{10}\psi_{02} + Er^*\psi_{01}\right)\psi_{01}. \end{aligned}$$

(The "U" in SU refers to τ unknown.) The class of test statistics SU is a natural extension of the class of test statistics SK. Test statistics of the form (2.9), up to the considered order are, for example, the LMP test with the given τ replaced by the restricted or unrestricted MLE, the LR test, the test based on the MLE of θ in the unrestricted model, Wald's test and Rao's efficient score test. In view of the fact that our final results do not depend on q and r, we do not present the specific q and r of the before mentioned tests, but see Albers, Boon and Kallenberg (1997b) for more details.

Similarly, for testing $\bar{H}_0$: $\tau = 0$ (with θ unknown) we start with (2.9), where T and S are interchanged. Because here we have a two-sided testing problem, there is no need to correct the (small) bias in expectation. Therefore, the corresponding correction for r^* is given by

$$r = r^* - \tfrac{1}{2}\left\{E\psi_{10}\psi_{01}^2 + 2\left[Eq\psi_{10} + Er^*\psi_{01}\right]\right\}\psi_{01}$$

and the test statistic is called TU. Although we use the same notation, the functions q and r appearing in SU may be different from those in TU. Which q and r are meant, from SU or TU, is always clear from the context.

Conceptually, a straightforward approach for investigating the pretest procedure would be to derive an expression for its (asymptotic) power and to analyze that, in particular by comparing it with the (asymptotic) power of SU, which we should use when no pretest is applied. We do not follow this path, but take a more subtle approach. It is seen from (2.4) and (2.9) that SK and SU are not that much different if n is large and ρ is small: S is the leading term of SK and SU. One of the basic technical tools of this paper is to use this similarity from the very beginning. In this way we avoid a large number of terms which cancel afterwards anyhow. In a more abstract form, this argument is presented in the lemma below.

First, however, we introduce some notation. Let Φ denote the standard normal distribution function and $\Phi^{(j)}$ its jth derivative. Instead of $\Phi^{(1)}$ we also write φ for the standard normal density. Moreover, let $\Phi(\cdot, \cdot, \rho)$ be the distribution function of the bivariate normal $N(0, 0, 1, 1, \rho)$-distribution. [Note that here ρ is not the quantity defined in (2.2), but the usual name for the correlation coefficient in the bivariate normal distribution. The ρ occurring in Lemma 2.1 is in general simply the name of a variable and hence should not be identified with the quantity defined in (2.2), until Lemma 2.1 is applied in the proof of Theorem 3.2, when it is the quantity from (2.2). In view of this application it is convenient to use this notation in Lemma 2.1.]

LEMMA 2.1. *Let (U_{1n}, V_n) and (U_{2n}, V_n) be (standardized) sequences of r.v.'s (possibly depending on ρ) admitting Edgeworth expansions of the following form*:

$$Pr(U_{in} \le u, V_n \le v) = \Phi(u, v; \rho_i) + n^{-1/2} \sum_{j=0}^{3} c_{ij} \Phi^{(j)}(u) \Phi^{(3-j)}(v) + O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2}) \tag{2.11}$$

as $\rho \to 0$, $n \to \infty$ and $i = 1, 2$, uniformly for (u, v) in each compact set in $\mathbb{R}^2$, where c_{ij}, $i = 1, 2$, $j = 0, \dots, 3$, are constants and ρ_1, ρ_2 are functions of ρ and n. Let $\rho_1 - \rho_2 = O(\rho + n^{-1/2})$ and $\rho_1 = O(n^{-1/2})$. Then for $u_{1n} = u_0 + O(\rho + n^{-1/2})$, $v_n = v_0 + O(\rho^2 + n^{-1/2})$ and $u_{2n} = u_0 + O(\rho^2 + n^{-1/2})$,

$$\begin{aligned} &Pr(U_{1n} \le u_{1n}, V_n \le v_n) - Pr(U_{2n} \le u_{2n}, V_n \le v_n) \\ &\quad = \varphi(u_0) \big\{ (u_{1n} - u_{2n}) \Phi(v_0) - \tfrac{1}{2} (u_{1n} - u_{2n})^2 u_0 \Phi(v_0) \\ &\qquad + (\rho_1 - \rho_2) \varphi(v_0) - \tfrac{1}{2} (\rho_1 - \rho_2)^2 u_0 v_0 \varphi(v_0) \big\} \\ &\qquad + n^{-1/2} \sum_{j=1}^{3} (c_{1j} - c_{2j}) \Phi^{(j)}(u_0) \Phi^{(3-j)}(v_0) \\ &\qquad + O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2}). \end{aligned} \tag{2.12}$$

Before proving Lemma 2.1 we emphasize that $O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2})$ in expressions like (2.11) and (2.12) is understood as follows. There exist a constant C and a sequence $\{a_n\}$ with $\lim_{n \to \infty} a_n = 0$ such that the difference between the expression on the left-hand side and that of the right-hand side without the O- and o-terms is bounded by $C(\rho^3 + \rho n^{-1/2}) + a_n n^{-1/2}$ for all ρ and n. Hence, our results are *uniformly* valid in ρ and n, but of course they are only meaningful (e.g., as approximations) for $n \to \infty$ and $\rho \to 0$.

PROOF. Write $\Phi(u_{1n}, v_n; \rho_1) - \Phi(u_{2n}, v_n; \rho_2)$ as

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

For the first part we use

$$\begin{aligned}\Phi(u_1, v; \rho_1) - \Phi(u_2, v; \rho_1) = &(u_1 - u_2)\varphi(u_2)\{\Phi(v) + O(\rho_1)\}\\ &+ \tfrac{1}{2}(u_1 - u_2)^2\varphi(u_2)\{-u_2\Phi(v) + O(\rho_1)\}\\ &+O(|u_1 - u_2|^3),\end{aligned} \tag{2.13}$$

resulting in

$$\begin{aligned}&\Phi(u_{1n}, v_n; \rho_1) - \Phi(u_{2n}, v_n; \rho_1)\\ &\quad= \varphi(u_0)\{(u_{1n} - u_{2n})\Phi(v_0) - \tfrac{1}{2}(u_{1n} - u_{2n})^2 u_0\Phi(v_0)\}\\ &\qquad+ O(\rho^3 + n^{-1} + \rho n^{-1/2}).\end{aligned}$$

For the second part we get, with $\varphi(u, v; \rho)$ denoting the density of the bivariate normal $N(0, 0, 1, 1, \rho)$-distribution,

$$\begin{aligned}\Phi(u, v; \rho_1) - \Phi(u, v; \rho_2) = &\ (\rho_1 - \rho_2)\varphi(u, v; \rho_1)\\ &-\tfrac{1}{2}(\rho_1 - \rho_2)^2 uv\varphi(u)\varphi(v) + O(|\rho_1 - \rho_2|^3)\end{aligned} \tag{2.14}$$

and hence, using $\varphi(u, v; \rho_1) = \varphi(u)\varphi(v) + O(\rho_1)$, we obtain

$$\begin{aligned}\Phi(u_{2n}, v_n; \rho_1) - \Phi(u_{2n}, v_n; \rho_2) = &\ \varphi(u_0)\varphi(v_0)\{(\rho_1 - \rho_2) - \tfrac{1}{2}(\rho_1 - \rho_2)^2 u_0 v_0\}\\ &+ O(\rho^3 + n^{-1} + \rho n^{-1/2}).\end{aligned}$$

Note that by considering the marginal distribution of V_n it follows that $c_{10} = c_{20}$. The proof is now easily completed. □

One should think of U_{1n}, U_{2n} and V_n as the test statistics SK, SU and TU, respectively, standardized under local alternatives with u_{in} and v_n as their critical values, shifted and rescaled by the standardization of the test statistics. The resemblance of U_{1n} and U_{2n} in our application is represented in Lemma 2.1 by the closeness of u_{1n} and u_{2n} to each other.

The straightforward approach, mentioned before, would imply an expansion of $\Phi(u, v; \rho_i)$ for ρ_i around 0 in (2.11) and expansions of *all* the terms in (2.11) for u_{in} around u_0 and v_n around v_0, where the ρ-, ρ^2- and $n^{-1/2}$-terms of u_{1n}, u_{2n} and v_n are made explicit. This would give a great number of terms. Presumably, also the c_{ij} will be given explicitly in such an approach, which again gives a lot of terms. Carefully gathering all these terms would show that many of them are the same for U_{1n} and U_{2n} and hence cancel in taking the difference.

Replacing the expansion of $\Phi(u, v; \rho_i)$ for ρ_i around 0 and the expansion of all the terms in (2.11) for u_{in} around u_0 and v_n around v_0 by the expansions given in (2.13) and (2.14) yields an enormous reduction in the number of terms and shows moreover what is *really* needed: for u_{in} their first-order term u_0 and furthermore only the *difference* $u_{1n} - u_{2n}$, while for v_n even the first-order term v_0 suffices. This is due to the similarity between U_{1n} and U_{2n} and the small correlation of V_n and U_{in}. Finally, for the c_{ij} also only the differences $c_{1j} - c_{2j}$ are involved, which gives again a large reduction of terms. The application of Lemma 2.1 in Section 3 to SK, SU and TU clearly shows that the more subtle approach using (2.13) and (2.14) is very profitable compared to the straightforward one, sketched above.

In the application of Lemma 2.1 we need an Edgeworth expansion of the form (2.11). This is presented in Theorem 3.1.

3. Main results. Our main aim in this paper is to reveal the difference in size and power of the pretest procedure and the one-stage test SU. We consider local alternatives of the form (θ_n, τ_n) with

$$\theta_n = bI_{11}^{-1/2}n^{-1/2}, \qquad \tau_n = cI_{22}^{-1/2}n^{-1/2} \quad \text{with } b > 0 \text{ and } c \in \mathbb{R}. \tag{3.1}$$

(Note that θ_n and τ_n are described in $I_{11}^{-1/2}$ and $I_{22}^{-1/2}$ "units." In this way the redundant parameters I_{11} and I_{22} are absorbed and I_{12} reduces to ρ.) As usual in this kind of analysis the (composite) null hypothesis H_0 is represented by sequences $(0, cI_{22}^{-1/2}n^{-1/2})$, $\bar{H}_0$ by $(bI_{11}^{-1/2}n^{-1/2}, 0)$ and H_1 of course by $(bI_{11}^{-1/2}n^{-1/2}, cI_{22}^{-1/2}n^{-1/2})$.

We start with the Edgeworth expansions of the simultaneous distribution of (SK, TU) and (SU, TU) under (θ_n, τ_n).

THEOREM 3.1. *Suppose that the regularity conditions hold. Write*

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

with

$$\begin{aligned}
\mu_{1n}(b,c) &= b + c\rho + \tfrac{1}{2}\left\{b^2 E\psi_{10}\psi_{20} + 2bcE\psi_{10}\psi_{11} + c^2 E\psi_{10}\psi_{02}\right.\\
&\qquad\qquad \left. + 2(1+b^2)Ek\psi_{10} + 2bcEk\psi_{01}\right\} n^{-1/2},\\
\sigma_{1n}(b,c) &= 1 + \tfrac{1}{2}\left\{bE\psi_{10}^3 + cE\psi_{10}^2\psi_{01} + 4bEk\psi_{10} + 2cEk\psi_{01}\right\} n^{-1/2},\\
\mu_{2n}(b,c) &= b - \tfrac{1}{2}b\rho^2 + \tfrac{1}{2}\left\{b^2 E\psi_{10}\psi_{20} + 2bcE\psi_{10}\psi_{11} - bcE\psi_{10}^2\psi_{01}\right.\\
&\qquad\qquad \left. - E\psi_{10}\psi_{02} + 2(1+b^2)Eq\psi_{10}\right\} n^{-1/2},\\
\sigma_{2n}(b) &= 1 + \tfrac{1}{2}b(E\psi_{10}^3 + 4Eq\psi_{10})n^{-1/2},\\
\mu_n(b,c) &= c - \tfrac{1}{2}c\rho^2 + \tfrac{1}{2}\left\{c^2 E\psi_{01}\psi_{02} + 2bcE\psi_{01}\psi_{11} + b^2 E\psi_{01}\psi_{20}\right.\\
&\qquad\qquad - bcE\psi_{10}\psi_{01}^2 + 2(1+b^2)Er\psi_{10}\\
&\qquad\qquad \left. + 2(1+c^2)Eq\psi_{01}\right\} n^{-1/2},\\
\sigma_n(c) &= 1 + \tfrac{1}{2}c(E\psi_{01}^3 + 4Eq\psi_{01})n^{-1/2}.
\end{aligned} \tag{3.2}$$

Then, uniformly for $(u, v) \in \mathbb{R}^2$,

$$\begin{aligned}
P_{\theta_n,\tau_n}(U_{in} \le u, V_n \le v) &= \Phi(u, v; \rho_i) + n^{-1/2}\sum_{j=0}^{3} c_{ij}\Phi^{(j)}(u)\Phi^{(3-j)}(v)\\
&\quad + O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2})
\end{aligned} \tag{3.3}$$

with

$$\rho_1 = n^{-1/2} r_1(b,c), \qquad \rho_2 = -\rho + n^{-1/2} r_2(b,c)$$

(3.4) *for some functions* $r_1(b,c)$, $r_2(b,c)$ *satisfying*

$$r_2(b,c) - r_1(b,c) = -\tfrac{1}{2} b E\psi_{10}^2\psi_{01} - cE\psi_{10}\psi_{02} - bEk\psi_{01},$$

and

$$c_{10} = c_{20},\ c_{11} - c_{21} = -\tfrac{1}{2}E\psi_{10}\psi_{02},\ c_{12} - c_{22} = -\tfrac{1}{2}E\psi_{10}^2\psi_{01} - Ek\psi_{01},$$
$$c_{13} - c_{23} = Eq\psi_{10} - Ek\psi_{10}.$$

Hence (2.11) *holds.*

Since SU and TU are U-statistics and since the expansion is under local alternatives, the proof of Theorem 3.1 needs an Edgeworth expansion for bivariate U-statistics when dealing with a probability measure depending on n. Such a result is given in Götze [(1987), page 215]. [The formulation of Corollary 1.18 in Götze (1987) is not very transparent. A precise formulation, resulting from personal communication with Götze, is given in Albers, Boon and Kallenberg (1997b).]

To apply Götze's result a further (standard) regularity condition is needed. Because of lack of space we do not present this regularity condition and the proof of Theorem 3.1 here, but refer for more details to Albers, Boon and Kallenberg (1997b).

The test based on SK is meant for testing $\theta = 0, \tau = 0$ against $\theta > 0, \tau = 0$. Since $\mu_{1n}(0,0) = Ek\psi_{10}n^{-1/2}$ and $\sigma_{1n}(0,0) = 1$, it follows from Theorem 3.1 that

$$P_{0,0}(SK \le u) = \Phi(u - Ek\psi_{10}n^{-1/2}) + n^{-1/2}c_{13}\Phi^{(3)}(u) + o(n^{-1/2}).$$

It is easily seen that $c_{13} = -\frac{1}{6}E\psi_{10}^3 - Ek\psi_{10}$. Hence $\theta = 0, \tau = 0$ is rejected when

$$SK > u_\alpha + n^{-1/2}\left(\tfrac{1}{6}E\psi_{10}^3(u_\alpha^2 - 1) + Ek\psi_{10}u_\alpha^2\right)$$

with $u_\alpha = \Phi^{-1}(1-\alpha)$, giving size $\alpha + o(n^{-1/2})$ under (0,0).

The test based on SU is meant for testing H_0 against H_1. Since $\mu_{2n}(0,c) = (Eq\psi_{10} - \frac{1}{2}E\psi_{10}\psi_{02})n^{-1/2}$ for all c and $\sigma_{2n}(0) = 1$, it follows from Theorem 3.1 that [it turns out that the $O(\rho^3)$-term cancels under H_0]

$$P_{0,\tau_n}(SU \le u) = \Phi(u - (Eq\psi_{10} - \tfrac{1}{2}E\psi_{10}\psi_{02})n^{-1/2}) + n^{-1/2}c_{23}\Phi^{(3)}(u) + O(\rho n^{-1/2}) + o(n^{-1/2})$$

and hence H_0 is rejected when ($c_{23} = -\frac{1}{6}E\psi_{10}^3 - Eq\psi_{10}$)

$$SU > u_\alpha + n^{-1/2}\left(-\tfrac{1}{2}E\psi_{10}\psi_{02} + \tfrac{1}{6}E\psi_{10}^3(u_\alpha^2 - 1) + Eq\psi_{10}u_\alpha^2\right),$$

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

The test based on TU is meant for testing $\bar{H}_0$ against $\bar{H}_1$. Since $\mu_n(b,0) = O(n^{-1/2})$ and $\sigma_n(0) = 1$, it follows from Theorem 3.1 that [it turns out that the $O(\rho^3)$-term cancels under $\bar{H}_0$]

$$P_{\theta_n,0}(TU \le v) = \Phi(v - \mu_n(b,0)) + n^{-1/2} c_{10} \Phi^{(3)}(v) + O(\rho n^{-1/2}) + o(n^{-1/2}).$$

Hence rejecting $\bar{H}_0$ when

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

gives size $\delta + O(\rho n^{-1/2}) + o(n^{-1/2})$ under $(\theta_n, 0)$.

The pretest procedure is defined to reject H_0 if

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

or

$$SU > u_\alpha + n^{-1/2}\left(-\tfrac{1}{2} E\psi_{10}\psi_{02} + \tfrac{1}{6} E\psi_{10}^3 (u_\alpha^2 - 1) + Eq\psi_{10} u_\alpha^2\right) \quad \text{and} \quad |TU| > u_{\delta/2}.$$

Write the probability of rejection by this testing procedure under $(\theta_n, \tau_n) = (bI_{11}^{-1/2} n^{-1/2}, cI_{22}^{-1/2} n^{-1/2})$ as $\pi^* = \pi^*(b,c)$ and the probability of rejection by the test based on SU as $\tilde{\pi} = \tilde{\pi}(b,c)$. The next theorem is our main result. It gives an attractive expression for the difference $\pi^* - \tilde{\pi}$, from which much insight can be obtained for the comparison of the pretest procedure with the test based on SU.

Theorem 3.2 provides a good approximation of $\pi^* - \tilde{\pi}$ if n is large and ρ is small. Our results are *uniformly* valid in ρ and n, but of course they are only meaningful for $n \to \infty$ and $\rho \to 0$. The latter is no serious restriction, because if ρ is not small the situation is clear: in that case the pretest procedure is unacceptable (see Section 2). Moreover, Theorem 3.2 gives the rate of convergence, $O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2})$, which demonstrates the (high) accuracy of the approximation and its dependence on n and ρ. A further illustration of it is seen in the numerical results of Section 4, while for the technical meaning of $O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2})$ we refer to the remark just before the proof of Lemma 2.1. Note that for $\rho = 0$ the error term in Theorem 3.2 reduces to $o(n^{-1/2})$.

THEOREM 3.2. *Suppose that the regularity conditions hold. Then*

$$\begin{aligned} \pi^*(b,c) - \tilde{\pi}(b,c) = {} & \varphi(u_\alpha - b)\big\{h_1(c, u_{\delta/2})[\rho + \tfrac{1}{2}\{(b/c) + (u_\alpha - b)c\}\rho^2 + m(c, u_\alpha) n^{-1/2}] \\ & + h_2(c, u_{\delta/2})[\rho + \tfrac{1}{2}(u_\alpha - b)c\rho^2 + m(c, u_\alpha) n^{-1/2}] \\ & + h_3(c, u_{\delta/2})[(u_\alpha - b)\rho^2 - E\psi_{10}\psi_{02} n^{-1/2}]\big\} \\ & + O(\rho^3 + \rho n^{-1/2}) + o(n^{-1/2}) \quad \textit{as } \rho \to 0 \textit{ and } n \to \infty, \end{aligned}$$

where

$$h_1(x, y) = x\{\Phi(y-x) - \Phi(-y-x)\},$$
$$h_2(x, y) = \varphi(y+x) - \varphi(y-x),$$
$$h_3(x, y) = \tfrac{1}{2} y\{\varphi(y+x) + \varphi(y-x)\},$$
$$m(x, y) = \tfrac{1}{2}\left\{x E\psi_{10}\psi_{02} + y\left(E\psi_{10}^2\psi_{01} + 2Ek\psi_{01}\right)\right\}.$$

PROOF. Let $u_{1n} = \{u_\alpha + Ek\psi_{10}n^{-1/2} - c_{13}n^{-1/2}(u_\alpha^2 - 1) - \mu_{1n}(b, c)\}/\sigma_{1n}(b, c)$, $u_{2n} = \{u_\alpha + (Eq\psi_{10} - \frac{1}{2}E\psi_{10}\psi_{02})n^{-1/2} - c_{23}n^{-1/2}(u_\alpha^2 - 1) - \mu_{2n}(b, c)\}/\sigma_{2n}(b)$, $v_n^U = \{u_{\delta/_2} - \mu_n(b, c)\}/\sigma_n(c)$, $v_n^L = \{-u_{\delta/_2} - \mu_n(b, c)\}/\sigma_n(c)$.
Then

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

By Theorem 3.1 we get (2.11). It is easily seen that the other conditions of Lemma 2.1 are satisfied with $u_0 = u_\alpha - b$, $v_0 = -u_{\delta/_2} - c$ when v_n^L is used and $v_0 = u_{\delta/_2} - c$ in case of v_n^U. Noting that

$$\begin{aligned}u_{1n} - u_{2n} = {} & -c\rho - \tfrac{1}{2}b\rho^2 + (u_0^2 - 1)(Ek\psi_{10} - Eq\psi_{10})n^{-1/2}\\ & -cm(c, u_\alpha)n^{-1/2} + O(n^{-1} + \rho n^{-1/2})\end{aligned}$$

and

$$\rho_1 - \rho_2 = \rho + n^{-1/2}\{\tfrac{1}{2}bE\psi_{10}^2\psi_{01} + cE\psi_{10}\psi_{02} + bEk\psi_{01}\},$$

straightforward calculation gives the result. □

It is remarkable that $\pi^*(b, c) - \tilde{\pi}(b, c)$ does not depend (up to the considered order) on the q's and r's occurring in SU and TU and only through $Ek\psi_{01}$ on k. This property is related to the phenomenon that "first-order efficiency implies second-order efficiency" [cf. Bickel, Chibisov and van Zwet (1981)]. Note that due to this phenomenon for testing $\theta = 0$ the power at $(bI_{11}n^{-1/2}, 0)$ of SK is up to order $n^{-1/2}$ the same, irrespective of the choice of k.

For the tests mentioned in Section 2 we get, ignoring ρ-terms [cf. (2.5)],

$$\begin{aligned}&\text{LMP}: Ek\psi_{01} = 0,\\ (3.5)\qquad &\text{LR}: Ek\psi_{01} = \tfrac{1}{2}(E\psi_{20}\psi_{01} - E\psi_{10}^2\psi_{01}),\\ &\text{MLE and Wald}: Ek\psi_{01} = E\psi_{20}\psi_{01} - E\psi_{10}^2\psi_{01}.\end{aligned}$$

A special case which is of particular interest is when $\rho = 0$. The following corollary gives this as an immediate consequence of Theorem 3.2.

COROLLARY 3.3. *Suppose that the regularity conditions hold and* $\rho = 0$. *Then*

$$\pi^*(b,c) - \tilde{\pi}(b,c)$$
$$= \varphi(u_\alpha - b)[h(c, u_{\delta/_2})m(c, u_\alpha) - h_3(c, u_{\delta/_2})E\psi_{10}\psi_{02}]n^{-1/2} + o(n^{-1/2})$$

as $n \to \infty$ *where* $h = h_1 + h_2$.

4. Consequences for the actual size and power. Under the null hypothesis H_0: $\theta = 0$, the difference $\pi^* - \tilde{\pi}$ reduces to $\pi^*(0,c) - \alpha + o(n^{-1/2})$. Hence the departure from the nominal level of the pretest procedure follows from Theorem 3.2 and approximately equals

$$\varphi(u_\alpha)\left[h(c, u_{\delta/_2})\left\{\rho + \tfrac{1}{2}u_\alpha c\rho^2 + m(c, u_\alpha)n^{-1/2}\right\} + h_3(c, u_{\delta/_2})\left\{u_\alpha\rho^2 - E\psi_{10}\psi_{02}n^{-1/2}\right\}\right], \tag{4.1}$$

where $h = h_1 + h_2$. Firstly, we note that the family of distributions and classes of tests are involved through only four parameters: ρ, $E\psi_{10}\psi_{02}$, $E\psi_{10}^2\psi_{01}$ and $Ek\psi_{01}$. If ρ tends to 0 and $n \to \infty$, the error $\pi^*(0,c) - \alpha$ tends to 0. Also for $\alpha \to 0$ the error tends to 0.

Secondly, we analyze the behavior of h and h_3. As $h(c, u_{\delta/_2})$ is odd in c and $h_3(c, u_{\delta/_2})$ is even in c, only consider $c \geq 0$. Note that h increases in $u_{\delta/_2}$ and that $h(0, u_{\delta/_2}) = h(c, 0) = 0$ for all c and $u_{\delta/_2}$. This implies, among others, $h(c, u_{\delta/_2}) \geq 0$ for all $c \geq 0$ and $u_{\delta/_2} \geq 0$. Since $\lim_{c\to\infty} ch(c, u_{\delta/_2}) = 0$ and $\lim_{c\to\infty} h_3(c, u_{\delta/_2}) = 0$, there exists $c^* = c^*(u_{\delta/_2})$ for which the error, given by (4.1), is maximal.

If we ignore the ρ^2-terms (which includes of course the case $\rho = 0$ presented in Corollary 3.3), and assume that $E\psi_{10}\psi_{02} = 0$ (which holds in many examples; see Sections 4.1 and 4.2), then (4.1) reduces to

$$\varphi(u_\alpha)h(c, u_{\delta/_2})\left\{\rho + u_\alpha\left(\tfrac{1}{2}E\psi_{10}^2\psi_{01} + Ek\psi_{01}\right)n^{-1/2}\right\}. \tag{4.2}$$

Expression (4.2) can be interpreted as follows. If *SK* is used without a pretest, we have $\delta = 0$ and $h(c, u_{\delta/2}) = c$ and hence (4.2) becomes

$$\varphi(u_\alpha)c\left\{\rho + u_\alpha\left(\tfrac{1}{2}E\psi_{10}^2\psi_{01} + Ek\psi_{01}\right)n^{-1/2}\right\}, \tag{4.3}$$

which means that the error $\pi^*(0,c) - \alpha$ grows linearly in that case. For $\delta > 0$, if there were no dependence between the preliminary test and the main tests, the deviation from the nominal level would be $\{\pi(0,c) - \alpha\}\{1 - \bar{\pi}(0,c)\}$, with

$$\bar{\pi}(0,c) = P_{(0, cI_{22}^{-1/2}n^{-1/2})}\left(|TU| > u_{\delta/_2}\right).$$

Noting that $\pi(0,c) - \alpha$ is given by (4.3), it is seen that only the first-order approximation of $1 - \bar{\pi}(0,c)$ is needed to get

$$\varphi(u_\alpha)h_1(c, u_{\delta/_2})\{\rho + u_\alpha\left(\tfrac{1}{2}E\psi_{10}^2\psi_{01} + Ek\psi_{01}\right)n^{-1/2}\}$$

as approximation for $\{\pi(0,c) - \alpha\}\{1 - \bar{\pi}(0,c)\}$. The function c is now replaced by the redescending $h_1(c, u_{\delta/_2})$, sketched for $\delta = 0.05$ in Figure 1.

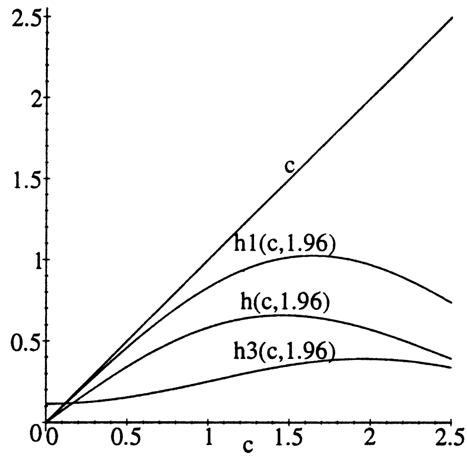

FIG. 1.

Finally, the effect of taking the dependence into account is a further reduction of the error through replacement of h_1 by h. (That indeed $0 \le h \le h_1$ is seen by noting that $h_2 \le 0$.)

Again ignoring terms of order ρ^2, it is immediately seen from Theorem 3.2 that the power difference is nothing but the size difference, inflated by a factor $\varphi(u_\alpha - b)/\varphi(u_\alpha)$. For $\alpha = 0.05$ this factor runs from 1 at $b = 0$ to its maximal value 3.9 at $b = u_\alpha$ and, being a multiple of φ, it then decreases. Here, it is seen that the idea mentioned in the introduction of getting higher power due to "knowing the value of the nuisance parameter" does not come true. If there is a gain in power, it is due to the difference between size and level, possibly blown up by some factor. Noting that $h(c, u_{\delta/2})$ is odd in c, a gain in power at c will as a rule imply a loss in power at $-c$. An exception should be made for (very) small values of c, because in that case ρ^2-terms are dominant and should not be ignored. Consider the special case $c = 0$. The approximation given by Theorem 3.2 now reads as

$$\begin{aligned} \varphi(u_\alpha - b)\big[\tfrac{1}{2}b\rho^2\left\{\Phi(u_{\delta/_2}) - \Phi(-u_{\delta/_2}) - 2u_{\delta/_2}\varphi(u_{\delta/_2})\right\} \\ + u_{\delta/_2}\varphi(u_{\delta/_2})\{u_\alpha\rho^2 - E\psi_{10}\psi_{02}n^{-1/2}\}\big]. \end{aligned} \tag{4.4}$$

Since the coefficient of $b\rho^2$ is positive for all $u_{\delta/_2}$, the term with $b\rho^2$ gives some (small) gain in power not due to the difference between size and level.

In the remainder of this section we consider some examples, gathered together in two important classes: two-parameter exponential families and symmetric location-scale families.

4.1. *Two-parameter exponential families* Let μ be some probability measure on $\mathbb{R}^2$ and assume that the moment generating function of μ exists in some open neighborhood Ω_0 of $(0, 0)$. Define

$$f(x; \theta, \tau) = \exp\left\{\theta x^{(1)} + \tau x^{(2)} - \omega(\theta, \tau)\right\}$$

with $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$ and ω a normalizing constant.

By direct calculation we have $\psi_{02} = \psi_{01}^2 - 1$ and hence $E\psi_{10}\psi_{02} = E\psi_{10}\psi_{01}^2$. Similarly, $E\psi_{20}\psi_{01} = E\psi_{10}^2\psi_{01}$ and therefore $Ek\psi_{01} = 0$ for LMP, LR, MLE and Wald's test; compare (3.5).

We consider some special cases. In Example 4.1 numerical calculations are performed, while in the other examples simulation results are presented. Each of the simulations in this paper is based on 100,000 repetitions. Hence, the standard deviations of the simulated power differences are at most $(100{,}000)^{-1/2} = 0.0032$. This reduces to 0.0019 if at least one of the "powers" π^* or $\tilde{\pi}$ is at most 0.05 and to 0.001 if both "powers" are at most 0.05. In Example 4.1–4.3 we use LMP tests with nuisance parameters, if present, estimated by the MLE in the unrestricted model.

EXAMPLE 4.1. Suppose we have a sample from a normal distribution with expectation θ and variance 1. Further, a second sample is available from a normal distribution, with variance w^2, but we are not sure whether this second sample has the same expectation as the first one. We denote this expectation by $\theta - \tau$. If the second sample has the same mean, we might want to use it together with the first sample. A preliminary test is performed to investigate the equality of the means.

Let Y be a r.v. with a standard normal distribution and let Z be independent of Y, having a normal distribution with expectation 0 and variance w^2. Let the probability measure μ correspond to $(X^{(1)}, X^{(2)}) = (Y + Z/w^2, -Z/w^2)$. The density $f(x; \theta, \tau)$ represents the distribution of two independent normally distributed r.v.'s with expectation θ, $\theta - \tau$ and variance 1, w^2, respectively. Here we have $\rho = -1/\sqrt{w^2 + 1}$ and hence the larger w, the smaller $|\rho|$. Furthermore, $E\psi_{10}^2\psi_{01} = E\psi_{10}\psi_{01}^2 = 0$ and $SU = n^{-1/2}\sum_{i=1}^n\{X_i^{(1)} + X_i^{(2)}\}$. This means that SU is the standardized sample mean of the first sample, as it should be, since for unknown τ the second sample is of no use for testing about θ.

Because SK, SU and TU are exactly normally distributed, it is no surprise that the $n^{-1/2}$-terms cancel in this case. Application of Theorem 3.2 yields

$$\begin{aligned}\pi^*(b, c) - \tilde{\pi}(b, c) = \varphi(u_\alpha - b)\big[&h(c, u_{\delta/2})\left\{\rho + \tfrac{1}{2}(u_\alpha - b)c\rho^2\right\} \\ &+ h_1(c, u_{\delta/2})\tfrac{1}{2}(b/c)\rho^2 + h_3(c, u_{\delta/2})(u_\alpha - b)\rho^2\big] \\ &+ O(\rho^3).\end{aligned} \tag{4.5}$$

Suppose that we want to know the error of the size of the pretest procedure if the nominal level $\alpha = 0.05$, $c = 1$, $w = 4$ and $\delta = 0.05$. Approximation (4.5) with $b = 0$ yields -0.0092. The numerical value of $\pi^*(0, c) - \alpha$ equals -0.0099. Furthermore, consider the power difference $\pi^* - \tilde{\pi}$ if the nominal level $\alpha = 0.05$, $c = -1$, $b = 1$, $w = 5$ and $\delta = 0.05$. Approximation (4.5) yields 0.0466. The numerical value of $\pi^*(1,-1)-\tilde{\pi}(1,-1)$ equals 0.0473. This illustrates the accuracy of the approximations.

In the next example the main testing problem concerns the scale parameter in an exponential distribution. However, we are not sure about the model and therefore the idea is to perform a preliminary test to decide whether the exponential distribution is appropriate against the alternative of gamma distributions.

EXAMPLE 4.2. Let Z be exponentially distributed with parameter 1. Let the probability measure μ correspond to the distribution of $(X^{(1)}, X^{(2)}) = (Z, \log Z)$ on $\mathbb{R}^2$. The density $f(x;\theta,\tau)$ gives the gamma-density with "scale" parameter $1-\theta$ and second parameter $\tau+1$. Here

$$\rho = \operatorname{cov}(Z, \log Z)/\sqrt{\operatorname{var} Z \operatorname{var}(\log Z)} = \sqrt{6}/\pi = 0.78.$$

As argued in Section 2 this indicates that the pretest procedure is unacceptable. Indeed, the simulated value of $\pi^*(0,1)$ if $\alpha = 0.05$, $n = 25$ and $\delta = 0.05$ equals 0.1808, which differs too much from the prescribed level.

EXAMPLE 4.3. Let Z be standard-normally distributed. Let the probability measure μ correspond to the distribution of $(X^{(1)}, X^{(2)}) = (Z, Z^2)$ on $\mathbb{R}^2$. The density $f(x;\theta,\tau)$ gives the normal distribution with expectation $\theta/(1-2\tau)$ and variance $(1-2\tau)^{-1}$. Since $EZ^3 = 0$, we have $\rho = 0$ and moreover, $E\psi_{10}\psi_{01}^2 = 0$, because $EZ^5 = 0$. Further, $E\psi_{10}^2\psi_{01} = EZ^2(Z^2-1)/\sqrt{2} = \sqrt{2}$.

The main testing problem is to test expectation 0 against a positive expectation. A pretest is presented for $\bar{H}_0$: $\tau = 0$ against $\bar{H}_1$: $\tau \neq 0$, which means variance equal to one against variance not equal to one. Test statistic SK (with known variance) yields the Gauss-test, while SU gives (an asymptotically equivalent form of) the t-test. Test statistic TU is the well-known χ^2-test statistic for testing the variance.

Application of Corollary 3.3 gives

$$\pi^*(b,c) - \tilde{\pi}(b,c) = \varphi(u_\alpha - b)h(c, u_{\delta/2})u_\alpha/\sqrt{2n} + o(n^{-1/2}), \tag{4.6}$$

which approximation was also presented in Theorem 2.1 of Albers, Boon and Kallenberg (1997a).

Suppose we want to know the error in size if $\alpha = 0.05$, $n = 50$, $c = -1$ and $\delta = 0.05$. Application of (4.6) with $b = 0$ yields -0.0099 while the simulated value equals -0.0113. Next consider the power if $\alpha = 0.05$, $n = 50$, $c = 1$, $b = 2$ and $\delta = 0.05$. Approximation (4.6) gives 0.0359 and the simulated value equals 0.0462.

Finally, suppose that we want to know how wrong the size can be if $\alpha = 0.05$, $n = 50$ and $\delta = 0.05$. Then we have to deal with $\max_c h(c, 1.96)$. The maximum equals 0.6581 and is attained at $c = 1.4583$ (cf. also Figure 1). Therefore, according to (4.6), the error maximized over c equals 0.0112. The simulated value of $\pi^*(0,c) - \tilde{\pi}(0,c)$ at $c = 1.4583$ is 0.0157, while the simulated value of $\pi^*(0,c)-\alpha$ at $c=1.4583$ equals 0.0024. Consequently, the simulated value of $\tilde{\pi}(0,c)$ is $0.0524-0.0157=0.0367$, which is rather far from $\alpha=0.05$. Note that $c=1.4583$ corresponds to 1.4118 as value of the variance of the normal distribution. Hence the variance is rather far away from 1. The conservatism of the test based on SU is also the reason for the difference between the approximated and simulated value of $\pi^*(0,c)-\tilde{\pi}(0,c)$.

If we replace SU by the t-test in the pretest procedure, $\pi^*(0,c)-\alpha$ can be calculated numerically. The result is 0.0119 and the approximation according to (4.6), 0.0112, is quite close to it.

It is seen from the preceding examples that many interesting situations can be written in the form of a two-parameter exponential family. This way of presenting makes application very easy.

4.2. *Symmetric location-scale families.* Let f_0 be a given probability density w.r.t. the Lebesgue measure on $\mathbb{R}$ with $f_0(x)>0$ for all $x\in\mathbb{R}$. Consider the location-scale family defined by

$$f(x;\theta,\tau)=\frac{1}{1+\tau}f_0\left(\frac{x-\theta}{1+\tau}\right)$$

with (θ,τ) in some open neighborhood of $(0,0)$. Suppose in addition that f_0 is symmetric: $f_0(x)=f_0(-x)$ for all $x\in\mathbb{R}$.

Since $\psi_{10}^*=-f_0'/f_0$ and $\psi_{01}^*=-1-x(f_0'/f_0)$, and symmetry implies that f_0'/f_0 is odd, we get that ψ_{10}^* is odd, ψ_{01}^* is even and hence $\rho=0$. Direct calculation gives $\psi_{02}^*=2+4x(f_0'/f_0)+x^2(f_0''/f_0)$, which is even. Therefore, we have $E\psi_{10}\psi_{02}=0$. Application of Corollary 3.3 gives

$$\begin{aligned}\pi^*(b,c)-\tilde{\pi}(b,c)&=\tfrac{1}{2}\varphi(u_\alpha-b)h(c,u_{\delta/2})u_\alpha\left(E\psi_{10}^2\psi_{01}+2Ek\psi_{01}\right)n^{-1/2}\\&\quad+o(n^{-1/2}).\end{aligned}\tag{4.7}$$

This means that the approximation for the difference in a general symmetric location-scale family is the same as that for the normal case, corresponding to $f_0(x)=\exp(-\frac{1}{2}x^2)/\sqrt{2\pi}$, except for the multiplicative constant $E\psi_{10}^2\psi_{01}+2Ek\psi_{01}$, which may differ from family to family. For the normal distribution $E\psi_{10}^2\psi_{01}=\sqrt{2}$.

In the following examples we use LMP tests where nuisance parameters, if present, are estimated by the MLE in the unrestricted model.

EXAMPLE 4.4. Let f_0 be the logistic distribution, that is,

$$f_0(x)=e^{-x}(1+e^{-x})^{-2}.$$

We get

$$\begin{aligned}\psi_{10}(x)&=\sqrt{3}(1-e^{-x})(1+e^{-x})^{-1},\\ \psi_{01}(x)&=3(\pi^2+3)^{-1/2}\{x-1-(x+1)e^{-x}\}(1+e^{-x})^{-1},\\ E\psi_{10}^2\psi_{01}&=3(\pi^2+3)^{-1/2}\end{aligned}\tag{4.8}$$

and hence the difference in the logistic case is $3(\pi^2+3)^{-1/2}/\sqrt{2} = 0.591$ times the difference in the normal case.

Suppose we want to know how large we should take δ in order that the relative error of the size $(\pi^*(0,c)-\alpha)/\alpha$ is at most ε for some given $\varepsilon > 0$ (e.g., $\varepsilon = 0.2$ leading to 2% or 3% if $\alpha = 0.025$) for all c. Remember that $c^* = c^*(u_{\delta/_2})$ is the c that maximizes the approximation of $\pi^*(0,c)-\alpha$, given in (4.1). In view of (4.7) c^* is the c that maximizes $h(c, u_{\delta/_2})$ for given δ. Let $h^*(u_{\delta/_2}) = h(c^*, u_{\delta/_2})$ be the maximum value. Note that, as h itself decreases in δ, so does h^*. By (4.7) and (4.8), δ should be sufficiently large to ensure

$$h^*(u_{\delta/_2}) \le \frac{2\varepsilon\alpha(\pi^2+3)^{1/2}n^{1/2}}{3u_\alpha\varphi(u_\alpha)}. \tag{4.9}$$

To evaluate $h^*(u_{\delta/_2})$ we can use the following further approximation. For $u_{\delta/_2} \ge 1$ we have $c^* \ge 1.1$ and hence we ignore $c\Phi(-u_{\delta/_2}-c)$ and $\varphi(u_{\delta_2}+c)$. Then $c^* = u_{\delta/2} - g(1/u_{\delta/_2})$ with $g = (\varphi/\Phi)^{-1}$ and $h^*(u_{\delta/_2}) = \varphi(g(1/u_{\delta/_2}))\{u_{\delta/_2}[u_{\delta/_2} - g(1/u_{\delta/_2})]-1\}$. It turns out that for $1 \le u_{\delta/_2} \le 2.5$ a good approximation is obtained by taking $g(x) = \frac{3}{2}(1-\frac{5}{4}x)$ and $h^*(x) = \frac{3}{5}x - \frac{1}{2}$. Taking $n = 25$, $\alpha = 0.025$ and $\varepsilon = 0.2$, (4.9) reads as $\frac{3}{5}u_{\delta/_2} - \frac{1}{2} \le 0.5220$, yielding $\delta \ge 0.089$. The simulated value of $\pi^*(0,c)$ at $\delta = 0.089$ and $c = u_{\delta/_2} - \frac{3}{2}\{1-5/(4u_{\delta/_2})\} = 1.3041$ equals 0.0303, just as it should be.

EXAMPLE 4.5. Let f_0 be a mixture of two normal distributions, as is often used, for example, in robustness studies [cf. Huber (1981), page 2],

$$f_0(x) = 0.95\varphi(x) + 0.05\varphi(x/3)/3.$$

We get

$$E\psi_{10}^2\psi_{01} = 1.104$$

and hence the difference in this mixture model is 0.781 times the difference in the normal case.

Suppose we want to know how large we should take n in order that the (absolute) error of the size is at most 0.01 when $c = 1.5$, $\alpha = 0.05$ and $\delta = 0.05$. Inserting this in (4.7) with $b = 0$ we get $0.0616\ n^{-1/2} \le 0.01$ implying $n \ge 37.9$. The simulated value for $n = 38$ of $\pi^*(0, 1.5)$ equals 0.0673 and that of $\pi^*(0,1.5) - \tilde{\pi}(0,1.5)$ is 0.0111, which is close to the required 0.01.

It is seen from the examples that quite good and very useful answers are achieved for many questions in an easy way using the approximations given in Theorem 3.2 and Corollary 3.3.

Acknowledgment. The authors thank F. Götze for helpful discussion on Corollary 1.18 in Götze (1987).

REFERENCES

[1] ALBERS, W., BOON, P. C. and KALLENBERG, W. C. M. (1997a). The asymptotic behavior of tests for normal means based on a variance pretest. Memorandum 1403, Univ. Twente.

[2] ALBERS, W., BOON, P. C. and KALLENBERG, W. C. M. (1997b). Size and power of pretest procedures. Memorandum 1421, Univ. Twente.

[3] ALBERS, W., BOON, P. C. and KALLENBERG, W. C .M. (1998). Testing equality of two normal means using a variance pretest. *Statist. Probab. Lett.* **38** 221–227.

[4] BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

[5] BICKEL, P. J., CHIBISOV, D. M. and VAN ZWET, W. R. (1981). On efficiency of first and second order. *Internat. Statist. Rev.* **49** 169–175.

[6] GÖTZE, F. (1987). Approximations for multivariate U-statistics. *J. Multivariate Anal.* **22** 212–229.

[7] HOEL, P. G. (1984). *Introduction to Mathematical Statistics*, 5th ed. Wiley, New York.

[8] HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

[9] LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

[10] MARKOWSKI, C. A. and MARKOWSKI, E. P. (1990). Conditions for the effectiveness of a preliminary test of variance. *Amer. Statist.* **44** 322–326.

[11] MOSER, B. K., STEVENS, G. R. and WATTS, C. L. (1989). The two-sample t test versus Satterthwaite's approximate F test. *Comm. Statist. Theory Methods* **18** 3963-3975.

FACULTY OF MATHEMATICAL SCIENCES
UNIVERSITY OF TWENTE
P.O. BOX 217, 7500 AE ENSCHEDE
THE NETHERLANDS
E-MAIL: W.C.M.Kallenberg@math.utwente.nl