

# Testing for Multivariate Equivalence with Random Quadratic Forms

Rafael Weißbach

Universität Dortmund, Fachbereich Statistik,  
Institut für Wirtschafts- und Sozialstatistik, 44221 Dortmund, Germany

Correspondence: Rafael Weißbach, Institut für Wirtschafts- und Sozialstatistik,  
Fachbereich Statistik, Universität Dortmund, 44221 Dortmund, Germany  
E-Mail: Rafael.Weissbach@uni-dortmund.de, Fax: +49/231/755-5284.

## Abstract

Multivariate equivalence testing becomes necessary whenever the similarity rather than a difference between several treatment groups with multiple endpoints has to be shown. This problem occurs in various applications, including bioequivalence or the comparison of dissolution profiles. Therefore, several tests have been suggested during the last decade for the assessment of multivariate equivalence. Recently Munk & Pflüger (1999) proposed to test ellipsoidal instead of rectangular hypotheses as it is current practice in many applications. In this paper we provide several asymptotic tests for ellipsoidal equivalence which are compared numerically with competitors suggested by Brown, Cassella & Hwang (1995) and Munk & Pflüger (1999). We find that the proposed tests are superior (up to 90%) to both tests with respect to power. In addition, a simulation study reveals the suggested tests as robust against violation of normality. These tests are very simple to apply, because inversion of confidence regions is avoided. Asymptotic formulas for the power function and sample size determination are given. Finally, all procedures are compared in two data examples.

*keywords and phrases:* hypotheses testing, multivariate bioequivalence, multiple endpoints, confidence inclusion rules, equivalence, Hotelling's  $T^2$  test, equivariance, bootstrap, quadratic forms, power function.

## 1 Introduction

Recently, testing of multivariate equivalence hypotheses has become a field of certain interest, mainly motivated by the problem of assessing multivariate *bioequivalence*, i.e. the assessment of a similar rate and extend of absorption in the blood circulation (cf. CHOW & LIU (1992) for a survey) of several pharmacokinetic characteristics of two different formulations of a drug. We refer to BROWN, CASSELLA & HWANG (1995), HSU, LU & CHAN (1995), CHINCHILLI & ELSWICK (1997) or WANG, DASGUPTA & HWANG (1999) for a discussion and several tests for the multivariate bioequivalence problem. However, the multivariate equivalence problem is not solely restricted to bioequivalence assessment as the following example from neurophysiology shows. For further applications in environmental or managerial science cf. ERICKSON & McDONALD (1995), DIXON (1998) or MCBRIDE (1999).

**Example 1.** Exteroceptive suppression (cf. SCHOENEN (1993)), ES, of temporalis muscle activity is the electrophysiological correlate of the jaw-opening reflex. ES consists itself in two further components ESM and ESP, the monosynaptic and polysynaptic suppression. STEINHOFF ET AL. (1996) investigated the influence of epilepsy on ES in a study with 20 healthy volunteers and 31 patients with epilepsy. The conjecture, that ES could be possibly influenced by epilepsy was essentially based on the observation that Parkinson's disease and chronic headache seem to cause an effect on ES (NAKASHIMA ET AL. (1990)). This analogy, however, was never based on solid physiological grounds. Therefore, it was the aim of STEINHOFF ET AL.'S (1996) study to show that this conjecture cannot be supported by the data. To this end several two sided  $t$  tests for the null hypotheses of no difference were performed, each resulting in  $p$ -values larger than

0.8. However, this does not imply the rigorous assessment of similarity at a controlled error rate between the control and epileptic group as pointed out by STEINHOFF ET AL. (1996). An explorative data analysis suggests to assume a bivariate (ESM, ESP) normal model with the same covariance structure for the epileptic (E) and control group (C), respectively (cf. MUNK & PFLÜGER (1999) for the raw data)

$$X_{ij} \sim N_2(\theta_i, \tilde{\Sigma}) \quad i = C, E \quad \text{and} \quad j = 1, \dots, n_i \quad \text{with} \quad n_C = 19, n_E = 31.$$

The following (with pooled variance estimator for the ESP and ESM values, respectively) estimators for the mean difference and covariance matrix are obtained

$$X := \bar{X}_C - \bar{X}_E = \begin{pmatrix} -0.0586 \\ -0.3325 \end{pmatrix} \quad \text{and} \quad \hat{\Sigma} = \begin{pmatrix} 0.8526 & 0.0188 \\ 0.0188 & 1.3525 \end{pmatrix}.$$

In the following we present tests in order to assess that there is no 'relevant difference' of the means between the reference group and the epilepsy group, for *ESM* and *ESP*, respectively. Hence, it follows that *ES* is not an appropriate diagnostic tool for epilepsy or the success of a particular therapeutic method as conjectured by STEINHOFF ET AL. (1996). To this end it is required to show at a controlled error rate that  $\theta_E \approx \theta_C$ .

According to various definitions of similarity several tests for multivariate equivalence were suggested in the literature during the last 5 years. Most of these tests are designed for rectangular hypotheses under a normal assumption (cf. BERGER (1982), BERGER & HSU (1996), ROY (1996) or WANG, DASGUPTA & HWANG (1999)). These authors provide several tests in order to assess that a multivariate mean  $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$  is contained in a hyperrectangle, i.e. for the hypotheses

$$H : \theta \in \Theta_H = \{\theta : \exists i : |\theta_i| > \Delta\} \quad \text{vs.} \quad K : \theta \in \Theta_K = \{\theta : \max_{i=1, \dots, p} |\theta_i| \leq \Delta\}. \quad (1)$$

The most popular tests for (1) are based on the intersection union principle applied to confidence interval inclusion rules (BERGER & HSU (1996)). These tests are simple to perform and well established in applications. They may have the disadvantage to be rather conservative (resulting in a low power) when the number of endpoints  $p$  increases as pointed out by HWANG (1996) and MUNK & PFLÜGER (1999). The definition of equivalence in terms of the maximum norm is, of course, not the only possibility to formalize similarity. Therefore, MUNK & PFLÜGER (1999) suggested to consider ellipsoidal hypotheses

$$H : \theta' A \theta > \Delta \quad \text{vs.} \quad K : \theta' A \theta \leq \Delta, \quad (2)$$

for a positive definite matrix  $A > 0$ . This is, e.g., in accordance with a measure of equivalence recommended by the FDA (1997) for testing equivalence of dissolution profiles, i.e. the proportion of a tablet dissolved against time. For a discussion of recent methodology in dissolution profile testing we refer to O'HARA ET AL. (1997), SHAH ET AL. (1998) or SIERRA-CAVAZOS & BERGER (1999). Because these hypotheses cannot be represented as the intersection resp. union of several one-dimensional hypotheses the intersection union principle (cf. BERGER & HSU (1996)) or related methods cannot be applied and different testing methodology is required. Here, the general testing principle of Aitchison (1964) by inverting  $1 - \alpha$  confidence regions is applicable. This leads, however, in general to very conservative tests (MUNK (1994), BERGER & HSU (1996)). For general hypotheses this technique was improved by BROWN ET AL. (1995) and for convex alternatives by MUNK & PFLÜGER (1999).

This paper is organized as follows. In Section 2 we state the precise model for the test. In Section 3 we review on existing methods for the assessment of multivariate equivalence and discuss various new tests for the hypotheses (2). Our approach is based on the asymptotic normality of the centered quadratic form  $X'AX - \theta' A \theta$  (after normalizing by the sample size). In Section 4 we present the results of a Monte Carlo study on size and power which

shows that the normal approximation yields rather liberal tests, i.e. the nominal level is exceeded. In order to improve the accuracy of the normal approximation several finite sample approximations are investigated including various Box - type  $\chi^2$  approximations and bootstrap tests. We found that the bias corrected accelerated bootstrap method (EFRON & TIBSHIRANI, 1993) represents a good compromise between a very powerful method and a method which maintains its nominal level with high accuracy. In particular, differences in power up to 90% compared to BROWN, CASELLA & HWANG'S (1995) and MUNK & PFLÜGER'S (1999) test were observed. Furthermore, simulations show that the new test is very robust against violation of normality.

Finally, in Section 5 we compare all methods by reanalyzing the data set in Example 1 and an example from CHINCHILLI & ELSWICK (1997), who investigated the multivariate bioequivalence of two different formulations of Ibuprofen.

The paper closes with a discussion and summary section where also formulae for computing sample size and power are given.

## 2 Model and testing problem

We restrict ourselves for the moment to i.i.d. observations coming from a normal model. Generalizations to the two sample comparison and other designs will be discussed later. We mention also that our results are asymptotically valid under very weak assumptions on the error distribution which is in contrast to existing methods for testing multivariate equivalence. This will be made precise in the next section.

Let throughout the following  $X_1, \dots, X_n \sim N_p(\theta, \tilde{\Sigma})$  independent random variables, which follow a  $p$ -variate normal distribution with mean  $\theta \in \mathbb{R}^p$  and covariance  $\tilde{\Sigma} \in \mathbb{R}^{p \times p}$  which is assumed to be positive definite,  $\tilde{\Sigma} > 0$ . Sufficient statistics are the mean  $\frac{1}{n} \sum_{i=1}^n X_i =: \bar{X}$  and the covariance estimator  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^t$ . It follows that  $\bar{X} \sim N_p(\theta, \Sigma)$ , and  $\hat{\Sigma} \sim W_d(\Sigma)$  (here  $\Sigma = \frac{1}{n} \tilde{\Sigma}$  and  $d = n - 1 \geq p$ ) are independently distributed

according to a multivariate normal distribution with unknown mean  $\theta \in \mathbb{R}^p$  and unknown covariance structure  $\Sigma$  and a Wishart distribution with  $d$  degrees of freedom, respectively (ANDERSON (1984)). Throughout the following we may consider w.l.o.g. instead of the testing problem in (2) the standardized problem

$$H : \theta^t \theta > 1 \quad \text{vs.} \quad K : \theta^t \theta \leq 1. \quad (3)$$

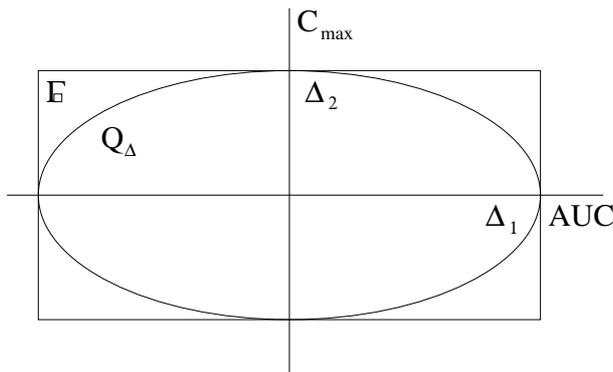
This can be achieved in (2) for any  $A > 0$  by a transformation of the data  $X_i$  into

$$X_i^* = \frac{1}{\sqrt{\Delta}} A^{\frac{1}{2}} X_i \quad (4)$$

because  $A > 0$  allows a unique decomposition as  $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$ ,  $A^{1/2}$  a symmetric matrix. In order to keep notation simple we will write in the following  $X_i$  instead of  $X_i^*$ .

Before we consider the construction of several tests for (3), we find it pertinent to recall briefly the discussion in (MUNK & PFLÜGER (1999)) concerning the choice of ellipsoidal hypotheses instead of rectangular hypotheses as in (1). In most applications  $\theta$  will be the difference of two means  $\theta = \theta_1 - \theta_2$  which have to be shown to be equivalent (cf. Example 1). For example, drug authorities, such as the FDA, currently require that in a single dose bioequivalence study of oral drug formulation *average bioequivalence* is shown with respect to both, AUC and  $C_{\max}$ . We mention that the criterion of average bioequivalence which focuses on the comparison of means solely, has been criticised by various authors (cf. HAUCK & ANDERSON (1992) or SCHALL (1995) among many others) and different bioequivalence criteria (various types of population and individual bioequivalence) have been suggested. This is highlighted in a recent draft guidance for industry entitled "Average, Population and Individual Approaches to Establishing Bioequivalence" (U.S. DEP. OF HEALTH AND HUMAN SERVICES, FOOD AND DRUG ADMINISTRATION, CDER, ROCKVILLE, MD (1999)). In the present paper, however,

we focus only on average bioequivalence, simply denoted as bioequivalence. According to the FDA guidance, bioequivalence may be claimed if every univariate comparison of the pharmacokinetic parameters  $AUC$ ,  $T_{\max}$ , or  $C_{\max}$  allows one to declare bioequivalence at level of significance  $\alpha = 0.05$  (e.g. GUIDANCE FOR INDUSTRY: IN VIVO PHARMACOKINETICS AND BIOAVAILABILITY STUDIES AND IN VITRO DISSOLUTION TESTING FOR LEVOTHYROXINE SODIUM TABLETS, U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, CDER (1999)). To our knowledge up to now all approaches focus on hypotheses as in (1), i.e. bioequivalence is declared if each of the components of the difference of the mean vector is close to zero. We believe, however, that often the consideration of quadratic forms  $Q_A(\theta) = \theta^t A \theta$  as a distance measure is more appropriate (cf. MUNK & PFLÜGER (1999) for a discussion). Here the region of equivalence is the ellipsoid  $Q_\Delta = \{\theta : Q_A(\theta) \leq \Delta\}$  where  $\Delta$  denotes a fixed tolerance bound (cf. Figure 1).



**Figure 0:** *Ellipsoidal ( $Q_\Delta$ ) and rectangular hypotheses ( $\Gamma_\square$ ) in the case  $p = 2$*

Interestingly, for the analysis of dissolution profiles, the FDA (1997), recommends ellipsoidal hypotheses completely analogous to (3). Methods for this problem are discussed in MOORE & FLANNER (1996), O'HARA ET AL. (1997) and SHAH, TSONG, SATHE & MIA (1998). For example, the  $f_2$  method proposed in MOORE & FLANNER (1996) requires weights  $\omega_k$  which corresponds to the eigenvalues of the matrix  $A$  in (2). MA ET AL. (2000) performed a simulation study for bootstrapping the  $f_2$ -statistic with the

percentile method and found quite reasonable performance of the resulting test.

### 3 Tests

In this Section we will discuss several testing procedures for the multivariate equivalence problem (3). This includes tests already suggested in the literature and various new tests. We restrict our representation to the case of an *i.i.d.* sample  $X_1, \dots, X_n$ . Generalizations to the two sample comparison with independent samples, such as required in Example 1, are straight forward and discussed in Example 1 in Section 5.

#### 3.1 Confidence inclusion rules

A classical principle due to AITCHISON (1964) states that for any confidence set  $C_{1-\alpha}(X, \hat{\Sigma})$  for  $\theta$  at level  $1 - \alpha$  the region

$$A(\Theta_H) = \{X : \Theta_H \cap C_{1-\alpha}(X, \hat{\Sigma}) \neq \emptyset\}$$

defines the acceptance region of a test for  $H : \theta \in \Theta_H$  against  $K : \theta \in \Theta_K$ ,  $\Theta_H \cap \Theta_K = \emptyset$ , at level  $\alpha$  because

$$P_\theta(\theta \notin C_{1-\alpha}(X, \hat{\Sigma})) \leq \alpha.$$

Hence, if  $\Theta_K := \{\theta \in \mathbb{R}^p : \theta^t \theta \leq 1\}$ , the  $p$ -dimensional sphere, the following rule constitutes a test at level  $\alpha$  for the hypothesis (3):

$$\psi_\alpha(X, \hat{\Sigma}) = \begin{cases} 1 & : C_{1-\alpha}(X, \hat{\Sigma}) \subset \Gamma_K \\ 0 & : \text{else} \end{cases}.$$

MUNK & PFLÜGER (1999) showed, that if the confidence set obeys certain equivariance properties and the alternative  $K$  is convex (which is, of course, the case in the present setting) it is possible to improve AITCHISON'S (1964) result because then the resulting level of the test is  $\alpha/2$ . The level  $\alpha/2$  is sharp, i.e. it cannot be improved in general. This leads to the following test which results from inverting Hotelling's confidence set.

### 3.1.1 Hotelling's $T^2$ confidence set

In a first step we invert Hotelling's  $T^2$  test (CF. ANDERSON, (1984)) for  $H : \theta = \theta_0$  with test statistic

$$T_{\theta_0}^2 := d(X - \theta_0)^t \hat{\Sigma}^{-1} (X - \theta_0) \quad (5)$$

which yields the  $(1 - \alpha)$  confidence set

$$C_{1-\alpha}^{T^2} := \{\theta_0 : T_{\theta_0}^2 \leq f_{\alpha,p,d}^2\}, \quad (6)$$

where  $f_{\alpha,p,d}^2 = \frac{dp}{d-p+1} F_{p,d-p+1}^{-1}(\alpha)$  and  $F_{\mu,\nu}$  denotes a central  $F$ -distribution with  $\mu$  and  $\nu$  degrees of freedom. Theorem 2.1 in MUNK & PFLÜGER (1999) yields that

$$\psi_{\alpha}^{T^2}(X, \hat{\Sigma}) = \begin{cases} 1 & : C_{1-2\alpha}^{T^2}(X, \hat{\Sigma}) \subset \Theta_K \\ 0 & : \text{else} \end{cases}$$

constitutes a test at level  $\alpha$ . Observe, that the confidence region required for the test  $\psi_{\alpha}^{T^2}$  has confidence coefficient  $1 - 2\alpha$  and not  $1 - \alpha$ .

### 3.1.2 Brown-Casella-Hwang's confidence set

Another candidate for the confidence inclusion rule is the confidence set constructed by BROWN ET AL. (1995). The reasoning behind their approach is to minimise the expected volume at a preassigned parameter point  $\theta_0$ , which can be assumed w.l.o.g. as  $\theta_0 = 0$ . Originally, this confidence set was constructed for the case of known variance  $\Sigma$ , resulting in the confidence set

$$C_{1-\alpha}^{BCH}(X, \Sigma) := \left\{ \theta : (\theta^t \Sigma^{-1} \theta)^{\frac{1}{2}} \leq z_{1-\alpha} + \frac{X \Sigma^{-1} \theta}{(\theta^t \Sigma^{-1} \theta)^{\frac{1}{2}}} \right\},$$

where  $z_{1-\alpha}$  is the upper  $\alpha$  quantile from a univariate standard normal distribution. For the case of unknown covariance we will throughout the following simply replace  $\Sigma$  by the estimator  $\frac{1}{n} \hat{\Sigma}$ . Hence denote

$$\psi_{\alpha}^{BCH}(X, \hat{\Sigma}) = \begin{cases} 1 & : C_{1-\alpha}^{BCH}(X, \frac{1}{n} \hat{\Sigma}) \subset \Theta_K \\ 0 & : \text{else} \end{cases}$$

as the according equivalence test. Note, that the *BCH*-confidence set does not share the equivariance property (like the Hotelling's  $T^2$  confidence set) required for application of MUNK & PFLÜGERS (1999) Theorem 2.1 and hence the  $1 - 2\alpha$ -adjustment is not valid. Other confidence regions and generalizations of this approach to nonnormal errors can be found in DASGUPTA, GOSH & ZEN (1995), WANG, DASGUPTA & HWANG (1999) or ROY (1996).

## 3.2 Testing with quadratic forms

### 3.2.1 The $\delta$ -method

The most direct way to construct a test for (3) is to estimate in a first step  $\theta^t \theta$  by  $\|X\|_p^2 = X^t X$  and then reject the hypothesis for  $\|X\|_p^2$  too small. The main difficulty

encountered with this approach is the effective determination of the distribution of  $\|X\|_p^2$ . Observe, that (cf. MATHAI & PROVOST (1992))

$$\|X\|_p^2 \stackrel{\mathcal{D}}{=} \sum_{j=1}^p \lambda_j (U_j + b_j)^2 \quad (7)$$

equals in distribution a sum of weighted noncentral  $\chi_1^2$  random variables where  $U_j \sim N_1(0, 1)$  are i.i.d.,  $j = 1, \dots, p$ . The weights  $\lambda_j$ ,  $j = 1, \dots, p$ , are the eigenvalues of  $\Sigma$  and

$$b = (b_1, \dots, b_p)^t = \left( \frac{1}{\sqrt{n}} P^t \tilde{\Sigma}^{\frac{1}{2}} \theta \right)^t, \quad (8)$$

where  $P$  denotes an orthogonal matrix such that  $\frac{1}{n} P^t \tilde{\Sigma} P = \text{diag}(\lambda_1, \dots, \lambda_p)$ . Here  $\text{diag}(\lambda_1, \dots, \lambda_p)$  denotes a  $p \times p$  diagonal matrix with diagonal elements  $\lambda_1, \dots, \lambda_p$ . This representation highlights the difficulty to determine the exact distribution of  $\|X\|_p^2$  because particularly the estimation of the eigenvalues of  $\Sigma$  causes serious problems. Therefore, we investigate in the following various approximations of the distribution of  $\|X\|_p^2$ .

The simplest method of approximation is the C.L.T.

**Theorem 1.** *Let  $X_1, \dots, X_n \sim F$  be an i.i.d. sample from a  $p$ -variate distribution  $F$ , s.t.  $EX_1 = \theta \in \mathbb{R}^p$ ,  $\tilde{\Sigma} = \text{Cov}[X_1] > 0$  and  $E\|X_1\|_p^2 < \infty$ . Then we have for any consistent estimator  $\hat{\Sigma}$  of the covariance  $\tilde{\Sigma}$*

$$T_n^\delta := \sqrt{n} \frac{X^t X - \theta^t \theta}{\sqrt{4X^t \hat{\Sigma} X}} \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{for } \theta \neq 0, \quad \text{as } n \rightarrow \infty.$$

**Proof.** Apply the  $\delta$ -method and the multivariate C.L.T. (cf. Serfling (1980)) in order to conclude that

$$\sqrt{n}(X^t X - \theta^t \theta) \xrightarrow{\mathcal{D}} N(0, 4\theta^t \tilde{\Sigma} \theta) \quad \text{for } \theta \neq 0.$$

Observe, that from Kolmogorov's strong law of large numbers it follows that  $X^t \hat{\Sigma} X \xrightarrow{a.s.} \theta^t \tilde{\Sigma} \theta$ . Now the assertion follows from Slutsky's Lemma (Chow & Teicher (1997)).  $\square$

From Theorem 1 we obtain the following asymptotic test

$$\psi_\alpha^\delta(X, \hat{\Sigma}) = \begin{cases} 1 & : \sqrt{n}(X^t X - 1) / \sqrt{4X^t \hat{\Sigma} X} \leq z_\alpha \\ 0 & : \text{else} \end{cases}$$

for the problem (3), where  $z_\alpha$  denotes the lower  $\alpha$ -quantile of the standard normal distribution. In the next section we will demonstrate in a Monte Carlo study that this tests improves essentially on the power of the confidence region inclusion rules discussed in the last subsection. However, it turns also out that the normal approximation yields a rather liberal test (i.e. the nominal level is exceeded) and hence we investigate in the following various finite sample approximations.

**Remark 1.** In particular, when  $\theta$  is close to zero, the distribution of  $T_n^\delta$  is skewed. For  $\theta = 0$  the asymptotic normality in Theorem 1 even fails to hold. Instead, the distribution, after multiplication by  $n^{1/2}$  once more, is a sum of weighted  $\chi^2$  distributions (cf. (7) again). To account for this skewness we used an idea dating back to Box (1954) where the approximation by a scaled  $\chi^2$ -distribution  $g\chi_f^2$  with random degrees of freedom  $f$  and scaling parameter  $g$  is suggested. However, simulations for the approach revealed sizes of up to three times the level for small sample sizes as 10. As a consequence, we refrain from displaying the results. Increasing the number of fitted moments by using a non-central  $\chi^2$ -distribution did not show to improve the size.

### 3.2.2 The bootstrap

Another option is, of course, to work with bootstrap approximations. The bias corrected and accelerated bootstrap ( $BC_\alpha$ ) is investigated with

$$\psi_\alpha^{bootbca}(X_1, \dots, X_n) = \begin{cases} 1 & : (X^{*t} X^*)_{[n^{boot} \times \alpha_2] + 1} \leq 1 \\ 0 & : \text{else} \end{cases}$$

where  $\alpha_2 = \Psi\left(\hat{z}_0 + \frac{\Psi^{-1}(1-\frac{\alpha}{2})}{1-\hat{a}(\hat{z}_0 + \Psi^{-1}(1-\frac{\alpha}{2}))}\right)$  with  $\hat{z}_0 = \#\left\{\frac{X^{*t} X^* \leq X^t X}{n^{boot}}\right\}$  and the acceleration  $\hat{a} = \frac{\sum_{i=1}^n (\vartheta_i - \bar{\vartheta})^2}{(\sum_{i=1}^n (\vartheta_i - \bar{\vartheta})^3)^{\frac{2}{3}}}$ . Here  $\vartheta_i$  is a jackknifed version of  $X^t X$ , s.t.  $X_i$  omitted. Further  $\bar{\vartheta}$  denotes the empirical mean of the  $\vartheta_i$ 's  $i = 1, \dots, n$  and  $\Psi$  the c.d.f. of a standard normal r.v.. For more details see EFRON & TIBSHIRANI (1993).

**Remark 2.** Additional to the bias corrected and accelerated bootstrap we used the percentile bootstrap (EFRON & TIBSHIRANI (1993)). As in the case of the  $\chi^2$ -approximations we found a sizes for our small sample situation of 10 observations which are clearly - up two-and-a-half times - above the level. Again, we refrain from displaying the results.

## 4 Simulation study

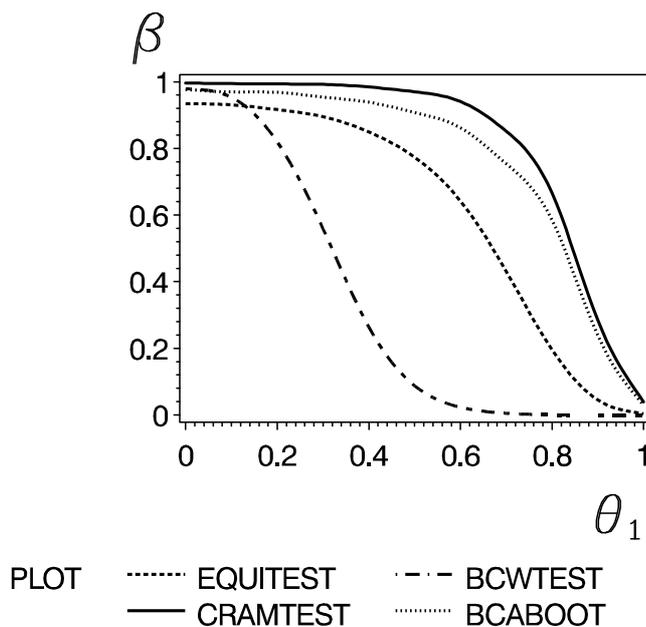
In this section we report on a Monte-Carlo study for the testing problem (3) which was conducted using the random generator of SAS/IML, internal programming language of the statistical software SAS, Version 6.12 and 8.03. In each scenario 10000 Monte-Carlo simulations were performed. The nominal level is always chosen as  $\alpha = 0.05$ . In our first scenario 20 bivariate normal random variables were generated where the covariance was assumed as

$$\tilde{\Sigma} = \begin{pmatrix} \frac{1}{8} & -\frac{3}{16} \\ -\frac{3}{16} & \frac{17}{16} \end{pmatrix}.$$

The power curve was obtained by cubic B-spline interpolation (in the alternative  $K$ ) between the grid points

$$\left\{ \theta : \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \frac{i}{10} \\ 0 \end{pmatrix}, \quad i = 1, \dots, 10 \right\}.$$

The number of bootstrap replications for the  $BC_a$  bootstrap was  $n^{boot} = 300$ , respectively. Power curves for all tests discussed in the last section are shown in Figure 1. Note, that  $\theta_1 = 1$  corresponds to the point  $(1, 0)^t$  on the boundary of the null hypotheses (3), whereas smaller values of  $\theta_1$  are in the alternative of equivalence. If for the  $BC_a$  bootstrap the argument of  $\Psi^{-1}(\cdot)$  in the calculation of  $\hat{z}_0$  was zero, we have dropped this value. This occurred in less than 1% of all cases.

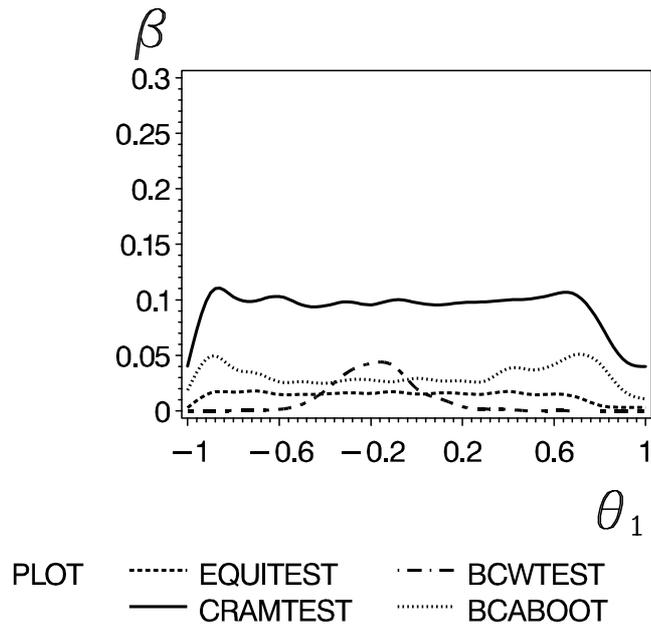


**Figure 1:** Power curves of the  $\delta$ -method (CRAMTEST), Munk & Pflüger’s (1999) test (EQUITEST), the Brown et al. (1995) test (BCWTEST), and the BCA bootstrap (BCABOOT) for  $n = 20$ , normal error and 10.000 simulations. The nominal level was  $\alpha = 0.05$ . The next simulation study was performed in order to investigate the actual size of all tests more detailed. Therefore in Figure 2 the actual level of all tests is displayed for the

following boundary points of the hypothesis  $H$  in (3)

$$\left\{ \theta : \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \frac{i}{10} \\ \sqrt{1 - (\frac{i}{10})^2} \end{pmatrix}, \quad i = -10, \dots, 10 \right\}.$$

under the same distributional assumptions as in Figure 1. Again cubic B-spline interpolation is used to interpolate between these grid points.



**Figure 2:** Actual size of the tests where  $\|\theta\| = 1$  and  $\theta_2 > 0$ .

$\tilde{\Sigma}$	$n$	T-Hot	Brown-CH	delta	$BC_a$
$\begin{pmatrix} \frac{1}{8} & -\frac{3}{16} \\ -\frac{3}{16} & \frac{17}{16} \end{pmatrix}$ $(\rho \approx -0.50)$	10	0.0005	0.0000	0.0364	0.0314
	20	0.0045	0.0000	0.0405	0.0293
	50	0.0200	0.0010	0.0431	0.0254
$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$ $(\rho = 0)$	10	0.0144	0.0385	0.0908	0.0505
	20	0.033	0.0500	0.0752	0.0382
	50	0.0492	0.0486	0.0586	0.0306
$\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$ $(\rho = 0.5)$	10	0.0136	0.0241	0.0984	0.0607
	20	0.0267	0.026	0.08	0.0409
	50	0.0507	0.0292	0.1514	0.0332

**Table 1:** Size of all tests for various correlations and sample sizes where  $\theta' = (0, 1)$ .

The numerical values can be found in In Table 1 where also additional sample sizes and covariance structures were investigated.

From these Figures the following results can be drawn: The normal approximation ( $\delta$ -method) is anti conservative in some of scenarios investigated. The  $BC_a$  bootstrap keeps its nominal level with high accuracy, even under the assumption of strong correlation. Note, that the actual level of BROWN, CASELLA & HWANG'S (1995) test may fall far below the nominal level, particularly when negative correlation is present. A similar observation holds for the test of MUNK & PFLÜGER (1999). As a main conclusion from Figure 1 we may draw that the  $BC_a$  bootstrap outperforms all the other competitors, having reasonable size and power.

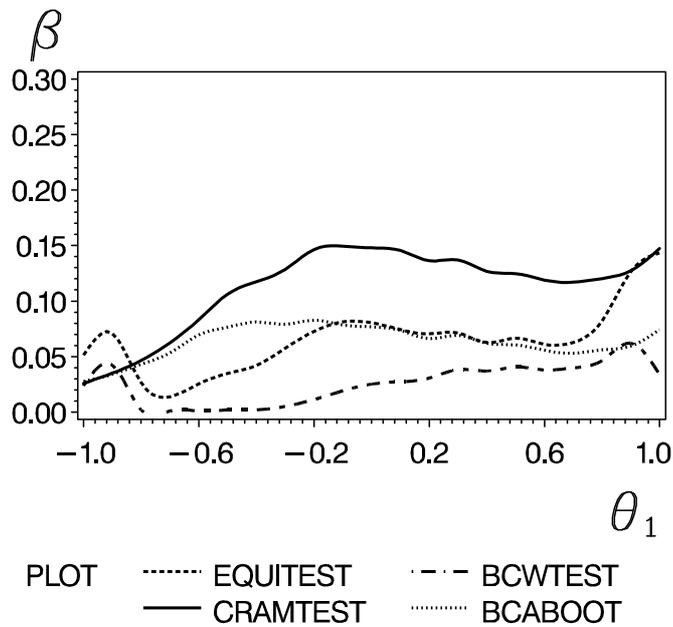
In what follows we would like to address briefly some robustness aspects of the above discussed tests. To this end we considered outcomes generated by two independent ( $\chi^2 -$

$1)/4$  distributions with additive location  $\theta$ . Note, that the expectation of these r.v.'s is  $\theta$  and the variance of each component equals  $\frac{1}{8}$  which reflects the first normal component in the previous normal simulation scenario.

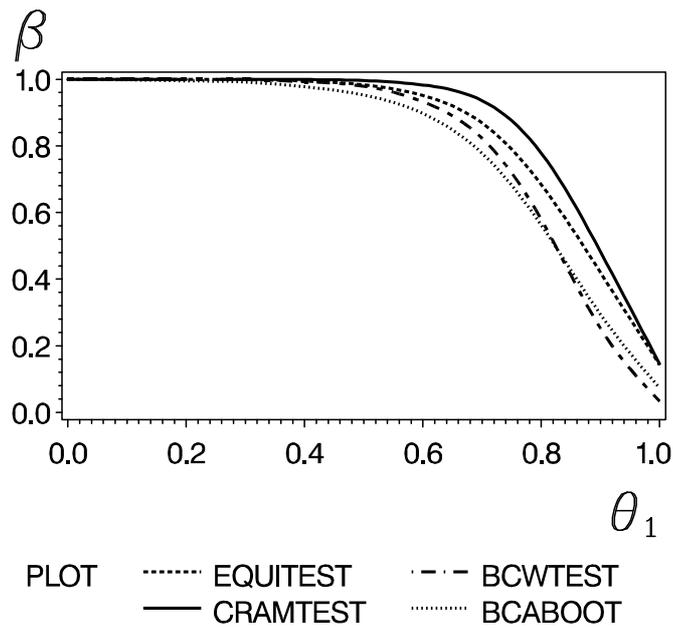
In Table 2 we have displayed a selection of numerical values for the boundary point  $\theta = (0, 1)$ . The corresponding curves of the actual level can be found in Figure 3.

Test	T-Hot	Brown-CH	delta	$BC_a$
size	0.1428	0.0334	0.1453	0.0732

**Table 2:** Size of tests for  $(\chi^2 - 1)/4$  error



**Figure 3:** Size of the tests on the boundary of the hypotheses, where  $\|\theta\| = 1$  and  $\theta_2 > 0$  for  $(\chi^2 - 1)/4$  error.



**Figure 4:** Power curves of the tests for  $n = 20$ ,  $(\chi^2 - 1)/4$  error and 10.000 simulations. The nominal level was  $\alpha = 0.05$ .

In Figure 4 the resulting power functions are displayed. In summary we may draw the following conclusions. The actual level is drastically exceeded by MUNK & PFLÜGER'S (1999) test as well as for the  $\delta$ -method. BROWN, CASELLA & HWANG'S (1995) method performs rather accurate albeit very conservative for negative values of  $\theta_1$ . In contrast the  $BC_a$  method yields a very accurate approximation of the nominal level  $\alpha = 0.05$  over abroad range of  $\theta_1$  values. The differences in power between the two last named tests are rather small. Note the the superior power of the other tests does not yield a fair comparison due to its exceedance of the nominal level. In summary, only BROWN, CASELLA & HWANG'S (1995) method and the  $BC_a$ -bootstrap are able to maintain the given level under skewed errors. Interestingly, the power of the BCH test is significantly increased by the  $\chi^2$ -error.

## 5 Examples

To demonstrate the practical relevance of the proposed methods we reanalyse the epilepsy data from the Introduction and an example of a multivariate bioavailability study from CHINCHILLI & ELSWICK (1997).

### 5.1 Example from neurology

In order to transfer the suggested methodology to the situation of Example 1 in the Introduction we require a generalization of Theorem 1 in Section 3 to the two sample case.

**Theorem 2.** *Let  $X_1, \dots, X_m \sim F$  be an i.i.d. sample from a  $p$ -variate distribution  $F$ , and independently  $Y_1, \dots, Y_n \sim G$  an i.i.d. sample from a  $p$ -variate distribution  $G$ , s.t.  $EX_1 - EY_1 = \theta_1 - \theta_2 = \theta \in \mathbb{R}^p$ ,  $\tilde{\Sigma}_1 = Cov[X_1]$ ,  $\tilde{\Sigma}_2 = Cov[Y_1] > 0$  and  $E\|X_1\|_p^2, E\|Y_1\|_p^2 < \infty$ . Then we have for  $X = \bar{X} - \bar{Y}$  and  $\hat{\Sigma} = (1 + \lambda^{-1})\hat{\Sigma}_1 + (1 + \lambda)\hat{\Sigma}_2$  with  $m/n \rightarrow \lambda \in (0, \infty)$ , that*

$$T_{(m,n)}^\delta := \frac{\sqrt{m+n}(X^t X - \theta^t \theta)}{\sqrt{4X^t \hat{\Sigma} X}} \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{for } \theta \neq 0.$$

**Proof.** The proof is similar to Theorem 1, where we take into account that

$$\sqrt{m+n} \{ \bar{X} - \bar{Y} - \theta \} \xrightarrow{\mathcal{D}} (1 + \lambda^{-1})^{1/2} Z_1 + (1 + \lambda)^{1/2} Z_2,$$

where  $Z_1$  and  $Z_2$  are independent  $p$ -variate normal r.v.s with covariances  $\tilde{\Sigma}_1$  and  $\tilde{\Sigma}_2$ , respectively.

Observe, that for the case of equal covariances  $\Sigma_1 = \Sigma_2$  (which we have assumed in this example) the asymptotic variance  $\Sigma = (\lambda + 1)^2/\lambda \Sigma_1$ . Observe further, that a  $BC_a$  bootstrap test can be immediately obtained by drawing separately subsamples from the samples  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$ .

In order to apply the afore mentioned equivalence tests we have to specify a bound  $\Delta$ . Here we have chosen 20% of the range of the control group, for  $ESM$  and  $ESP$ , respectively.

Hence we end up with the testing problem

$$H : \left\| \begin{pmatrix} \frac{1}{4}\theta_{ESM} \\ \frac{1}{8}\theta_{ESP} \end{pmatrix} \right\|_2^2 > 1 \quad \text{versus} \quad K : \left\| \begin{pmatrix} \frac{1}{4}\theta_{ESM} \\ \frac{1}{8}\theta_{ESP} \end{pmatrix} \right\|_2^2 \leq 1 .$$

Test	T-Hot	Brown-CH	delta	$BC_a$
$p$ -value	0.0101	0.0045	$7.610^{-9}$	0.00989

**Table 3:**  $p$ -values for the epilepsy data

We mention that for the inversion of Hotelling's confidence interval the pooled estimator

$$\hat{\Sigma} = \frac{1}{n + m - 2} \left( \sum_{i=1, \dots, n} (X_i - \bar{X})^t (X_i - \bar{X}) + \sum_{j=1, \dots, m} (Y_j - \bar{Y})^t (Y_j - \bar{Y}) \right)$$

has been used where the degrees of freedom of the  $F$ -distribution are now 2 and  $n + m - 2$ .

In summary we find that all tests decide significantly in favour of equivalence.

## 5.2 Bioequivalence study

The equivalence of availability of Ibuprofen in the blood of two drugs will be reinvestigated for data presented by CHINCHILLI & ELSWICK (1997). Following these authors it is to be

tested whether the ratio of the expected  $AUC$ 's (Area under the Curve) and the expected  $C_{\max}$ 's (maximal concentration) are between 0.8 and 1.25.

In detail, it is assumed that independent bivariate r.v.'s are observed in a  $2 \times 2$  crossover design without period effects. After a logarithmic transformation it is to assess whether

$$H_1 : |\eta_A^T - \eta_A^R| \leq \log 1.25 \quad \text{and} \quad |\eta_C^T - \eta_C^R| \leq \log 1.25 \quad (9)$$

where  $\eta_A^{(\cdot)}$  and  $\eta_C^{(\cdot)}$  denotes the corresponding expectation of  $\log AUC^{(\cdot)}$  and  $\log C_{\max}^{(\cdot)}$ . Taking

$$X_{1i} := \frac{1}{\log 1.25} (\log AUC_i^T - \log AUC_i^R), \quad X_{2i} := \frac{1}{\log 1.25} (\log C_{\max,i}^T - \log C_{\max,i}^R)$$

yields the model from Theorem 1 with  $p = 2$ . After dividing the logarithms of raw data by  $\frac{1}{\log 1.25}$  in order to transform the problem into the standardized form of the hypothesis in (3) we obtain the following summary statistics

$$\text{(Mean difference)} \quad \bar{X} = \begin{pmatrix} -0.121332 \\ 0.190746 \end{pmatrix}, \quad \text{(Covariance)} \quad \hat{\Sigma} = \begin{pmatrix} 0.1300518 & 0.1368549 \\ 0.1368549 & 0.8743755 \end{pmatrix}.$$

For the complete data see CHINCHILLI & ELSWICK (1997). In our previous notation, we have  $\theta = (\frac{\eta_A^T - \eta_A^R}{\log 1.25}, \frac{\eta_C^T - \eta_C^R}{\log 1.25})$ . In Table 2 we display the resulting  $p$ -values for the hypothesis (3). We find that all tests under consideration lead to a significant rejection of  $H$ , showing bivariate bioequivalence of the two ibuprofen formulations. Note, however, that the ordering of the  $p$ -values nicely reflects the superior power as found in the last section of the tests based on quadratic forms compared to the inclusion rules.

Test	T-Hot	BCH	delta	BC <sub><math>\alpha</math></sub>
$p$ -value	0.00050	0.00021	$\leq 10^{-15}$	$9.97710^{-11}$

**Table 4:**  $p$ -values for the Ibuprofen data.

## 6 Further Remarks and Extensions

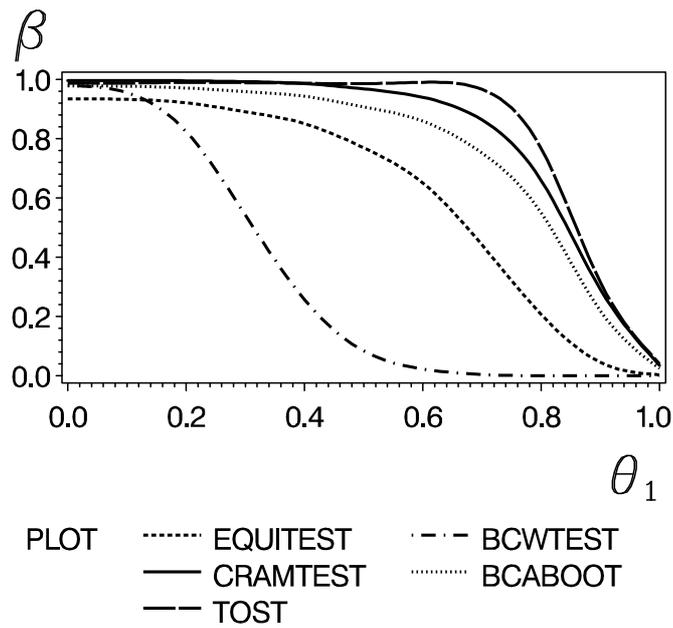
**Remark 3.** Some readers may find it unsuitable for many purposes to choose ellipsoidal hypotheses instead of the more common rectangular ones. Nevertheless, we would like to mention, that even for the case of testing a hyperrectangle as in (9) our tests for ellipsoidal alternatives could be used. To this end, the largest sphere within the rectangle has to be chosen as the alternative (cf. again Figure 0). Throughout the following we have always chosen  $\Delta = 1$  and  $p = 2$ .

In the following we investigate briefly how the ellipsoidal tests behave in comparison to the intersection union methods if they are used for rectangular hypotheses. Here, the intersection-union-test (TOST-test) has rejection region

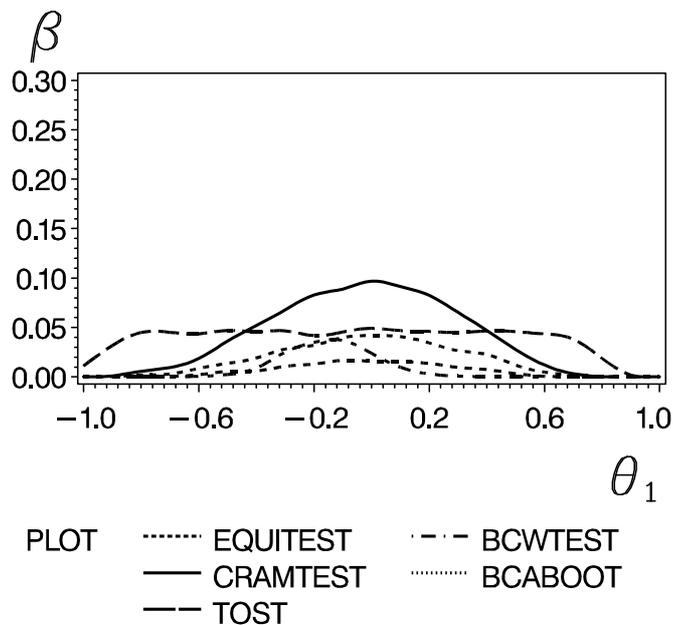
$$\bigcap_{i=1}^p \left\{ |X_i| < \Delta_i - t_{1-\alpha, d} \sqrt{\frac{\hat{\Sigma}_{ii}}{\alpha}} \right\}$$

where  $X_i$  denotes the  $i$ -th coordinate of  $X$  and  $t_{1-\alpha, d}$  the  $1 - \alpha$  quantile of a  $t$ -distribution with  $d$  degrees of freedom.

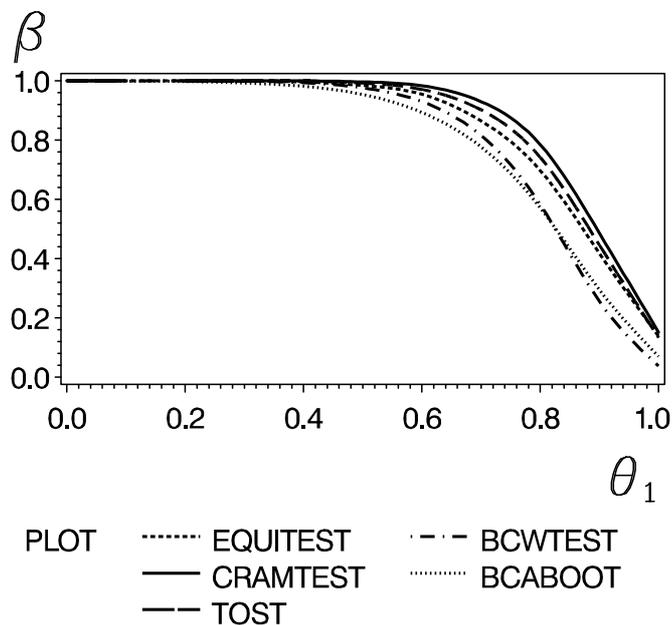
In Figure 5 the power curves of all tests are displayed under normal error assumption. As it was expected the IU-test TOST is most powerful due to its property to be exactly designed for a rectangular hypotheses. Note, that the BCH-test is significantly inferior with respect to power whereas all the other tests have reasonable power.



**Figure 5:** Power curves of the tost procedure (*TOST*),  $\delta$ -method (*CRAMTEST*), Munk & Pflüger's (1999) test (*EQUITEST*), the Brown et al. (1995) test (*BCWTEST*), and the BCA bootstrap (*BCABOOT*) for  $n = 20$ , normal error and 10.000 simulations. The nominal level was  $\alpha = 0.05$ .



**Figure 6:** Size of the tests on the boundary of the hypotheses, where  $\theta_2 = 1$  for normal error.



**Figure 7:** Power curves of the tests for  $n = 20$ ,  $(\chi^2 - 1)/4$  error and 10.000 simulations. The nominal level was  $\alpha = 0.05$ .

In Figure 6 the nominal level is investigated where  $\theta_1 \in [-1, 1]$  and  $\theta_2 = 1$ . It can be seen that again the  $BC_a$ -test keeps the nominal level albeit rather conservative. The IU-method has size close to  $\alpha = 0.05$  over the entire range of  $\theta_1$  values besides those values of the hypothesis where  $\theta_1$  is close to -1 or 1.

In Figure 7 and Table 5 a similar scenario is investigated for skewed error. Observe, that from Table 5 it follows that the IU-method now exceeds the nominal level whereas the  $BC_a$  method again keeps the nominal level rather accurate.

Test	TOST	T-Hot	Brown-CH	delta	BC <sub>a</sub>
size	0.1330	0.1390	0.0379	0.1516	0.0702

**Table 5:** Size of tests for  $(\chi^2 - 1)/4$  error

We mention that further simulations have shown (not displayed) that for increasing dimension  $p$  the difference in power between both tests becomes smaller. When  $p = 10$ , the power of the  $BC_a$  test is nearly the same as for the IU-method.

**Remark 4.** In practical applications often it is important to determine the required sample size to achieve a preassigned power of the asymptotic tests for ellipsoidal hypotheses. We will assume that  $X_1, \dots, X_n$  are independently normally distributed with expectation  $\theta$  and covariance matrix  $\tilde{\Sigma}$ . The most interesting case occurs if exact equivalence holds, i.e. where  $\theta = 0$ . In this case the power of the test  $\psi_\alpha^\delta$  (and hence asymptotically also of its  $BC_a$  bootstrap version) is given as

$$P_{\theta=0} \left( \sqrt{n} \frac{\|X\|_p^2 - 1}{2\sqrt{X'\hat{\Sigma}X}} < u_\alpha \right) = P_{\theta=0} \left( n\|X\|_p^2 < n^{1/2}u_\alpha 2\sqrt{X'\hat{\Sigma}X} + n \right).$$

Observe, that  $n\|X\|_p^2$  is distributed as  $\sum_{j=1}^p \tilde{\lambda}_j U_j^2$  where  $\tilde{\lambda}_j$  are the eigenvalues of  $\tilde{\Sigma}$  and  $U_j$  are i.i.d. standard normal r.v.'s. This is asymptotically a nondegenerate law whereas the r.h.s. expression  $n^{1/2}u_\alpha 2\sqrt{X'\hat{\Sigma}X} = 2u_\alpha \sqrt{(n^{1/2}X)'\hat{\Sigma}(n^{1/2}X)}$  is asymptotically distributed as the square root of  $\sum_{j=1}^p \zeta_j U_j^2$  with  $\zeta_j$  being the eigenvalues of  $\tilde{\Sigma}^2$  and the same r.v.'s  $U_j$  as before. The resulting law is rather complicated and hence we suggest to simulate the resulting power for given  $\tilde{\Sigma}$  in order to obtain the required sample size. The case of independent components is particularly simple and we give in Table 6 some values of the required sample size in order to achieve a power of 0.8 and 0.9, respectively.

dimension $p$	$\sigma^2 = 0.25$	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
$p = 2$	3(4)	6(8)	11(14)	20(26)
$p = 3$	4(5)	6(8)	13(16)	24(27)

**Table 6:**  $BC_a$ -Bootstrap simulated sample sizes to achieve 0.8 (0.9) power for perfect equivalence  $\theta = 0$  for two and three dimensions for various variances of the independent univariate normal variables (300 bootstrap replications)

We mention finally, that the above formula can be generalized of course to the two sample case, as treated in Theorem 4.2.

## 7 Conclusions

We have compared several tests for ellipsoidal hypothesis  $H : \theta' A \theta \geq \Delta$  versus  $K : \theta' A \theta \leq \Delta$ . It was shown that the  $BC_a$ -bootstrap version of the test based on the statistic  $X'AX$  yields satisfactory results with respect to power and size. Even when this test is applied to rectangular hypotheses its power comes close to the intersection union test (which is only applicable for rectangular hypotheses), particularly as the dimension  $p$  of the parameter  $\theta$  increases. Moreover, the suggested  $BC_a$  bootstrap test is very robust against violation of the normality assumption of the error.

Finally we have proved asymptotic normality of the test statistic  $X'AX$  and the validity of the bootstrap principle for testing multivariate equivalence. This gives a theoretical justification for recent approaches suggested for dissolution profile testing where a similar test statistic was suggested and numerically investigated.

### Acknowledgements

The author is indebted to A. Munk for raising the general question and constant guidance in the analysis. R. Berger and G. Freitag pointed out some references. The financial

support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

## REFERENCES

- AITCHINSON, J.: Confidence-region tests, *Journ. Roy. Statist. Soc. B*, **26**, 1964, 462-476.
- ANDERSON, T.W.: *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, 1984, second edition.
- BERGER, R.L.: Multiparameter hypothesis testing and acceptance sampling, *Technometrics*, **24**, 1982, 295-300.
- BERGER, R.L., HSU, J.C.: Bioequivalence trials, intersection-union tests, and equivalence confidence sets (with discussion), *Statistical Science*, **11**, 1996, 283-319.
- BOX, G. E. P.: Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. *Ann. Math. Stat.*, **25**, 1954 290–302.
- BROWN, L.D., CASELLA, G., HWANG, J.T.G.: Optimal confidence sets, bioequivalence, and the limaçon of Pascal, *Journ. Americ. Statist. Assoc.*, **90**, 1995, 880-889.
- CHINCHILLI, V.M., ELSWICK, R.K.JR.: The multivariate assessment of bioequivalence, *Journ. Biopharmac. Statist.*, **7**, 1997, 113-123.
- CHOW, S.-C., LIU, J.-P.: *Design and Analysis of Bioavailability and Bioequivalence Studies*, Marcel Dekker, 1992.
- CHOW, Y.S., TEICHER, H.: *Probability Theory: Independence, Interchangeability, Martingales*. 3rd ed. Springer Texts in Statistics. New York, NY: Springer, 1997.
- DASGUPTA, A., GOSH, J.K., ZEN, M.M., A new general method for constructing confidence sets in arbitrary dimensions: with applications. *Ann. Stat.* **23**, 1995, 1408-1432.
- DIXON, P.M.: Assessing effect and no effect with equivalence tests. *Risk Assessment: Logic and Measurement*, eds. M. Newman & C. Strojan., Chelsea, MI: Ann Arbor Press., 1998.
- EFRON, B., TIBSHIRANI, R.J.: *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

- ERICKSON, W.P., McDONALD, L.L., Tests for bioequivalence of control media and test media in studies of toxicity. *Environ. Toxic. Chem.* **14**, 1995, 1247-1256.
- UNITED STATES FOOD AND DRUG ADMINISTRATION (FDA) Guidance for Industry: Dissolution Testing of Immediate Release Solid Oral Dosage Forms, *U.S. Department of Health and Human Services, CDER*, 1997.
- UNITED STATES FOOD AND DRUG ADMINISTRATION (FDA): Bioavailability and bioequivalence requirements, *U.S. Government Printing Office*, 1992, U.S. Code of Federal Regulations, volume 21, chapter 320.
- HAUCK, W.W., ANDERSON, S.: Types of bioequivalence and related statistical considerations. *Intern. Journ. Clin. Pharmacol., Ther. Toxic.* **30**, 1992, 181-187.
- HSU, H.C., LU, H.L., CHAN, K.K.H.: A novel multivariate approach for estimating the degree of similarity in bioavailability between two pharmaceutical products, *Journ. Biopharmaceutic. Statist.*, **84**, 1995, 768-772.
- HWANG, J.T.G. Comment on “Bioequivalence trials, intersection union tests and equivalence confidence sets”. *Statistical Science*, **11**, 1996, 313-315.
- JOHNSON, L.J., KOTZ, S.: *Continuous Univariate Distributions II*, Houghton Mifflin Company, 1969.
- MA, M.-C. AND WANG, B. B. C. AND LIU, J.-P. AND TSONG, Y., Assessment of similarity between dissolution profile. *J. Biopharmaceut. Statist.*, **10**, 2000, 229-249.
- MATHAI, A.M., PROVOST, S.B.: *Quadratic Forms in Random Variables*, Marcel Dekker, 1992.
- MCBRIDE, G.: Equivalence tests can enhance environmental statistics and management, *Austral. New Zealand Journ. Statist.*, **41**, 1999, 19-29.
- MOORE, J.W., FLANNER, H.H. Mathematical comparison of dissolution profiles. *Pharmaceut. Technology* **20**, 1996, 64-74.
- MUNK. A.: Correspondence: On a method of combining double t-test and Anderson-Hauck test, *Biometrics*, **50**, 1994, 884-886.
- MUNK. A., PFLÜGER, R.:  $1 - \alpha$  equivariant confidence rules for convex alternatives are  $\frac{\alpha}{2}$ -level tests – with applications to the multivariate assessment of bioequivalence, *Journ. Americ. Statist. Assoc.*, **94**, 1999, 1311-1320.

- NAKASHIMA, K., TAKAHASHI, K., AZUMI, T., ISHIDA, G.: Exteroceptive suppression of the masseter and temporalis muscles produced by electric stimulation of the mental nerve in patients with Parkinson's disease, *Acta Neuro. Scand.*, **81**, 1990, 407-410.
- O'HARA, T., DUNNE, A., KINAHAN, A., CUNNINGHAM, S., STRAK, P., DEVANE, J.. Review of methodologies for the comparison of dissolution profile data. IN: *In vitro in Vivo Correlations*, 1997, 167-171. Plenum Press, NY. Editors: Young et al.
- ROY, T., Calibrated nonparametric confidence sets, *Journ. Mathem. Chemistry* **21**, 1996, 103-109.
- SCHALL, R.: Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar, *Biometrics*, **51**, 1995, 615-626.
- SCHALL, R., GROENEWALD, P.C.N., POTGIETER, L.A., LUUS, H.G.: Multivariate assessment of bioequivalence, *Technical Report, University of the Orange Free State*, 1996.
- SCHOENEN, J.: Exteroceptive suppression of temporalis muscle activity: methodological and physiological aspects, *Cephalia*, **13**, 1993, 3-10.
- SERFLING, R., *Approximation Theorems of Mathematical Statistics*, Wiley, 1980.
- SHAH, V. P. AND TSONG, Y. AND SATHE, P. AND LIU, J.-P., In vitro dissolution profile comparison - statistics and analysis of the similarity factor,  $f_2$ , *Pharmres*, **15**, 1998, 889-896.
- SIERRA-CAVAZOS, J.H., BERGER, R.L. Intersection-union tests in dissolution profile testing. *Technical Report, NCSU Institute of Statistics Mimeo ser. No. 2521*, 1999.
- STEINHOFF, B.J., FANGMEIER, B., REIMERS, C.D., PAULUS, W.: Einfluß von Epilepsien und antiseptischer Medikation auf die exterozeptive Suppression der Temporalismuskelaktivität, *Zeitschrift für Elektroenzephalographie, Elektromyographie und verwandte Gebiete*, **2**, 1996, 65-69.
- WANG, W., DASGUPTA, A., HWANG, J.T.G.: Statistical tests for multivariate bioequivalence, *Biometrika* **86**, 1999, 395-402.