University of Alberta

Robustness of Testing Procedures for Behrens-Fisher Problem with Normal Mixture Data

by

Abdulkadir Ahmed Hussein ©

A thesis submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of the requirements for the Degree of Master of Science

in

Statistics

Department of Mathematical Sciences

Edmonton, Alberta

Fall 1999

0-612-47044-X

University of Alberta

Library Release Form

Name of Author:	Abdulkadir Ahmed Hussein
Title of Thesis:	Robustness of Testing Procedures for Behrens-Fisher Problem with Normal Mixture Data
Degree:	Master of Science
Year this Degree Granted:	1999

A. Hussein

3A 9014 112 St. Edmonton, Alberta
Canada, T6G 2C5

Date: July 20th, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and reccommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Robustness of Testing Procedures for Behrens-Fisher Problem with Normal Mixture Data* submitted by *Abdulkadir Ahmed Hussein* in partial fulfillment of the requirements for the degree of M.Sc. in Statistics.

K.C Carrière

Dr. K.C. Carrière (Supervisor)

Dr. Y. Wu

Stephen Newman

Dr. S. Newman

Date: July 8th 1999

To my Parents

Haji Ahmed Hussein & Dahira Abdi Ibrahim

Abstract

The problem of comparing the mean values of two populations has a great importance in applied statistics. Such comparison consists of either testing the hypothesis that the difference between the two means is equal to a given value or constructing a confidence interval for their difference. If in addition, the two populations are heteroscedastic (i.e. have different variances), then the problem becomes what is known as "*Behrens-Fisher problem*" named after its first two investigators. Furthermore, non-normality in the data will add further complications in testing the hypothesis.

The first aim of this thesis is to give a comprehensive review of literature of the currently available strategies for both univariate and multivariate Behrens-Fisher problem and summarize the related Monte Carlo studies as reported in the literature. The second aim is to carry out additional Monte Carlo comparisons of some of these procedures in order to shed light on their robustness when applied to data from normal mixture distributions.

Acknowledgements

I am very grateful to my Supervisor, Dr. K. C. Carrière, for leading me to the research field of this thesis, and for her many helpful and crucial comments and suggestions throughout the preparation of this thesis. Without her guidance, this thesis would have been impossible.

I express my gratitude to my thesis committee members: Dr. Y. Wu and Dr. S. Newman for their time spent in reading my thesis. In particular, I thank Dr. Y. Wu for his invaluable suggestions.

Special thanks go to all my friends, in particular, Alex, Bhushan, Eshetu and Rogemar for their unpayable help.

Thanks are also due to the Department of Mathematical Sciences for providing me the technical support needed for this thesis, especially to Marion Benedict for her kindness and patience.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter introduces the problem under consideration, known as the Behrens-Fisher problem. We state the hypothesis to be tested and discuss testing procedures under three situations that depend on the information about the population variance or covariance matrices. The notion of univariate and multivariate Behrens-Fisher problem is clearly identified. Finally, we give motivation for this thesis.

1.1 Univariate Populations

Suppose we have two independent random samples $(x_{11},...,x_{1n_1})$ and $(x_{21},...,x_{2n_2})$ or their observed values from two normal populations, $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$ respectively.

We denote the sample mean and sample variance of the i^{th} population by $\bar{x}_i$, s_i^2 respectively, for $i = 1,2$. We also define, $d = \bar{x}_1 - \bar{x}_2$, $f_1 = n_1 - 1$, $f_2 = n_2 - 1$ and $f = f_1 + f_2$.

The following properties of the sample mean and sample variances are well established (Bickel and Doksum 1977) :

(i) the pair $(\bar{x}_1, s_1^2)$ is independent of $(\bar{x}_2, s_2^2)$,

(ii) $\bar{x}_i$ is independent of s_i^2 for $i = 1,2$ and

(iii) $\bar{x}_i \sim N(\mu_i, \sigma_i^2/n_i)$ and $f_i s_i^2/\sigma_i^2 \sim \chi^2_{f_i}$ for $i = 1,2$.

Under this set-up, a test for $H_0 : \mu_1 - \mu_2 = \delta$ against any of the alternatives

(a) $\mu_1 - \mu_2 \neq \delta$,

(b) $\mu_1 - \mu_2 < \delta$ or

(c) $\mu_1 - \mu_2 > \delta$

will depend on the information about σ_1^2 and σ_2^2. Accordingly, we have three cases that will be considered in the next subsections. In what follows, without loss of generality, we will focus our attention on the alternative (a) $\mu_1 - \mu_2 \neq \delta$.

1.1.1 Known Variances

If σ_1^2 and σ_2^2 are known and we denote the variance of

$$d = \bar{x}_1 - \bar{x}_2$$

by

$$\sigma_d^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2, \tag{1.1}$$

then $(d-\delta)/\sigma_d \sim N(0,1)$ under H_0. In this case, a natural way of testing the null hypothesis is to use the statistic $(d-\delta)/\sigma_d$ and reject H_0 at level α if

$$|z| = \frac{|d-\delta|}{\sigma_d} > Z_{\alpha/2}, \tag{1.2}$$

where $Z_{\alpha/2}$ is the upper $100\alpha/2$ percentile of the standard normal distribution.

The square of $(d-\delta)/\sigma_d$ is distributed as a chi-square with one degree of freedom and it provides an equivalent test for rejecting H_0 at level α if

$$z^2 = \left(\frac{d-\delta}{\sigma_d}\right)^2 > \chi^2_{1,\alpha}, \tag{1.3}$$

where $\chi^2_{1,\alpha}$ is the 100α percentile of the chi-square on one degree of freedom.

1.1.2 Homoscedasticity

The homoscedasticity arises when the variances, σ_1^2 and σ_2^2, are unknown but have a common value σ^2. In this case, $d \sim N(\delta, \sigma^2(n_1+n_2)/n_1n_2)$ under the null hypothesis. By analogy to (1.2), (1.3), we may use the statistic

$$t = \frac{d-\delta}{\hat{\sigma}\sqrt{\dfrac{n_1+n_2}{n_1n_2}}} \tag{1.4}$$

where $\hat{\sigma}$ is the square root of an estimator $\hat{\sigma}^2$ for σ^2.

An unbiased estimator of σ^2 which uses the information contained in both s_1^2 and s_2^2 is the " *pooled sample variance* ",

$$s_p^2 = \frac{1}{f}(f_1s_1^2 + f_2s_2^2).$$

Replacing $\hat{\sigma}$ by s_p, (1.4) becomes

$$t = \frac{d - \delta}{s_p \sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}} \tag{1.5}$$

or equivalently,

$$t = \frac{d - \delta \Big/ \sqrt{\sigma^2 \left(\dfrac{n_1 + n_2}{n_1 n_2}\right)}}{\sqrt{s_p^2/\sigma^2}}. \tag{1.6}$$

The numerator of (1.6) is a standardized normal variate independent of the denominator which is the square root of $s_p^2/\sigma^2 \sim \chi_f^2/f$. This implies that the statistic in (1.5) has Student's t-distribution with f degrees of freedom (Bickel and Doksum 1977).

Now, using (1.5), H_0 is rejected at level α if

$$|t| = \frac{|d - \delta|}{s_p \sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}} > t_{f,\alpha/2}. \tag{1.7}$$

where $t_{f,\alpha/2}$ is the upper $100\alpha/2$ percentile of Student's t-distribution with f degrees of freedom.

An equivalent procedure can be obtained by using the square of (1.7) which follows the F-distribution with 1 and f degrees of freedom. Thus, H_0 is rejected if

$$t^2 > F_{1,f,\alpha}. \tag{1.8}$$

The above tests and those of the previous section can be derived rigorously by the likelihood ratio technique (Bickel and Doksum 1977) .

1.1.3 Heteroscedasticity

Now we consider the situation where variances are unknown and unequal. The problem of testing $H_0 : \mu_1 - \mu_2 = \delta$ or equivalently, constructing a confidence interval for δ under variance heterogeneity (i.e. $\sigma_1^2 \neq \sigma_2^2$) is called the "*Behrens-Fisher problem*" (Bickel and Doksum 1977, Kendall and Stuart 1969). Let σ_d^2, the variance of $d = \bar{x}_1 - \bar{x}_2$, be defined by $\sigma_d^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2$ as before so that $d \sim N(\delta, \sigma_d^2)$ under H_0. Analogy to the earlier discussion suggests the use of $d - \delta/\sigma_d$, with σ_d replaced by the square root of the unbiased estimator of σ_d.

$$s_d^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}. \tag{1.9}$$

This leads to

$$t' = \frac{d-\delta}{s_d} = \frac{(d-\delta)/\sigma_d}{\sqrt{s_d^2/\sigma_d^2}} \tag{1.10}$$

for testing the null hypothesis.

Unfortunately, due to the variance heterogeneity, the quantity in the square root (s_d^2/σ_d^2) is not exactly a chi-square divided by its degrees of freedom, as was the case in (1.6) . Therefore, the statistic $t' = (d-\delta)/s_d$, known as the *Fisher-Behrens statistic*, does not have an exact t-distribution. Its exact distribution, called *Fisher-Behrens distribution*, depends on the unknown ratio $\theta = \sigma_1^2/\sigma_2^2$ and has no simple form (Kendall and Stuart 1969).

When the difference between σ_1^2 and σ_2^2 can be ignored, one may use the homoscedastic procedures of the previous section. However, violating the underlying variance homogeneity condition renders the Type I error rate of the test much higher than any nominal α, especially when n_1 and n_2 are unequal and when the sample with a smaller sample size is associated with the larger of the two variances (Keselman, Carrière and Lix 1993 and 1995). Furthermore, the test is asymptotically incorrect for large unequal sample sizes (Scheffé 1970).

Variety of solutions have been proposed for the Behrens-Fisher problem. These will be discussed in Chapter 2 in detail where we will notice that most of these solutions are based on approximating the distribution of the t' given in (1.10).

1.2 Multivariate Populations

Let ($\mathbf{x}_{11},\ldots,\mathbf{x}_{1n_1}$) and ($\mathbf{x}_{21},\ldots,\mathbf{x}_{2n_2}$) be two independent random samples or their observed values from the two p-variate normal distributions, $N(\boldsymbol{\mu}_1,\Sigma_1)$ and $N(\boldsymbol{\mu}_2,\Sigma_2)$ respectively. Let $\bar{\mathbf{x}}_i,\mathbf{S}_i$ denote, respectively, the sample mean and sample variance-covariance (v-c) matrix of the i^{th} population ($i = 1,2$). Let also $\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$, $f_1 = n_1 - 1$, $f_2 = n_2 - 1$ and $f = f_1 + f_2$ as before.

As in the univariate case, we have:

(i) $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$, $\mathbf{S}_1$ and $\mathbf{S}_2$ are independently distributed.

(ii) $\bar{\mathbf{x}}_i \sim N(\boldsymbol{\mu}_i,\Sigma_i/n_i)$ and $f_i\mathbf{S}_i \sim W_p(f_i,\Sigma_i)$ for $i = 1,2$.

For the multivariate set-up, we will consider testing

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta} \qquad vs \qquad H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \boldsymbol{\delta}, \tag{1.11}$$

similar to the univariate case.

1.2.1 Known v-c Matrices

If the v-c matrices, Σ_1 and Σ_2, are known, then $\mathbf{d} \sim N(\boldsymbol{\delta},\Sigma_d)$ under H_0, with $\Sigma_d = \Sigma_1/n_1 + \Sigma_2/n_2$. A generalized form of the univariate z^2 statistic, (1.3), is given by

$$z^2 = (\mathbf{d} - \boldsymbol{\delta})^T\boldsymbol{\Sigma}_d^{-1}(\mathbf{d} - \boldsymbol{\delta}),$$

which is distributed as a χ^2 with p degrees of freedom. Based on this statistic, H_0 is rejected at level α if

$$z^2 = (\mathbf{d} - \boldsymbol{\delta})^T\boldsymbol{\Sigma}_d^{-1}(\mathbf{d} - \boldsymbol{\delta}) > \chi^2_{p,\alpha}. \tag{1.12}$$

1.2.2 Homoscedasticity

In the homoscedastic case, the two p-variate normal distributions have common but unknown v-c matrix $\mathbf{\Sigma}$. Again, a generalization of the univariate t^2 statistic is given by

$$T^2 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} (\mathbf{d} - \boldsymbol{\delta})^T \mathbf{S}_p^{-1} (\mathbf{d} - \boldsymbol{\delta}) \tag{1.13}$$

where

$$\mathbf{S}_p = \frac{f_1 \mathbf{S}_1 + f_2 \mathbf{S}_2}{f} \tag{1.14}$$

is the *"pooled sample covariance estimator"*.

Since $f_1\mathbf{S}_1 \sim W_p(f_1, \Sigma_1)$ independently of $f_2\mathbf{S}_2 \sim W_p(f_2, \Sigma_2)$, we have (Johnson and Wichern 1998)

$$f\mathbf{S}_p = f_1\mathbf{S}_1 + f_2\mathbf{S}_2 \sim W_p(f, \Sigma).$$

Also, under H_0,

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\mathbf{d} - \boldsymbol{\delta}) \sim N(\mathbf{0}, \Sigma).$$

Hence, (1.13) follows the Hotelling's T^2 distribution with a dimension p and f degrees of freedom which is equivalent to $\{fp/(f-p+1)\}F_{p,f-p+1}$. Then, we reject H_0 at level α if

$$T^2 > \frac{fp}{f-p+1} F_{p,f-p+1,\alpha}. \tag{1.15}$$

1.2.3 Heteroscedasticity

The heteroscedasticity in multivariate data arises when the v-c matrices of the data from the two populations are unknown and unequal. The problem of testing $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}$ or constructing a confidence interval for the vector $\boldsymbol{\delta}$ under the v-c heterogeneity becomes the *"Multivariate Behrens-Fisher problem."*

In this case, we have $\mathbf{d} \sim \mathcal{N}(\boldsymbol{\delta}, \Sigma_d)$ under H_0 with

$$\Sigma_d = \frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}.$$

If we estimate Σ_d by its uniformly minimum variance unbiased estimator,

$$\mathbf{S}_d = \frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}, \tag{1.16}$$

then by analogy to (1.13), we may consider a test statistic of the form,

$$T'^2 = (\mathbf{d} - \boldsymbol{\delta})^T \mathbf{S}_d^{-1} (\mathbf{d} - \boldsymbol{\delta}) \tag{1.17}$$

for testing H_0. However, the distribution of this statistic is not the Hotelling's T^2-distribution. Its distribution depends on the unknown parameters, Σ_1 and Σ_2.

In situations where Σ_1 and Σ_2 are not very different, we can still use the Hotelling's T^2 procedure for the homoscedastic case. Many investigators have shown via limited simulation studies that mild departures from homoscedasticity do not inflate the Type I error (Keselman *et al.* 1993, Keselman *et al.* 1995, Algina *et al.* 1995). However, one should exercise caution as such a procedure may suffer from inflated Type I error rate similar to the univariate case.

1.3 Motivation and Summary

In this chapter, we have defined the problem that will be investigated in this thesis. In the subsequent chapters we will attempt to identify the best strategy, when the existing procedures are applied to the data arising from mixture distributions.

In reality, data do not always satisfy all classical statistical assumptions such as normality, homogeneity of variances, linearity, etc. This thesis will devote its attention to the problem of heterogeneity and non-normality.

Many investigators have studied the Behrens-Fisher problem for these situations (Keselman *et al.* 1993, Keselman *et al.* 1995). However, testing procedures with data from mixture distributions are not yet known. Our aim is to examine whether the currently available tests can be used in such situations. To accomplish our aim, we first review and compare various available solutions to the Behrens-Fisher problem in Chapter 2. In Chapter 3, we formally define the type of mixture distributions we are concerned with and carry out simulation study to investigate robustness of various existing testing procedures when sampling from mixture distributions. Chapter 4 will give examples, concluding remarks and directions for further research.

Chapter 2

Review of the Literature

This chapter will review various strategies that have been developed since 1935 to address the Behrens-Fisher problem. We review univariate and multivariate solutions, followed by comparative evaluations based on what is reported in the literature. We will then identify several competitive methods to be investigated thoroughly for empirical robustness with mixture data in the next chapter.

2.1 Solutions for the Univariate Behrens-Fisher Problem

We will review an important class of solutions with rejection region of the form,

$$|t| = \frac{|d-\delta|}{s} > v_{\alpha/2} \tag{2.1}$$

for testing $H_0 : \mu_1 - \mu_2 = \delta$ *vs* $H_1 : \mu_1 - \mu_2 \neq \delta$, where $d = \bar{x}_1 - \bar{x}_2$. $v_{\alpha/2}$ is a critical value (either constant or a function of the data) and s is a function of the data. From here on, we shall use the term "d-solution" to refer to the members of this class. The notation s_d^2 will be as defined in (1.9). In the following sections we will review the proposed forms for s and $v_{\alpha/2}$ in (2.1).

2.1.1 Fisher-Behrens (F-B)

Fisher (1935) constructed a fiducial interval for δ using the equation

$$\frac{d-\delta}{\sqrt{s_d}} = \frac{\bar{x}_1 - \mu_1}{s_1/\sqrt{n_1}} \sin\theta + \frac{\bar{x}_2 - \mu_2}{s_2/\sqrt{n_2}} \cos\theta$$

provided by Behrens, where θ is such that $\tan\theta = (s_1/\sqrt{n_1})/(s_2/\sqrt{n_2})$. The testing procedure obtained from this fiducial interval is a d-solution with $v_{\alpha/2}$ obtained from tables that were first calculated by Sukhatme (1938). The tables provide critical values only for $\alpha = 0.05$ and for selected values of f_1, f_2 and θ. A more detailed discussion of the fiducial argument and in particular this procedure can be found in Kendall and Stuart (1969). However, due to the limited tabulation of this Fisher-Behrens statistic, many investigators turned to develop approximate solutions.

2.1.2 Welch's approximate degrees of freedom solution (W)

Welch (1936) approximated the distribution of the random variable s_d^2/σ_d^2 by $e\chi^2_{f_W}/f_W$ by equating their first two moments to solve for e and f_W and found that,

$$e = 1, \quad f_W^{-1} = c^2 f_1^{-1} + (1-c)^2 f_2^{-1}$$

with

$$c = \frac{s_1^2/n_1}{s_d^2}. \tag{2.2}$$

Hence, the Fisher-Behrens statistic, t', will have an approximate t-distribution with f_W degrees of freedom and the Welch's approximate degrees of freedom procedure (APDFP) rejects H_0 at level α if

$$|t'| = \frac{|d-\delta|}{s_d} > v_{\alpha/2} = t_{f_W,\alpha/2}. \tag{2.3}$$

This d-solution is generally recommended in most introductory statistical text books. See for example, Bickel and Doksum (1977).

2.1.3 Scheffé (S)

Scheffé (1943) investigated a ratio of the form

$$\frac{L-\delta}{\sqrt{Q/k}} \equiv \frac{(L-\delta)/\sqrt{V}}{\sqrt{(Q/V)/k}} \tag{2.4}$$

such that, for all values of σ_1^2 and σ_2^2,

i) L is a linear function of the data and Q is a quadratic function of the data.

ii) $L \sim N(\delta, V)$ independently of $Q/V \sim \chi_k^2$.

Clearly, any such ratio has an exact t-distribution with k degrees of freedom, and can be used to construct an exact confidence interval for δ. On the other hand, for

$n_1 \leq n_2$. Scheffé proved a *per se* important result which states that no ratio of the form (2.4) can have an exact t-distribution with more than $k = n_1 - 1$ degrees of freedom. Therefore, it is sufficient to construct $d_1, \ldots, d_{n_1} \overset{iid}{\sim} N(\delta, V)$ linear functions of the data, and define,

$$L = \sum_{i=1}^{n_1} \frac{d_i}{n_1}, \qquad Q = \sum_{i=1}^{n_1} (d_i - L)^2. \tag{2.5}$$

Scheffé's choice of the d_i was,

$$d_i = x_{1i} - \sum_{j=1}^{n_2} c_{ij} x_{2j},$$

with c_{ij} to be determined. After imposing the conditions that all d_i have the same mean and the same variance, (δ, V), and the expected length of the confidence interval associated with the statistic (2.4) is minimum, Scheffé determined the following particular set of c_{ij}:

$$c_{ij} = \begin{cases} \delta_{ij}\sqrt{\frac{n_1}{n_2}} - \sqrt{\frac{1}{n_1 n_2}} + \frac{1}{n_2}; & j \leq n_1 \\ \frac{1}{n_2}; & j > n_1 \end{cases}$$

where δ_{ij} is the Kronecker delta. Hence,

$$d_i = x_{1i} - \sqrt{\frac{n_1}{n_2}} x_{2i} + \sqrt{\frac{1}{n_1 n_2}} \sum_{j=1}^{n_1} x_{2j} + \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j}.$$

These d_i placed in (2.5) will give,

$$L = \bar{x}_1 - \bar{x}_2, \quad Q = \frac{1}{n_1} \sum_{i=1}^{n_1} (u_i - \bar{u})^2 \tag{2.6}$$

where,

$$\begin{aligned} u_i &= x_{1i} - \sqrt{\frac{n_1}{n_2}} x_{2i}, \quad i = 1, \ldots, n_1 \\ \bar{u} &= \sum_{i=1}^{n_1} u_i / n_1. \end{aligned} \tag{2.7}$$

Now, using (2.4) with the above L and Q. Scheffé's method rejects H_0 at level α if

$$\left|\frac{d-\delta}{\sqrt{Q/f_1}}\right| > t_{f_1,\alpha/2}. \tag{2.8}$$

where $f_1 = n_1 - 1$. The procedure is, therefore, a d-solution with $s = \sqrt{Q/f_1}$ and $v_{\alpha/2} = t_{f_1,\alpha/2}$.

One practical difficulty that led Scheffé himself to renounce this method is that in defining Q in equations (2.6) and (2.7), only n_1 randomly chosen elements of the large sample are involved. That is, the method may give different results to different analysts even with the same set of data.

2.1.4 Welch-Aspin (W-A)

This solution (also known as, Welch's series solution) was first proposed by Welch (1938, 1947) and its critical values tabulated later by Aspin (1948).

Welch's basic idea for the two-sample problem consisted in constructing a confidence interval for δ by finding a series solution, $h(s_1^2, s_2^2, \alpha)$, for the integral equation,

$$P\left[|t'| < h(s_1^2, s_2^2, \alpha)\right] = 1 - \alpha \tag{2.9}$$

where t' is the Fisher-Behrens statistic defined in (1.10).

Using Taylor's series expansion, he found

$$\begin{aligned} h(s_1, s_2, \alpha) = \xi \; &+ \; \frac{\xi(1+\xi^2)}{4}\left(\sum_1^2 c_i^2/f_i\right) \\ &- \; \frac{\xi(1+\xi^2)}{2}\left(\sum_1^2 c_i^2/f_i^2\right) \\ &+ \; \frac{\xi(3+5\xi^2+\xi^4)}{3}\left(\sum_1^2 c_i^3/f_i^2\right) + \ldots \end{aligned} \tag{2.10}$$

where $\xi = Z_{\alpha/2}$ is the upper $100\alpha/2$ percentile of the standard normal distribution, and $c_i = (s_i^2/n_i)/s_d^2$ for $i = 1, 2$.

The W-A test procedure is then obtained by inverting the inequality in (2.9). It is a d-solution and rejects the null hypothesis at level α, if

$$|t'| = \frac{|d-\delta|}{s_d} > v_{\alpha/2} = h(s_1, s_2, \alpha) \tag{2.11}$$

where $h(s_1, s_2, \alpha)$ is the same in (2.10).

2.1.5 Banerjee (BS)

Banerjee (1961) constructed a confidence interval for δ with confidence coefficient not less than α by proving that,

$$P\left[|d-\delta| \le \sqrt{\frac{s_1^2}{n_1}t_{f_1,\alpha/2}^2 + \frac{s_2^2}{n_2}t_{f_2,\alpha/2}^2}\right] \ge 1-\alpha. \tag{2.12}$$

Dividing by s_d and inverting inequalities in (2.12), we have

$$P\left[|t'| \ge \sqrt{\frac{s_1^2/n_1}{s_d^2}t_{f_1,\alpha/2}^2 + \frac{s_2^2/n_2}{s_d^2}t_{f_2,\alpha/2}^2}\right] \le \alpha.$$

Using this latter inequality, Banerjee's test rejects H_0 at level α if

$$|t'| = \frac{|d-\delta|}{s_d} > \sqrt{ct_{f_1,\alpha/2}^2 + (1-c)t_{f_2,\alpha/2}^2} \tag{2.13}$$

where c is defined in (2.2) for method (W) and $t_{f_i,\alpha/2}$ is the upper $100\alpha/2$ percentile of the Student's t-distribution with f_i degrees of freedom.

2.1.6 Denish Bhoj (DB1 and DB2)

It can be easily seen that the Fisher-Behrens statistic, $t' = (d-\delta)/s_d$, satisfies,

$$t' = \frac{d-\delta}{s_d} = \frac{(d-\delta)/\sigma_d}{\sqrt{\frac{1}{\sigma_d^2}\frac{s_1^2}{n_1} + \frac{1}{\sigma_d^2}\frac{s_1^2}{n_1}}} \sim \frac{N(0,1)}{\sqrt{b_1\chi_{f_1}^2 + b_2\chi_{f_2}^2}} \tag{2.14}$$

with

$$b_i = \frac{\sigma_i^2/n_i}{\sigma_d^2}, \qquad i = 1,2 \tag{2.15}$$

where,

$$\sigma_d^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Bhoj (1993) constructed a random variable whose first three moments match those of $b_1\chi_{f_1}^2 + b_2\chi_{f_2}^2$ by generalizing a result of Gabler and Wolff (1987) in which they approximated the distribution of a linear combination of chi-square random variables each with one degree of freedom. As a consequence, he obtained an approximate density of the form

$$b_1 t_{f_1/b_1} + b_2 t_{f_2/b_2} \tag{2.16}$$

for the above t' statistic, where the parameters b_i are to be replaced by their unbiased estimators, $\hat{b}_1 = c = \frac{s_1^2/n_1}{s_d^2}$ and $\hat{b}_2 = 1 - c = \frac{s_2^2/n_2}{s_d^2}$.

Thus, the first test of Bhoj, denoted from here on by DB1, rejects H_0 at level α if

$$|t'| = \frac{|d-\delta|}{s_d} > c t_{f_1/c,\alpha/2} + (1-c) t_{f_2/(1-c),\alpha/2}. \tag{2.17}$$

The quantity (2.16), with b_is defined by (2.15), is itself a linear combination of t-distributions and can be approximated by $e\ t_{f_{DB2}}$ (a weighted Student's t-variate with degrees of freedom f_{DB2}) by matching their first two moments. In this way, Bhoj determined,

$$f_{DB2} = \frac{4B - 2A^2}{B - A^2}, \qquad \epsilon = \left(\frac{AB}{2B - A^2}\right)^2,$$

with

$$B = \frac{f_1^2/c}{(\frac{f_1}{c} - 2)(\frac{f_2}{c} - 4)} + \frac{f_2^2/(1-c)}{(\frac{f_1}{1-c} - 2)(\frac{f_2}{1-c} - 4)},$$

and

$$A = \frac{f_1}{(\frac{f_1}{c} - 2)} + \frac{f_2}{(\frac{f_2}{1-c} - 2)}.$$

Hence, Bhoj's second method (DB2) rejects H_0 at level α if

$$|t'| = \frac{|d - \delta|}{s_d} > e t_{f_{DB2},\alpha/2} \quad (2.18)$$

with the above defined e and f_{DB2}.

2.2 Solutions for the Multivariate Behrens-Fisher Problem

A class of solutions similar to that of Section 2.1 has also been proposed for the multivariate Behrens-Fisher problem. Most of the solutions of this class are extensions of their univariate counterparts. The rejection region of these solutions is of the form,

$$T^* = (\mathbf{d} - \delta)^T \mathbf{S}^{-1} (\mathbf{d} - \delta) > v_\alpha \quad (2.19)$$

where $\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$, $\mathbf{S}$ is a $p \times p$ symmetric and positive definite matrix obtained from the data and v_α is a critical value which is either a constant or a function of the data.

2.2.1 Bennett (B)

Bennett (1951) was probably the first to investigate the multivariate Behrens-Fisher problem. He extended Scheffé's univariate method S for $n_1 \leq n_2$ by defining the multivariate versions of (2.6) and (2.7) so that

$$
\begin{aligned}
\mathbf{u}_i &= \bar{\mathbf{x}}_{1i} - \sqrt{\left(\frac{n_1}{n_2}\right)}\bar{\mathbf{x}}_{2i} \qquad i = 1, \ldots, n_1 \\
\bar{\mathbf{u}} &= \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbf{u}_i \\
\mathbf{S} &= \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T. \qquad (2.20)
\end{aligned}
$$

Using these definitions, Benett's test rejects H_0 at level α if

$$
T^* = (\mathbf{d} - \delta)^T \mathbf{S}^{-1}(\mathbf{d} - \delta) > v_\alpha = T^2_{p,n_1-1,\alpha} \qquad (2.21)
$$

with $T^2_{p,n_1-1,\alpha}$ being the 100α percentile of the Hotelling's T^2.

2.2.2 James' Series Solution (JS)

This method is an extension of the univariate W-A method. Similar to (2.10), James (1954) derived a series function $h(\mathbf{S}_1, \mathbf{S}_2, \alpha)$ whose first order terms have the form.

$$
h(\mathbf{S}_1, \mathbf{S}_2, \alpha) = \chi^2_{p,\alpha}\left[1 + \frac{1}{2}\left(\frac{k_1}{p} + \frac{k_2\chi^2_{p,\alpha}}{p(p+1)}\right)\right] \qquad (2.22)
$$

where

$$
\begin{aligned}
k_1 &= \sum_{i=1}^{2} f_i^{-1}\left(\mathrm{tr}\mathbf{S}^{-1}\frac{\mathbf{S}_i}{n_i}\right)^2 \\
k_2 &= \sum_{i=1}^{2} f_i^{-1}\left[\left(\mathrm{tr}\mathbf{S}^{-1}\frac{\mathbf{S}_i}{n_i}\right)^2 + 2\mathrm{tr}\left(\mathbf{S}^{-1}\frac{\mathbf{S}_i}{n_i}\mathbf{S}^{-1}\frac{\mathbf{S}_i}{n_i}\right)\right]. \qquad (2.23)
\end{aligned}
$$

The James' method then, rejects H_0 at level α if (2.19) is satisfied with $v_\alpha = h(\mathbf{S}_1, \mathbf{S}_2, \alpha)$ and $\mathbf{S} = \mathbf{S}_d$ defined in (1.16).

To improve its performance, he extended this solution to second order series solution, which has not been in much use in practice due to its relative complexity.

2.2.3 Yao's Approximate Degrees of Freedom Solution (Y)

Yao (1965) suggested that an extension of Welch's approximate degrees of freedom solution (W) to the multivariate case will involve $T^2_{p.f_Y}$ (Hotelling's T^2-distribution with dimension p and f_Y degrees of freedom) in place of the Student's t-distribution. The approximate degrees of freedom, f_Y, which were determined by Yao by formally extending Welch's f_W, are given by

$$f_Y^{-1} = \sum_{i=1}^{2} f_i^{-1} \left(\frac{\mathbf{d}^T\mathbf{S}^{-1}\left(\frac{\mathbf{S}_i}{n_i}\right)\mathbf{S}^{-1}\mathbf{d}}{\mathbf{d}^T\mathbf{S}^{-1}\mathbf{d}} \right)^2 . \tag{2.24}$$

The method of Yao, therefore, rejects H_0 at level α if (2.19) is satisfied with $v_\alpha = T^2_{p.f_Y,\alpha}$ and $\mathbf{S} = \mathbf{S}_d$.

2.2.4 Johansen (JH)

Johansen's procedure (Johansen 1980) is a refinement of JS and rejects H_0 at level α if (2.19) is satisfied with

$$\mathbf{S} = \frac{\mathbf{S}_1/n_1 + \mathbf{S}_2/n_2}{[p - 2A - 6A]/[p(p-1)+2)]} \tag{2.25}$$

and

$$v_\alpha = F_{p.f_{JH},\alpha} \tag{2.26}$$

where $A = \sum_{i=1}^{2} \frac{1}{2(n_i - 1)} (\mathrm{tr}(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i)^2 + (tr(\mathbf{I} - \mathbf{V}^{-1}\mathbf{V}_i))^2)$ with $\mathbf{V}_i = (\mathbf{S}_i/n_i)^{-1}$, $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$ and $f_{JH} = p(p+2)/3A$.

Due to its flexibility in specifying the test statistics and relative simplicity, it has been adopted for use in repeated measures and factorial designs (For example, Keselman *et al.* 1993 and 1995).

2.2.5 Nel and Van der Merwe (NV)

Nel and Van der Merwe (1986) presented a method similar to that of Yao in the sense that it extends Welch's univariate approximate degrees of freedom method. They approximated the distribution of the $\mathbf{S}_d$ statistic, which is a weighted sum of Wishart variates, by a Wishart distribution. For this procedure, $\mathbf{S} = \mathbf{S}_d$ and

$$v_\alpha = T^2_{p,f_{NV},\alpha} \tag{2.27}$$

where the approximate degrees of freedom, f_{NV}, is given by

$$f_{NV} = \frac{\mathrm{tr}(\mathbf{S}^2) + [\mathrm{tr}(\mathbf{S})]^2}{\sum_{i=1}^2 \frac{1}{n_i - 1}\left[\mathrm{tr}(\mathbf{S}_i/n_i)^2 + [\mathrm{tr}(\mathbf{S}_i/n_i)]^2\right]}. \tag{2.28}$$

2.2.6 Kim (K)

Kim's procedure (Kim 1992) has $\mathbf{S}$ defined by

$$\mathbf{S} = \mathbf{d}^T\mathbf{Q}(\Lambda^{1/2} + r\mathbf{I})^{-2}\mathbf{Q}^T\mathbf{d}. \tag{2.29}$$

The matrix Λ is a diagonal matrix containing the generalized eigenvalues λ_k of $\frac{\mathbf{S}_1}{n_1}\mathbf{x} = \lambda\frac{\mathbf{S}_2}{n_2}\mathbf{x}$, for $k = 1, \ldots, p$, and $\mathbf{Q}$ is a non-singular matrix whose columns are the corresponding eigenvectors $\mathbf{q}_k$ and $r = (\Pi_{k=1}^p \lambda_k)^{1/2p}$. The critical value of the test is given by

$$v_\alpha = \frac{c f f_Y}{f_Y - p + 1} F_{f, f_Y - p + 1, \alpha}, \tag{2.30}$$

where $c = \sum l_k^2 / \sum l_k$, $f = (\sum l_k)^2 / \sum l_k^2$, $l_k = (\lambda_k + 1)/(\lambda_k^{1/2} + r)^2$ and f_Y being the approximate degrees of freedom of Yao's procedure as given in (2.24).

2.2.7 Jordan and Krishnamoorthy (JK)

Jordan and Krishnamoorthy (1995), while constructing an exact confidence region for the common mean vector of several multivariate populations, provided a conservative test for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ in the Behrens-Fisher problem when k groups are involved.

Considering the case $k = 2$ and following the remark of the authors, the confidence region for the common mean is nonempty provided that,

$$\sum_{i=1}^{k} \left(\sum_{j \neq i}^{k} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)^T \mathbf{W}_j^{-1} \right) \mathbf{V}^{-1} \mathbf{W}_i^{-1} \mathbf{V}^{-1} \sum_{i=1}^{k} \left(\sum_{j \neq i}^{k} \mathbf{W}_j^{-1} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) \right) \leq a \qquad (2.31)$$

holds with probability $1 - \alpha$, where the c_i are positive weights satisfying $\sum c_i = 1$, $\mathbf{W}_i^{-1} = c_i n_i \mathbf{S}_i^{-1}$, $\mathbf{V} = \mathbf{W}_1^{-1} + \mathbf{W}_2^{-1}$ and a is such that $P(c_1 T_1^2 + c_2 T_2^2 \leq a) = 1 - \alpha$ with $T_i^2 \sim \frac{f_i p}{f_i - p + 1} F_{p, f_i - p + 1}$ being independent Hotelling's T^2-variates for $i = 1, 2$.

A test that rejects $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ if (2.31) does not hold, would be at the level of α. Further simplifying (2.31), the null hypothesis is rejected at level α if

$$T^* = \mathbf{d}^T \mathbf{S}^{-1} \mathbf{d} > v_\alpha$$

with

$$\mathbf{S} = \mathbf{W}_1^{-1} \mathbf{V}^{-1} \mathbf{W}_2^{-1} \mathbf{V}^{-1} \mathbf{W}_1^{-1} + \mathbf{W}_2^{-1} \mathbf{V}^{-1} \mathbf{W}_1^{-1} \mathbf{V}^{-1} \mathbf{W}_2^{-1} \qquad (2.32)$$

and $v_\alpha = a$. The authors suggested the use of

$$c_i = \frac{[\mathrm{Var}(T_i^2)]^{-1}}{[\mathrm{Var}(T_1^2)]^{-1} + [\mathrm{Var}(T_2^2)]^{-1}} \qquad (2.33)$$

as the positive weights, where $\mathrm{Var}(T_i^2) = 2pf_i^2(f_i - 1)/(f_i - p - 1)^2(f_i - p - 3)$. However, there could be other choices of the c_i that might improve the test.

They also provided the exact critical values, $v_\alpha = a$, corresponding to $\alpha = 0.05, 0.01$ and $p = 2, 3, 4$.

2.3 Summary Review

2.3.1 The Univariate Case

Scheffé (1970) considered some of the solutions of Section 2.1 and compared their significance levels and powers. He disqualified his own procedure because, as was previously mentioned, it may give different results to different analysts who are dealing with the same data set.

In addition to the difficulties mentioned in Section 2.1.1, Fisher-Behrens solution, F-B, is rather liberal and always has higher Type I error and lower power than W-A. Banerjee (1961) compared the critical value of his solution SB with those of W-A for different values of c (see (2.2)). He considered the situations where $n_1 = 9$ $n_2 = 9, 13, \infty$ with $\alpha = 0.05$ and $n_1 = 13$, $n_2 = 13, \infty$ with $\alpha = 0.01$. The SB was found to have larger critical points for the same values of the null statistic, which means inflated Type I error, leading to liberal rejections of the correct null hypothesis. Bhoj (1993) compared the significance levels and powers of his solutions, DB1 and DB2, with those of method W. He based his comparison on sizes and powers of the method W reported by Davenport and Webster (1975) for the cases: $n_1 = n_2 = 3$, $n_1 = n_2 = 18$ and $n_1 = 3, n_2 = 19$, all with $\alpha = 0.05$. Bhoj recommended the use of his solutions because they are simple and have better control on the empirical size than Welch's approximate degrees of freedom solution.

2.3.2 The Multivariate Case

Many Monte Carlo studies were carried out by several investigators in order to study and compare the methods developed for the multivariate Behrens-Fisher problem. In this section we summarize the results of some of these studies done recently.

Yao (1965) used uncorrelated data (i.e. Σ_1 and Σ_2 were diagonal matrices) and compared the significance level of her method with that of James, JS. She considered two nominal significance levels, $\alpha = 0.01$ and $\alpha = 0.05$. For each nominal α, she carried out the simulation for various combinations of covariance matrices and for $(p, f_1, f_2) = (2, 6, 12), (2, 12, 12), (2, 8, 18)$. From Yao's study it seems that the method Y is superior to JS in terms of controlling the Type I error.

Subrahmaniam and Subrahmaniam (1973) compared the significance levels and powers of the procedures B, JS and Y. They considered the combinations:
(i) $p = 2$ and $p = 3$ each with $(f_1, f_2) = (6, 12), (12, 12), (6, 18)$
(ii) $p = 4$ and $p = 5$ each with $(f_1, f_2) = (12, 12), (6, 18)$
(iii) $p = 10$ with $(f_1, f_2) = (15, 15), (15, 30), (20, 20)$
In each case they varied the eigenvalues of the matrices Σ_1, Σ_2 (assumed diagonal) and a noncentrality parameter ϕ and they confirmed that method B achieves the exact level α as it should. The Y method is more conservative than JS although neither of them protects the nominal α. Power of JS is the highest but Y has a power that is only slightly lower than the power of JS. Method B's power is alarmingly low. The power of all methods tends to 1 as the noncentrality parameter ϕ increases. Also, the power improves as f_1 increases, holding fixed all other parameters. On the other hand, the power decreases for a given value of ϕ as p increases.

They concluded that method Y is superior to JS and B in terms of Type I error and empirical powers.

de la Rey and Nel (1993) compared the Type I error and powers of B, JS, Y and NV using diagonal covariance matrices. They considered the situations:
(i) $p = 2$ with $(f_1, f_2) = (6, 12), (12, 12), (6, 18)$
(ii) $p = 3$ with $(f_1, f_2) = (12, 12)$
(iii) $p = 4$ and $p = 5$ each with $(f_1, f_2) = (12, 12), (8, 16)$

(iv) $p = 10$ with $(f_1, f_2) = (15, 15), (15, 30)$.

In all situations, the nominal significance level α was 0.05 and they varied the eigenvalues of Σ_1, Σ_2 and the non-centrality parameter ϕ. Their conclusion was that NV has the lowest Type I error followed by Y (note that B achieves the exact nominal α). The power of JS was always the highest and B's power was the lowest. The powers of Y and NV compete except when the heteroscedasticity (i.e. the difference between Σ_1 and Σ_2) was large and $n_1 << n_2$, in which case NV was superior to Y.

Kim (1992) reported Monte Carlo powers and significance levels of his method K and Yao's Y for the cases:

(i) $p = 2$ and $(f_1, f_2) = (5, 11), (5, 17)$

(ii) $p = 5$ and $(f_1, f_2) = (7, 15), (7, 23)$.

In each case, the eigenvalues of Σ_1 and Σ_2 (both assumed diagonal), and the non-centrality parameter ϕ were varied, whereas the nominal level was held at $\alpha = .05$. Kim concluded that the level of K varied in [.026, .062] whereas Y's level varied in [.038, .172] reaching the value .172 for $p = 5$. Moreover, if the sample with a smaller size is associated with the larger variance, then K has better power than Y, otherwise Y has better power.

Wilcox (1995) studied the effect of non-normality on the power and significance levels of six methods. The six methods were: K, JH, modified JH (the modification consisted of using $\mathbf{S} = \mathbf{S}_1/f_1 + \mathbf{S}_2/f_2$ in (2.25)) and a trimmed mean version of each one of these (i.e. modifying them for comparing trimmed means instead of means). He also mentioned that the method NV performed very poorly in non-normality situations and therefore it was excluded from the simulation study. The non-normality was introduced using the g-and-h distribution proposed in Hoaglin, Mosteller and Tukey (1985), where g and h are parameters controlling, respectively, the heaviness of the tail and the skewness of the distribution. Simulation were carried out for each combination of $p = 4$, $\alpha = .05$ and $(f_1, f_2) = (11, 17), (19, 19), (11, 23), (5, 11)$. In each situation, the non-centrality parameter ϕ, the correlation matrices (considering all possible combinations of four correlation matrices) and the parameters g and h were varied. According to Wilcox's simulation, all six methods appear to behave well in terms of Type I errors provided that sample sizes are not too small. Under non-normality, methods based on trimmed means have better power except for small unequal sample sizes, in which case the trimmed K is less satisfactory than all others. In general, the method JH with trimmed means is the best "in terms of power" when both n_1 and $n_2 \geq 20$, and the modified JH with trimmed means seems to be the best if $10 \leq n_1, n_2 \leq 20$ whereas, for both n_1 and $n_2 \leq 10$ the trimmed K is preferable.

2.3.3 Concluding Remarks

From the Monte Carlo studies reported in the literature, it seems that the methods, Y, K, JH and NV are quite competitive and perform reasonably well in keeping the empirical level close to the nominal value in many situations. But according to Wilcox (1995), under nonnormality (like the mixture models we intend to study), NV would perform very badly. Therefore, we shall consider adopting the methods of Johansen, Kim, Yao and Jordan & Krishnamoorthy to a mixture data set. We shall also include Hotelling's T^2 method for comparative purposes.

Chapter 3

Mixture Distributions and Simulation Study

This Chapter is concerned with mixture data and the performance of the currently available procedures for the Behrens-Fisher problem that were found to work well under variance heterogeneity when coupled with mixture data. We first introduce notations and related parameters for mixture populations followed by our simulation study results.

3.1 Introduction

Mixture distributions are common models in many applications ranging from fisheries, taxonomy, medical sciences and satellite imaging to robustness studies of certain statistical procedures (Averitt and Hand 1981). It arises when the underlying population is composed of several subpopulations and the sampled data is not easily classifiable.

In general, a mixture of K p-variate normals has the density,

$$f(\mathbf{x}) = \sum_{1}^{K} \pi_k \phi(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k). \quad (3.1)$$

where $\phi(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$ is the density function of a p-variate normal distribution with mean $\boldsymbol{\mu}_k$ and v-c matrix $\boldsymbol{\Sigma}_k$ and π_ks are mixing proportions satisfying $\sum_k \pi_k = 1$ and $0 < \pi_k < 1$ for all $k = 1, \ldots, K$.

When $K = 2$ for comparing two populations, (3.1) reduces to

$$f_i(\mathbf{x}) = \pi_{i1}\phi(\mathbf{x}; \boldsymbol{\mu}_{i1}, \Sigma) + \pi_{i2}\phi(\mathbf{x}; \boldsymbol{\mu}_{i2}, \Sigma), \qquad i = 1, 2 \quad (3.2)$$

for $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$.

Each population ($i = 1, 2$) is a mixture of two bivariate normal populations having the same v-c matrix Σ, but different means $\boldsymbol{\mu}_{i1}$ and $\boldsymbol{\mu}_{i2}$, and mixing proportions π_{i1} and $\pi_{i2} = 1 - \pi_{i1}$.

The exact distributions of sample statistics from this type of mixtures are functions of $\{\pi_{i1}, \Delta_i, p, n_i;\ i = 1, 2\}$, where

$$\Delta_i^2 = (\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2}), \quad (3.3)$$

is the squared Mahalanobis distance associated with the i^{th} mixture (Srivastava 1982).

In particular, the quadratic forms, $(\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i2})^T \mathbf{S}^{-1} (\bar{\mathbf{x}}_{i1} - \bar{\mathbf{x}}_{i2})$, that are involved in the methods considered here will depend on the parameters listed above. Thus, by varying the components in the set $\{\pi_{i1}, \Delta_i, p, n_i;\ 1 = 1, 2\}$, we have different mixture distributions.

It can be seen (Johnson 1987) that mean value and v-c matrices of the population with density (3.2) are given by ,

$$\begin{aligned} \boldsymbol{\mu}_i^* &= \pi_{i1}\boldsymbol{\mu}_{i1} + \pi_{i2}\boldsymbol{\mu}_{i2} \\ \Sigma_i^* &= \Sigma + \pi_{i1}\pi_{i2}(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})^T, \quad i = 1, 2. \end{aligned} \tag{3.4}$$

For equal mixing proportions, i.e. $\pi_{i1} = \pi_{i2} = .5$, we see that the mean of the mixed data $\boldsymbol{\mu}_i^*$ is

$$\boldsymbol{\mu}_i^* = \frac{\boldsymbol{\mu}_{i1} + \boldsymbol{\mu}_{i2}}{2}$$

and the v-c matrices of the mixed data Σ_i^* is

$$\boldsymbol{\Sigma}_i^* = \Sigma + \frac{(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})^T}{4}.$$

whereas for $\pi_{i1} = .9$ and $\pi_{i2} = .1$ they are

$$\boldsymbol{\mu}_i^* = .9\boldsymbol{\mu}_{i1} + .1\boldsymbol{\mu}_{i2}$$

and

$$\Sigma_i^* = \boldsymbol{\Sigma} + .09(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})(\boldsymbol{\mu}_{i1} - \boldsymbol{\mu}_{i2})^T,$$

respectively and the bivariate normal data with a higher mixing proportion dominate the eventual density and the mean vector and v-c matrices. In particular, if the data is composed of two normal variates with the same mean, $\boldsymbol{\Sigma}_i^* = \boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_i^* = \boldsymbol{\mu}_i$ regardless of the mixing proportions.

In this chapter, we carry out Monte Carlo evaluations to investigate significance levels of the methods of Kim, Yao, Johansen and Jordan & Krishnamoorthy using the density (3.2) with $p = 2$. The use of such a mixture of two multivariate normal distributions enables us to evaluate robustness under a wide range of departures from normality (e.g, bimodality, non-normal skewness and kurtosis) and heteroscedasticity.

3.2 Simulation Methods

Without loss of generality, we test the null hypothesis

$$H_0 : \boldsymbol{\mu}_1^* = \boldsymbol{\mu}_2^* = \mathbf{0}$$

and we assume that the common v-c matrix is given by

$$\Sigma = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\boldsymbol{\mu}_{i1} = \begin{pmatrix} 0 \\ \mu_{i1}^{(2)} \end{pmatrix}; \quad \boldsymbol{\mu}_{i2} = \begin{pmatrix} 0 \\ \mu_{i2}^{(2)} \end{pmatrix},$$

where $\mu_{ik}^{(2)}$ are chosen as functions of Δ_i and π_{ik}. Then, from (3.4) with the condition that $\boldsymbol{\mu}_1^* = \boldsymbol{\mu}_2^* = \mathbf{0}$, we have

$$\boldsymbol{\mu}_{i1} = \begin{pmatrix} 0 \\ -\pi_{i2}\Delta_i \end{pmatrix}; \quad \boldsymbol{\mu}_{i2} = \begin{pmatrix} 0 \\ \pi_{i1}\Delta_i \end{pmatrix} \tag{3.5}$$

and

$$\Sigma_i^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 + \pi_{i1}\pi_{i2}\Delta_i^2 \end{pmatrix}, \tag{3.6}$$

so that,

$$\boldsymbol{\mu}_i^* = \mathbf{0}. \tag{3.7}$$

In order to have an idea of the type of mixtures we are dealing with, we portray some of these mixtures in Figures 3.1-3.2 with $\pi_{i1} = .5$, $\pi_{i1} = .9$ and for Mahalanobis distances $\Delta_i = 0, 2, 3$.

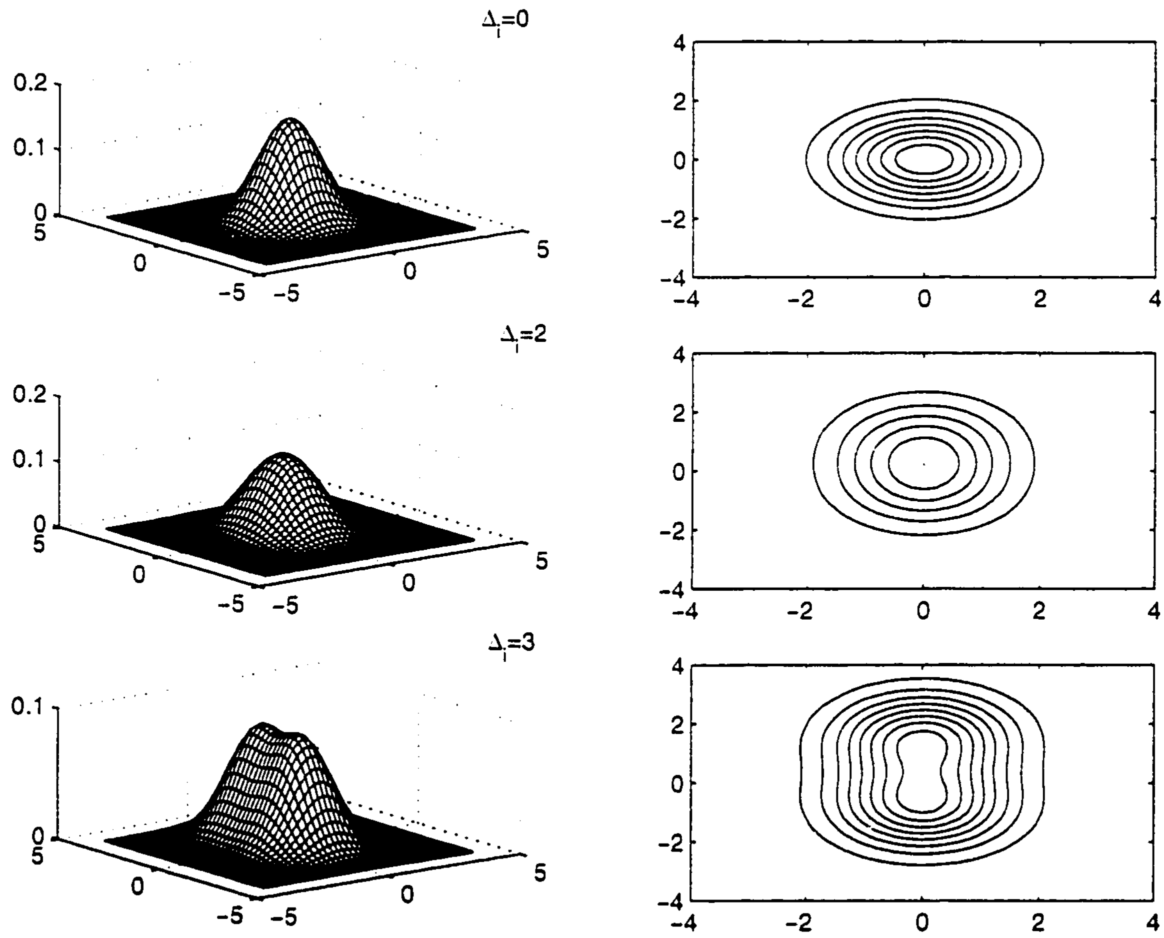

Figure 3.1: Bivariate two-component normal mixture densities and their contours with proportions $\pi_{i1} = \pi_{i2} = .5$ and Mahalanobis distances $\Delta_i = 0, 2, 3$

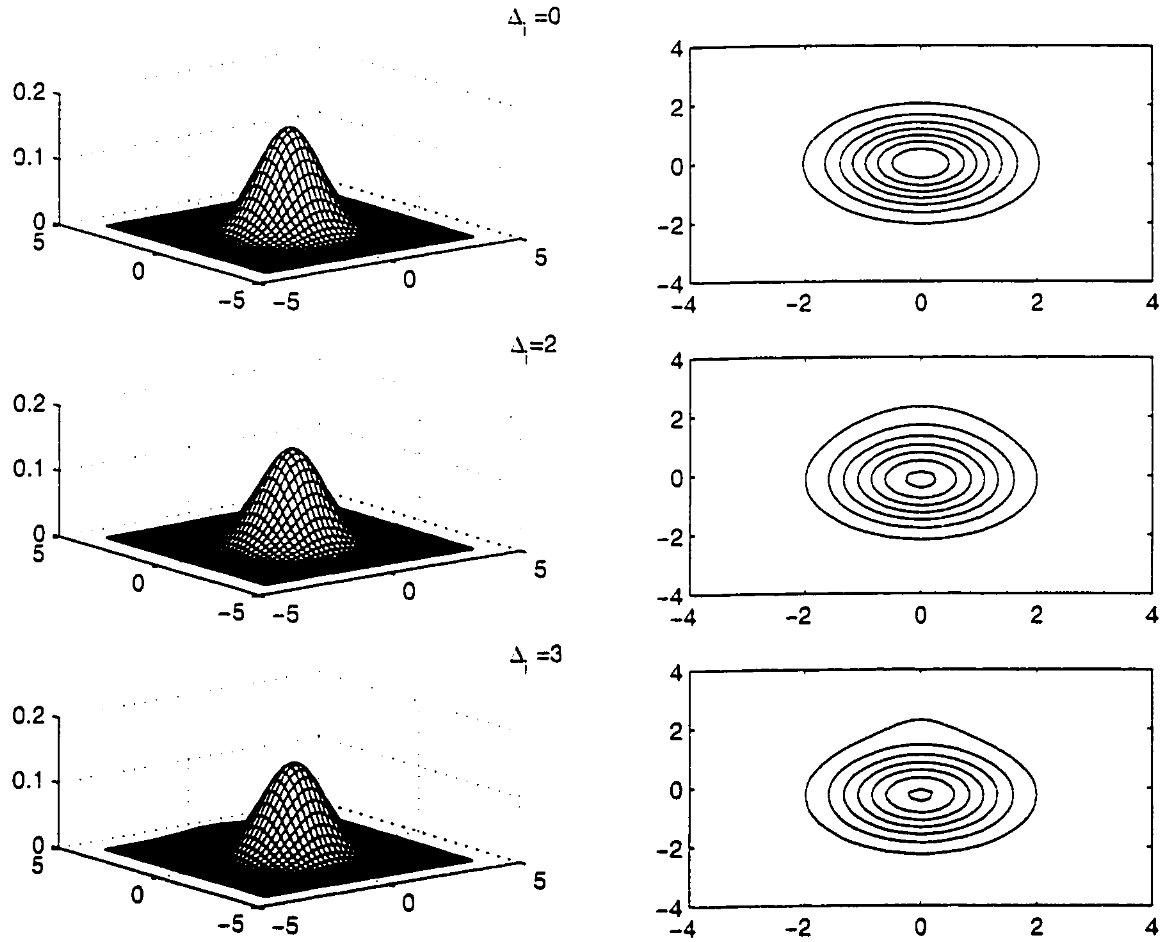

Figure 3.2: Bivariate two-component normal mixture densities and their contours with proportions $\pi_{i1} = .9$, $\pi_{i2} = .1$ and Mahalanobis distances $\Delta_i = 0, 2, 3$

For a given set of parameters $\{\pi_{i1}, \Delta_i, n_i: \ i = 1, 2\}$, we generate 10000 samples from each of the two mixture populations (3.2) according to the following scheme:

Algorithm:

Compute $\boldsymbol{\mu}_{i1}$, $\boldsymbol{\mu}_{i2}$ *from (3.5).*

For $j = 1, ..., n_i$

Generate $\mathbf{x}_{ij}$ *from* $N(\mathbf{0}, I_2)$

Generate u_j *from* $Uniform(0, 1)$

If $u_j \leq \pi_{i1}$

$\mathbf{x}_{ij} = \mathbf{x}_{ij} + \boldsymbol{\mu}_{i1}$

otherwise,

$\mathbf{x}_{ij} = \mathbf{x}_{ij} + \boldsymbol{\mu}_{i2}$.

end

Output $\mathbf{x}_{i1}, ..., \mathbf{x}_{in_i}$

These samples are then analyzed using procedures of Hotelling, Kim, Yao, Johansen and Jordan & Krishnamoorthy with nominal significance level $\alpha = 0.05$. The percentage of times in which the null hypothesis is rejected is reported as the Monte Carlo significance level. For the empirical power, we add $(1\ 1)^T$ to one of the two samples generated. That is, we consider the alternative

$$H_1 : \boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and again we account the number of times the null hypothesis is rejected as the empirical power of the test. Results for

$$H_1 : \boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

were also obtained but not reported in this thesis. However, these results confirm the power increase which is usually expected.

Due to extremely conservative preliminary results of Jordan & Krishnamoorthy's procedure, we decided not to pursue it any further in the following analysis. Please refer to the Appendix A for detailed results of this procedure.

3.3 Results for the Case of $\pi_{11} = \pi_{21} = 0.5$

When mixing proportions are equal for both populations, the resulting mixture populations are symmetric. We recall that the null hypothesis of interest is $H_0 : \boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = \mathbf{0}$ with nominal significance level set at $\alpha = .05$. We also assume, without loss of generality, that $\Delta_2 > \Delta_1$. The Tables A1-A4 in Appendix A contain detailed simulation results for $(n_1, n_2) = (7,7), (7,14), (14,7), (21,21)$ for this case.

Tables 3.2-3.5 display the performance of these selected procedures in keeping the nominal level. If the empirical sizes are within [.045, .054], the corresponding procedures shall be deemed adequate and satisfactory for $\alpha = .05$.

We shall distinguish the simulated cases as being

(a) Homogeneous populations or

(b) Heterogeneous populations.

These distinctions are based on the variance heterogeneity factor (VHF) of the two populations. That is, if $\Delta_2 - \Delta_1 > 0$ they are called heterogeneous populations because the resulting v-c matrix for $i = 1$ differs from that for $i = 2$ as shown in (3.6). Even with the equal mixing proportions, if there is a sizable difference in their Mahalanobis distances, the [2,2] entry of $\boldsymbol{\Sigma}_i^*$ in (3.6) will increase as a function of the difference $\Delta_2 - \Delta_1$, resulting in heterogeneous populations. If $\Delta_2 - \Delta_1 = 0$, then the populations are homogeneous. However, due to other problems affecting its performance such as bimodality, we summarized the results based on the $\Delta_2 - \Delta_1$ rather than variance heterogeneity factor.

The heterogeneous populations in (b) can also be divided into marginally heterogeneous population if $0 < \Delta_2 - \Delta_1 \leq 2$ and severely heterogeneous population if $\Delta_2 - \Delta_1 > 2$.

To ease the interpretation of the simulation results and have an idea of the amount of heterogeneity in the variances of the two populations being compared in this case, we report in Table 3.1 the variance heterogeneity factor

$$VHF = \frac{1 + \pi_{21}\pi_{22}\Delta_2^2}{1 + \pi_{11}\pi_{12}\Delta_1^2} \tag{3.8}$$

obtained from (3.6).

Table 3.1: Variance heterogeneity factor for $\pi_{11} = \pi_{21} = 0.5$.

	Δ_2					
Δ_1	0	1	2	3	5	6
0	1	1.25	2	3.25	7.25	10
1		1	1.6	2.6	5.8	8
2			1	1.625	3.625	5
3				1	2.23	3.08
5					1	1.38
6						1

The above table exhibits gross heterogeneity of variances as $\Delta_2 - \Delta_1$ increases.

Studies have shown that in the Behrens-Fisher problem with unequal sample sizes, the performance of various testing procedures depends on where the larger variance lies (Keselman *et al.* 1993). Recognizing this, we define a " *positive correlation* " in sample size and sample variance if a larger variance is associated with a larger sample size and a " *negative correlation* " if otherwise.

Figures 3.3-3.6 plot the summary empirical levels (averaged over corresponding cases of $\Delta_2 - \Delta_1$) against the values of the difference $\Delta_2 - \Delta_1 = 0, 1, 2, 3, 5, 6$.

On the average, when $n_1 = n_2$, the Hotelling's T^2 performs quite competitively, specially when the sample sizes are reasonably large at $n_1 = n_2 = 21$. In fact, there were no discernible discrepancies among the procedures we considered in this case.

However, its performance deteriorates as $\Delta_2 - \Delta_1$ exceeds 2. In general, it appears that Kim's procedure is the best.

Clearly, the Hotelling's T^2 is not an option when the sample sizes are unequal and the populations are heterogeneous. Figures 3.5-3.6 indicate that its performance is rather sensitive and deteriorates rapidly with $\Delta_2 - \Delta_1 > 0$. All tests had a tendency to be liberal when smaller sample sizes are associated with larger Σ_i^*. This tendency is reversed when larger sample sizes are associated with larger Σ_i^* and hence with the Mahalanobis distance Δ_i.

3.3.1 Homogeneous Populations

Overall, all procedures perform quite well even with mixture data when the Mahalanobis distances are equal and the sample sizes are equal. When the sample sizes are unequal, HT appears to be the best followed by Johansen. This is also evident in Tables 3.2-3.5 in the portion headed by $\Delta_2 - \Delta_1 = 0$.

3.3.2 Marginally Heterogeneous Populations

Again when sample sizes are equal, all procedures perform reasonably well. The simple HT works rather well here. When sample sizes are unequal and smaller sample sizes are associated with larger Δ_i and hence larger Σ_i^*, none of the methods was able to keep the empirical level at the nominal value although Kim's procedure was close enough to $\alpha = .05$. However, this was not the case when the larger sample sizes were associated with larger Σ_i^* and Δ_i where all except Hotelling's T^2 performed rather well.

3.3.3 Severely Heterogeneous Populations

When the sample sizes are equal and large, there were very minor differences among the procedures in maintaining the nominal level. Kim's procedure was the best followed by Yao. When the sample sizes are unequal and negatively associated with sample variance and Mahalanobis distance, none of the procedures was able to maintain the nominal level although Kim's method appeared to be better, followed by Yao. All procedures were positively biased when the sample size is negatively correlated with Σ_i^*.

3.3.4 Monte Carlo Power

Overall, in terms of sizes, Kim's procedure seems to be the best in case of negative association of small sample size and large variance. However, Kim's procedure tends to be rather conservative for small sample sizes. It is recommended to use Kim's for the case of negative association between sample size and variance and Johansen's method otherwise.

To confirm this performance, we also examined the empirical power. We note here that all tests tend to be conservative except when $\Delta_1 - \Delta_2 > 2$ in which case all of them are liberal. HT is the least conservative (most liberal) in terms of empirical significance levels, a fact that is reflected in Table 3.6 of Powers.

Because HT has always relatively larger sizes than the other methods, our Table unfairly depicts HT as being powerful (see also Figures 3.7-3.10). However, other than the case $\Delta_1 - \Delta_2 = 0$, HT should not be recommended as it does not maintain the empirical sizes.

It is interesting to observe that HT gives satisfactory results for the type of mixture data we considered as long as there is no heteroscedasticity in the two populations being compared.

In our power calculations. we conclude that JH is the best overall closely followed by Y. However, note that only Kim's procedure can be used for the case of negative association between sample size and variance.

Table 3.2: Summary empirical levels for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .5$ and nominal $\alpha =$ .05. This table contains the results of Table A.1 summarized over the Mahalanobis distance differences of the two populations being compared. $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0481	0.0526	0.0546	0.0025
Y	0.0419	0.0453	0.0477	0.0025
JH	0.0457	0.0496	0.0523	0.0028
K	0.0419	0.0454	0.0481	0.0028
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0481	0.0529	0.0560	0.0029
Y	0.0423	0.0464	0.0496	0.0024
JH	0.0457	0.0496	0.0523	0.0023
K	0.0407	0.0453	0.0481	0.0025
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0573	0.0635	0.0683	0.0040
Y	0.0506	0.0548	0.0595	0.0028
JH	0.0533	0.0581	0.0618	0.0029
K	0.0476	0.0503	0.0533	0.0019

Table 3.3: Summary empirical levels for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .5$ and nominal $\alpha = .05$. This table contains the results of Table A.2 summarized over the Mahalanobis distance differences of the two populations being compared, $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
	$\Delta_2 - \Delta_1 = 0$			
HT	0.0480	0.0513	0.0536	0.0020
Y	0.0507	0.0570	0.0625	0.0039
JH	0.0492	0.0553	0.0595	0.0036
K	0.0469	0.0547	0.0596	0.0045
	$0 < \Delta_2 - \Delta_1 \leq 2$			
HT	0.0586	0.0688	0.0795	0.0082
Y	0.0574	0.0608	0.0656	0.0029
JH	0.0550	0.0579	0.0620	0.0024
K	0.0509	0.0545	0.0584	0.0029
	$\Delta_2 - \Delta_1 > 2$			
HT	0.0925	0.1121	0.1358	0.0161
Y	0.0633	0.0657	0.0678	0.0016
JH	0.0609	0.0643	0.0656	0.0015
K	0.0514	0.0539	0.0587	0.0024

Table 3.4: Summary empirical levels for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .5$ and nominal $\alpha = .05$. This table contains the results of Table A.3 summarized over the Mahalanobis distance differences of the two populations being compared, $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0476	0.0504	0.0517	0.0015
Y	0.0522	0.0571	0.0645	0.0049
JH	0.0501	0.0546	0.0603	0.0040
K	0.0481	0.0535	0.0612	0.0051
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0339	0.0407	0.0490	0.0057
Y	0.0453	0.0508	0.0595	0.0050
JH	0.0459	0.0503	0.0583	0.0045
K	0.0444	0.0503	0.0596	0.0048
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0298	0.0331	0.0358	0.0020
Y	0.0470	0.0504	0.0542	0.0023
JH	0.0493	0.0519	0.0563	0.0023
K	0.0462	0.0496	0.0527	0.0022

Table 3.5: Summary empirical levels for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .5$ and nominal $\alpha = .05$. This table contains the results of Table A.4 summarized over the Mahalanobis distance differences of the two populations being compared. $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0472	0.0499	0.0552	0.0029
Y	0.0465	0.0494	0.0542	0.0028
JH	0.0466	0.0495	0.0546	0.0029
K	0.0466	0.0493	0.0550	0.0031
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0463	0.0502	0.0534	0.0027
Y	0.0454	0.0492	0.0519	0.0024
JH	0.0453	0.0492	0.0520	0.0024
K	0.0458	0.0493	0.0523	0.0024
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0512	0.0542	0.0569	0.0019
Y	0.0500	0.0514	0.0528	0.0013
JH	0.0498	0.0515	0.0527	0.0011
K	0.0462	0.0491	0.0507	0.0014

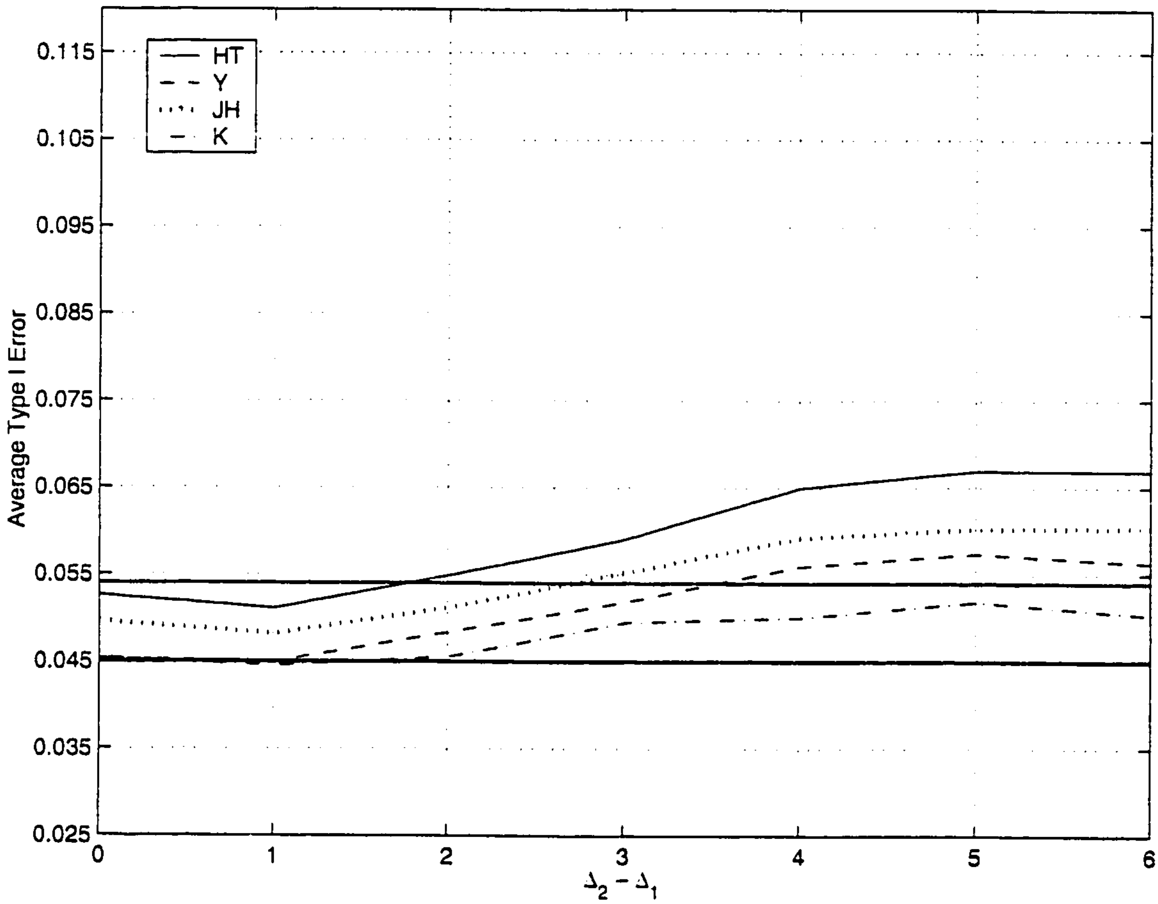

Figure 3.3: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .5$.

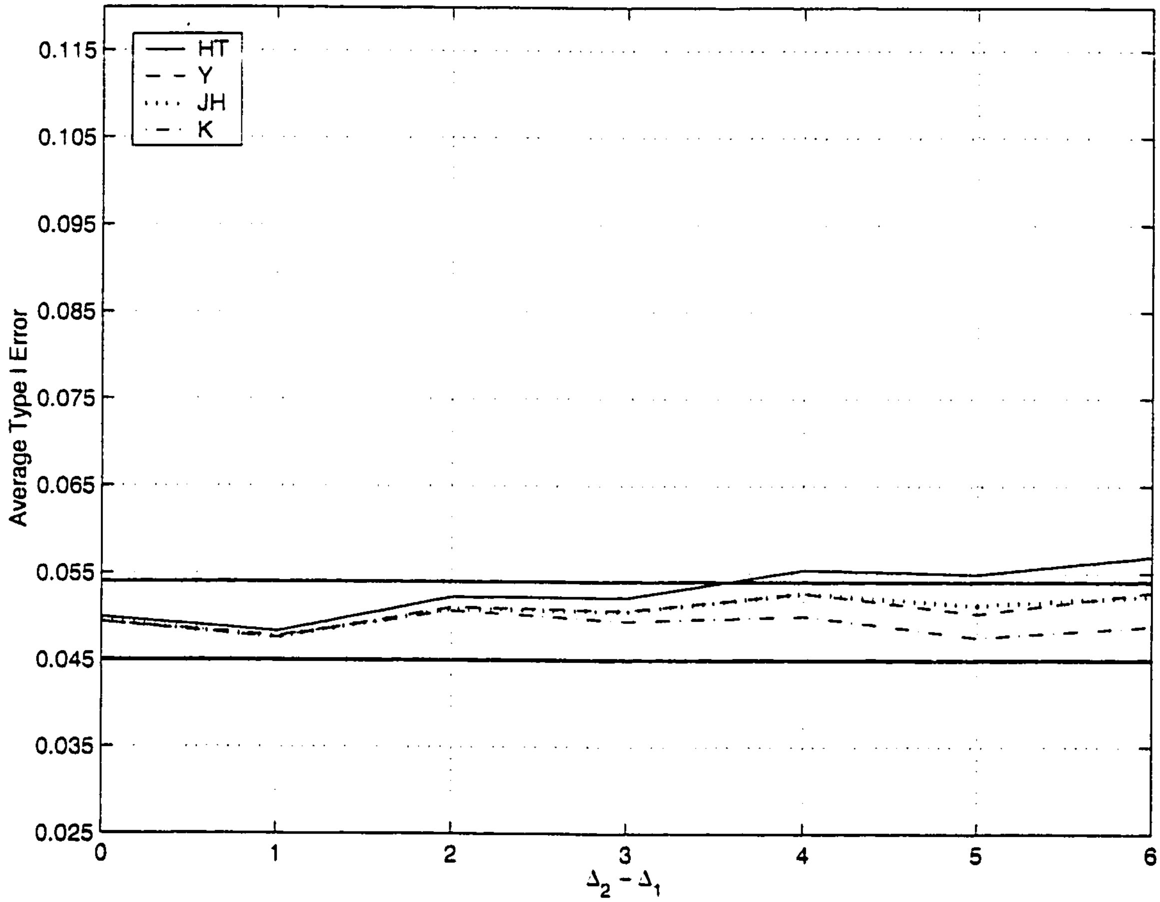

Figure 3.4: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .5$

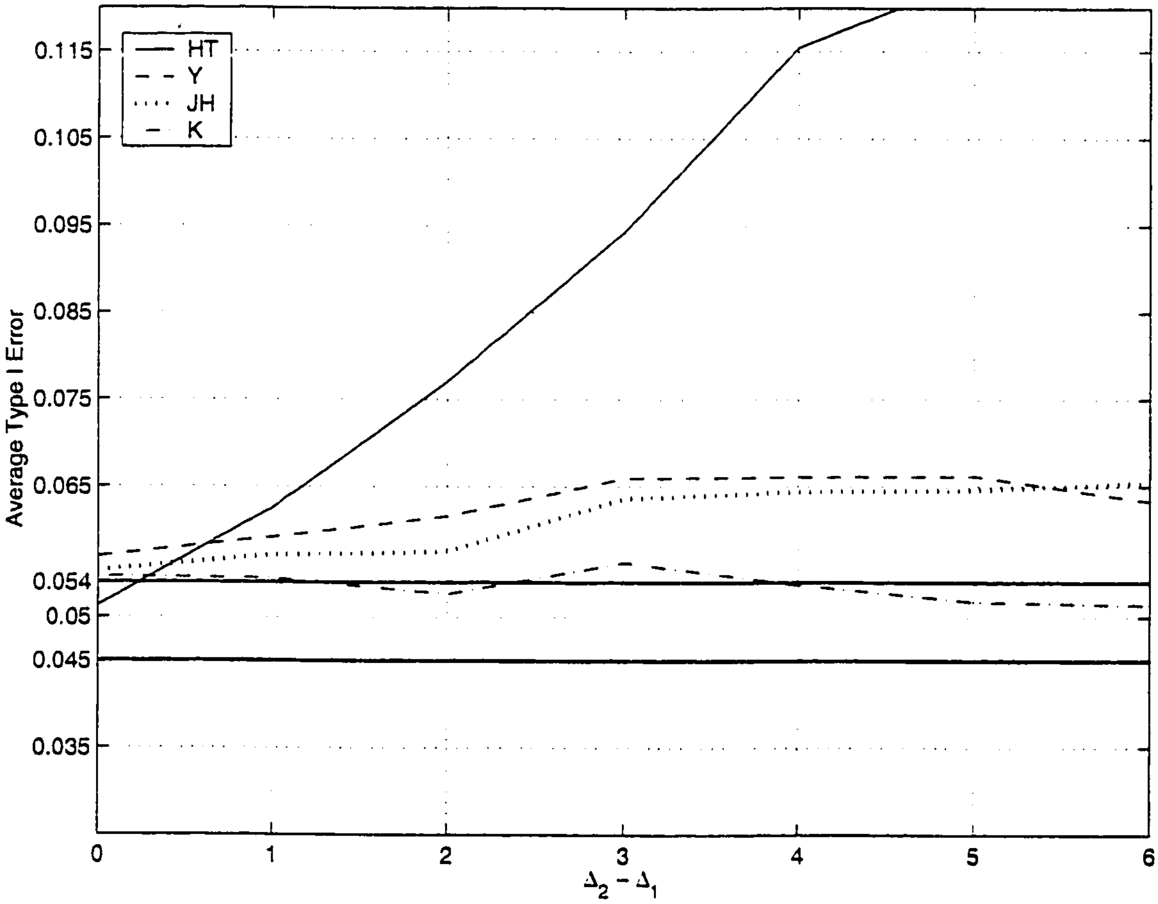

Figure 3.5: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .5$.

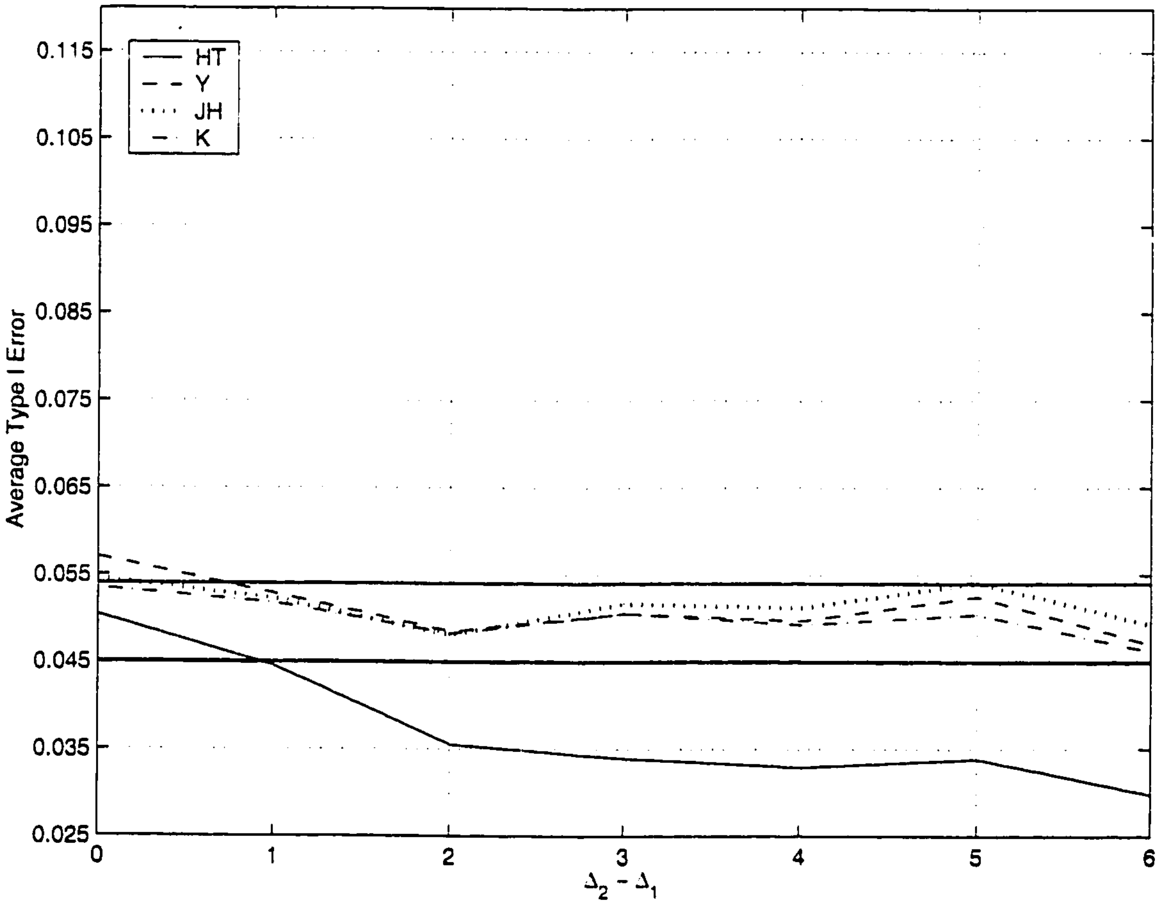

Figure 3.6: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .5$

Table 3.6: Empirical power averaged over Mahalanobis distance differences with nominal $\alpha = 0.05$ and $\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = 1$

	HT	Y	JH	K
	$n_1 = n_2 = 7,\ \pi_{11} = \pi_{21} = .5$			
$\Delta_2 - \Delta_1 = 0$	0.3947	0.3677	0.3850	0.3702
$0 < \Delta_2 - \Delta_1 \leq 2$	0.4101	0.3818	0.3970	0.3748
$\Delta_2 - \Delta_1 > 2$	0.3476	0.3184	0.3246	0.2763
	$n_1 = n_2 = 21,\ \pi_{11} = \pi_{21} = .5$			
$\Delta_2 - \Delta_1 = 0$	0.9142	0.9136	0.9137	0.9134
$0 < \Delta_2 - \Delta_1 \leq 2$	0.9216	0.9207	0.9208	0.9193
$\Delta_2 - \Delta_1 > 2$	0.8769	0.8741	0.8724	0.8376
	$n_1 = 14,\ n_2 = 7,\ \pi_{11} = \pi_{21} = .5$			
$\Delta_2 - \Delta_1 = 0$	0.5639	0.5195	0.5265	0.5136
$0 < \Delta_2 - \Delta_1 \leq 2$	0.5791	0.5014	0.5009	0.4781
$\Delta_2 - \Delta_1 > 2$	0.5417	0.4336	0.4099	0.3422
	$n_1 = 7,\ n_2 = 14,\ \pi_{11} = \pi_{21} = .5$			
$\Delta_2 - \Delta_1 = 0$	0.5628	0.5175	0.5225	0.5113
$0 < \Delta_2 - \Delta_1 \leq 2$	0.5421	0.5420	0.5529	0.5409
$\Delta_2 - \Delta_1 > 2$	0.4551	0.4826	0.4999	0.4632

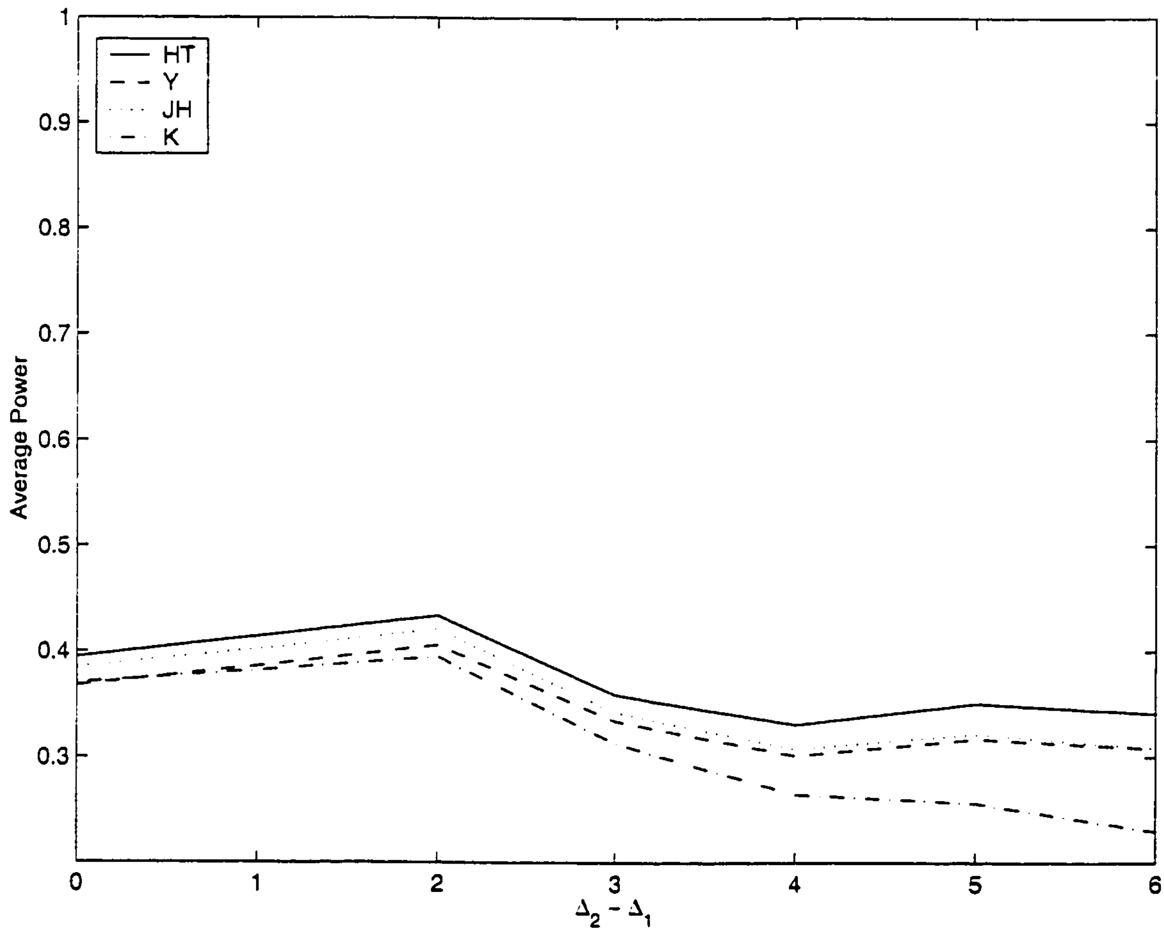

Figure 3.7: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .5$, $\mu_1^* - \mu_2^* = 1$

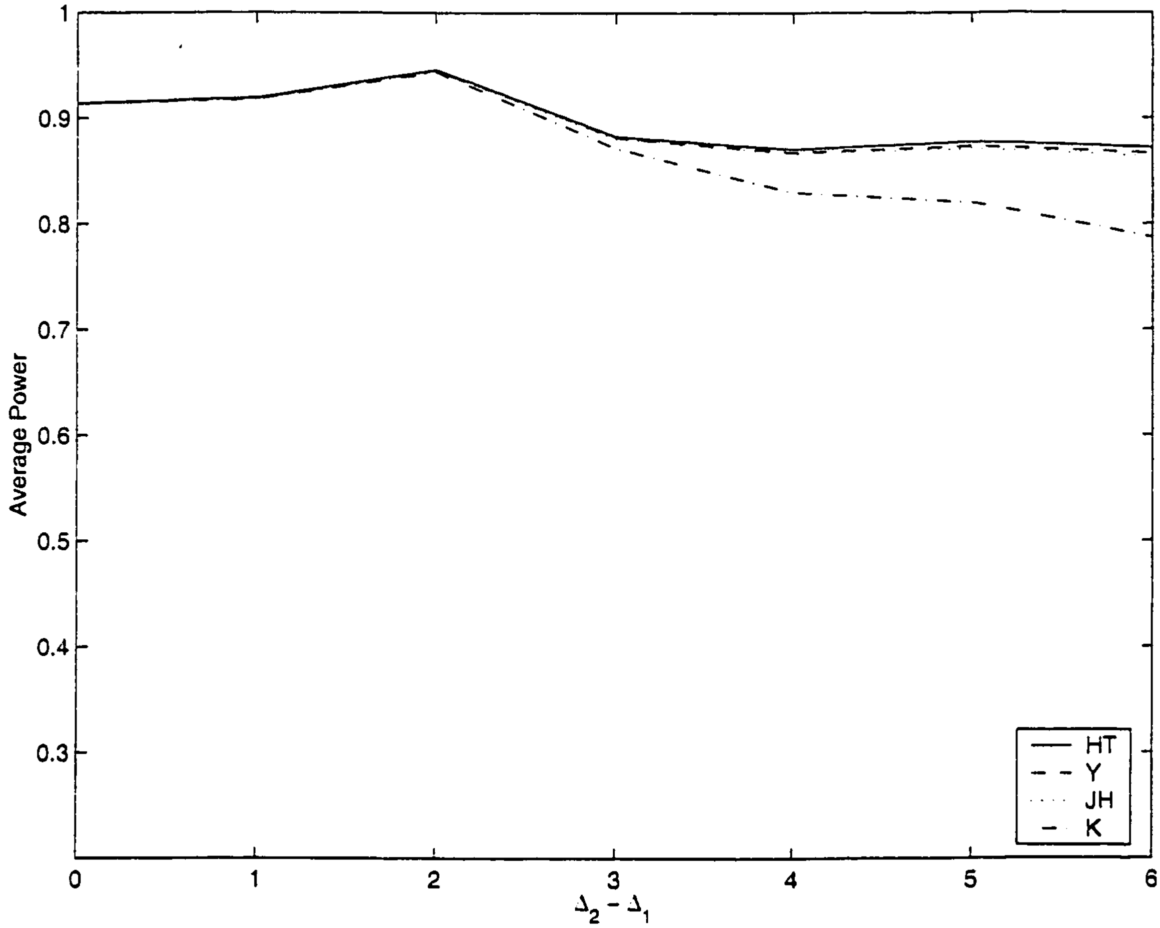

Figure 3.8: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .5$, $\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = 1$

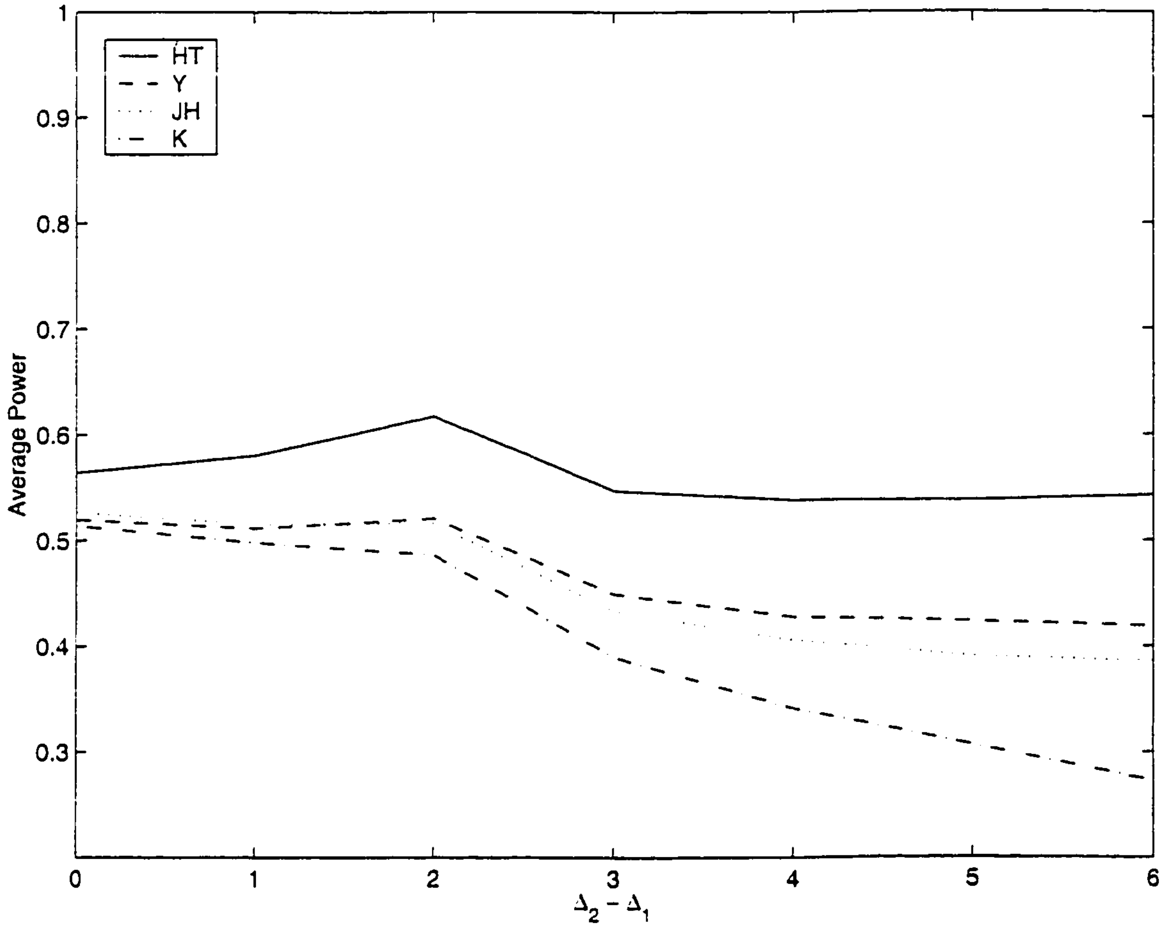

Figure 3.9: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .5$, $\mu_1^* - \mu_2^* = 1$

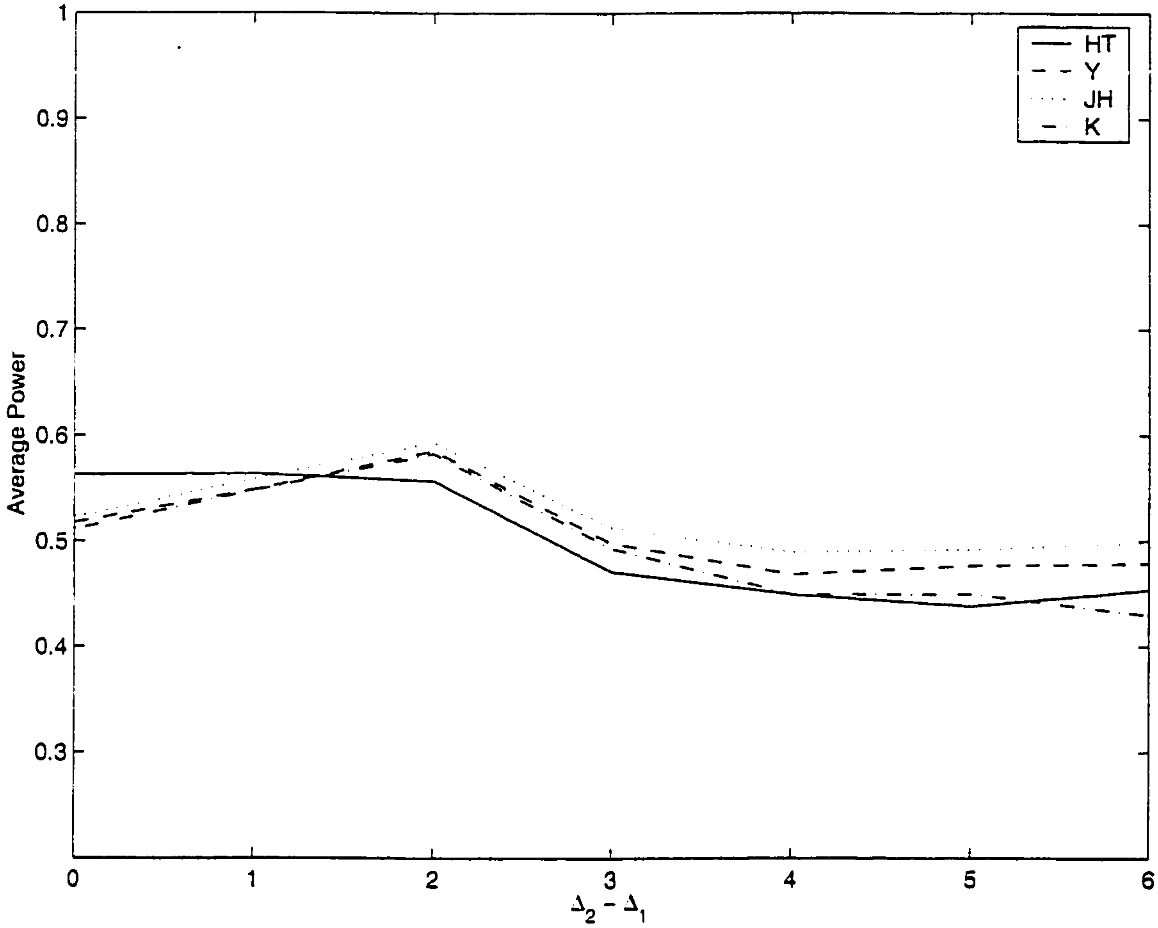

Figure 3.10: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .5$, $\mu_1^* - \mu_2^* = 1$

3.4 Results for the Case $\pi_{11} = \pi_{21} = .9$

This is the situation when each population is 10% contaminated by another normal population with different location parameter. Unlike the previous case, these populations are not symmetric.

To assist interpretation of the results we also give in Table 3.7 the variance heterogeneity factor defined in (3.8) for this case.

Table 3.7: Variance heterogeneity factor for $\pi_{11} = \pi_{21} = 0.9$.

	Δ_2					
Δ_1	0	1	2	3	5	6
0	1	1.09	1.36	1.81	3.25	4.24
1		1	1.25	1.66	2.98	3.89
2			1	1.33	2.39	3.12
3				1	1.793	2.34
5					1	1.3
6						1

Compared to Table 3.1, the VHF is not grossly large when the populations are contaminated only by 10%. However, our simulation results are still classified according to the Mahalanobis distance differences rather than VHF.

Figures 3.11-3.14 and Tables 3.8-3.11 show the performance of the selected procedures in keeping the nominal level $\alpha = .05$. On the average, when $n_1 = n_2$, Hotelling's T^2 outperformed the other methods even when the sample size is as small as 7. However, when $\Delta_2 - \Delta_1 > 2$ none were able to perform well. The situation becomes worse if there is a negative association between sample size and variance (See Figure 3.13). All procedures seem to be rather sensitive as $\Delta_2 - \Delta_1$ increases.

3.4.1 Homogeneous Populations

For this rather asymmetric case, Hotelling's T^2 performed exceptionally well even at such small sample size as $n_1 = n_2 = 7$. Its performance was independent of unequal or equal sample sizes. It is worth noting that all other procedures tend to be extremely conservative when sample size is small. Johansen's procedure was a little better than the others, while Kim's procedure tends to be the most conservative.

3.4.2 Marginally Heterogeneous Populations

In general, an overall increase of the empirical significance level as a function of $\Delta_2 - \Delta_1$ is noted (see Figures 3.11-3.14). Kim's method continues to be the most conservative when sample sizes are small. Hotelling's T^2 becomes noncompetitive when sample sizes are negatively correlated with the Σ_i^*. Johansen's procedure remains to be an excellent method.

3.4.3 Severely Heterogeneous Populations

From Figures 3.11-3.14, it is clear that none of the procedures appear appropriate when the non-symmetric populations are extremely heterogeneous. This was especially more so when the sample sizes are negatively correlated with the Mahalanobis distance. Tables 3.8-3.11 show the overall performance for this case (the bottom portion). It appears that one should increase sample size for the population with larger variance in which case Hotelling's T^2 appears to do reasonably well.

3.4.4 Monte Carlo Power

When the mixture data are contaminated only slightly, we observed that the best strategy was to increase the sample size for the population with larger variance. If this strategy is achievable, Table 3.12 and Figures 3.15-3.18 indicate that Johansen is the most powerful procedure to use. The other alternative is to make sample size reasonably large. However, the large sample size effect appears diminished as $\Delta_2 - \Delta_1$ increases, leading to the strategy observed above for the unequal sample size case, i.e. taking more samples from the population with a larger variance.

Table 3.8: Summary empirical levels for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .9$ and nominal $\alpha = .05$. This table contains the results of Table A.5 summarized over the Mahalanobis distance differences of the two populations being compared, $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0419	0.0486	0.0515	0.0041
Y	0.0349	0.0413	0.0459	0.0043
JH	0.0365	0.0448	0.0491	0.0051
K	0.0307	0.0392	0.0445	0.0057
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0414	0.0492	0.0529	0.0040
Y	0.0345	0.0425	0.0462	0.0040
JH	0.0361	0.0455	0.0495	0.0046
K	0.0303	0.0409	0.0456	0.0050
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0518	0.0637	0.0775	0.0092
Y	0.0437	0.0557	0.0693	0.0092
JH	0.0471	0.0590	0.0724	0.0092
K	0.0424	0.0521	0.0641	0.0082

Table 3.9: Summary empirical levels for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .9$ and nominal $\alpha = .05$. This table contains the results of Table A.6 summarized over the Mahalanobis distance differences of the two populations being compared. $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0453	0.0480	0.0535	0.0030
Y	0.0478	0.0515	0.0553	0.0026
JH	0.0459	0.0499	0.0530	0.0026
K	0.0464	0.0487	0.0518	0.0020
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0494	0.0568	0.0627	0.0049
Y	0.0555	0.0587	0.0654	0.0033
JH	0.0528	0.0561	0.0625	0.0033
K	0.0513	0.0536	0.0602	0.0031
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0635	0.0825	0.1079	0.0153
Y	0.0578	0.0791	0.0966	0.0122
JH	0.0543	0.0753	0.0948	0.0126
K	0.0506	0.0697	0.0870	0.0112

Table 3.10: Summary empirical levels for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .9$ and nominal $\alpha = .05$. This table contains the results of Table A.7 summarized over the Mahalanobis distance differences of the two populations being compared, $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0441	0.0485	0.0512	0.0025
Y	0.0464	0.0514	0.0539	0.0026
JH	0.0450	0.0498	0.0519	0.0025
K	0.0430	0.0486	0.0512	0.0029
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0432	0.0458	0.0488	0.0019
Y	0.0451	0.0496	0.0537	0.0033
JH	0.0446	0.0484	0.0523	0.0028
K	0.0437	0.0477	0.0511	0.0030
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0459	0.0514	0.0593	0.0041
Y	0.0467	0.0551	0.0625	0.0054
JH	0.0481	0.0560	0.0643	0.0055
K	0.0453	0.0536	0.0601	0.0050

Table 3.11: Summary empirical levels for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .9$ and nominal $\alpha = .05$. This table contains the results of Table A.8 summarized over the Mahalanobis distance differences of the two populations being compared, $\Delta_2 - \Delta_1$.

Test	Min	Mean	Max	STD
		$\Delta_2 - \Delta_1 = 0$		
HT	0.0464	0.0487	0.0528	0.0025
Y	0.0448	0.0476	0.0522	0.0029
JH	0.0449	0.0479	0.0525	0.0029
Y	0.0460	0.0485	0.0525	0.0027
		$0 < \Delta_2 - \Delta_1 \leq 2$		
HT	0.0471	0.0504	0.0522	0.0018
Y	0.0467	0.0492	0.0508	0.0017
JH	0.0467	0.0493	0.0509	0.0016
K	0.0464	0.0495	0.0519	0.0018
		$\Delta_2 - \Delta_1 > 2$		
HT	0.0495	0.0618	0.0720	0.0075
Y	0.0487	0.0605	0.0708	0.0073
JH	0.0485	0.0604	0.0704	0.0072
K	0.0480	0.0581	0.0673	0.0067

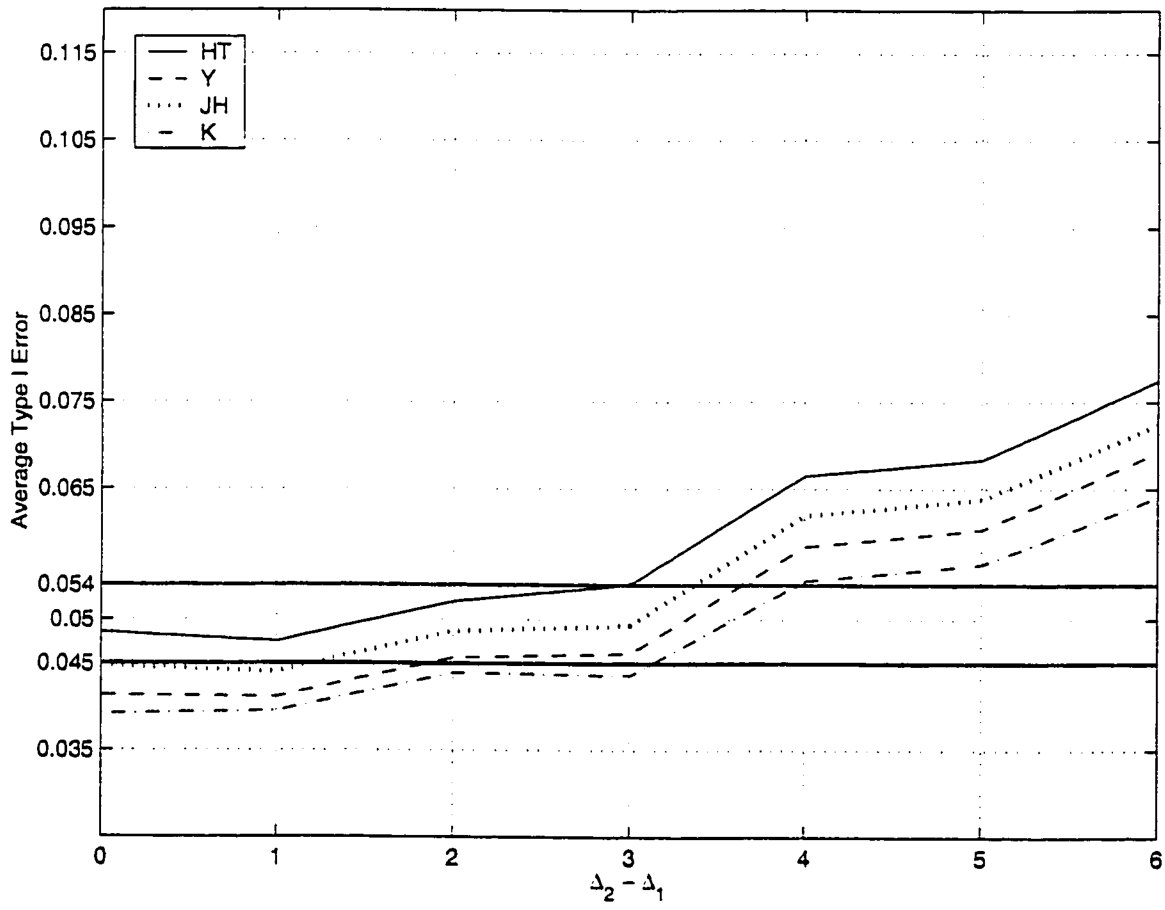

Figure 3.11: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .9$

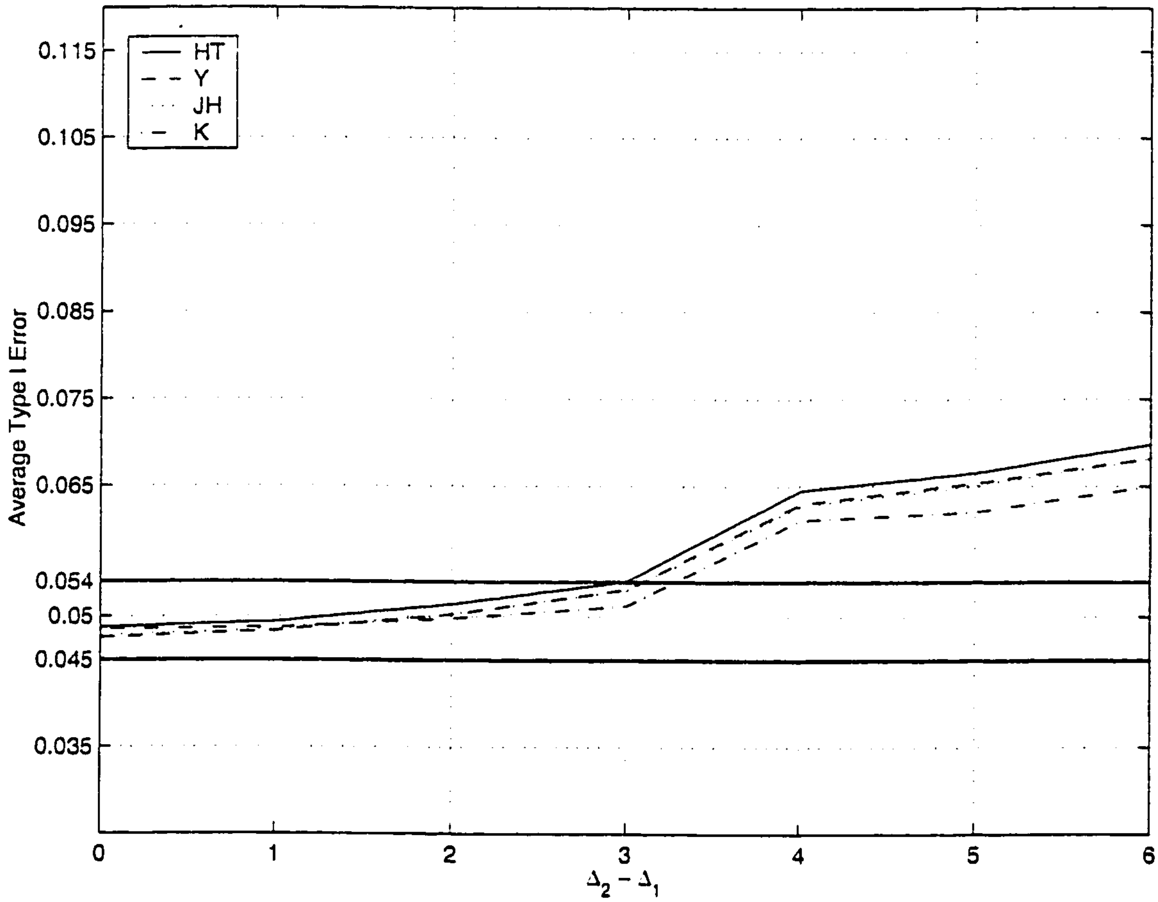

Figure 3.12: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .9$

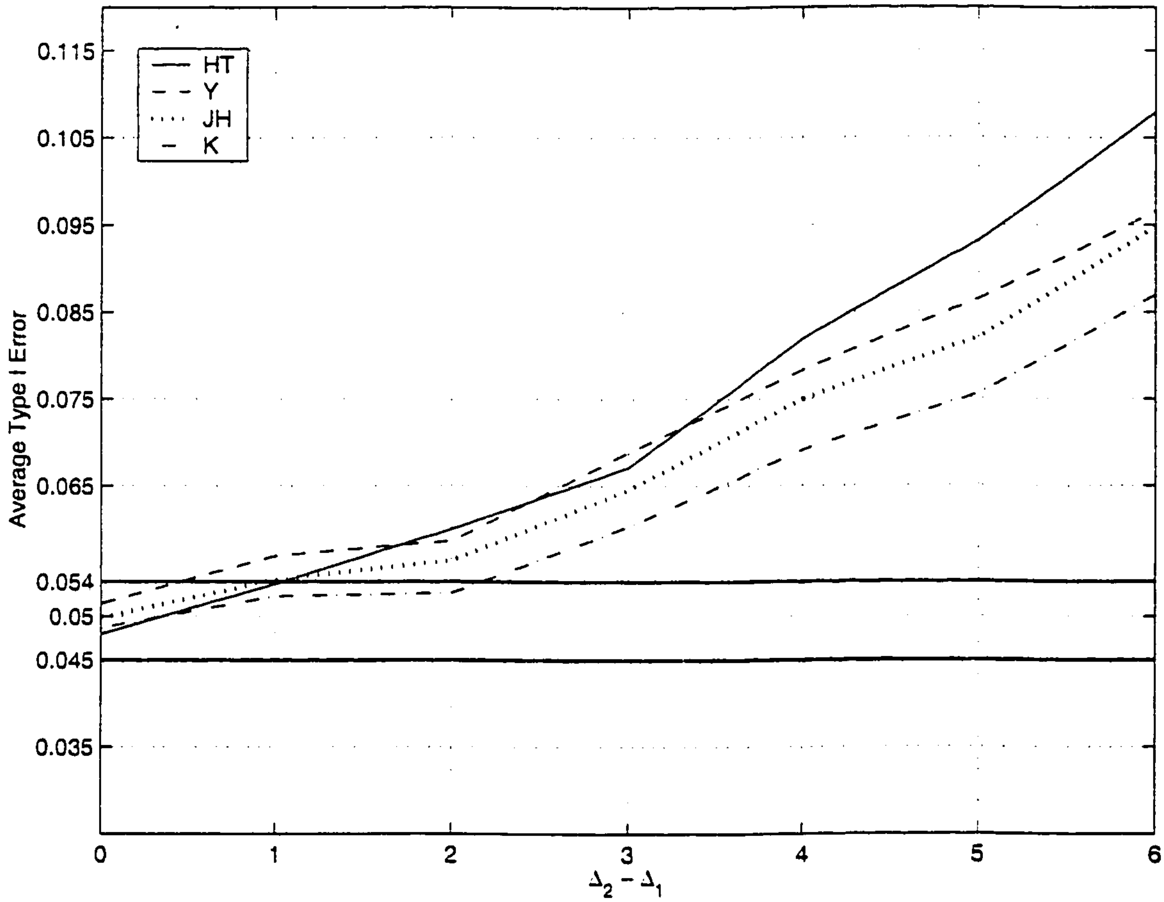

Figure 3.13: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .9$

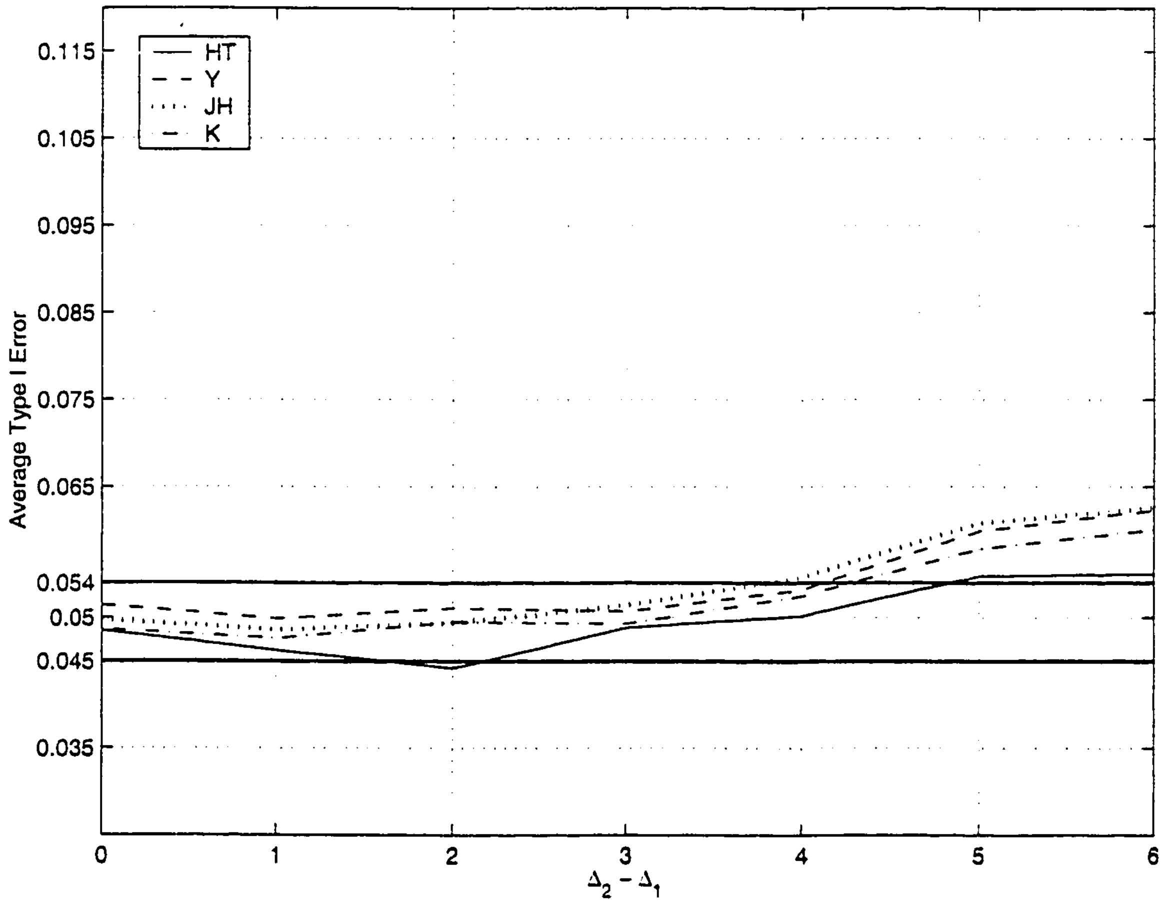

Figure 3.14: Empirical levels as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .9$

Table 3.12: Empirical power averaged over Mahalanobis distance differences with nominal $\alpha = 0.05$ and $\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = 1$

	HT	Y	JH	K
	$n_1 = n_2 = 7,\ \pi_{11} = \pi_{21} = .9$			
$\Delta_2 - \Delta_1 = 0$	0.4533	0.4212	0.4351	0.4116
$0 < \Delta_2 - \Delta_1 \leq 2$	0.4843	0.4564	0.4706	0.4463
$\Delta_2 - \Delta_1 > 2$	0.4990	0.4691	0.4807	0.4433
	$n_1 = n_2 = 21,\ \pi_{11} = \pi_{21} = .9$			
$\Delta_2 - \Delta_1 = 0$	0.9462	0.9452	0.9450	0.9442
$0 < \Delta_2 - \Delta_1 \leq 2$	0.9534	0.9525	0.9521	0.9501
$\Delta_2 - \Delta_1 > 2$	0.9244	0.9229	0.9216	0.9103
	$n_1 = 14,\ n_2 = 7,\ \pi_{11} = \pi_{21} = .9$			
$\Delta_2 - \Delta_1 = 0$	0.6171	0.6018	0.6034	0.5868
$0 < \Delta_2 - \Delta_1 \leq 2$	0.6484	0.6039	0.6041	0.5827
$\Delta_2 - \Delta_1 > 2$	0.6541	0.6046	0.5958	0.5522
	$n_1 = 7,\ n_2 = 14,\ \pi_{11} = \pi_{21} = .9$			
$\Delta_2 - \Delta_1 = 0$	0.6182	0.5560	0.5702	0.5502
$0 < \Delta_2 - \Delta_1 \leq 2$	0.6271	0.5906	0.6025	0.5863
$\Delta_2 - \Delta_1 > 2$	0.5711	0.5697	0.5882	0.5619

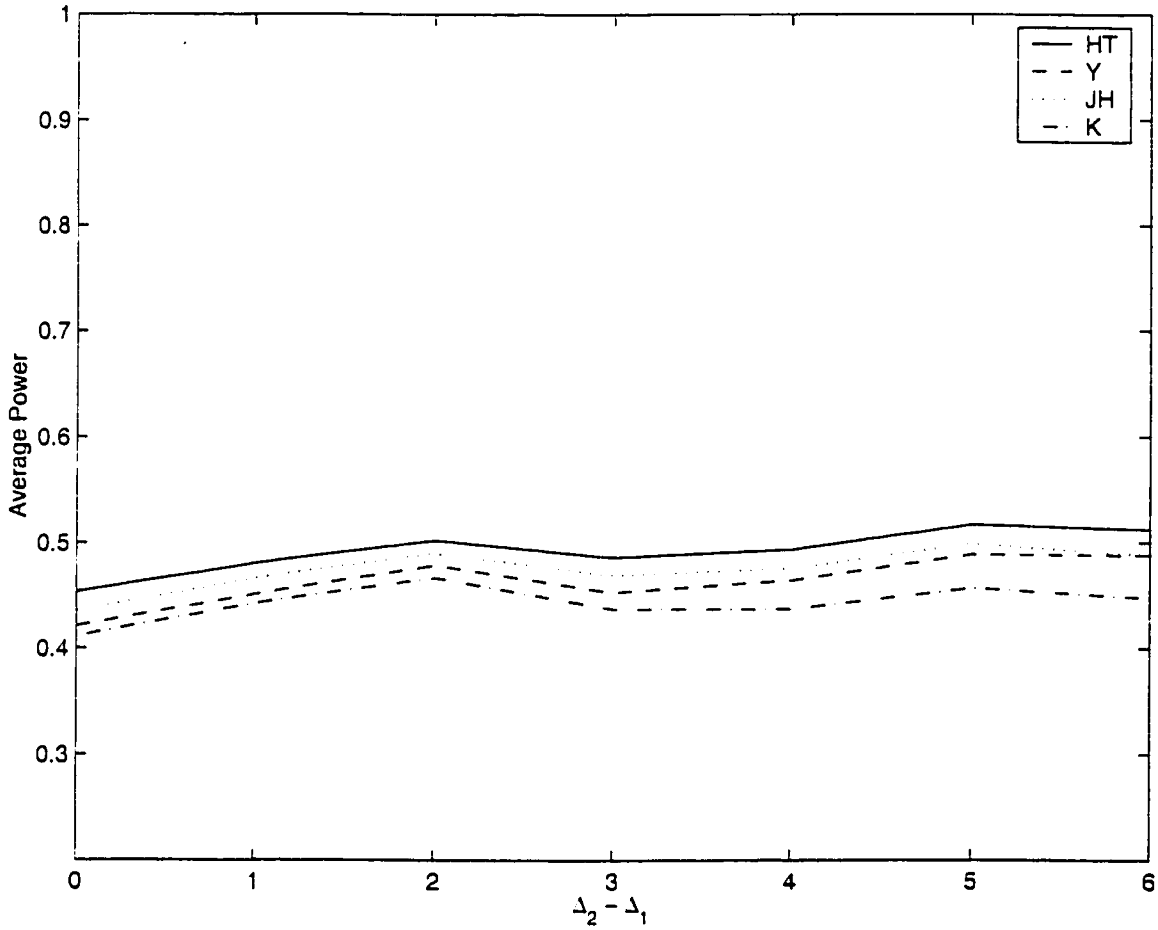

Figure 3.15: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/7$, $\pi_{11} = \pi_{21} = .9$, $\mu_1^* - \mu_2^* = 1$

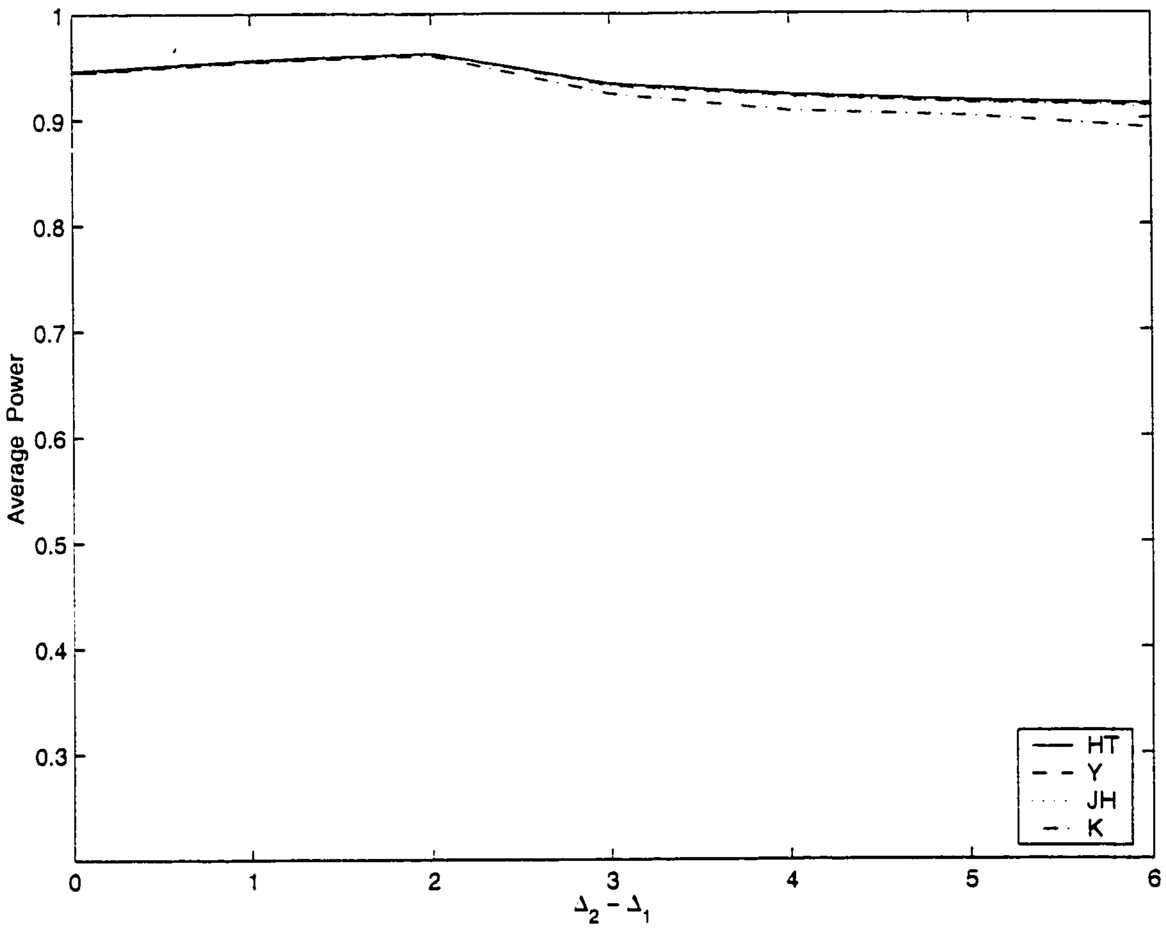

Figure 3.16: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 21/21$, $\pi_{11} = \pi_{21} = .9$, $\mu_1^* - \mu_2^* = 1$

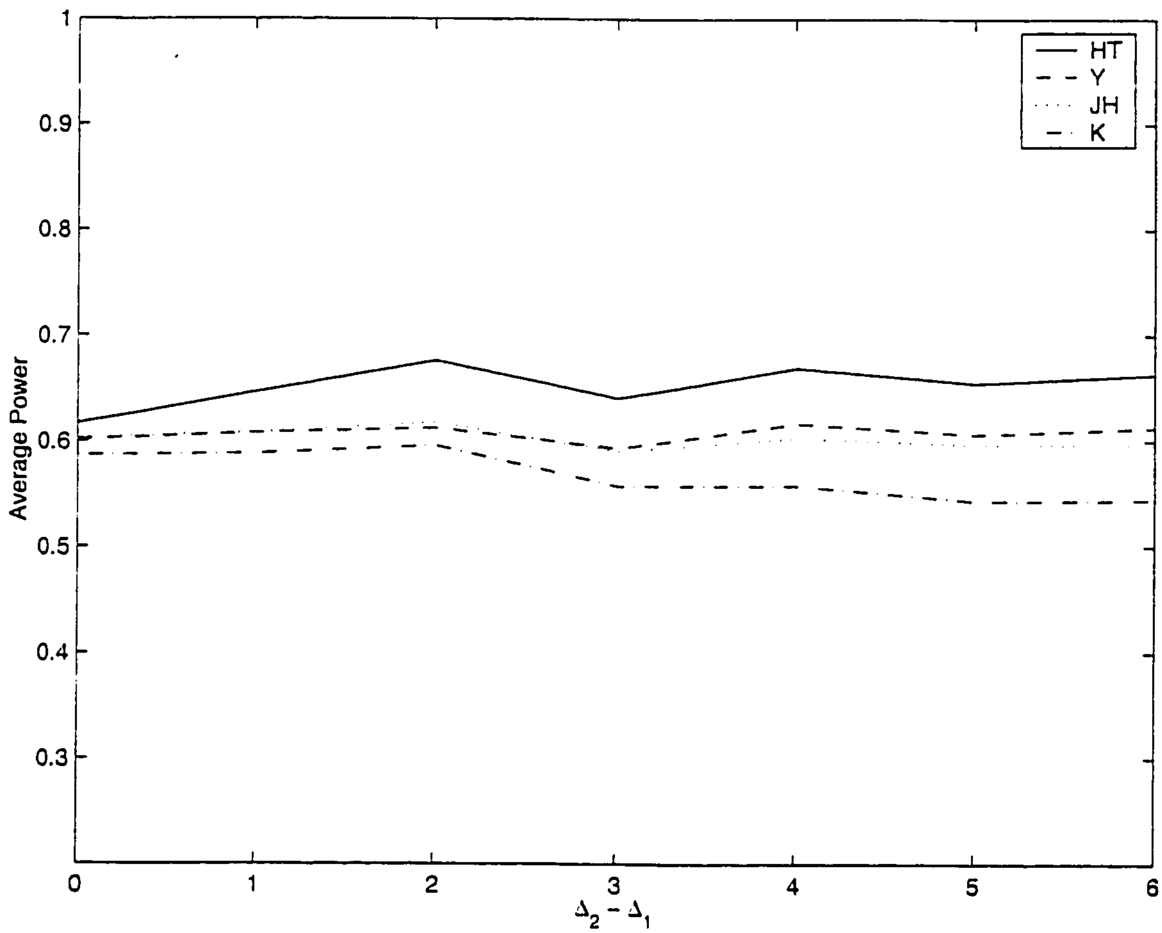

Figure 3.17: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 14/7$, $\pi_{11} = \pi_{21} = .9$, $\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = 1$

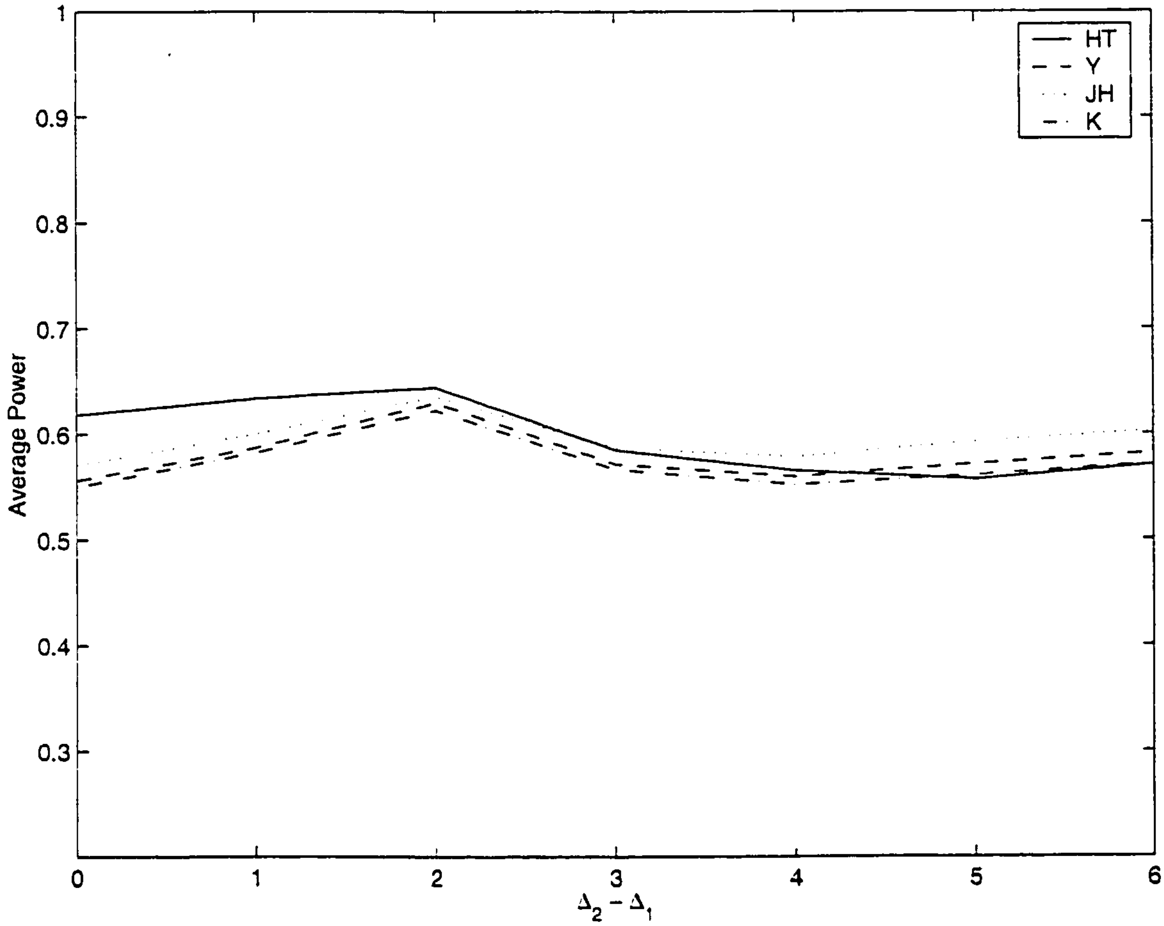

Figure 3.18: Empirical powers as functions of $\Delta_2 - \Delta_1$ for $n_1/n_2 = 7/14$, $\pi_{11} = \pi_{21} = .9$, $\mu_1^* - \mu_2^* = 1$

3.5 Overall Summary

The case of $\Delta_2 - \Delta_1 = 0$ corresponds to the situation where the data do not have Behrens-Fisher problem although they may consist of 2 different populations. Regardless this fact, the conventional procedures developed originally for normally distributed data appear to work well. On the average, all satisfied the usual criteria by staying within the confidence regions with an exception of the cases when the mixed populations are not symmetric and sample sizes are both small or populations are symmetric but samples are unbalanced.

When $\Delta_2 - \Delta_1 \neq 0$. we considered two situations. A moderately heterogeneous and thus nearly homogeneous if $\Delta_2 - \Delta_1 \leq 2$. Overall. Hotelling's procedure was not to be recommended. Kim's procedure tends to be extremely conservative with small samples. Johansen's procedure appears to be the best except when the sample size has negative correlation with the Mahalanobis distance. However. all procedures are rather liberal in this case.

When the data are extremely heterogeneous with $\Delta_2 - \Delta_1 > 2$, some of these procedures still work if the mixing proportions are equal leading to symmetric distributions. In such cases, Johansen's test was excellent followed by Kim and Yao. But this situation requires moderately large sample sizes and when the sample sizes are unequal, they have to be positively correlated with the Mahalanobis distance.

Chapter 4

Concluding Remarks

This thesis is concerned with testing hypothesis for data arising from a mixture of two normal distributions that are likely to have the Behrens-Fisher type problems. To our knowledge, no studies have been done to accommodate such testing situations of empirical data, although it is very plausible having data with mixture distributions. In this chapter, we summarize our findings and give several practical examples where our investigation will find its usefulness and real applications.

The literature that is available for finite mixture models is mainly concerned with the estimation of parameters involved in the mixture (e.g, mixing proportions, the number of components, location and scale parameters etc.) using estimation methods such as the Maximum Likelihood with the EM algorithm or some graphical methods (Averitt and Hand 1981). To our knowledge, there are no methods devised for testing hypothesis of equality of the location parameters of two populations, each of which is a mixture. For the perfectly normal populations, such methods exist and some of them have been extensively investigated and are known to perform well. Therefore, for the time being, a researcher dealing with hypothesis testing in mixture populations has no choice other than to use the available methods developed for normal distributions with the Behrens-Fisher problem, hoping that they perform reasonably well.

In this thesis, we investigated the four possibly better procedures based on the literature. These methods were, Johansen's, Yao's, Kim's and Jordan and Krishnamoorthy's, and for comprehensive comparison purposes, we also included the Hotelling's two-sample T^2-test procedure as well. Our overall conclusion was that Johansen's procedure appears to be the best for the small sample data we considered. Based on the performance with respect to the empirical size and power in small samples, we recommend the Johansen's procedure. However, the noted performance is not uniform for all cases. Therefore, the recommendation for investigators who are still in planning stage is that they try to increase the sample size if equal mixing proportions of both components are suspected. If the data are expected to be contaminated only slightly, the best strategy will be to increase the sample size for the component that will have disproportionately large variance.

Next, we give few examples where univariate and bivariate normal mixtures are used to model actual data sets.

Example1: Gage and Therriault (1998) fitted a two component univariate normal mixture to the human birth-weight data for different sex and ethnic groups. The authors assumed a general case in which the two components have different location and scale parameters. Using the Maximum Likelihood estimation procedure with the EM algorithm, they found that the male African American's birth-weight distribution is of the form,

$$F(x) = 0.926N(3337, 465) + 0.074N(2662, 1090).$$

Similar mixture distributions were found for European, Asian and Hispanic Americans's birth-weight data. Here we have 4 populations to compare, where each population is a mixture of two univariate normal components (i.e, $K=2$). This is an example of a situation where one would desire to compare mean birth-weights for different ethnic groups. If any of the univariate procedures for the Behrens-Fisher problem that were discussed in Chapter 2 is known to perform satisfactorily on mixture data, then one could use them to test such hypothesis without even requiring the estimation of the parameters of each group. This is extremely desirable especially when sample sizes are only moderately large (i.e, not greater than 30) in which case many of the parameter estimation methods do not give satisfactory results (Titterington *et al.* 1985). That is, large samples are required for such estimation methods to be stable and robust.

<u>**Example2:**</u> Another example is provided by a study conducted by Friedlander, Kark, Kidron and Bar-On (1995) where the authors used bivariate mixtures with normal components in order to model fasting glucose concentration and the glucose concentration two hours after an oral glucose load for the Jewish population of Jerusalem.

According to (Friedlander *et al.* 1995), from a previous study there was no evidence of sex influence on these glucose concentrations. Therefore, in their study, male data and female data were combined.

In this situation, one could employ the procedure developed by Johansen (1980), which was found in this thesis to perform quite well and was the best overall procedure, to test the hypothesis of no difference between male and female mean glucose concentrations.

Other examples of situations where bivariate normal mixture models where fitted to data can be found in Namboodiri *et al.* (1975).

Having recognized the the demand for mixture models and hypothesis testing there in, especially in human life matters, some possible directions for future investigations are:

1- There is a need to develop robust methods and procedures for the Behrens-Fisher problem in order to better accommodate analyses of data from the mixture distributions such as the situations considered in this thesis.

2- There is a need to develop a new procedure for testing the null hypothesis of no difference between two groups, each of which is a finite mixture with components having the same variance-covariance matrices. This is the case that does not suffer from identifiability problems (McLachlan 1988).

3- The method of Jordan & Krishnamoorthy was not considered further in this thesis due to its extreme conservativeness. However, there could be a possibility of improvement. Use of weights other than the c_i's suggested by the authors might lead to a better solution.

4- Further Monte Carlo study is needed to investigate and compare the univariate procedures such as Welch's approximate degrees of freedom. Banerjee's procedure and Bhoj's two procedures. The latter two procedures are quite recent methods and not considered in any previous extensive studies. In fact, Lee and D'Agostino (1976) considered two component normal mixtures and investigated the performance of the two-sample Student's t, Welch's approximate degrees of freedom method and the Mann-Whitney test under such mixtures. Therefore, an extension of the results of Lee and D'Agostino (1976) to new solutions of the univariate Behrens-Fisher problem will be useful in the literature.

5- The effects that other values of p have on the empirical sizes and powers need to be investigated.

Appendix A

Tables of Empirical Significance Levels

This Appendix contains tables of empirical significance levels as functions of Δ_1 and Δ_2 (the Mahalanobis distances associated with the two bivariate normal mixture populations being compared, see Section 3.1). The procedures of Hotelling, Yao, Johansen, Kim and Jordan & Krishnamoorthy were used with nominal $\alpha = .05$, mixing proportions $\pi_{11} = \pi_{21} = .5, .9$ and various sample sizes. The Mahalanobis distances were varied in the set $\{0, 1, 2, 3, 5, 6\}$. In each table, each cell contains empirical levels for the procedures HT, Y, JH, K, JK respectively.

Table A.1: Empirical significance levels for $n_1 = 7$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.5$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.0514	0.0481	0.0539	0.0588	0.0683	0.0669
	0.0427	0.0423	0.0471	0.0518	0.0595	0.0563
	0.0466	0.0457	0.0506	0.0553	0.0618	0.0604
	0.0419	0.0407	0.0446	0.0492	0.0533	0.0503
	0.0086	0.0081	0.0085	0.0133	0.0197	0.0230
1		0.0481	0.0501	0.0558	0.0658	0.0656
		0.0419	0.0447	0.0496	0.0558	0.0553
		0.0457	0.0476	0.0517	0.0580	0.0587
		0.0419	0.0438	0.0464	0.0483	0.0504
		0.0083	0.0090	0.0114	0.0183	0.0213
2			0.0541	0.0536	0.0573	0.0642
			0.0466	0.0470	0.0506	0.0561
			0.0512	0.0498	0.0533	0.0605
			0.0463	0.0468	0.0476	0.0518
			0.0102	0.0095	0.0150	0.0183
3				0.0546	0.0560	0.0610
				0.0477	0.0484	0.0531
				0.0520	0.0523	0.0569
				0.0481	0.0481	0.0516
				0.0102	0.0143	0.0160
5					0.0545	0.0525
					0.0477	0.0458
					0.0523	0.0496
					0.0481	0.0469
					0.0107	0.0112
6						0.0528
						0.0452
						0.0499
						0.0462
						0.0105

Table A.2: Empirical significance levels for $n_1 = 14$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.5$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 0	1	2	3	5	6
0	0.0504	0.0586	0.0747	0.0925	0.1257	0.1358
	0.0562	0.0582	0.0619	0.0639	0.0676	0.0633
	0.0543	0.0566	0.0583	0.0609	0.0654	0.0655
	0.0547	0.0536	0.0544	0.0551	0.0514	0.0514
	0.0008	0.0003	0.0011	0.0014	0.0065	0.0089
1		0.0480	0.0647	0.0795	0.1165	0.1213
		0.0507	0.0574	0.0613	0.0661	0.0648
		0.0492	0.0550	0.0568	0.0646	0.0638
		0.0469	0.0523	0.0509	0.0533	0.0521
		0.0007	0.0005	0.0018	0.0065	0.0075
2			0.0529	0.0659	0.0981	0.1145
			0.0570	0.0586	0.0662	0.0661
			0.0554	0.0566	0.0644	0.0643
			0.0533	0.0534	0.0550	0.0541
			0.0007	0.0019	0.0054	0.0095
3				0.0509	0.0770	0.0925
				0.0563	0.0656	0.0678
				0.0554	0.0620	0.0656
				0.0553	0.0584	0.0587
				0.0010	0.0048	0.0071
5					0.0536	0.0610
					0.0625	0.0625
					0.0595	0.0603
					0.0596	0.0584
					0.0035	0.0049
6						0.0520
						0.0590
						0.0582
						0.0583
						0.0042

Table A.3: Empirical significance levels for $n_1 = 7$, $n_2 = 14$, $\pi_{11} = \pi_{21} = 0.5$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.0505 0.0535 0.0527 0.0492 0.0008	0.0490 0.0546 0.0535 0.0511 0.0009	0.0371 0.0503 0.0496 0.0495 0.0007	0.0347 0.0510 0.0511 0.0511 0.0010	0.0327 0.0508 0.0519 0.0482 0.0012	0.0298 0.0470 0.0493 0.0462 0.0011
1		0.0507 0.0522 0.0501 0.0481 0.0007	0.0399 0.0453 0.0459 0.0444 0.0010	0.0339 0.0467 0.0465 0.0470 0.0011	0.0333 0.0520 0.0531 0.0508 0.0013	0.0350 0.0542 0.0563 0.0527 0.0013
2			0.0476 0.0533 0.0515 0.0508 0.0007	0.0421 0.0520 0.0516 0.0520 0.0012	0.0313 0.0500 0.0515 0.0495 0.0009	0.0325 0.0474 0.0493 0.0477 0.0021
3				0.0502 0.0585 0.0548 0.0542 0.0014	0.0359 0.0474 0.0470 0.0488 0.0009	0.0358 0.0506 0.0525 0.0510 0.0020
5					0.0516 0.0604 0.0581 0.0575 0.0036	0.0472 0.0595 0.0583 0.0596 0.0027
6						0.0517 0.0645 0.0603 0.0612 0.0046

Table A.4: Empirical significance levels for $n_1 = 21$, $n_2 = 21$, $\pi_{11} = \pi_{21} = 0.5$. Each cell contains empirical significance levels of the methods. HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.0503 0.0500 0.0502 0.0496 0.0011	0.0473 0.0470 0.0467 0.0462 0.0011	0.0529 0.0518 0.0513 0.0511 0.0014	0.0529 0.0518 0.0513 0.0507 0.0015	0.0546 0.0507 0.0508 0.0462 0.0024	0.0569 0.0528 0.0524 0.0489 0.0023
1		0.0477 0.0472 0.0474 0.0466 0.0013	0.0463 0.0454 0.0453 0.0458 0.0007	0.0516 0.0503 0.0504 0.0502 0.0009	0.0550 0.0527 0.0527 0.0503 0.0022	0.0552 0.0500 0.0518 0.0490 0.0015
2			0.0552 0.0542 0.0546 0.0550 0.0011	0.0497 0.0486 0.0491 0.0492 0.0012	0.0512 0.0500 0.0498 0.0486 0.0015	0.0558 0.0527 0.0527 0.0498 0.0015
3				0.0485 0.0480 0.0480 0.0482 0.0011	0.0534 0.0519 0.0520 0.0523 0.0018	0.0524 0.0502 0.0505 0.0490 0.0017
5					0.0472 0.0465 0.0466 0.0469 0.0013	0.0499 0.0493 0.0494 0.0500 0.0010
6						0.0506 0.0503 0.0503 0.0498 0.0016

Table A.5: Empirical significance levels for $n_1 = 7$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.9$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.0514	0.0462	0.0529	0.0518	0.0636	0.0775
	0.0427	0.0411	0.0462	0.0437	0.0556	0.0693
	0.0466	0.0437	0.0495	0.0471	0.0588	0.0724
	0.0419	0.0405	0.0456	0.0424	0.0515	0.0641
	0.0086	0.0082	0.0084	0.0095	0.0107	0.0169
1		0.0515	0.0507	0.0513	0.0641	0.0731
		0.0449	0.0450	0.0451	0.0554	0.0651
		0.0491	0.0482	0.0479	0.0593	0.0688
		0.0445	0.0438	0.0422	0.0516	0.0612
		0.0078	0.0082	0.0077	0.0126	0.0151
2			0.0512	0.0517	0.0554	0.0689
			0.0459	0.0440	0.0479	0.0614
			0.0487	0.0482	0.0507	0.0646
			0.0443	0.0435	0.0440	0.0573
			0.0084	0.0085	0.0095	0.0128
3				0.0504	0.0499	0.0550
				0.0422	0.0418	0.0468
				0.0472	0.0449	0.0503
				0.0397	0.0403	0.0444
				0.0073	0.0088	0.0089
5					0.0449	0.0414
					0.0373	0.0345
					0.0405	0.0361
					0.0339	0.0303
					0.0069	0.0056
6						0.0419
						0.0349
						0.0365
						0.0307
						0.0056

Table A.6: Empirical significance levels for $n_1 = 14$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.9$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2: 0	1	2	3	5	6
0	0.0484	0.0521	0.0590	0.0635	0.0914	0.1079
	0.0532	0.0557	0.0578	0.0578	0.0820	0.0966
	0.0518	0.0528	0.0550	0.0543	0.0760	0.0948
	0.0501	0.0515	0.0513	0.0506	0.0712	0.0870
	0.0011	0.0011	0.0006	0.0013	0.0014	0.0027
1		0.0482	0.0586	0.0611	0.0793	0.0950
		0.0514	0.0587	0.0596	0.0728	0.0910
		0.0513	0.0551	0.0579	0.0698	0.0881
		0.0486	0.0530	0.0541	0.0641	0.0800
		0.0012	0.0006	0.0011	0.0016	0.0022
2			0.0463	0.0549	0.0719	0.0846
			0.0500	0.0581	0.0710	0.0839
			0.0490	0.0555	0.0670	0.0802
			0.0471	0.0534	0.0623	0.0742
			0.0008	0.0008	0.0017	0.0015
3				0.0535	0.0627	0.0661
				0.0553	0.0654	0.0778
				0.0530	0.0625	0.0724
				0.0518	0.0602	0.0682
				0.0006	0.0017	0.0013
5					0.0462	0.0494
					0.0512	0.0555
					0.0482	0.0536
					0.0482	0.0515
					0.0009	0.0013
6						0.0453
						0.0478
						0.0459
						0.0464
						0.0005

Table A.7: Empirical significance levels for $n_1 = 7$, $n_2 = 14$, $\pi_{11} = \pi_{21} = 0.9$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.0491 0.0522 0.0499 0.0491 0.0009	0.0461 0.0494 0.0474 0.0462 0.0008	0.0432 0.0495 0.0480 0.0482 0.0006	0.0459 0.0524 0.0516 0.0503 0.0011	0.0502 0.0574 0.0574 0.0565 0.0015	0.0550 0.0623 0.0626 0.0601 0.0013
1		0.0512 0.0539 0.0515 0.0506 0.0004	0.0437 0.0511 0.0502 0.0496 0.0005	0.0452 0.0527 0.0509 0.0508 0.0005	0.0497 0.0545 0.0566 0.0545 0.0014	0.0593 0.0625 0.0643 0.0593 0.0011
2			0.0503 0.0517 0.0511 0.0495 0.0005	0.0463 0.0537 0.0523 0.0511 0.0008	0.0490 0.0531 0.0546 0.0523 0.0009	0.0507 0.0519 0.0526 0.0505 0.0010
3				0.0479 0.0517 0.0493 0.0485 0.0006	0.0470 0.0451 0.0456 0.0444 0.0005	0.0517 0.0467 0.0481 0.0453 0.0008
5					0.0485 0.0527 0.0519 0.0512 0.0006	0.0488 0.0454 0.0446 0.0437 0.0006
6						0.0441 0.0464 0.0450 0.0430 0.0007

Table A.8: Empirical significance levels for $n_1 = 21$, $n_2 = 21$, $\pi_{11} = \pi_{21} = 0.9$. Each cell contains empirical significance levels of the methods, HT, Y, JH, K, JK respectively.

Δ_1 \ Δ_2	0	1	2	3	5	6
0	0.0528 0.0522 0.0525 0.0525 0.0013	0.0471 0.0467 0.0467 0.0464 0.0009	0.0507 0.0498 0.0497 0.0493 0.0009	0.0495 0.0487 0.0485 0.048 0.0011	0.0611 0.0599 0.0597 0.0568 0.003	0.0698 0.0682 0.0682 0.065 0.006
1		0.047 0.0465 0.0467 0.0469 0.0006	0.0515 0.0508 0.0505 0.0502 0.0007	0.0522 0.0507 0.0509 0.0503 0.0013	0.0638 0.0618 0.0617 0.0601 0.0034	0.072 0.0708 0.0704 0.0673 0.005
2			0.0505 0.0501 0.0501 0.0506 0.0009	0.0493 0.0485 0.0484 0.0479 0.0009	0.056 0.0555 0.0552 0.0528 0.0013	0.0652 0.0641 0.0637 0.0621 0.0043
3				0.047 0.0458 0.0458 0.046 0.0011	0.0522 0.0507 0.0507 0.0519 0.0011	0.0567 0.0552 0.0554 0.0529 0.0026
5					0.0464 0.0448 0.0449 0.0461 0.0007	0.0496 0.0474 0.0482 0.0504 0.0006
6						0.0488 0.0461 0.0471 0.0491 0.0012

Appendix B

Tables of Empirical Powers

This appendix contains tables of empirical powers for all the methods analysed in Chapter 3 except Jordan & Krishnamoorthy's procedure. The empirical powers are considered as functions of $\Delta_1, \Delta_2 \in \{0, 1, 2, 3, 5, 6\}$ (the Mahalanobis distances associated with the two bivariate normal mixture populations being compared, see Section 3.1) under the alternative $H_1 : \mu_1^* - \mu_2^* = 1$. Various sample sizes are used with mixing proportions $\pi_{11} = \pi_{21} = .5, .9$. Each cell contains empirical powers of HT, Y, JH, K respectively.

Table B.1: Empirical powers for $n_1 = 7$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.9$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 0	1	2	3	5	6
0	0.5225	0.5185	0.5055	0.5010	0.5050	0.5125
	0.4920	0.4930	0.4795	0.4745	0.4735	0.4880
	0.5100	0.5065	0.4935	0.4895	0.4855	0.4880
	0.4880	0.4865	0.4690	0.4550	0.4465	0.4475
1		0.5015	0.5000	0.4990	0.4960	0.5290
		0.4695	0.4690	0.4775	0.4685	0.5040
		0.4845	0.4885	0.4865	0.4770	0.5125
		0.4635	0.4630	0.4645	0.4410	0.4675
2			0.4570	0.4645	0.4950	0.4910
			0.4285	0.4365	0.4595	0.4600
			0.4405	0.4520	0.4770	0.4745
			0.4245	0.4305	0.4495	0.4330
3				0.4305	0.4635	0.4625
				0.3960	0.4340	0.4250
				0.4135	0.4480	0.4415
				0.3920	0.4200	0.4065
5					0.4060	0.4390
					0.3735	0.4055
					0.3835	0.4190
					0.3565	0.3905
6						0.4025
						0.3675
						0.3785
						0.3450

Table B.2: Empirical powers for $n_1 = 7$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.5$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.5215	0.5035	0.4565	0.4105	0.3525	0.3410
	0.5000	0.4755	0.4310	0.3840	0.3205	0.3080
	0.5120	0.4905	0.4475	0.3935	0.3285	0.3085
	0.4975	0.4675	0.4195	0.3595	0.2685	0.2295
1		0.4730	0.4555	0.4110	0.3325	0.3490
		0.4435	0.4255	0.3805	0.2960	0.3140
		0.4605	0.4380	0.3950	0.3050	0.3155
		0.4410	0.4110	0.3700	0.2595	0.2445
2			0.4035	0.3755	0.3225	0.3290
			0.3715	0.3485	0.3005	0.3070
			0.3930	0.3685	0.3050	0.3110
			0.3765	0.3505	0.2760	0.2695
3				0.3640	0.3470	0.3435
				0.3370	0.3175	0.3175
				0.3560	0.3280	0.3295
				0.3365	0.3060	0.3035
5					0.3085	0.3220
					0.2830	0.2940
					0.2995	0.3115
					0.2900	0.2990
6						0.2975
						0.2715
						0.2890
						0.2800

Table B.3: Empirical powers for $n_1 = 21$, $n_2 = 21$, $\pi_{11} = \pi_{21} = 0.9$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

	Δ_2					
Δ_1	0	1	2	3	5	6
0	0.9840 0.9835 0.9835 0.9845	0.9835 0.9835 0.9835 0.9835	0.9705 0.9705 0.9705 0.9700	0.9690 0.9690 0.9690 0.9665	0.9280 0.9260 0.9260 0.9230	0.9140 0.9125 0.9105 0.8910
1		0.9780 0.9780 0.9780 0.9780	0.9740 0.9725 0.9730 0.9735	0.9545 0.9540 0.9540 0.9510	0.9355 0.9345 0.9335 0.9250	0.9060 0.9045 0.9035 0.8815
2			0.9660 0.9655 0.9650 0.9655	0.9650 0.9635 0.9630 0.9630	0.9255 0.9235 0.9230 0.9175	0.9105 0.9090 0.9060 0.8900
3				0.9505 0.9485 0.9500 0.9510	0.9240 0.9215 0.9185 0.9120	0.9065 0.9045 0.9015 0.8880
5					0.9055 0.9045 0.9025 0.8985	0.9025 0.9020 0.9020 0.8980
6						0.8930 0.8910 0.8910 0.8880

Table B.4: Empirical powers for $n_1 = 21$, $n_2 = 21$, $\pi_{11} = \pi_{21} = 0.5$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.9830	0.9765	0.9645	0.9275	0.8860	0.8720
	0.9825	0.9755	0.9640	0.9260	0.8805	0.8665
	0.9820	0.9760	0.9640	0.9260	0.8780	0.8635
	0.9815	0.9760	0.9645	0.9225	0.8320	0.7870
1		0.9695	0.9555	0.9275	0.8850	0.8685
		0.9690	0.9550	0.9270	0.8805	0.8660
		0.9690	0.9555	0.9265	0.8800	0.8635
		0.9695	0.9550	0.9230	0.8410	0.8080
2			0.9385	0.9060	0.8710	0.8555
			0.9380	0.9050	0.8695	0.8540
			0.9385	0.9045	0.8675	0.8530
			0.9380	0.9035	0.8530	0.8180
3				0.9015	0.8820	0.8495
				0.9010	0.8795	0.8495
				0.9010	0.8800	0.8475
				0.9015	0.8750	0.8390
5					0.8440	0.8395
					0.8440	0.8390
					0.8435	0.8390
					0.8420	0.8380
6						0.8490
						0.8470
						0.8480
						0.8480

Table B.5: Empirical powers for $n_1 = 14$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.9$ and $\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^* = \mathbf{1}$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

	Δ_2					
Δ_1	0	1	2	3	5	6
0	0.7025	0.7100	0.6900	0.6865	0.6605	0.6630
	0.6555	0.6480	0.6135	0.6040	0.6020	0.6125
	0.6685	0.6540	0.6250	0.6105	0.5960	0.5965
	0.6565	0.6355	0.6045	0.5835	0.5495	0.5440
1		0.6865	0.6755	0.6615	0.6745	0.6490
		0.6470	0.6145	0.6110	0.6135	0.6105
		0.6555	0.6215	0.6100	0.6010	0.5975
		0.6370	0.6045	0.5865	0.5620	0.5370
2			0.6430	0.6410	0.6270	0.6640
			0.6095	0.5915	0.5910	0.6195
			0.6105	0.5900	0.5860	0.6050
			0.5935	0.5690	0.5505	0.5535
3				0.6200	0.6020	0.6085
				0.6010	0.5725	0.5840
				0.6025	0.5615	0.5740
				0.5885	0.5350	0.5380
5					0.5530	0.5585
					0.5640	0.5765
					0.5605	0.5665
					0.5380	0.5440
6						0.4975
						0.5335
						0.5230
						0.5070

Table B.6: Empirical powers for $n_1 = 14$, $n_2 = 7$, $\pi_{11} = \pi_{21} = 0.5$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.7065	0.6765	0.6535	0.6085	0.5405	0.5425
	0.6600	0.6045	0.5490	0.4745	0.4215	0.4190
	0.6700	0.6150	0.5485	0.4630	0.3915	0.3860
	0.6545	0.5935	0.5205	0.4260	0.3210	0.2725
1		0.6735	0.6395	0.5810	0.5390	0.5345
		0.6285	0.5465	0.4925	0.4275	0.4250
		0.6390	0.5555	0.4840	0.4080	0.3890
		0.6250	0.5370	0.4520	0.3380	0.2920
2			0.5885	0.5490	0.5305	0.5370
			0.5280	0.4805	0.4380	0.4280
			0.5385	0.4800	0.4180	0.4040
			0.5230	0.4645	0.3645	0.3445
3				0.5115	0.4985	0.5015
				0.4680	0.4235	0.4355
				0.4750	0.4130	0.4200
				0.4585	0.3825	0.3795
5					0.4465	0.4560
					0.4100	0.4135
					0.4135	0.4105
					0.4035	0.3965
6						0.4570
						0.4225
						0.4230
						0.4170

Table B.7: Empirical powers for $n_1 = 7$, $n_2 = 14$, $\pi_{11} = \pi_{21} = 0.9$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.7135 0.6740 0.6840 0.6695	0.7195 0.6680 0.6875 0.6700	0.6615 0.6460 0.6540 0.6395	0.6320 0.6280 0.6365 0.6250	0.5715 0.5905 0.6110 0.5810	0.5705 0.5820 0.6030 0.5715
1		0.6740 0.6175 0.6315 0.6135	0.6510 0.6090 0.6225 0.6105	0.6270 0.6135 0.6175 0.6055	0.5680 0.5775 0.5930 0.5705	0.5420 0.5520 0.5740 0.5400
2			0.6610 0.5975 0.6100 0.5955	0.6235 0.5900 0.5975 0.5845	0.5775 0.5695 0.5875 0.5650	0.5620 0.5415 0.5635 0.5335
3				0.6180 0.5580 0.5700 0.5545	0.5665 0.5280 0.5460 0.5295	0.5450 0.5165 0.5375 0.5090
5					0.5405 0.4665 0.4870 0.4635	0.5405 0.4800 0.4925 0.4645
6						0.5025 0.4225 0.4385 0.4045

Table B.8: Empirical powers for $n_1 = 7$, $n_2 = 14$, $\pi_{11} = \pi_{21} = 0.5$ and $\mu_1^* - \mu_2^* = 1$. Each cell contains empirical powers of the methods, HT, Y, JH, K respectively.

Δ_1	Δ_2 = 0	1	2	3	5	6
0	0.7145	0.6850	0.5940	0.5195	0.4520	0.4535
	0.6650	0.6525	0.6230	0.5600	0.4900	0.4785
	0.6740	0.6575	0.6300	0.5685	0.5070	0.4985
	0.6605	0.6510	0.6225	0.5535	0.4700	0.4295
1		0.6600	0.5915	0.5195	0.4560	0.4240
		0.6155	0.5910	0.5480	0.4820	0.4630
		0.6270	0.6035	0.5565	0.5015	0.4765
		0.6155	0.5855	0.5425	0.4615	0.4290
2			0.5830	0.5165	0.4440	0.4425
			0.5280	0.5100	0.4775	0.4545
			0.5295	0.5210	0.4965	0.4775
			0.5135	0.5095	0.4690	0.4370
3				0.5145	0.4265	0.4490
				0.4745	0.4285	0.4555
				0.4800	0.4495	0.4730
				0.4685	0.4305	0.4565
5					0.4560	0.4615
					0.4050	0.4410
					0.4065	0.4525
					0.3985	0.4450
6						0.4490
						0.4170
						0.4180
						0.4115

Appendix C

Matlab Code for the Simulations

```
function comb1=comb1(file,n1,n2,pi1,pi2,cvjk)
%cvjk represents  the critical values of  JK's procedure.
%These critical values are tabulated
%in Jordan and Krishnamoorthy (1995).
DELTA=[0 1 2 3 5 6];
fid = fopen(file,'w');
fprintf(fid,'this file contains results of case: ...
n1=%2.1f n2= %2.1f',n1,n2);
fprintf(fid,' pi1= %3.2f pi2= %3.2f\n\n',pi1,pi2);
for r=1:6
    fprintf(fid,'%5.4f ',DELTA(1,r));
end
fprintf(fid,'\n');
status = fclose(fid);
cor=[1 0;0 1];

for i=1:6
```

```
    result1=zeros(5,6);
    DELTA1=DELTA(i);
    mu11=[0;(pi1-1)* DELTA1];
    mu12=[0;pi1* DELTA1];
    for k=i:6

        DELTA2=DELTA(k);
mu21=[0;(pi2-1)* DELTA2];
 mu22=[0;pi2* DELTA2];

result=sig_levs(mu11,mu12,mu21,mu22,cor,pi1,pi2,...
n1,n2,8.373);

  result1(:,k) =result
    end



fid = fopen(file,'a');

for f=1:5
   for l=1:6
if (l<i )
fprintf(fid,'%s','&')
elseif (l>=i & l~=6)
       fprintf(fid,'%3.3f' ,result1(f,l));
       fprintf(fid,'%s','&');
else
fprintf(fid,'%3.3f' ,result1(f,l));
```

```
fprintf(fid,'%s','\\');
      end
   end
fprintf(fid,'\n');
end
fprintf(fid,'\n');
status = fclose(fid);

end
```

```
function sig_levs=sig_levs(mu11,mu12,mu21,mu22,cor,pi1,...
pi2,n1,n2,cvjk)
p=2;
ccor=chol(cor);

sig_lev=[0;0; 0; 0;0];
for k=1:10000

X1=randn(p,n1);
X2=randn(p,n2);
        for j=1:n1
            if (pi1>= rand(1))
              X1(1:p,j)=ccor*X1(1:p,j)+mu11;
            else
              X1(1:p,j)=ccor*X1(1:p,j)+mu12;
            end
        end

        for j=1:n2
             if (pi2>=rand(1))
                  X2(1:p,j)=ccor*X2(1:p,j)+mu21;
             else
                   X2(1:p,j)=ccor*X2(1:p,j)+mu22;
             end
        end

f1=n1-1; f2=n2-1;
f=f1+f2;
```

```
X1bar=mean(X1')';

X2bar=mean(X2')';

S1=cov(X1');

S2=cov(X2');

d=X1bar-X2bar;

Sp=((f1.*S1) + (f2.*S2))./(f1+f2);

Su=(S1./n1) + (S2./n2);


%%%%%%%%%%%%%%% HOTELLING'S T2 CALCULATIONS %%%%%%%%%%%%%%


fhot=fcv(p,f-p+1);

cvhot=((f*p)/(f-p+1))*fhot*((n1+n2)/(n1*n2));

Thot=d'*(inv(Sp))*d;

p=2;


if Thot > cvhot

   hot=1;

else

   hot=0;

end


%%%%%%%%%%%%%%%%%%% YAO'S APDF CALCULATIONS %%%%%%%%%%%%%%%%%%


yapdf1= (( (d'*(inv(Su))*(S1./n1)*(inv(Su))*d)...

 /(d'*(inv(Su))*d) )^2)/f1;

yapdf2= (((d'*(inv(Su))*(S2./n2)*(inv(Su))*d)/...

(d'*(inv(Su))*d))^2)/f2;
```

```
yapdf=1/(yapdf1+yapdf2);

fyao=fcv(p, yapdf-p+1);
cvyao=fyao*(p*yapdf)/(yapdf-p+1);
Tyao=d'*inv(Su)*d;
if Tyao > cvyao
   yao=1;
else
   yao=0;
end


%%%%%%%%%%%%%%%%%%%% JOHANSEN'S PROC. CALCULATIONS %%%%%%%%%%%%%%%%
V1=inv(S1./n1); V2=inv(S2./n2);
V=V1+V2;
B1=eye(p,p)-(inv(V)*V1); B2=eye(p,p)-(inv(V)*V2);
A1= (1/(2*f1))* ( trace(B1*B1) + (trace(B1))^2 ) ;
A2=  (1/(2*f2))* (trace(B2*B2)+(trace(B2))^2) ;
A = A1 + A2;

Sjd=p+(2*A)-( (6*A)/(p*(p-1) +2) ) ;
%Sjd=p+2*A-6*A*(p+2);
japdf=(p*(p+2))/(3*A);
fjohan = fcv(p,japdf);
cvjohan = fjohan;
Tjohan=(1/Sjd)*(d'*(inv(Su))*d);

if Tjohan > cvjohan
```

```
   johan=1;
else
    johan=0;
end

%%%%%%%%%%%%%%%%%%%%%% KIM'S PROC. CALCULATIONS%%%%%%%%%%%%%
[Q,LAMDA]=eig(S1./n1, S2./n2);
r=det(LAMDA)^(1/(2*p));
S1S2=sqrtm(S2./n2)*sqrtm(sqrtm(V2)*(S1./n1)*sqrtm(V2))*...
sqrtm(S2./n2);
Sk=(S1./n1)+(r^2.*(S2./n2))+((2*r).*S1S2);
lamda=diag(LAMDA);
l=(lamda+ones(p,1)) ./ ((lamda.^(1/2) + (r*ones(p,1)) ).^2);
c=(l'*l)/sum(l); f=(sum(l))^2 /(l'*l);
fkim= fcv(f,yapdf-p+1);
cvkim=fkim*c*f*yapdf/(yapdf-p+1);
Tkim=d'*inv(Sk)*d;
if Tkim > cvkim
   kim=1;
else
   kim=0;
end
%%%%%%%%% JORDAN & KRISHNAMOORTHY PROC. CALCULATION%%%%%%%%%%%%
b1=1/(2*p*(f1^2)*(f1-1)/((f1-p-1)^2*(f1-p-3)));
b2=1/(2*p*(f2^2)*(f2-1)/((f2-p-1)^2*(f2-p-3)));
c1=b1/(b1+b2); c2=b2/(b1+b2);
W1=c1*V1; W2=c2*V2;
```

```
W=W1+W2;
SJK=( W2*inv(W)*W1*inv(W)*W2 )+(W1*inv(W)*W2*inv(W)*W1);
Tjk=d'*SJK*d;
%cvjk=cvjk;
if Tjk > cvjk
   jk=1;
else
   jk=0;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

sig_lev=sig_lev+[hot; yao; johan; kim;jk ];
end

% Returning a vector  containing Empirical significance
% levels of HT, Y, JH, K and JK respectively.

sig_levs=sig_lev./10000
```

Bibliography

Algina, J., T. Oshima and K. L. Tang (1995). Robustness of Yao's James' and Johansen's tests under variance-covariance heteroscedasticity and nonnormality. *Journal of Educational Statistics* **16**. 125–139.

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comapring two mean values. *Biometrika* **35**, 88–96.

Averitt, B. and D. J. Hand (1981). *Finite Mixture Distributions*. Chapman and Hall.

Banerjee, S. (1961). On confidence interval for two-mean problem based on separate estimates of the variances and tabulated values of t-table. *Sankhya* **23**. 359–78.

Bennett, B. M. (1951). Note on a solution of the generalized Behrens-Fisher problem. *Ann. Inst. Stat. Math.* **2**, 87–90.

Bhoj, D. S. (1993). An approximate solution to the Behrens-Fisher problem. *Biometrical J.* **35**, 635–40.

Bickel, P. J. and K. A. Doksum (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Inc.

Davenport, J. M. and J. T. Webster (1975). The Behrens-Fisher problem, an old solution revisited. *Metrika* **22**, 47–54.

de la Rey, N. and D. G. Nel (1993). A comparison of the significance levels and power functions of several solutions to the multivariate Behrens-Fisher problem. *South African Statist. J.* **27**, 129–148.

Fisher, R. A. (1935). Fiducial argument in statistical inference. *Annals of Eugenics* **6**, 391–8.

Friedlander, Y., J. D. Kark, M. Kidron and H. Bar-On (1995). Univariate and bivariate admixture analyses of serumglucose and glycated hemoglobin distributions in a Jerusalem population sample. *Human Biology* **67**(1), 151–170.

Gabler, S. and C. Wolff (1987). A quick and easy approximation to the distribution of a sum of weighted chi-square variables. *Statische Hefte* **28**, 317–325.

Gage, T. B. and G. Therriault (1998). Variabity of birth-weight distributions by sex and ethnicity: Analysis using mixture methods. *Human Biology* **70**(3), 517–534.

Hoaglin, D. C., F. Mosteller and J. W. Tukey (1985). *Exploring data tables, trends, and shapes.* New York: Wiley.

James, G. S. (1954). Tests of linear hypothesis in univariate and multivariate analysis when the ratios of the variances are unknown. *Biometrika* **41**, 19–43.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika* **67**, 85–92.

Johnson, M. E. (1987). *Multivariate Statistical Simulation.* New York: Wiley.

Johnson, R. A. and D. W. Wichern (1998). *Applied Multivariate Statistical analysis.* 4 ed., Printice Hall.

Jordan, S. M. and K. Krishnamoorthy (1995). Confidence regions for the common mean vector of several multivariate mormal polpulations. *The Canadian Journal of Statistics* **23**(3), 283–297.

Kendall, M. and A. Stuart (1969). *The Advanced Theory of Statistics*. Vol. 3. Hafner Publishers, New York, N. Y.

Keselman, H. J., K. C. Carrière and L. Lix (1993). Testing repeated measures hypothesis when covariance matrices are heterogeneous. *J. Ed. Stat.* **18**(4), 305–319.

Keselman, H. J., K. C. Carrière and L. Lix (1995). Robust and powerful nonorthogonal analysis. *Psychometrika* **60**(3), 395–418.

Kim, S-J. (1992). A practical solution to the Behrens-Fisher problem. *Biometrika* **1**(79), 171–6.

Lee, F. S. and R. B. D'agostino (1976). Levels of significance of some two-sample tests when observations are from compound normal distributions. *Commun. Statis. - Theor. Meth.* **A5**(4), 325–342.

McLachlan, G. J. (1988). *Mixture models: inference and applications to clustering.* New York, N.Y.: M. Dekker.

Namboodiri, K. K., R. C. Elson, C. J. Gueck *et al* (1975). Bivariate analyses of cholesterol and triglyceride levels in families in which probands have Type IIb lipoprotein phenotype. *Am. J. Hum. Genet.* **27**, 454–471.

Nel, D. G. and C. A. Van der Merwe (1986). A solution to the multivariate Behrens-Fisher problem. *Commun. Statis. -Theor. and Meth.* **15**, 3719–3735.

Scheffé, H. (1943). On the solutions of Behrens-Fisher problem, based on the t-distribution. *The Annals of Mathematical Statistics* **14**, 34–44.

Scheffé, H. (1970). Practical solutions of the Behrens-Fisher problem. *JASA* **65**, 1501–1509.

Srivastava, M. S. (1982). On the robustness of Hotelling's T^2-test and distribution of linear and quadratic forms in sampling from a mixture of two multivariate normal populations. *Commun. Statist.-Theor. Meth.* **11**(1), 81–107.

Subrahmaniam, K. and K. Subrahmaniam (1973). On the multivariate Behrens-Fisher problem. *Biometrika* **60**(1), 107–112.

Sukhatme, P. V. (1938). On fisher and behrens test of significance for the difference in means of two normal samples. *Sankhya* **1**, 39–48.

Titterington, D. M., A. F. M. Smith and U. E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester, New York: Wiley.

Welch, B. L. (1936). Specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. *Journal of the Royal Statistical Society* **Sup. 3**, 29–48.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 350–62.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika* **34**, 128–35.

Wilcox, R. R. (1995). Simulation results on solutions to the multivariate Behrens-Fisher problem via trimmed means. *The Statistician* **44**(2), 213–225.

Yao, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika* **52**, 139–47.