t-Statistics for Weighted Means with Applications to Credit Risk Models

Lisa R. Goldberg[*]
MSCI Barra, Berkeley, CA

Alec N. Kercheval[†]
Dept. of Mathematics
Florida State University

Kiseop Lee[‡]
Dept. of Mathematics
University of Louisville

September 10, 2004

[*]lisa.goldberg@barra.com
[†]kercheval@math.fsu.edu
[‡]kiseop.lee@louisville.edu

Abstract

We present a generalization of the two-sample t-test for equality of the means to the case where the sample values are to be given unequal weights. This is a natural situation in financial risk modelling where some samples are considered more reliable than others in predicting a common mean. We describe pooled and unpooled weighted t-tests, and show with an example of real credit data that using the standard unweighted t-test can lead to the wrong statistical conclusion.

1 Introduction

A common statistical question arises when two random samples, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are drawn from two different normal populations, possibly with different variances, and we wish to know whether or not the means of the two populations are equal. This is a classical problem known as the Behrens-Fisher problem, reviewed in section 2 below.

In this paper we address the following more general situation: it may happen that the sample X_i's, while independent and having the same mean, are not identically distributed because they have different variances:

$$X_i \sim N(\mu_X, \sigma_i^2), \qquad i = 1, \ldots, n$$

where $N(a, b)$ denotes the normal distribution with mean a and variance b. This may also be true of the Y_j's,

$$Y_j \sim N(\mu_Y, \sigma'^2_j), \qquad j = 1, \ldots, m.$$

We don't expect to decide whether $\mu_X = \mu_Y$ at this level of generality because the number of unknown parameters exceeds the size of the data. However, in some cases the context may indicate the individual variances, up to an overall scale factor, as in the following situation that inspired this study.

1.1 A Credit Risk Model

One way to model credit risk for bond portfolios is to use a multiple factor risk model as described in Grinold and Kahn (2000), chapter 3, in which the correlations among individual bond returns are assumed to be explained by a relatively small set of common factors.

A standard choice for the factors is simply the set of the various combinations of sector and rating, e.g. AA Financial. We would proceed as follows.

At the end of each month, the AA Financial bonds are priced and an option adjusted spread (OAS) is computed for each bond. The OAS measures the additional yield the bond is paying to compensate investors for the risk of default.

The sample measurement for each bond is the change in the value of this OAS from that of the previous month; this change is called the *spread return* of the bond.

Suppose there are n AA Financial bonds, and let X_i denote the spread return, in a given month, of the ith bond. Our modelling assumption is that spread returns of different AA Financial bonds are correlated only through linear exposure to the AA Financial factor return, which is to be computed as an average OAS across all the AA Financial bonds.

Since the actual spread returns vary from bond to bond, we suppose that the returns are sampled from normal distributions with mean μ, and we then wish to estimate μ from the sample data.

For various reasons, we expect some bonds to be more reliable than others in predicting the common average spread return. This reliability may depend on known features of the bond, such as duration, market capitalization, traded volume, or benchmark status. This means that we wish to give some bonds greater weight than others in computing the average, and is equivalent to the hypothesis that the distributional variance of the bond's spread return depends on the bond.

Suppose the mean OAS is to be computed according to the weighted average

$$\bar{X} = \sum_{i=1}^{n} w_i X_i,$$

where the w_i are known positive numbers with $\Sigma w_i = 1$. We assume the X_i are independent normal with a common mean. As described in Lemma 1, this is the best estimator of the mean if and only if the variance of the ith spread return is proportional to $1/w_i$:

$$X_i \sim N(\mu, \alpha/w_i),$$

where α is any positive constant.

In this situation, the specified weights will come from our financial views about which bonds should play the most important roles in determining the estimated sector-rating mean spread return. The constant α is probably unknown.

Now suppose we have a second collection of AA Financial bonds, but issued in a different country and denominated in a different currency. An important question is whether the mean spread return of the first group is equal to that of the second (Breger et. al., 2003). A positive answer would indicate the existence of global credit risk factors while a negative answer would indicate that different markets are driven by different credit risk factors. Because the bonds are drawn from distributions with different variances, the standard t-test methods don't apply. The purpose of this paper is to generalize the two-sample t-test method to this situation.

In the remainder of this section, we summarize the main results. In section 2 we recall for comparison the classical situation, and in section 3 we describe the weighted t-statistics. At the end of this section we illustrate their use with some credit spread data. The second example shows that using a t-test that ignores the variable weights can lead to an erroneous failure to reject the null hypothesis.

1.2 Main Results

For $i = 1, \ldots, n$, and $j = 1, \ldots, m$, let w_i and w'_j be known positive numbers (weights) with $\sum w_i = 1 = \sum w'_j$. Let X_i, Y_j (our sample values) be independent normal random variables, with

$$X_i \sim N(\mu_X, \sigma_i^2), \quad Y_j \sim N(\mu_Y, {\sigma'}_j^2).$$

The question is to examine the hypothesis test

$$H_0 : \mu_X = \mu_Y \quad \text{vs.} \quad H_1 : \mu_X \neq \mu_Y.$$

Lemma 1 *The weighted means*

$$\bar{X} = \sum_{i=1}^{n} w_i X_i, \quad \bar{Y} = \sum_{j=1}^{m} w'_j Y_j \tag{1}$$

are the best (UMVUE) estimators of μ_X and μ_Y, respectively, if and only if there exist positive constants α_X and α_Y such that for $i = 1, \ldots, n$ and $j = 1, \ldots, m$,

$$\sigma_i^2 = \frac{\alpha_X}{w_i} \quad \text{and} \quad {\sigma'}_j^2 = \frac{\alpha_Y}{w'_j}. \tag{2}$$

Hence, specifying the weights that define the weighted means is equivalent to stipulating the *relative* variances σ_i/σ_1 and σ'_j/σ'_1. Depending on how much is known about the proportionality constants α_X and α_Y, there are three different test statistics to use.

Case I (the Normal test): The values α_X and α_Y are known, i.e. all the variances are known. We may use the test statistic

$$W = \frac{\bar{X} - \bar{Y}}{\sqrt{\alpha_X + \alpha_Y}},$$

which is standard Normal by Lemma 2 in section 3.1.

Case II (Pooled two-sample test: the Weighted t-statistic): The ratio α_X/α_Y is known. Then we may use the test statistic

$$T_p = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X/\alpha_X + S_Y/\alpha_Y}{n+m-2}}\sqrt{\alpha_X + \alpha_Y}}, \tag{3}$$

where

$$S_X = \sum w_i (X_i - \bar{X})^2 \quad \text{and} \quad S_Y = \sum w'_j (Y_j - \bar{Y})^2. \tag{4}$$

Case III (Unpooled two-sample test): The ratio α_X/α_Y is unknown. Then we may use the test statistic

$$T_u = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{\alpha_X} + \hat{\alpha_Y}}}, \tag{5}$$

where

$$\hat{\alpha_X} = \frac{S_X}{n-1} \quad \text{and} \quad \hat{\alpha_Y} = \frac{S_Y}{n-1}. \tag{6}$$

In Case III, as in the classical case, the statistic (5) does not follow any t distribution exactly, but may be approximated by one, as discussed in Section 3.4, via a generalization of the Welch-Satterthwaite approximation (described below).

Theorem 1 *Let $T(d)$ denote the Student's t random variable with d degrees of freedom.*
(a) $T_p \sim T(n+m-2)$.
(b) T_u may be approximated by a t-distributed $T(d)$ with

$$d = \frac{(\hat{\alpha}_X + \hat{\alpha}_Y)^2}{\frac{\hat{\alpha}_X^2}{(n-1)} + \frac{\hat{\alpha}_Y^2}{(m-1)}}.$$

As an alternative, a simpler and more conservative choice is

$$d = \min(m-1, n-1).$$

(c) T_u, along with $T(d)$, tends in law to standard normal as $n, m \to \infty$.

Theorem 1 means that in either case, a t distribution may be used to reject the null hypothesis, just as in the unweighted case. See Appendix for all proofs.

2 Review: The classical case

In this section, for the convenience of the reader, we review some well-known methods for two-sample tests for equality of means.

2.1 Uniformly Minimum Variance Unbiased Estimators

Suppose that θ is our parameter to estimate and $\hat{\theta}$ is an estimator. If $E(\hat{\theta}) = \theta$, we call $\hat{\theta}$ an unbiased estimator (UE) of θ.

When we have more than one unbiased estimator, we prefer one with smaller variance; hence we seek the minimum variance unbiased estimator (MVUE). One difficulty is that the MVUE may depend not just on the sample values but also on θ itself. If it does not, then it is called the uniformly minimum variance estimator (UMVUE). If the UMVUE exists, it is unique. There are many ways to find the UMVUE, for example, via the Rao-Cramer rule or using exponential family properties. The reader can consult Hogg and Craig (1995) or Lehmann (1983) for detailed discussion of the UMVUE.

2.2 Two-Sample t-tests

Suppose we have a sequence of independent samples from a normal distribution with mean μ_X and variance σ_X^2. Denote the sample values by $X_1, X_2, \ldots, X_n$, so $X_i \sim N(\mu_X, \sigma_X^2)$.

The UMVUE of the mean μ is the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

If $Y_1, Y_2, \ldots, Y_m$ is another group of independent samples with $Y_i \sim N(\mu_Y, \sigma_Y^2)$,we could ask whether or not $\mu_X = \mu_Y$. We take the *null hypothesis* to be the statement that this equality is true. In other word, we are interested in the following hypothesis test:

$$H_0 \,:\, \mu_X = \mu_Y \quad \text{vs.} \quad H_1 \,:\, \mu_X \neq \mu_Y.$$

Given our sample data, we cannot determine the truth or falsity of the null hypothesis, but we can determine the likelihood of the realized sample values assuming the null hypothesis. To accomplish this, we look for a test statistic T where we can determine the probability, given H_0, that $|T|$ is greater than or equal to the magnitude of the realized value. Typically, if this probability (the "p-value") is below 5% or 1%, we reject H_0 in favor of H_1.

There are three cases. First, if the variances σ_X and σ_Y are known, we can use the normal test with

$$W = \frac{\bar{X} - \bar{Y}}{\sqrt{\alpha_X + \alpha_Y}}.$$

More commonly we do not know the variances, but we may know their ratio (e.g. they may be equal). Then we can use the pooled two-sample t-statistic (Moore and McCabe, 1999)

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)\hat{\sigma}_X^2 + (m-1)\alpha\hat{\sigma}_Y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \tag{7}$$

as a test statistic, where

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is the sample variance of X, and similarly for Y. The test statistic T has a (Student's) t-distribution with $n + m - 2$ degrees of freedom.

The third case is when we have no information on σ_X, σ_Y. This is the case of the well-known Behrens-Fisher problem (see, e.g. Duong and Shorrock, 1996; Scheffe, 1970). The most popular of many approaches to this problem is the Welch-Satterthwaite approximation.

In the Welch-Satterthwaite method, The test statistic is given by

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{\sigma_X}^2}{n} + \frac{\hat{\sigma_Y}^2}{m}}} \tag{8}$$

In this case, (8) does not follow a t-distribution exactly, but may be approximated by a formula of Satterthwaite (1946) for the optimal number of degrees of freedom. It is common to use the more conservative choice d.f.$=\min(n-1, m-1)$ for convenience.

3 The Weighted t-tests

3.1 Set-Up

We formalize our statistical set-up as follows.

Standing assumptions:

Let α_X and α_Y be fixed positive numbers. For $i = 1, \ldots, n$, and $j = 1, \ldots, m$, let w_i and w'_j be positive numbers and X_i, Y_j independent random variables such that

- $\sum_{i=1}^n w_i = 1$ and $\sum_{j=1}^m w'_j = 1$, and
- for each i, j, $X_i \sim N(\mu, \alpha_X/w_i)$ and $Y_j \sim N(\mu, \alpha_Y/w'_j)$.

Notation:

- $\bar{X} = \sum w_i X_i$ and $\bar{Y} = \sum w'_j Y_j$
- $S_X = \sum w_i (X_i - \bar{X})^2$ and $S_Y = \sum w'_j (Y_j - \bar{Y})^2$

Lemma 2 *With notation and assumptions as above,*

1. $\bar{X} \sim N(\mu, \alpha_X)$ *and* $\bar{Y} \sim N(\mu, \alpha_Y)$.
2. $\bar{X}, \bar{Y}, S_X$, *and* S_Y *are mutually independent.*
3. $S_X/\alpha_X \sim \chi^2(n-1)$ *and* $S_Y/\alpha_Y \sim \chi^2(m-1)$, *where* $\chi^2(k)$ *denotes the chi-squared distribution with* k *degrees of freedom.*
4. $$\hat{\alpha}_X = \frac{1}{n-1} \sum_{i=1}^n w_i (X_i - \bar{X})^2 = \frac{1}{n-1} S_X, \quad \hat{\alpha}_Y = \frac{1}{m-1} \sum_{i=1}^m w_i (Y_j - \bar{Y})^2 = \frac{1}{m-1} S_Y \tag{9}$$
   are the UMVUEs of the α*'s.*

3.2 When α_X and α_Y are known

In the easiest case, both α_X and α_Y are known. Then we can use

$$W = \frac{\bar{X} - \bar{Y}}{\sqrt{\alpha_X + \alpha_Y}}$$

as our test statistic. From Lemma 1, $\bar{X} \sim N(\mu, \alpha_X)$, $\bar{Y} \sim N(\mu, \alpha_Y)$ and they are independent. Therefore $\bar{X} - \bar{Y} \sim N(0, \alpha_X + \alpha_Y)$ under H_0 and W follows the standard Normal distribution under H_0.

3.3 When the α ratio is known

The question of interest is how to proceed with incomplete information about the variances of the sample distributions. Suppose we know only the ratio of α's.

The natural generalization of the classical pooled t-test is

$$T_p = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X/\alpha_X + S_Y/\alpha_Y}{n+m-2}}\sqrt{\alpha_X + \alpha_Y}}. \tag{10}$$

Setting $w_i = 1/n$, $w_i' = 1/m$, and $\alpha_Y = (n/m)\alpha_X$ reduces this expression to equation (7).

Note that T_p is independent of the scale of (α_X, α_Y), so depends only on the ratio α_X/α_Y. Theorem 1 says that it is a true t-statistic with $n+m-2$ degrees of freedom.

If we believe *a priori* that $\frac{\alpha_X}{\alpha_Y} = r$, we may wish to confirm this with a separate test. In this case, the following F test is useful. Our hypotheses are

$$H_0 : \frac{\alpha_X}{\alpha_Y} = r \quad v.s. \quad H_1 : \frac{\alpha_X}{\alpha_Y} \neq r.$$

Here is the test.

Theorem 2 $F = \frac{\hat{\alpha_X}}{r\hat{\alpha_Y}}$ *follows the F-distribution $F(n-1, m-1)$ under H_0.*

If the F-test is not significant, we may safely assume the ratio is r, and use the statistic T_p in equation (10). Otherwise we should use the following test.

3.4 When the α ratio is unknown

This is a generalization of the Behrens-Fisher problem. The natural extension of Satterthwaite's approximate t-distribution method is to use

$$T_u = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{\alpha_X} + \hat{\alpha_Y}}}$$

as our test statistic.

Observe that

$$T_u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y}}\sqrt{\alpha_X + \alpha_Y}} = \frac{W}{\sqrt{\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y}}}, \tag{11}$$

where W follows $N(0,1)$.

If T_u were to follow a t-distribution, then $\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y}$ would be of the form $\frac{V}{r}$, where V follows a chi-square distribution with r degree of freedom.

By Lemma 2, we see that

$$\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y} \backsim \frac{\alpha_X}{\alpha_X+\alpha_Y}\frac{1}{n-1}\chi^2(n-1)+\frac{\alpha_Y}{\alpha_X+\alpha_Y}\frac{1}{m-1}\chi^2(m-1),$$

which is clearly not of the form $\frac{V}{r}$ unless $n=m$. However, following Satterthwaite we can use an approximating chi-square distribution in which the number of degrees of freedom is chosen to provide good agreement with the exact distribution in the sense that the variances agree.

Observe that the variance of $\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y}$ is given by

$$\begin{aligned} Var(\frac{\hat{\alpha_X}+\hat{\alpha_Y}}{\alpha_X+\alpha_Y}) &= 2(\frac{\alpha_X}{\alpha_X+\alpha_Y}\frac{1}{n-1})^2(n-1)+2(\frac{\alpha_Y}{\alpha_X+\alpha_Y}\frac{1}{m-1})^2(m-1) \\ &= \frac{2}{(\alpha_X+\alpha_Y)^2}\{\frac{{\alpha_X}^2}{n-1}+\frac{{\alpha_Y}^2}{m-1}\}. \end{aligned}$$

Since Var$(\frac{V}{r})=\frac{2}{r}$, we conclude that the d.f. of the approximating chi-squared V should have

$$\text{d.f.} = \frac{(\alpha_X+\alpha_Y)^2}{\frac{{\alpha_X}^2}{n-1}+\frac{{\alpha_Y}^2}{m-1}}.$$

Since we don't know the true values of the α's, following Satterthwaite we replace them with $\hat{\alpha}$'s. Hence our approximation of T_u is the t-distribution with number of degrees of freedom equal to

$$\frac{(\hat{\alpha_X}+\hat{\alpha_Y})^2}{\frac{\hat{\alpha_X}^2}{n-1}+\frac{\hat{\alpha_Y}^2}{m-1}}. \tag{12}$$

Note that this quantity is not necessarily an integer. In a classical two sample t-test case, it's common to use the more convenient expression d.f.=min$(n-1,m-1)$ instead. If $n>m$,

$$\frac{(\alpha_X+\alpha_Y)^2}{\frac{{\alpha_X}^2}{n-1}+\frac{{\alpha_Y}^2}{m-1}} > \frac{(\alpha_X+\alpha_Y)^2}{\frac{{\alpha_X}^2+{\alpha_Y}^2}{m-1}} = \frac{(\alpha_X+\alpha_Y)^2}{{\alpha_X}^2+{\alpha_Y}^2}(m-1) > m-1.$$

Since the the tail of the t-distribution gets fatter tail as the number of degrees of freedom gets smaller, if we use $m-1$ instead of (12), we have more conservative test. This conclusion is consistent with classical theory.

3.5 Sample Results

3.5.1 cross-sectional data

To illustrate the use of these statistics, we test the difference of weighted means of one-month spread returns for a basket of Euro-denominated (X) and Sterling-denominated (Y) Financial AA bonds for each of three months: September, October, and November 2000. Each bond's weight is taken to be it's duration. The null hypothesis is that the weighted mean spread return of the Euro bonds is equal to that of the Sterling bonds.

Table 1 summarizes the results. The columns are: the values of the weighted mean spread returns in basis points, with the sample size in parentheses, and the sample (weighted) standard deviations; the value of the unpooled test statistic T_u; the p-value determined by using an approximate t-distribution with degrees of freedom computed via (12) and indicated in parentheses; the pooled test statistic T_p computed using a hypothesized alpha ratio shown in parentheses; and the p-value for the t-statistic T_p.

It is clear that by either the pooled or the unpooled method we may easily reject the null hypothesis in September and October, but not in November. (Examining longer history shows low correlation between between the two monthly means; see Breger et. al. (2003)). The full data are presented graphically in Figure 1.

Month	$\bar{X}$ (#)	S_X	$\bar{Y}$ (#)	S_Y	T_u	p-value (df)	T_p ($\frac{\alpha_X}{\alpha_Y}$)	p-value
Sep 00	-4.31 (51)	34.4	7.23 (36)	201	-4.59	$< 10^{-4}$ (43)	-4.50 ($\frac{1}{9}$)	$< 10^{-4}$
Oct 00	-0.86 (53)	1.2	7.02 (31)	21.1	-9.25	$< 10^{-10}$ (40)	-9.32 ($\frac{1}{30}$)	$< 10^{-10}$
Nov 00	2.09 (58)	2.6	1.30 (33)	42.8	0.67	0.508 (34)	0.66 ($\frac{1}{30}$)	0.507

Table 1: Sample Bond Return Data Summarized

3.5.2 Time series data

The following example shows that use of the weighted t-statistic will sometimes prevent an erroneous conclusion caused by inappropriately applying the usual unweighted test. We use a ten year time series (1991 – 2001) of monthly factor return data for US bonds, comparing a basket of BB rated Energy bonds with BB rated Transportation bonds. The factor return for a given month is an average of the spread returns over that month for all the bonds in the basket. This time, weighting is exponential with a half-life of 24 months, where the most recent data is weighted highest. We compare our weighted, unpooled test using T_u with a naive classical test using (8), where the weights are

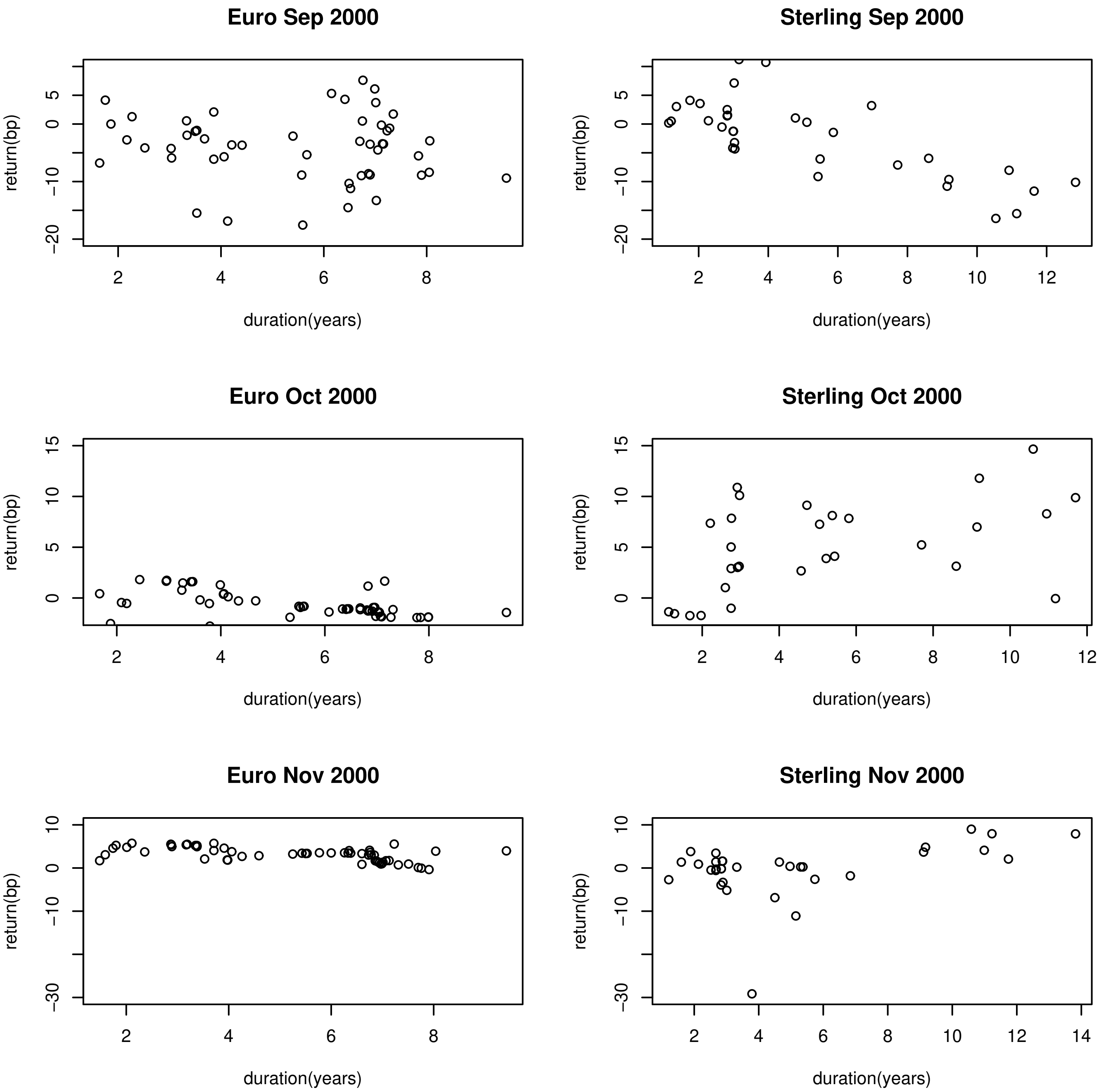

Figure 1: Spread return data for a selection of Financial AA bonds

ignored. In this case applying the weighted test allows us to reject the null hypothesis (95% confidence), while the naive test fails to reject the null hypothesis. The data are presented in Figure 2, and summarized in Tables 2 and 3, where X denotes Energy BB, and Y denotes Transportation BB. The number of degrees of freedom used is the best approximate value as given by (12).

factor	no. months	weighted time average	weighted variance
Energy BB	138	-5.83	$S_X = 1053$
Transportation BB	125	13.94	$S_Y = 9628$

Table 2: Time series average factor return for two US bond baskets

test method	approx. d.f.	p-value	conclusion
weighted $T_u = 2.13$	148	0.0348	reject H_0
unweighted $T = 1.52$	151	0.1287	accept H_0

Table 3: Weighted vs Unweighted t-test results

4 Conclusions

We have provided test statistics for testing the equality of means when the intrasample variances are not necessarily all equal, which corresponds to the case of unequally weighted means.

It can happen, as illustrated above, that incorrectly using the classical equal-weighted test statistic leads to the wrong statistical conclusion. Hence investigators should take care to use the weighted statistics T_p or T_u described here.

Typically practitioners will use the unpooled statistic T_u, which in practice means using a t-approximation with appropriately chosen number of degrees of freedom as described in Theorem 1.

Our sample results suggest it may not matter too much whether T_p or T_u is being used. This is because of the general fact that when the number of degrees of freedom is not too small (greater than 10 or so, say), the distributions of T_u and the t-distributions are all very close to Standard Normal.

To illustrate the quality of the approximation, we use the data from September 2000 in the cross-sectional sample described above to compare the density functions of T_u and $T(d)$ in Figure 3. Here the distribution of T_u computed from (11) by simulation, and the distribution of $T(d)$ is plotted for $d = 43$, the Satterthwaite approximate number of degrees of freedom. Notice that the two density functions are almost identical to the naked eye (and very close to standard normal). The sample size for the simulation was 10^6 samples.

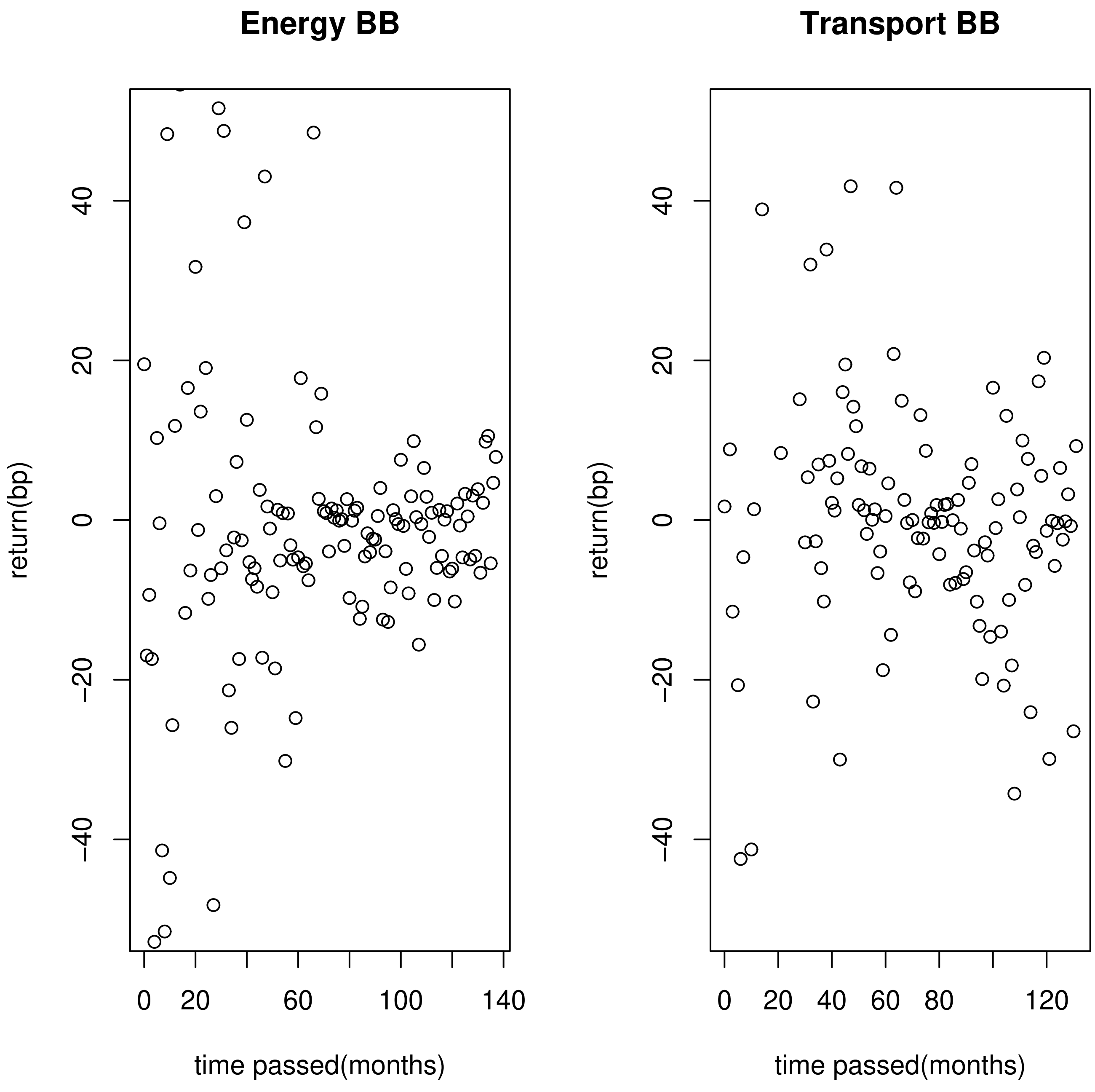

Figure 2: Monthly factor returns for two US bond factors 1991–2001. The horizontal axes measure months into the past, starting with the latest returns.

We can also compare the 95th percentile values of the two distributions, which we computed as 1.678 for T_u and 1.681 for $T(43)$. These are more than sufficiently close for purposes of statistical inference.

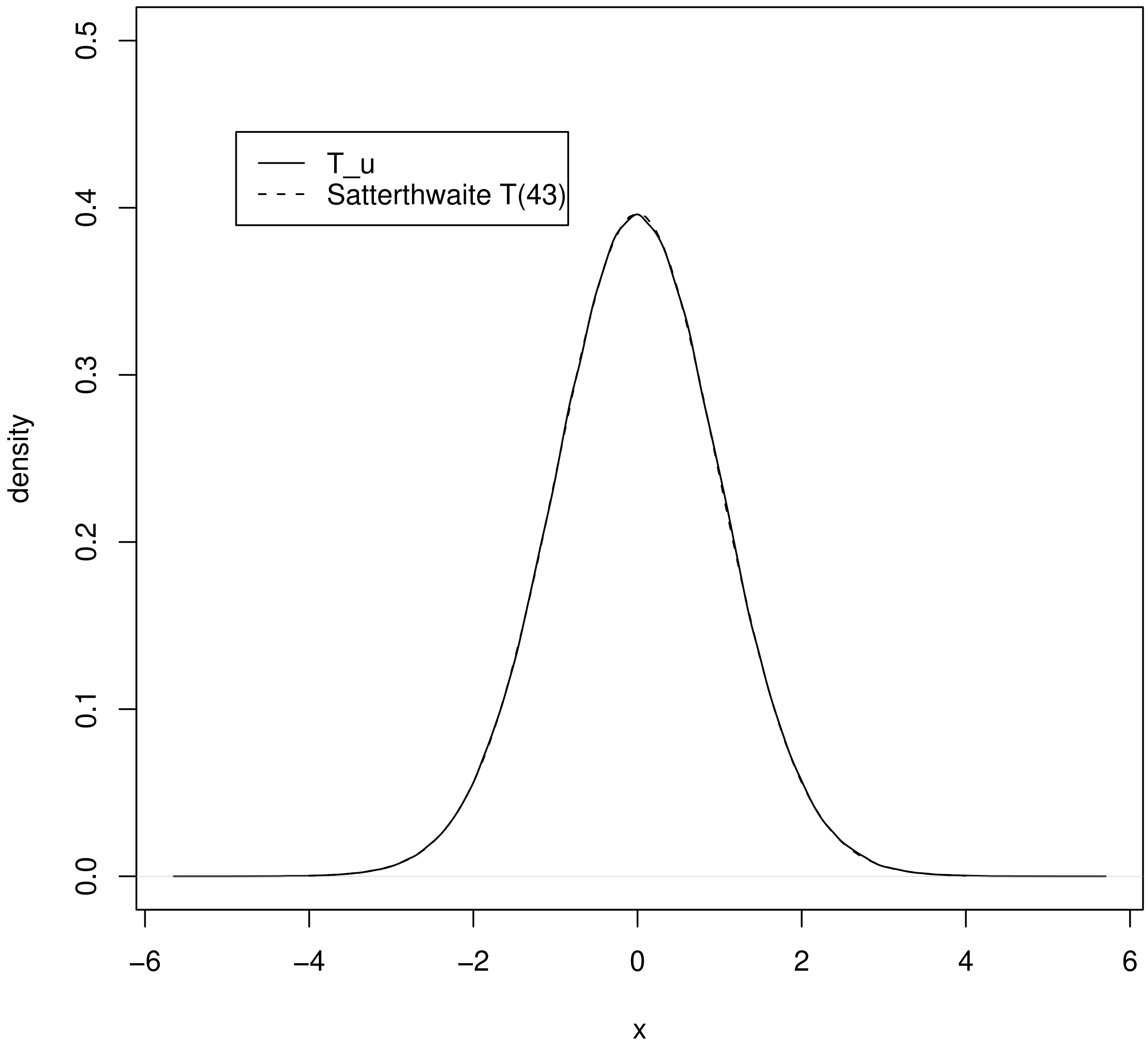

Figure 3: Comparison of densities for T_u and $T(43)$ using Sept 2000 data

In practice, for at least moderate sample sizes, practitioners may find it convenient to simply use T_u as if it followed a t-distribution with d.f. equal to $\min(m-1, n-1)$. This is no more difficult computationally than using the classical statistic (8).

We remark that Satterthwaite's approximate t-distribution method for the unpooled case is not the only way to handle the situation. There is a large literature on the Behrens-Fisher problem, see e.g. Scheffe (1970), Wang (1971), Yuen (1974).

A Appendix: Proofs

A.1 Proof of Theorem 1

(a) Let

$$W = \frac{\bar{X} - \bar{Y}}{\sqrt{\alpha_X + \alpha_Y}}$$

and

$$V = S_X/\alpha_X + S_Y/\alpha_Y.$$

By Lemma 2, W and V are independent, W is a standard normal random variable, and $V \sim \chi^2(n+m-2)$. Hence

$$T = \frac{W}{\sqrt{V/(n+m-2)}}$$

has the Student's t distribution with $n+m-2$ degrees of freedom.

(b) This part is proved in the discussion of Section 3.4.

(c) From equation 11 and the following discussion, we see that T_u is distributed as the ratio of a standard normal to an independent random variable

$$\sqrt{\frac{\hat{\alpha}_X + \hat{\alpha}_Y}{\alpha_X + \alpha_Y}}$$

whose variance tends to zero as $n, m \to \infty$. Hence the distribution of T_u tends to that of a standard normal, just as does a t-distributed random variable. ∎

A.2 Proof of Lemma 1

We consider only the μ_X and α_X cases; the proof is the same for μ_Y and α_Y.

First, suppose $\bar{X} = \sum w_i X_i$ is the UMVUE. Write $X_i = \mu + \epsilon_i$, where ϵ_i has mean 0 and variance σ_i^2. Since, by assumption, the variables ϵ_i are independent, the variance of $\bar{X}$ is given by

$$\begin{aligned} E[(\bar{X} - \mu)^2] &= E[(\textstyle\sum w_i e_i)^2] \\ &= \textstyle\sum E[w_i^2 \epsilon_i^2] \\ &= \textstyle\sum w_i^2 \sigma_i^2. \end{aligned}$$

It is straightforward to check via Lagrange multipliers that this quantity is minimized, under the constraint $\sum w_i = 1$, when

$$w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)}.$$

Equivalently,

$$\sigma_i^2 = \alpha / w_i \tag{13}$$

for all i and some positive constant α.

Conversely, suppose, for some positive α, that (13) holds for each i.

The joint p.d.f. of X_i's is given by

$$\begin{aligned} f(x_1, ..., x_n) &= \prod_{i=1}^{n} \sqrt{\frac{w_i}{2\pi\alpha_X}} \exp\{-\frac{1}{2}\frac{w_i}{\alpha_X}(x_i - \mu)^2\} \\ &= \exp\{\frac{1}{2}\sum_{i=1}^{n} \log(\frac{w_i}{2\pi\alpha_X}) - \frac{1}{2}\sum_{i=1}^{n}\frac{w_i}{\alpha_X}(x_i - \mu)^2\} \\ &= \exp\{-\frac{1}{2\alpha_X}\sum_{i=1}^{n} w_i x_i^2 + \frac{\mu}{2\alpha_X}\sum_{i=1}^{n} w_i x_i - \frac{\mu^2}{2\alpha_X} + \frac{1}{2}\sum_{i=1}^{n}\log(\frac{w_i}{2\pi\alpha_X})\}. \end{aligned}$$

This defines an Exponential family.(See Section 1.4 of Lehmann, 1983 for details.)

By Theorems 5.2 and 5.6 in Section 1.5 of Lehman (1983), $(\sum_{i=1}^{n} w_i x_i^2, \sum_{i=1}^{n} w_i x_i)$ become complete sufficient statistics for (μ_X, α_X).

Notice that

$$E(\bar{X}) = E(\sum_{i=1}^{n} w_i X_i) = \sum_{i=1}^{n} w_i \mu = \mu,$$

which implies $\bar{X}$ is an unbiased estimator of μ_X. Since $\bar{X}$ is a function of complete sufficient statistics, by Theorem 1.2 (ii) in Section 2.1 of Lehmann (1983), $\bar{X}$ is a UMVUE of μ_X.

A.3 Proof of Lemma 2

(1) $\bar{X} \sim N(\mu, \alpha_X)$ *and* $\bar{Y} \sim N(\mu, \alpha_Y)$.

This is a straightforward computation using the fact that a sum of independent normals is normal and variances add.

(2) $\bar{X}, \bar{Y}, S_X,$ *and* S_Y *are mutually independent.*

Clearly $\bar{X}$ and $\bar{Y}$ are independent, and similarly for S_X and S_Y. We show that $\bar{X}$ is independent of S_X, and the same argument works for Y. The argument is a direct generalization of the proof for the equal weighted case found, e.g., in Hogg and Craig (1995, ch. 4), which we include here for the reader's convenience. Write $\alpha = \alpha_X$ and denote the variance of X_i by ${\sigma_i}^2$ $(= \alpha/w_i)$. The joint pdf of $X_1, X_2, \ldots, X_n$ is

$$f(x_1,\ldots,x_n)=\frac{1}{(\prod_{i=1}^n\sqrt{2\pi}\sigma_i)}\exp[-\sum_{i=1}^n\frac{(x_i-\mu)^2}{2\sigma_i^2}]$$

Our strategy is to change variables in such a way that the independence of $\bar{X}$ and S_X will be evident.Letting $\bar{x}=\sum w_ix_i$, straightforward computation verifies that

$$\alpha=\frac{1}{\sum_{i=1}^n 1/\sigma_i^2}$$

and

$$\sum_{i=1}^n\frac{(x_i-\mu)^2}{\sigma_i^2}=\sum_{i=1}^n\frac{(x_i-\bar{x})^2}{\sigma_i^2}+(\bar{x}-\mu)/\alpha \tag{14}$$

Hence

$$f(x_1,\ldots,x_n)=\frac{1}{(\prod_{i=1}^n\sqrt{2\pi}\sigma_i)}\exp[-\sum_{i=1}^n\frac{(x_i-\bar{x})^2}{2\sigma_i^2}-\frac{(\bar{x}-\mu)^2}{2\alpha}] \tag{15}$$

Consider the linear transformation $(u_1,\ldots,u_n)=L(x_1,\ldots,x_n)$ defined by$u_1=\bar{x},u_2=x_2-\bar{x},\ldots,u_n=x_n-\bar{x}$, with inverse transformation

$$\begin{aligned}
x_1 &= u_1-(\frac{{\sigma_1}^2}{\sigma_2^2})u_2-(\frac{\sigma_1^2}{\sigma_3^2})u_3-\cdots-(\frac{\sigma_1^2}{\sigma_n^2})u_n,\\
x_2 &= u_1+u_2,\\
\ldots\\
x_n &= u_1+u_n
\end{aligned}$$

Likewise define new random variables $U_1=\bar{X},U_2=X_2-\bar{X},\ldots,U_n=X_n-\bar{X}$.If J denotes the Jacobian of L, then the joint pdf of $U_1,\ldots,U_n$ is

$$\frac{J}{(\prod_{i=1}^n\sqrt{2\pi}\sigma_i)}\exp[-\frac{(-(\frac{\sigma_1^2}{\sigma_2^2})u_2-(\frac{\sigma_1^2}{\sigma_3^2})u_3-\cdots-(\frac{\sigma_1^2}{\sigma_n^2})u_n)^2}{2\sigma_1^2}-\sum_{i=2}^n\frac{u_i^2}{2\sigma_i^2}-\frac{(u_1-\mu)^2}{2\alpha}]$$

This now factors as a product of the pdf of U_1 and the joint pdf of $U_2,\ldots,U_n$. Hence $U_1=\bar{X}$is independent of $U_2,\ldots,U_n$, and hence also independent of

$$\alpha[(-(\frac{\sigma_1^2}{\sigma_2^2})U_2-(\frac{\sigma_1^2}{\sigma_3^2})U_3-\cdots-(\frac{\sigma_1^2}{\sigma_n^2})U_n)^2+\sum_{i=2}^n\frac{U_i^2}{\sigma_i^2}]$$

$$= \alpha \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma_i^2} = S_X$$

(3) $S_X/\alpha_X \sim \chi^2(n-1)$ *and* $S_Y/\alpha_Y \sim \chi^2(m-1)$*, where* $\chi^2(k)$ *denotes the chi-squared distribution with* k *degrees of freedom.*

The proofs for X and Y are similar. Let

$$A = \sum_{1}^{n} \frac{(X_i - \mu_X)^2}{\sigma_i^2},$$

$$B = \sum_{1}^{n} \frac{(X_i - \bar{X})^2}{\sigma_i^2},$$

and

$$C = \frac{(\bar{X} - \mu_X)^2}{\alpha_X}.$$

Then by equation (14), $A = B + C$. Since $X_i \sim N(\mu_X, \sigma_i^2)$, $A \sim \chi^2(n)$. Similarly $C \sim \chi^2(1)$. This implies that $B = S_X/\alpha_X \sim \chi^2(n-1)$ provided that B and C are independent, which follows from the proof of part (i).

(4) As in the proof of Lemma 1,

$$E(\frac{1}{n-1} S_X) = \alpha_X$$

shows that $\frac{1}{n-1} S_X$ is an unbiased estimator of α_X. In addition,

$$\frac{1}{n-1} S_X = \frac{1}{n-1} \{\sum_{i=1}^{n} w_i x_i^2 - (\sum_{i=1}^{n} w_i x_i)^2\}$$

is a function of complete sufficient statistics, which implies $\frac{1}{n-1} S_X$ is a UMVUE of α_X.

A.4 Proof of Theorem 2

Recall that the $F(r_1, r_2)$ distribution is defined by $F = \frac{\frac{V_1}{r_1}}{\frac{V_2}{r_2}}$, where $V_1 \backsim \chi^2(r_1)$ and $V_2 \backsim \chi^2(r_2)$

are independent. By Lemma 2 and (9), we see that

$$\frac{\hat{\alpha_X}}{\alpha_X} = \frac{1}{n-1} \frac{S_X}{\alpha_X} \backsim \frac{\chi^2(n-1)}{n-1},$$

and

$$\frac{\hat{\alpha_Y}}{\alpha_Y} = \frac{1}{m-1}\frac{S_Y}{\alpha_Y} \backsim \frac{\chi^2(m-1)}{m-1}.$$

Since S_X and S_Y are independent, we have

$$\frac{\frac{\hat{\alpha_X}}{\alpha_X}}{\frac{\hat{\alpha_Y}}{\alpha_Y}} \backsim F(n-1, m-1). \tag{16}$$

If H_0 is true, then $\alpha_X = r\alpha_Y$ and (16) becomes

$$F = \frac{\hat{\alpha_X}}{r\hat{\alpha_Y}},$$

which follows $F(n-1, m-1)$ only when H_0 is true.

References

L. Breger, L. Goldberg, and A. Kercheval, "Modelling Credit Risk", *J. Portfolio Management*, 29, no. 2 (2003) 90–100.

Q.P.Duong and R.Shorrock, "On Behrens-Fisher solutions", *The Statistician*, 45, (1), 1996.

R. Grinold and R. Kahn, *Active Portfolio Management*, McGraw-Hill, 2000.

R.V. Hogg and A.T. Craig, *Introduction to Mathematical Statistics, 5th Edition*, Prentice-Hall, 1995.

E.L.Lehmann, *Theory of Point Estimation*, John Willey & Sons, Inc, 1983

Moore, David S. and G.P. McCabe, *Introduction to the Practice of Statistics*, 3rd Edition, Freeman, 1999.

F.E.Satterthwaite, "An approximate distribution of estimates of variance components", *Biometrics Bulletin*, 2, (6), Dec 1946

H. Scheffe, "Practical Solutions of Behren-Fisher Problem", *J. Amer. Stat. Assoc.*, Dec. 1970.

B.L. Welch, "The Significance of the Difference Between Two Means When the Population Variances are Unequal", *Biometrika*, 29 (1938), 29–48.

Y.Y. Wang, "Probabilities of Type I Errors of the Welch Tests for the Behrens-Fisher Problem", *J. Amer. Stat. Assoc.*, 66 (1971), 605–608.

K. K. Yuen, "The Two-sample Trimmed t for Unequal Population Variances", *Biometrika*, 61 (1974), 165–170.