Two-sample comparisons for serially correlated data

M. B. Seitshiro

Hons. B.Com.

Dissertation submitted in partial fulfilment of the requirements for the degree

Master of Science in Statistics

in the

School of Computer, Statistical and Mathematical Sciences

at the

North-West University (Potchefstroom Campus)

Supervisor:	**Prof. C.J. Swanepoel**
Co - supervisor:	**Prof. J.W.H. Swanepoel**

Potchefstroom

December 2006

TO MY FAMILY

ABSTRACT

The purpose of this study is to derive new tests for the equality of the means in two independent or dependent stationary time series, based on bootstrap critical values. Required properties of these tests include satisfactory probability of Type I errors, and high power. It is shown how critical points for various sample sizes and significance levels can be obtained by applying the parametric bootstrap. A limited Monte Carlo simulation study is conducted to illustrate the validity of the bootstrap approximation of the exact critical values, by producing satisfactory probability of Type I errors. It also shows that the newly proposed tests compare favourably with standard two-sample tests in the absence of serial correlation, under the null hypothesis of equal means, but are more powerful than the well-known $t-$test if small and moderate correlation structures are present, for a wide range of parameter values. All findings and conclusions of the Monte Carlo simulations are reported.

OPSOMMING

Die doel van hierdie studie is die ontwikkeling van nuwe twee-steekproeftoetse vir die vergelyking van twee gemiddeldes in twee onafhanklike óf afhanklike stasionêre tydreekse, deur gebruik te maak van skoenlus-gebaseerde kritieke waardes. Vooropgestelde eienskappe van die nuwe toetse sluit bevredigende waarskynlikhede op Tipe I foute in, asook hoë onderskeidingsvermoë. Daar word aangetoon hoe skoenlus-gebaseerde kritieke waardes bereken kan word deur gebruik te maak van die parametriese skoenlusmetode. 'n Beperkte Monte Carlo simulasiestudie word uitgevoer om aan te toon dat die skoenlusbenadering van die eksakte kritieke waardes bevredigende waarskynlikhede van Tipe I foute lewer. Dit is ook duidelik uit die studies dat die nuwe voorgestelde toetse gunstig vergelyk met standaard twee-steekproeftoeste in die afwesigheid van seriese korrelasie, onder die nulhipotese van gelyke gemiddeldes, maar dat die nuwe toetse oor hoër onderskeidingsvermoë beskik as die welbekende t – toets as klein en matige korrelasiestrukture teenwoordig is, oor 'n wye reeks parameterwaardes. Alle uitkomstes en gevolgtrekkings uit die Monte Carlo simulasies word gerapporteer.

PREFACE

The purpose of the study is to develop new tests for the equality of means in two independent or dependent stationary time series, based on bootstrap critical values, so that the new tests will have satisfactory probability of Type I errors and high power. Monte Carlo simulation studies are performed to illustrate the validity of the bootstrap approximation of the exact critical values and to determine the power of the newly proposed testing procedures.

Chapter 1 presents a brief review of popular two sample tests for independent and dependent samples. Chapter 2 describes basic concepts of the bootstrap methodology. In Chapter 3 a brief review of time series theory is presented. Chapter 4 introduces the newly developed test procedures. Chapter 5 gives the results of the Monte Carlo studies and conclusions.

In Chapter 1 the traditional parametric and non-parametric tests used for comparison of both independent and dependent data are discussed, regarding the means and medians of the populations concerned.

Chapter 2 describes important concepts regarding the bootstrap method, together with examples of the bootstrap applied to time series data and related fields, illustrating the skills that will be used in developing the new tests.

Chapter 3 summarizes time series theory and the concepts needed for developing the new tests, such as *autocovariance*, *autocorrelation function and partial autocorrelation function*, together with related examples, illustrating the concepts needed for the new tests.

The new test procedures are developed in Chapter 4, involving general formulations of time series, as well as the simple AR(1)-model. Terms such as *the bootstrap distribution of the test statistic*, *dependence/independence*, *equality of trends and serially correlated data* form part of this section, as it is applied in the new procedures.

In Chapter 5, the results of the Monte Carlo studies are discussed, together with the conclusions and recommendations. The first focus of the Monte Carlo studies is to examine the reliability of the bootstrap critical values of the tests concerned, with regard to the achieved probabilities of Type I errors. The performance of the bootstrap critical values is therefore also compared with the behaviour of standard two-sample tests described in Chapter 1. The final focus of the Monte Carlo studies centres around power calculations.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor, Prof. C.J. Swanepoel. Her vast experience, enthusiasm towards research and positive attitude have made it very inspiring to work under her guidance. I am deeply indebted to her for her encouragement and kind support, which have made this study possible.

I also wish to express my sincere gratitude towards Prof. J.W.H. Swanepoel, for his guidance and advice which made it possible to complete this study.

My warmest thanks go to my brother Dikgang Seitshiro and his wife Phindile Seitshiro and the rest of the family for always being there and for giving me all their support. I thank my parents Kabelo Seitshiro and Maele Seitshiro for their encouragement and belief in me and for letting me do things in my own way.

Finally, I would like to thank God, and express my deep appreciation for His help, guidance and blessings throughout my life.

TABLE OF CONTENTS

CHAPTER 2: BOOTSTRAP METHODOLOGY 28

CHAPTER 1

REVIEW OF TWO SAMPLE TESTS

During the past four decades fast and cheap computation became available, which enabled new procedures for application to areas in statistics which previously had to be considered impractical. It caused the replacement of theoretical analysis by computational methods. The new methods, however, have limitations as well, but should be evaluated for their usability when they fail. One such method is the bootstrap method, which was first introduced by Efron (1979). The bootstrap method will be introduced in Chapter 2, and will be used to develop new two-sample test procedures. However, a review of standard methods will first be presented below.

In Chapter 1 below a brief review of popular two sample tests is given. These tests are generally used in statistical practice, tests for independent and dependent samples are discussed.

Section 1.1 starts off with parametric and nonparametric two sample tests for the equality of means and medians. In Section 1.2 parametric two sample tests for the equality of variances for both the independent and dependent cases are discussed respectively. In Section 1.3 two sample tests for the equality of distributions are considered, for two independent samples and for two dependent samples. In Section 1.4 standard ways of dealing with problems encountered in the data are discussed, and in Section 1.5 a brief summary on the extension of two sample tests to more dimensions is given.

1.1 Two sample tests for the equality of means and medians

According to Sheskin (2000:245-508), one field of application of the two-sample *t*-test is to determine if two parameters, such as two population means, are equal or not. For example, this test is needed for verifying the equality of the mean production of a factory that is using a new process, to the mean production when an old process is used, or to find out if a new procedure is superior to a current process.

We now consider comparison tests in the *parametric* set up, applied to two *independent* samples.

1.1.1 Parametric tests for two independent samples

In the following section we discuss two parametric tests, i.e. the *z*-test and the *t*-test for two independent samples.

1.1.1.1 The *z*-test for two independent samples

In some occasions one would want to compare the means of two independent samples if the variances of the two underlying populations are *known.* In such situations the *z*-test for two independent samples should be employed, if normally distributed populations are available for sampling. The *z*-test assumes that the two samples are randomly selected from populations with normal distributions, with known variances, in an independent way. If large samples are taken, the effect on the test statistic if the normality assumption is violated, is concealed. For the *z*-test, the assumption of homogeneity of variance is not needed.

Regarding hypotheses for the population means, Sheskin (2000:248) states the following elementary hypotheses to consider which will be used throughout this chapter:

- The null hypothesis is $H_0 : \mu_X - \mu_Y = 0$.

One of the following alternative hypotheses is suggested.

- $H_1 : \mu_X - \mu_Y \neq 0$
- $H_1 : \mu_X - \mu_Y > 0$
- $H_1 : \mu_X - \mu_Y < 0$.

The *z*-test statistic, for two independent samples, is applicable when the number of subjects in each sample is either equal or unequal. If the two samples are denoted by $\mathbf{X} = \left\{X_1, X_2, \ldots, X_{n_X}\right\}$ and $\mathbf{Y} = \left\{Y_1, Y_2, \ldots, Y_{n_Y}\right\}$, with the sample sizes denoted by n_X and n_Y respectively, then the *z*-statistic is defined by:

$$z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}, \tag{1.1}$$

where σ_X^2 and σ_Y^2 denote the underlying population variances with the sample means denoted by $\bar{X} = n_X^{-1}\sum_{i=1}^{n_X} X_i$ and $\bar{Y} = n_Y^{-1}\sum_{j=1}^{n_Y} Y_j$ respectively. If $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, (1.1) simplifies to:

$$z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}. \tag{1.2}$$

The z critical value is evaluated with normal quantiles, because the z-test statistic follows a $N(0,1)$ distribution. The following decision rules are applied:

a) If the nondirectional alternative hypothesis $H_1 : \mu_X - \mu_Y \neq 0$ is of concern, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $|z| > z(\alpha/2)$ at the specified α-level of significance, where $z(\alpha/2)$ represents the tabled critical two-tailed value in the z-distribution.

b) If the directional alternative hypothesis $H_1 : \mu_X - \mu_Y > 0$ is of concern, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $z > z(\alpha)$, where the value of z should be positive and $z(\alpha)$ denotes the $(1-\alpha)^{\text{th}}$ quantile of $N(0,1)$ distribution.

c) Similarly, if for the alternative hypothesis $H_1 : \mu_X - \mu_Y < 0$, $H_0 : \mu_X - \mu_Y = 0$ is rejected if the value z is negative and $|z| > z(\alpha)$.

Confidence intervals for $\mu_X - \mu_Y$ can be computed to identify a range of values within which one can be confident to a specified degree that the true difference lies within the region specified by the interval, by using the normal percentiles. This confidence interval is specified by:

$$\left(\bar{X} - \bar{Y}\right) \pm z\left(\alpha/2\right)\sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}. \tag{1.3}$$

Since σ^2 is generally not known, it is estimated from the sample data, by calculating the pooled sample variance:

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}, \tag{1.4}$$

where $S_X^2 = \dfrac{\sum_{i=1}^{n_X} X_i^2 - \left(\left(\sum_{i=1}^{n_X} X_i\right)^2 \Big/ n_X\right)}{n_X - 1}$ and $S_Y^2 = \dfrac{\sum_{j=1}^{n_Y} Y_j^2 - \left(\left(\sum_{i=1}^{n_Y} Y_j\right)^2 \Big/ n_Y\right)}{n_Y - 1}$ are the sample variances of the two samples $\{X_i, i = 1, 2, ..., n_X\}$ and $\{Y_j, j = 1, 2, ..., n_Y\}$ respectively. Then the $t_{n_X+n_Y-2}$-distribution should be used, unless n_X and n_Y are large, as is explained in the next paragraph.

1.1.1.2 The *t*-test for two independent samples

The *t*-test for two independent samples is a parametric test to determine whether the two independent samples represent two populations with different mean values, if the sample sizes available are small. The *t*-test for two independent samples is based on the following assumptions:

- Each sample has been randomly selected from the population it represents.
- The distribution of data in the underlying population distribution from which each of the samples is obtained, is normal.
- Homogeneity of variance is also assumed.

According to Rice (1995:389-400), the *t*-test statistic is applicable when the number of subjects in each sample is either equal or unequal, and the values of the underlying population variances are unknown. The *t*-test statistic is defined by:

$$t = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}, \qquad (1.5)$$

where S_p denotes the pooled sample standard deviation, which is the square root of pooled variance formulated in (1.4) above. The denominator of the *t*-test statistic for two independent samples is referred to as the standard error of the difference, and is denoted by $S_{\bar{X}-\bar{Y}}$.

When using the *t*-test statistic, the following results are valid for evaluating the null hypothesis:

a) If the alternative hypothesis $H_1 : \mu_X - \mu_Y \neq 0$ is employed, the decision rule is that $H_0 : \mu_X - \mu_Y = 0$ is rejected if $|t| > t_{n_X+n_Y-2}(\alpha/2)$ at the specified α-level of significance, where $t_n(\alpha/2)$ represents the tabled two-tailed value in the t-distribution with n degrees of freedom.

b) If the alternative hypothesis $H_1 : \mu_X - \mu_Y > 0$ is of interest, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $t > t_{n_X+n_Y-2}(\alpha)$, whereby the value of t should be positive.

c) If the alternative hypothesis $H_1 : \mu_X - \mu_Y < 0$ is of concern, $H_0 : \mu_X - \mu_Y = 0$ is rejected if the value of t is negative and $|t| > t_{n_X+n_Y-2}(\alpha)$.

The confidence interval corresponding to the t-test for the difference between the two population means ($\mu_X - \mu_Y$) is denoted by:

$$(\bar{X} - \bar{Y}) \pm t(\alpha/2) S_{\bar{X}-\bar{Y}}, \qquad (1.6)$$

where $S_{\bar{X}-\bar{Y}} = S_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$ and $t(\alpha/2)$ represents the tabled critical two-tailed value of the t-distribution, for $n_X + n_Y - 2$ degrees of freedom, above which a proportion $(1-\alpha/2)$ of the distribution falls. This interval contains the possible differences between μ_X and μ_Y, for which $H_0 : \mu_X - \mu_Y = 0$ will not be rejected at α-level of significance, if the interval includes zero.

We have used the assumption that the two populations have the same variance. If the two variances are not assumed to be equal, a natural estimate of $\mathrm{Var}(\bar{X} - \bar{Y})$ is $\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}$, according to Rice (1995: 395). By using this expression in the denominator of the t statistic, causes the test statistic to have a distribution which is no longer a t-distribution. This distribution, however, can be closely approximated by a t-distribution with degrees of freedom calculated by the following expression, and rounded to the nearest integer:

$$df = \frac{\left[\left(S_X^2 / n_X\right) + \left(S_Y^2 / n_Y\right)\right]^2}{\frac{\left(S_X^2 / n_X\right)^2}{n_X - 1} + \frac{\left(S_Y^2 / n_Y\right)^2}{n_Y - 1}}.$$

1.1.2 Nonparametric test for two independent samples

In this Section we summarise a nonparametric test for testing the difference of means for two independent samples. This test is referred to as the Mann – Whitney *U* test. Nonparametric tests are also referred to as distribution-free tests. These tests have the advantage of not requiring population assumptions such as the assumption of normality or homogeneity of variance. The nonparametric tests are more robust, i.e., if the data have one or two outliers for example, the influence of the outliers is restricted.

1.1.2.1 Mann - Whitney *U* test

The Mann – Whitney *U* test is a nonparametric test executed on rank order data, involving a design with two independent samples. The test is used to determine measures of effect usually defined in terms of the median values, denoted by θ in the discussions below. The Mann – Whitney *U* test is based on the following assumptions:

- Each sample has been randomly selected from the population it represents.
- The two samples are independent of one another.
- The original variable observed is a continuous random variable.
- The underlying distributions from which the samples are derived are identical in shape.

The hypotheses of interest for illustrating the Mann – Whitney *U* test, comprises of the following elementary cases suggested by Sheskin (2000:290-291):

- The null hypothesis is $H_0: \theta_1 = \theta_2$, where θ_i denotes the median of the i^th^ population. This involves the quantities $\sum_{i=1}^{n_X} R_i = \sum_{j=1}^{n_Y} R_j$ obtained from the samples, where $n_X \equiv$ the first sample size and $n_Y \equiv$ the second sample size, and R_i and R_j denote the ranks given to the jointly ranked observations in the first and second samples, $i = 1, 2, \ldots, n_X$ and $j = 1, 2, \ldots, n_Y$.

If one of the following alternative hypotheses is supported by the data, then the null hypothesis will not be supported. Furthermore, if $n_X = n_Y = n$, the following intuitive argument illustrates the relationship between the possible alternative hypotheses and ranks given to the two sample's elements:

- $H_1 : \theta_1 \neq \theta_2$: With respect to the sample data, we expect that $\sum_{i=1}^{n} R_i \neq \sum_{j=1}^{n} R_j$.

- $H_1 : \theta_1 > \theta_2$: With respect to the sample data, we expect that $\sum_{i=1}^{n} R_i > \sum_{j=1}^{n} R_j$.

- $H_1 : \theta_1 < \theta_2$: With respect to the sample data, we expect that $\sum_{i=1}^{n} R_i < \sum_{j=1}^{n} R_j$.

The guidelines for ranking the values of two independent samples are as follows:

(a) The values of both the samples are ranked together.

(b) A rank of 1 is assigned to the lowest value in the joint sample, if there are no ties. Then a rank of 2 is assigned to the second lowest value, and so on, until the highest value is assigned a rank equal to N (where, $N = n_X + n_Y$).

(c) When there are tied (similar) values present, the average of the ranks involved is assigned to all tied values.

If we set $U_1 = n_X n_Y + \frac{n_X(n_X+1)}{2} - \sum_{i=1}^{n_X} R_i$ and $U_2 = n_X n_Y + \frac{n_Y(n_Y+1)}{2} - \sum_{j=1}^{n_Y} R_j$, (1.7)

where $\sum_{i=1}^{n_X} R_i$ and $\sum_{j=1}^{n_Y} R_j$ are the sum of the ranks for the two samples respectively, then the value of $U = \min(U_1, U_2)$ is referred to as the obtained Mann - Whitney U statistic.

The null hypothesis is evaluated with a table of critical values for the Mann – Whitney U statistic at the specified α -level of significance, with the following results:

(a) To support the alternative hypothesis $H_1 : \theta_1 \neq \theta_2$, the condition that $U < u_{n_X, n_Y}(\alpha/2)$ should hold at α -level of significance, irrespective of whether $\sum_{i=1}^{n_X} R_i > \sum_{j=1}^{n_Y} R_j$ or $\sum_{i=1}^{n_X} R_i < \sum_{j=1}^{n_Y} R_j$, where $u_{n_X, n_Y}(\alpha/2)$ is the lower $\alpha/2$ critical value for the Mann – Whitney U distribution with degrees of freedom n_X and n_Y.

(b) For the alternative hypothesis $H_1 : \theta_1 > \theta_2$ to be supported, the conditions that $\overline{R}_1 > \overline{R}_2$ and $U < u_{n_X, n_Y}(\alpha)$ should hold, where $\overline{R}_1 = \left(\sum_{i=1}^{n_X} R_i \middle/ n_X \right)$ and

$\bar{R}_2 = \left(\sum_{j=1}^{n_Y} R_j \Big/ n_Y \right)$ are the means of the ranks from the first and the second samples respectively.

(c) For the alternative hypothesis $H_1 : \theta_1 < \theta_2$ to be supported, both the conditions that $\bar{R}_1 < \bar{R}_2$ and $U < u_{n_X, n_Y}(\alpha)$ should hold.

Regarding the power efficiency of the Mann – Whitney U test, when the underlying population distributions are normal, the asymptotic relative efficiency of the Mann – Whitney U test can be constructed with regard to the t-test for two independent samples. For population distributions that are not normal, the asymptotic relative efficiency is generally equal to or greater than unity. Another nonparametric rank-order procedure for evaluating a design involving two independent samples is a computer intensive test based on the Bootstrap Method. A bootstrap test is illustrated in Chapter 2 below. We will now turn to tests developed for the case when the two samples are dependent.

1.1.3 Parametric test for two dependent samples

The dependent data imply that there is a one-to-one correspondence between the values in the two samples. That is, if $\{X_i, i = 1,2,...,n\}$ and $\{Y_i, i = 1,2,...,n\}$ are the two samples, then $\{X_i\}$ corresponds to $\{Y_i\}$ in some way. According to Rice (1995:410-413), for paired samples, the analysis is based on the difference $\{D_i = X_i - Y_i,\ i = 1,2,...,n\}$. In some applications, one may want to adopt a new process only if it exceeds the other current process by a specified amount, c. In this case, the null hypothesis can be stated to indicate that the difference between the two population means is equal to some constant ($\mu_X - \mu_Y = c$, where $c \equiv$ some constant). In this Section we study a parametric test for two dependent samples, referred to as the t-test for two dependent samples.

1.1.3.1 The *t*-test for two dependent samples

The parametric t-test for two dependent or paired samples is another inferential statistical test that is based on the t-distribution. According to Snedecor and Cochran (1980:83–89), the paired t-test is conducted when, for given paired samples (X_i, Y_i), $i = 1,2,...,n$, the means of the two samples (denoted by $\bar{X}$ and $\bar{Y}$ as before), are used to estimate the values of the population means μ_X and μ_Y, and it is assumed that the values of the population variances, σ_X^2 and σ_Y^2 are unknown. The t-test for two dependent samples is based on the following assumptions:

- The sample of n subjects has been randomly selected from the population it represents.
- The distributions of the data in the underlying populations are normal.

Regarding the hypotheses for the population means, the following elementary cases are considered:

- The null hypothesis of no effect is $H_0 : \mu_X - \mu_Y = 0$.
- The possible alternative hypotheses are similar to those defined for the independent case.

The t-test statistic for two dependent samples is referred to as the direct-difference equation (Sheskin, 2000:434-442) and it is defined by:

$$t = \frac{\bar{D}}{S_{\bar{D}}}, \tag{1.8}$$

where $\bar{D} = \bar{X} - \bar{Y}$ denotes the mean of the difference between the first sample and the second sample and $S_{\bar{D}}$ denote the standard error of the mean difference. One way of calculating $S_{\bar{D}}$ is by first calculating $\tilde{S}_D$, which is the estimated population standard deviation of the difference defined by:

$$\tilde{S}_D = \sqrt{\frac{\sum_{i=1}^{n} D_i^2 - \left(\sum_{i=1}^{n} D_i\right)^2 \Big/ n}{n-1}}, \tag{1.9}$$

where n is the number of paired subjects, and $D_i = X_i - Y_i$ with $i = 1, 2, \ldots, n$.

The standard error of the mean difference is then defined by:

$$S_{\bar{D}} = \frac{\tilde{S}_D}{\sqrt{n}}. \tag{1.10}$$

The following decision rules are applicable for the evaluation of the null hypothesis:

(a) If the alternative hypothesis $H_1 : \mu_X - \mu_Y \neq 0$ is employed, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $|t| > t_{n-1}(\alpha/2)$, where $t_{n-1}(\alpha/2)$ is the upper $\alpha/2$ critical value of the t-distribution, with $n-1$ degrees of freedom and α-level of significance.

(b) For the alternative hypothesis $H_1 : \mu_X - \mu_Y > 0$, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $t > t_{n-1}(\alpha)$.

(c) For the alternative hypothesis $H_1 : \mu_X - \mu_Y < 0$, $H_0 : \mu_X - \mu_Y = 0$ is rejected if $t < -t_{n-1}(\alpha)$.

According to Rice (1995:450-451), the confidence interval for $E(\bar{D}) = \mu_X - \mu_Y = \mu_D$ for two dependent samples can be computed for $\mu_X - \mu_Y$ as follows:

$$\begin{aligned}
&(\bar{X}-\bar{Y}) \pm t_{n-1}(\alpha/2)\frac{1}{\sqrt{n}}\left(\sqrt{\frac{\sum_{i=1}^{n} D_i^2 - \left(\sum_{i=1}^{n} D_i\right)^2 \Big/ n}{n-1}}\right) \\
&= (\bar{X}-\bar{Y}) \pm t_{n-1}(\alpha/2)\left(\frac{s_D}{\sqrt{n}}\right) \\
&= \bar{D} \pm t_{n-1}(\alpha/2)(s_{\bar{D}}). \qquad (1.11)
\end{aligned}$$

1.1.4 Nonparametric test for two dependent samples

The nonparametric test for two dependent samples is referred to as the Wilcoxon matched-pairs signed-ranks test.

1.1.4.1 Wilcoxon matched-pairs signed-ranks test

The Wilcoxon matched-pairs signed-ranks test is a nonparametric test that is often regarded as being similar to a matched-pairs *t*-test. The Wilcoxon matched-pairs signed-ranks test is used to test for differences between groups of paired data when the data do not meet the assumptions associated with parametric tests. It is a nonparametric procedure employed in a hypothesis testing situation involving a design with two dependent samples. According to Sheskin (2000:467-476), we wish to evaluate a hypothesis with a matched-pairs test to determine whether or not, in the underlying populations represented by the samples, a measure of difference in the location of the population equals zero. The Wilcoxon matched-pairs signed-ranks test is defined in terms of the medians of the populations, represented by θ_1 and θ_2 for the two populations. If a significance difference is indicated, it implies that it is very possible that the two samples represent two different populations. The Wilcoxon matched-pairs signed-ranks test determines whether or not the medians differ or not and is based on the following assumptions:

a) The sample of n subjects has been randomly selected from the population it represents.
b) The distribution of the difference in the populations represented by two samples is symmetric about the median of the population of the differences.

Let $\{(X_i, Y_i),\ i = 1,2,...,n\}$ represent pairs of subjects, then $\{D_i = X_i - Y_i,\ i = 1,2,...,n\}$ is the difference between the first sample and the second sample. Let R_i represent the rank of the absolute value of the i^{th} sample difference. Let θ_D represent the population median of the differences.

The null hypothesis $H_0 : \theta_D = 0$ implies that the difference of the population medians is zero. With respect to the sample data, this means that, if H_0 is true, we can expect that $\sum_{k=1}^{m_1} R_k + = \sum_{l=1}^{m_2} R_l -$, where $R_j +$ denotes the rank number of the j^{th} positive difference, $R_l -$ denotes the rank number of the l^{th} negative difference, $m_1 \equiv$ the number of positive differences and $m_2 \equiv$ the number of negative differences, such that $m_1 + m_2 = n$. $\sum_{l=1}^{m_2} R_l -$ therefore denotes the sum of the ranks of the absolute values of the differences with negative sign, and $\sum_{k=1}^{m_1} R_k +$ is similarly defined for differences with positive sign.

One of the following alternative hypotheses is applicable: $H_1 : \theta_D \neq 0$, $H_1 : \theta_D > 0$ or $H_1 : \theta_D < 0$. If the chosen alternative hypothesis is supported by the data, then the null hypothesis is rejected. Again, the signed rank numbers given to the differences D_i, can heuristically be associated with the alternative hypothesis, in the following way:

- $H_1 : \theta_D \neq 0$: With respect to the sample data, we expect that $\sum_{k=1}^{m_1} R_k + \neq \sum_{l=1}^{m_2} R_l -$.

- $H_1 : \theta_D > 0$: With respect to the sample data, we expect that $\sum_{k=1}^{m_1} R_k + > \sum_{l=1}^{m_2} R_l -$.

- $H_1 : \theta_D < 0$: With respect to the sample data, we expect that $\sum_{k=1}^{m_1} R_k + < \sum_{l=1}^{m_2} R_l -$.

The Wilcoxon test gives the following rules that should be adhered to when ranking the difference for the Wilcoxon matched-pairs signed-ranks test:

- The absolute values of the differences $\{|D_i|, i = 1, 2, ..., n\}$ are ranked.
- Any difference that is equal to zero is not ranked.
- When there are tied values in the differences, the average of the ranks involved is assigned to all values tied for a given rank.
- A rank of 1 should be assigned to the differenced value with the lowest absolute value, and a rank of n should be assigned to the differenced value with the highest absolute value.

$$\left(\sum_{k=1}^{m_1} R_k +\right) + \left(\sum_{l=1}^{m_2} R_l -\right) = \frac{n(n+1)}{2}. \tag{1.12}$$

The latter equation allows one to check the accuracy of the values $\sum_{k=1}^{m_1} R_k +$ and $\sum_{l=1}^{m_2} R_l -$. If the relationship indicated by (1.12) is not obtained, it indicates that an error has been made in the calculations.

- After the ranks have been given, the sign of the difference is attached to the rank number.

The value $W = \min\left(\left|\sum_{k=1}^{m_1} R_k +\right|, \left|\sum_{l=1}^{m_2} R_l -\right|\right)$ is designated as the Wilcoxon matched-pairs signed-ranks test statistic, with the following results used to evaluate the null hypothesis:

(a) To support $H_1 : \theta_D \neq 0$, the Wilcoxon test statistic (W) should be less or equal to the tabled critical two tailed value at α -level of significance (i.e., $W < W_n(\alpha/2)$, where $W_n(\alpha/2)$ is the lower $\alpha/2$ critical value from the table of the Wilcoxon matched-pairs signed-ranks test), irrespective of whether $\sum_{k=1}^{m_1} R_k + < \sum_{l=1}^{m_2} R_l -$ or $\sum_{k=1}^{m_1} R_k + > \sum_{l=1}^{m_2} R_l -$.

(b) To support $H_1 : \theta_D > 0$, the conditions that $W < W_n(\alpha)$ and $\sum_{k=1}^{m_1} R_k + > \sum_{l=1}^{m_2} R_l -$ must hold.

(c) To support $H_1 : \theta_D < 0$, the conditions that $W < W_n(\alpha)$ and $\sum_{k=1}^{m_1} R_k + < \sum_{l=1}^{m_2} R_l -$ must be satisfied.

For large sample sizes, the sampling distribution is approximately normally distributed with $\mathrm{E}(W) = \left(n(n+1) \big/ 4 \right)$ and $\mathrm{Var}(W) = \left(n(n+1)(2n+1) \big/ 24 \right)$.

The normal approximation of the Wilcoxon statistic is then given by:

$$z = \frac{W - \mathrm{E}(W)}{\sqrt{\mathrm{Var}(W)}}. \tag{1.13}$$

The z value is evaluated with a table of the normal distribution with the following results:

(a) As before, the nondirectional alternative hypothesis, $H_0 : \theta_D = 0$ is rejected if $|z| \geq z(\alpha/2)$.

(b) For a directional alternative hypothesis, $H_0 : \theta_D = 0$ is rejected if $|z| \geq z(\alpha)$.

Remarks regarding hypothesis testing:

- **p-value:** The p-value is the estimated probability that the test statistic is at least as extreme as the one that is observed, given that the null hypothesis is true (i.e., it is the probability of wrongly rejecting the null hypothesis if it is in fact true). A small p-value is an indication that the null hypothesis is likely to be false.
- **Power:** The power of a test is the estimated probability of rejecting the null hypothesis when it is actually false. Because the estimated probability of failing to reject a false null hypothesis is defined as Type II error and denoted by β, the power of a test can be denoted by $1 - \beta$. Rice (1995:400) states that the power of

the test depends on the size of the difference $|\mu_X - \mu_Y|$, the significance level α, the population standard deviation σ, and the sample size.

- **Effect size:** It is important to determine how large sample sizes should be to ensure great power of the test. According to Sheskin (2000:260-263) the equation below defines effect size (*d*).

$$d = \frac{|\mu_X - \mu_Y|}{\sigma}. \tag{1.14}$$

This index was developed by Cohen (1977, 1988) and was discussed by Howell (1992:204–217) and is known as Cohen's *d* index. The numerator represents the hypothesized difference between the two population means. This expression can be estimated. The mean difference $(\bar{X} - \bar{Y})$ is employed as an estimate of $(\mu_X - \mu_Y)$ and $\tilde{S}_D$ denotes the pooled sample standard deviation previously defined, which is an estimator of σ. Since the effect size *d* is based on population parameters, it is necessary to convert the value of effect size *d* into a measure that takes into account the size of the sample.

The measure δ is referred to as the non-centrality parameter, and is defined by:

$$\delta = d\sqrt{\frac{n}{2}}, \tag{1.15}$$

if the sample sizes are equal (i.e., $n = n_X = n_Y$). The value δ is evaluated with the power curves for student's *t*-distribution. When the sample sizes are unequal, the value of n will be represented by the harmonic mean of the sample size denoted by:

$$\bar{n}_h = \frac{2}{1/n_1 + 1/n_2}. \tag{1.16}$$

The proposed *d* values as criteria for identifying the magnitude of effect size are:

- if $0.2 < d < 0.5$, small effect is indicated,
- if $0.5 < d < 0.8$, medium effect is indicated,
- if $d > 0.8$, large effect is indicated.

The general rule is either to look for big effects or to use large samples.

1.2 Two sample tests for the equality of variances

In this section we look at the parametric test for equality of variances for two independent and dependent sample tests.

1.2.1 Parametric test for two independent samples

According to Sheskin (2000:253–258), a parametric test in this case for homogeneity of variance, determines whether there is evidence that inequality exist between the variances of the populations represented by the two independent samples.

1.2.1.1 Homogeneity of variance for two independent samples

Usually, methods of calculating confidence limits and executing tests of significance for the difference between the means of two independent samples, are based on the assumption that the two population variances are the same. The hypotheses of interest are:

- $H_0 : \sigma_X^2 = \sigma_Y^2$. The alternative hypotheses are:
- $H_1 : \sigma_X^2 \neq \sigma_Y^2$, which is evaluated with a two–tailed test.
- $H_1 : \sigma_X^2 > \sigma_Y^2$, which is evaluated with a one–tailed test.
- $H_1 : \sigma_X^2 < \sigma_Y^2$, which is also evaluated with a one–tailed test.

Hartley's $F_{\max}$ test (Hartley (1940,1950)) is one of the applicable procedures and it is applied if the two samples are independent.

The so-called $F_{\max}$-statistic is defined by:

$$F_{\max} = \frac{\tilde{S}_L^2}{\tilde{S}_S^2}, \tag{1.17}$$

where $\tilde{S}_L^2$ and $\tilde{S}_S^2$ are the largest and the smallest of the two estimated population variances (unbiased) respectively.

The $F_{\max}$-statistic for two population variances is associated with the F-distribution. The degrees of freedom for the numerator of the F-ratio is $v_L = n_L - 1$ and for the denominator is $v_S = n_S - 1$, where n_L and n_S represent the sample sizes in the groups with larger and smaller estimated population variances, respectively. Table A9 in

Sheskin only listed two-tailed values to be used for critical points and the null-hypothesis is rejected if $F_{\max}$ exceeds the critical value.

If we refer to the two samples as $\mathbf{X}$ and $\mathbf{Y}$, the test statistic can be defined as $F = \frac{\tilde{S}_X^2}{\tilde{S}_Y^2}$ in terms of the samples and the level $\left(\alpha/2\right)$ instead of α must be used in Table A10 of Sheskin for two sided alternatives. Table A10 is the usual F-table of critical values. Denote the samples size of $\mathbf{X}$ and $\mathbf{Y}$ by n_X and n_Y, respectively, as before. Also, the degrees of freedom are $v_X = n_X - 1$ and $v_Y = n_Y - 1$. Then the tabled $F_{1-\alpha}$ value is used to evaluate directional alternative hypothesis at α-level of significance. The following general decision rules are applicable:

a) For the alternative hypothesis $H_1 : \sigma_X^2 \neq \sigma_Y^2$, in order to reject $H_0 : \sigma_X^2 = \sigma_Y^2$, F must be greater than the tabled upper critical value at level $(\alpha / 2)$ or smaller than the lower critical value at level $(\alpha / 2)$.

b) For the alternative hypothesis $H_1 : \sigma_X^2 > \sigma_Y^2$, the condition that F is greater than the upper critical value at level α must hold, in order to reject the null hypothesis.

c) For the alternative hypothesis $H_1 : \sigma_X^2 < \sigma_Y^2$, the condition that F is less than the lower critical value at level α must hold, in order to reject the null hypothesis.

1.2.2 Parametric test for two dependent samples

We will now discuss a test for the homogeneity of variance for two dependent samples, according to Sheskin (2000:442–445).

1.2.2.1 Homogeneity of variance for two dependent samples

The following hypotheses are of interest:

- The null hypothesis is $H_0 : \sigma_X^2 = \sigma_Y^2$.
- The alternative hypotheses applied for the independent case previously will be the same as for the dependent case.

The *t*-statistic will be used and is defined by:

$$t=\frac{\left(\tilde{S}_X^2-\tilde{S}_Y^2\right)\sqrt{n-2}}{\sqrt{4\tilde{S}_X^2\tilde{S}_Y^2\left(1-r_{XY}^2\right)}}, \tag{1.18}$$

where $\tilde{S}_X^2$ and $\tilde{S}_Y^2$ are the values of the first and second estimated population variances respectively. r_{XY} below is the estimated population correlation coefficient between matched pairs of subjects on the dependent variables.

$$r_{XY}=\frac{\sum_{i=1}^{n}X_iY_i-\left(\sum_{i=1}^{n}X_i\right)\left(\sum_{i=1}^{n}Y_i\right)\Big/n}{\sqrt{\left(\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2\Big/n\right)\left(\sum_{i=1}^{n}Y_i^2-\left(\sum_{i=1}^{n}Y_i\right)^2\Big/n\right)}}. \tag{1.19}$$

An alternative but equivalent form of the t-test statistic for homogeneity of variance for two dependent samples is denoted by:

$$t=\frac{(F-1)\sqrt{n-2}}{2\sqrt{F\left(1-r_{XY}^2\right)}}, \tag{1.20}$$

where $F=\frac{\tilde{S}_X^2}{\tilde{S}_Y^2}$. The degree of freedom associated with the test statistic is $n-2$.

The decision rules for the dependent case using the t-distribution are equivalent to those in Section 1.1.3.1.

1.3 Two sample tests for the equality of distributions

In this section we look at parametric and nonparametric tests for equality of distributions for two independent and dependent samples. The usual chi-square test for n rows and c columns can be applied to test if the null hypotheses formulated in terms of distributions, should be rejected or not.

1.3.1 Nonparametric test for two independent samples

One nonparametric test for two independent samples is referred to as the Kolmogorov-Smirnorv test.

1.3.1.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test for two independent samples is a nonparametric test, which is meant to determine whether the two independent samples represent two different populations. The test is categorised as a test of at least ordinal data involving cumulative frequency distributions, i.e. which requires scores in each distribution, arranged in order of magnitude. When the two-tailed alternative hypothesis is evaluated, the Kolmogorov-Smirnorv test for two independent samples becomes sensitive to any kind of distribution difference. According to Sheskin (2000:319–328), the Kolmogorov-Smirnov test for two independent samples compares the cumulative frequency distributions of two independent samples. The test protocol is based on the principle that if there is a significant difference at any point along the two cumulative frequency distributions, then there is a high likelihood that the samples are derived from different populations. The Kolmogorov-Smirnov test assumes the following:

(a) All of the observations in the two samples are randomly selected and they are independent of one another.

(b) The scale of measurement is at least ordinal.

The construction of the hypotheses for the Kolmogorov-Smirnov test requires that a cumulative probability distribution should be computed for each of the samples. We will now adapt to the notation of Sheskin (2000: 320–325). $F_j(X)$ represents the population distribution function from which the j^{th} sample is drawn, $j = 1,2$. The null hypothesis is $H_0 : F_1(X) = F_2(X)$ for all values of X.

One of the alternative hypotheses below can be of interest.

- $H_1 : F_1(X) \neq F_2(X)$ for at least one value of X
- $H_1 : F_1(X) > F_2(X)$ for at least one value of X
- $H_1 : F_1(X) < F_2(X)$ for at least one value of X.

The test procedure is given below:

- Record the cumulative frequencies for $\left(F_{n_1}(.)\text{ and } F_{n_2}(.)\right)$ the ordered sample-values, for each possible value of the data, for each of the two samples X_1 and X_2.

- For example, if the data consist of $0\ 0\ 0\ 1\ 2\ 2\ 4\ 7\ 8\ 9$, for sample X_1, then the data and cumulative frequencies will be presented by:

X_1	$F_{n_1}(X)$
000	$3/10 = 0.3$
1	$4/10 = 0.4$
22	$6/10 = 0.6$
4	$7/10 = 0.7$
7	$8/10 = 0.8$
8	$9/10 = 0.9$
9	$10/10 = 1.0$

Similarly for X_2 and $F_{n_2}(X)$.

In other words: determine the cumulative proportions $F_{n_1}(X)$ and $F_{n_2}(X)$ by dividing the cumulative frequencies for each subject with the total sample size for each sample.

- Determine the absolute differences $\left(|d_i|, i = 1,2,\ldots,n\right)$, where $d_i = F_{n_1}(X) - F_{n_2}(X)$, for each possible data points in the two samples.
- Identify $D_{\max} = \max\left(|d_i|, i = 1,2,\ldots,n\right)$.
- This value is the test statistic, i.e.

$$KS = D_{\max}. \qquad (1.21)$$

If, at any point along the two cumulative probability distributions, the greatest distance is equal to or greater than the tabled Kolmogorov-Smirnov critical value recorded, then the null hypothesis is rejected. The critical values in the table of Kolmogorov-Smirnov test are listed with reference to the sample sizes. The following results are employed for evaluating the null hypothesis for the Kolmogorov-Smirnov test for two independent samples, according to Sheskin (2000: 324):

(a) If $H_1 : F_1(X) \neq F_2(X)$ is employed, then $H_0 : F_1(X) = F_2(X)$ can be rejected only when $KS \geq M_{n_1,n_2}(\alpha/2)$, where $M_{n_1,n_2}(\alpha/2)$ is the tabled (see Table A23 in Sheskin) critical (two-tailed) value. n_1 and n_2 are the sample sizes corresponding to the two independent samples.

(b) If $H_1 : F_1(X) > F_2(X)$ is employed, then $H_0 : F_1(X) = F_2(X)$ can be rejected when $KS \geq M_{n_1,n_2}(\alpha)$, where $M_{n_1,n_2}(\alpha)$ is the tabled critical (one-tailed) value. Additionally, the difference between the two cumulative probability distributions must be such that in reference to the point that represents the test statistic, the cumulative probability for Sample 1 must be larger than the cumulative probability for Sample 2.

(c) If $H_1 : F_1(X) < F_2(X)$ is employed, then $H_0 : F_1(X) = F_2(X)$ can be rejected when $KS \geq M_{n_1,n_2}(\alpha)$. Additionally, the difference between the two cumulative probability distributions must be such that in reference to the point that represents the test statistic, the cumulative probability for Sample 1 must be less than the cumulative probability for Sample 2.

Remarks:

- The statistic $M = 4(KS)^2 (n_1 n_2 / n_1 + n_2)$ follows an approximately chi-square distribution with 2 degrees of freedom if the sample sizes are large, and a one-tailed alternative is of importance.
- Seigel and Castellan (1988) reports that the Kolmogorov-Smirnov test in this case is more powerful than the chi-square $r \times c$-table test. For smaller samples it has higher power efficiency than the Mann-Whitney U test, but the reverse is true for larger samples.

1.3.2 Nonparametric test for two dependent samples

In this section we look at nonparametric tests for two dependent samples, referred to as the binomial sign test, as well as to the McNemar test for two dependent samples.

1.3.2.1 The binomial sign test for two dependent samples

According to Sheskin (2000:479), to employ the binomial sign test for two dependent samples, it is required that each of *n* subjects (or *n* pairs of matched subjects) has two scores (each score having been obtained under one of the two experimental conditions). The two scores are represented by the notations X_1 and X_2. For each subject (or pair of matched subjects), a determination is made with respect to whether a subject obtains a higher score in Condition 1 or Condition 2. Based on the latter, a signed difference $(D+ \text{ or } D-)$ is assigned to each pair of scores. The sign of the difference assigned to a pair of scores will be positive if a higher score is obtained in Condition 1 (i.e., $D+$ if $X_1 > X_2$), whereas the sign of the difference will be negative if a higher score is obtained in Condition 2 (i.e., $D-$ if $X_2 > X_1$). The hypothesis that is evaluated with the binomial sign test for two dependent samples, is whether or not in the underlying population represented by the sample, the proportion of subjects who obtain a positive signed difference (i.e., obtain a higher score in Condition 1) is some value other than 0.5. If the proportion of subjects in the sample who obtain a positive signed difference (which, for the underlying population, is represented by the notation $\pi+$) is some value that is either significantly above or below 0.5, it indicates there is a high likelihood the two dependent samples represent two different populations.

The objective of the binomial sign test for two dependent samples is to determine whether the two dependent samples represent two different populations. The binomial sign test is based on the following assumptions:

a) The sample of *n* paired subjects has been randomly selected from the population it represents.
b) The format of the data is such that, within each pair of scores the two scores can be rank-ordered.

The hypotheses for the binomial sign test comprises of the following:

- The null hypothesis is $H_0 : \pi+ = 0.5$, which means that in the underlying populations of concern, the proportion of subjects obtaining a positive sign difference, will be 0.5.

One of the following alternative hypotheses should be of interest:

- $H_1 : \pi+ \neq 0.5$
- $H_1 : \pi+ > 0.5$
- $H_1 : \pi+ < 0.5$.

The guidelines for executing the binomial sign test statistic are therefore as follows:

- Find the difference between the samples $\{D_i = X_i - Y_i,\ i = 1,2,\dots,n\}$. Any subject that obtains a zero difference score is eliminated from the data analysis.
- Record the signs of the difference (SD_i) .

- Find the total number of positive sign difference $Pos = \left(\sum_{i=1}^{n} I\left(SD_i > 0\right) \right)$ and the total number of negative sign difference $Neg = \left(\sum_{i=1}^{n} I\left(SD_i < 0\right) \right)$ where I denotes the indicator function.
- Assuming that, if subjects that obtain a difference score of zero are eliminated from the analysis, and the remaining sign differences are randomly distributed, it is expected that one-half of the subjects should obtain a positive sign difference and another one-half of the subjects should obtain a negative sign difference. Therefore, the observed proportion of positive sign differences and the observed proportion of negative sign differences are given respectively as:

  $$p+ = \frac{Pos}{m} \text{ and } p- = \frac{Neg}{m}, \qquad (1.22)$$

  where $m \equiv$ the number of sign differences, excluding the $(n-m)$ score differences of zero.

Let x denote the observed number of positive sign differences obtained. Let N_1 be a random variable denoting the number of positive sign differences obtained in m independent differences. Then equation (1.23) is used to determine the probability, under H_0, of obtaining x or more positive sign differences in a set of m scores (Sheskin, 2000:482–484).

$$N_1 \sim \text{Binomial}\left(m, \pi+\right), \text{and } P\left(N_1 \geq x\right) = \sum_{r=x}^{m} \binom{m}{x} (\pi+)^x (\pi-)^{m-x}, \qquad (1.23)$$

where $\pi+$ and $\pi-$ respectively denotes the hypothesised values for the proportion of positive and negative sign differences.

Sheskin (2000:483) and Howell (1992:115–117) discuss the rules to be applied for the binomial sign test as follows:

(a) To support alternative hypothesis $H_1 : \pi+ \neq 0.5$ at α-level of significance, the probability obtained under H_0 (i.e., $P\left(N_1 \geq x\right)$) should be less than $\alpha/2$.

(b) To support alternative hypothesis $H_1 : \pi+ > 0.5$, the following conditions must hold:

- The proportion of positive sign differences must be greater than the value stipulated in the null hypothesis (i.e., $(p+) > \pi +$), and
- The probability of obtaining a value equal or greater than N_1 under H_0 is less or equal to pre-specified value of α (i.e., $P(N_1 \geq x) \leq \alpha$)

(c) To support alternative hypothesis $H_1 : \pi + < 0.5$, the following conditions must hold:

- The proportion of positive sign differences $(p+)$ must be less than the value stipulated in the null hypothesis (i.e., $(p+) < \pi +$), and
- The probability of obtaining a value equal or less than N_1 under H_0 is less or equal to pre-specified value of α (i.e., $P(N_1 \leq x) \leq \alpha$).

A normal approximation of the binomial sign test for two dependent samples is also available. It is used for large sample sizes, when the correction for continuity (0.5) is considered, where z are evaluated under H_0, where $\pi+$ and $\pi-$ are specified. The test statistic is defined by:

$$z = \frac{|x-(m)(\pi+)|-0.5}{\sqrt{(m)(\pi+)(\pi-)}}. \tag{1.24}$$

The z-test statistic values obtained from samples are evaluated with the table of normal distribution and the decision rules of rejecting the null hypothesis is the same as was given in Section 1.1.1.1.

1.3.2.2 The McNemar test for two dependent samples

The McNemar test is a nonparametric procedure to be applied to categorical data in a hypothesis testing situation involving a design with two samples. The McNemar's test assumes that each of the *n* subjects or *n* pairs of matched subjects contributes to scores on the dependent variable. According to Sheskin (2000:492–500), McNemar's test is commonly applied to analyse data derived from two types of experimental designs, which are as follows:

- To analyse categorical data obtained in a true experiment (i.e., an experiment involving a manipulated independent variable). Two scores of each subject (or pair of matched subjects) represent a subject's responses under the two levels of the independent variable.

- To analyse a before-after design. This means that, n subjects are administered a pretest and a variable is measured. After the pretest, all n subjects are exposed to the experimental treatment. Subjects are then administered a posttest and the same variable is measured. The samples are now dependent.

The model for the McNemar test is summarized by Table 1.1 below, where a, b, c, and d represent the possible counts of subjects in each of the four categories (often under restrictions, such as fixed total or margin totals, depending on the way sampling was done for example) that can be employed to summarize the two responses of a subject on a dichotomous dependent variable.

Table 1.1: Model for McNemar Test

		Posttest		
		Response Category 1	**Response Category 2**	**Row sums**
Pretest	**Response Category 1**	a	b	$a + b = n_1$
	Response Category 2	c	d	$c + d = n_2$
	Column sums	$a + c$	$b + d$	n

The McNemar test involves the following assumptions:

- A sample of n subjects has been randomly selected from the population it represents;
- Each of the n observations in the contingency table is independent of the other observations;
- The scores of subjects are in the form of a dichotomous categorical measure involving two mutually exclusive categories; and
- Extremely small sample sizes should not be used for McNemar test.

Regarding the hypotheses for the McNemar test, the assumption is that in the underlying population, π_b and π_c represents the following population proportions:

$$\pi_b = \frac{b}{b+c} \text{ and } \pi_c = \frac{c}{b+c}. \qquad (1.25)$$

If there is no difference between the two experimental conditions or between the pretest and posttest, then the following is true:

$$\pi_b = \pi_c = 0.5. \qquad (1.26)$$

- The null hypothesis is $H_0 : \pi_b = \pi_c$.

One of the following alternative hypotheses is of importance:

- $H_1 : \pi_b \neq \pi_c$
- $H_1 : \pi_b > \pi_c$
- $H_1 : \pi_b < \pi_c$.

The test statistic for the McNemar test, which is based on the chi-square distribution, is denoted by:

$$\chi^2 = \frac{(b' - c')^2}{b' + c'}, \qquad (1.27)$$

where b' and c' represents the number of observed observations from the sample, in cells *b* and *c* of the McNemar test. If $b' = c'$ then $\chi^2 = 0$. χ^2 should be positive at all times, and if χ^2 is negative, there is an error in the calculations.

The chi-square value χ^2 is evaluated with the table of the chi-square distribution with 1 degree of freedom making note of the following criteria:

(a) If the alternative hypothesis $H_1 : \pi_b \neq \pi_c$ is applied, then $H_0 : \pi_b = \pi_c$ is rejected when $\chi^2 \geq \chi_1^2(\alpha/2)$, where $\chi_1^2(\alpha/2)$ is the upper $(\alpha/2)$ percentile of the chi-squared distribution with 1 degree of freedom..

(b) For one of the alternative hypotheses $H_1 : \pi_b > \pi_c$ and $H_1 : \pi_b < \pi_c$ to be supported, $H_0 : \pi_b = \pi_c$ is rejected when $\chi^2 \geq \chi_1^2(\alpha)$.

In the next Section we briefly revise the procedure for identifying outliers in the data samples, as well as proposing remedial steps via transformations.

1.4 Outliers and transformation

In this section we briefly discuss the ways of dealing with problems encountered in the data when two-sample tests are done, and ways of transforming the data.

1.4.1 Procedure for identifying outliers

Outliers are observations in data that do not appear to be consistent with the rest of the data. Often inconsistency is reflected in the magnitude of observations. Barnett and

Lewis (1994) describe 48 tests for detecting outliers in data that are assumed to be drawn from a normal distribution.

According to Stevens (1996: 17), it has been demonstrated that regardless of the distribution of the data, at least $\left[1-\left(\frac{1}{k^2}\right)\right]100\%$ of the observations in a dataset should fall within k standard deviations from the mean of the data, where k is any value greater than one. Scores corresponding to relatively high standard deviation values must be considered to contain possible outliers. Accommodation of outliers involves the use of procedures which utilizes all the data and minimizes the influence of outliers. Two such options that reduce the impact of outliers are:

- The use of the median instead of the mean as a measure of central tendency.
- Employing an inferential statistical test that uses rank orders instead of interval or ratio data.

Since there are many ways of detecting outliers in the data, it is important for the practitioner to consult more detailed sources to determine the most appropriate test for detecting outliers in a given dataset.

1.4.2 Data transformation

Data transformation involves mathematical operations of converting the data into a new set of scores, which are analysed in order to reduce the impact of outliers in the original data. Data transformation can also be employed to equate heterogeneous group variances, as well as to normalise a non-normal distribution, or to move closer to normality.

There are numerous examples of transformations which are employed within various experimental setups according to Howell (1992:54–56). There are commonly employed data transformations which can reduce the impact of outliers, normalise skewed data or produce homogeneity of variance:

a) **Square-root transformation:** This transformation may be used when the mean is proportional to the variance, which implies that the ratio of the mean of a treatment group to the variance of the treatment group is approximately the same for all of the treatments;

b) **Logarithmic transformation:** It may be useful when the mean is proportional to the standard deviation;

c) **Reciprocal transformation:** It is also referred to as an inverse transformation, which may be useful when the square of the mean is proportional to the standard deviation;

d) **Arcsin transformation:** This kind of transformation is also referred to as an angular or inverse sine transformation. It involves the use of a trigonometric function which is able to transform a proportion between 0 and 1 into an angle expressed in radians.

1.5 Extension of two sample test to more dimensions

Techniques for two-sample comparisons tests, can often be extended to more dimensions.

The single-factor between-groups analysis of variance. This test is employed with interval or ratio data and it is a parametric test. The test is most basic of the analysis of variance procedures and employed in a hypothesis testing situation involving k independent samples. In conducting the single-factor between-groups analysis of variance, each of the k sample means is employed to estimate the value of the mean of the population the sample represent.

The Kruskal-Wallis one-way analysis of variance by ranks. The test is an extension of the Mann-Whitney U test (described in section 1.1.2.1. above), to a design involving more than two independent samples.

These are only examples of popular tests used by Statisticians, but there is a vast literature on this subject. This is, however, not part of this study and we will now proceed to a new chapter, where the bootstrap technique will be reviewed briefly, and the application of the bootstrap to two sample comparisons, will be discussed.

CHAPTER 2

BOOTSTRAP METHODOLOGY

Usually, the estimation of parameters is a concern in inferential statistics, and statistics, based on samples taken from the true population, are calculated as estimators and to use for inference purposes. Questions regarding the accuracy of these estimators and the inference have to be answered, which are often done in uncertainty and hesitantly. This is due to assumptions being made about essential matters such as the distribution of the population. Also, theoretical drawbacks occur due to various mathematical and computational difficulties and restrictions. To combat some of the problems, Efron (1979) introduced a resampling computer based method which can be non-parametric or parametric, and is known as "The Bootstrap – method". The non-parametric bootstrap provides answers to situations where distributions are unknown and where the statistics of interest are too complex to calculate theoretically, while in the parametric bootstrap more information is known and assumed about the underlying distribution and only parameters are to be estimated. The bootstrap is applied analytically in a wide variety of real data situations and it is flexible, quick and easy to use, and involves minimum mathematical assumptions in most cases. Applying the bootstrap, enables statisticians and other quantitative researchers to compute more complicated statistical estimates and measures of accuracy for the parameters of the population from which the observed data is obtained, and to make inferences about these estimates. This is possible without making any assumptions about the form and shape of the probability distribution.

Applying the bootstrap method to derive two-sample comparison tests, we now present a brief introduction to the basic ideas surrounding the bootstrap methods. In Section 2.1 the basic bootstrap method is discussed. In Section 2.2 we discuss one of the examples of the bootstrap application, referred to as the bootstrap estimate of the standard error. In Section 2.3, a two sample test for equality of means is discussed, using the bootstrap. Two sample tests for equality of distributions based on the permutation test and the bootstrap test are discussed in Section 2.4. Section 2.5 concludes the chapter with an example of the bootstrap applied to time series data.

2.1 The bootstrap method

The main ideas defining the bootstrap method are briefly introduced in the paragraphs below.

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a finite, random sample of size n, consisting of independent and identically distributed observations obtained from the unknown distribution function F. The statistic of interest that depends on F is denoted by $T_n(\mathbf{X}_n, F)$. The Empirical Distribution Function (EDF) F_n is one of the estimators that are used to approximate the unknown distribution function F, but other estimators exist in literature, which are also used regularly. The EDF is mathematically defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \tag{2.1}$$

where $I(\cdot)$ is the indicator function.

The bootstrap sample is denoted by $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$ and it has been drawn randomly and with replacement from the original sample $\mathbf{X}_n$. The bootstrap principle is to use the sample data $\mathbf{X}_n$ and the bootstrap method to compute an estimate for the distribution of $T_n(\mathbf{X}_n, F)$ under F. This estimation process produces the bootstrap distribution of $T_n(\mathbf{X}_n^*, F_n)$ under F_n. By replacing $\mathbf{X}_n$ from F by $\mathbf{X}_n^*$ from F_n, the probability distribution of $T_n(\mathbf{X}_n, F)$ is approximated by that of $T_n(\mathbf{X}_n^*, F_n)$. The bootstrap sample is created by drawing any element of the original sample (with replacement) by equal chance. We say that the probability under the EDF F_n is denoted by:

$$P^*(X_j^* = X_i) = \frac{1}{n} \text{ for } i, j = 1, 2, \ldots, n. \tag{2.2}$$

We also say that the plug-in estimate of $T_n(\mathbf{X}_n, F)$ is $T_n(\mathbf{X}_n^*, F_n)$. In other words, we estimate the function T_n of the probability distribution F by the same function of EDF F_n.

The Section that follows is based on excerpts from the books of Efron and Tibshirani (1993) and Davison and Hinkley (1997).

2.1.1 The bootstrap procedure

Let $T_n(\mathbf{X}_n, F)$ be some statistic of interest, which is dependent on the unknown distribution function F. Then the sampling distribution of $T_n(\mathbf{X}_n, F)$ under F can be approximated by the bootstrap distribution of $T_n(\mathbf{X}_n^*, F_n)$ under F_n, where $\mathbf{X}_n^* = (X_1^*, X_2^*, ..., X_n^*)$ denotes a random sample of size *n* from F_n, drawn in the way explained in the previous paragraph. In other words, each X_i^*, $i = 1,...,n$ is drawn independently with replacement and with equal probability n^{-1} from the set $\{X_1, X_2, ..., X_n\}$. To calculate the bootstrap probability distribution, the following Monte Carlo algorithm is used:

A. Generate a random sample (the first bootstrap sample) $\mathbf{X}_n^*(1) = (X_{11}^*, X_{12}^*, ..., X_{1n}^*)$ from F_n, where the random number generator is used to obtain the bootstrap sample.

B. Calculate $T_n(\mathbf{X}_n^*(1), F_n)$ from the first bootstrap sample.

C. Independently repeat the above two steps B times and obtain the following vectors for the bootstrap samples and the bootstrap statistics respectively:

$$\mathbf{X}_n^*(1), \mathbf{X}_n^*(2), ..., \mathbf{X}_n^*(B)$$

and

$$T_n(\mathbf{X}_n^*(1), F_n), T_n(\mathbf{X}_n^*(2), F_n), ..., T_n(\mathbf{X}_n^*(B), F_n),$$

where $B \geq 200$ is a usual choice.

The distribution (histogram) of these bootstrap statistics $T_n(\mathbf{X}_n^*(i), F_n)$ for $i = 1,2,...,B$, is an approximation to the true sampling distribution of the statistic $T_n(\mathbf{X}_n, F)$. According to Efron (1981:589), as $B \to \infty$ the bootstrap approximated distribution of $T_n(\mathbf{X}_n^*, F_n)$ should lead to a more accurate approximation of the distribution of $T_n(\mathbf{X}_n, F)$.

Some important examples of successful application of the bootstrap will now be discussed briefly.

2.2 The bootstrap estimate of the standard error

Suppose $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ is a random sample from an unknown probability distribution *F*, and $\hat{\theta}$ is an estimator based on $\mathbf{X}_n$ of some unknown parameter θ. The objective is to estimate the parameter of interest, θ, by using the random sample $\mathbf{X}_n$. Also, the standard error of the estimate $\hat{\theta}$ is defined as:

$$\sigma_F\left(\hat{\theta}\right) = \sqrt{\operatorname{Var}_F\left(\hat{\theta}\right)}. \tag{2.3}$$

Using the plug-in principle, the bootstrap estimate of the standard error is then defined as:

$$\sigma_{F_n}\left(\hat{\theta}^*\right) = \sqrt{\operatorname{Var}_{F_n}\left(\hat{\theta}^*\right)}. \tag{2.4}$$

Here $\hat{\theta}^*$ is the same estimator of θ as $\hat{\theta}$, but it is determined from the bootstrap sample $\mathbf{X}_n^* = \left(X_1^*, X_2^*, \ldots, X_n^*\right)$. The bootstrap estimate of the standard error of a statistic $\hat{\theta}$, is therefore a plug-in estimate that uses the empirical distribution function F_n in place of the unknown distribution F. The standard error can be calculated using the Monte-Carlo algorithm which is described by the following algorithm:

A. Generate the first bootstrap sample $\mathbf{X}_n^*(1) = \left(X_{11}^*, X_{12}^*, \ldots, X_{1n}^*\right)$ from the EDF, F_n, i.e. sample with replacement from $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ with probability n^{-1}, as before.

B. From this bootstrap sample calculate the statistic $\hat{\theta}^*(1) = \hat{\theta}\left(\mathbf{X}_n^*(1)\right)$.

C. Independently repeat the above two steps B times to obtain the bootstrap samples $\mathbf{X}_n^*(1), \mathbf{X}_n^*(2), \ldots, \mathbf{X}_n^*(B)$ and calculate the respective statistics $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$ on these bootstrap samples.

D. Then, calculate the Monte Carlo approximation of $\sigma_{F_n}\left(\hat{\theta}^*\right)$ in (2.4), i.e.,

$$\hat{\sigma}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)\right)^2}, \tag{2.5}$$

where $\hat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b)$.

According to Efron and Tibshirani (1993:47–52), $\sigma_{F_n}\left(\hat{\theta}^*\right)$ denotes the ideal bootstrap estimate of the standard error of $\hat{\theta}$ and $\hat{\sigma}_B$ denotes the Monte-Carlo bootstrap estimate of the standard error. It can be shown that $\hat{\sigma}_B \to \sigma_{F_n}\left(\hat{\theta}^*\right)$ as $B \to \infty$. Values of B such that $50 \le B \le 200$, give satisfactory results when estimating the standard errors.

The following paragraph is an extension of the bootstrap method of estimating the standard error of an estimator. The method is referred to as the double bootstrap method.

2.2.1 The double bootstrap estimation of the standard error

The question that is often asked when calculating the bootstrap standard error of some statistic is: 'How accurate is the bootstrap estimator of the standard error?' We are therefore interested in estimating the standard error of the bootstrap estimate of the standard error. This is a typical scenario where the double bootstrap can be employed. If the bootstrap estimator of standard error is denoted by $\hat{\sigma} = \sigma_{F_n}\left(\hat{\theta}^*\right)$, then the bootstrap estimate of the bootstrap standard error (i.e. the bootstrap estimate), is defined as:

$$\sigma_{F_n}\left(\sigma_{F_n^*}\left(\hat{\theta}^{**}\right)\right) = \sqrt{\mathrm{Var}_{F_n^*}\left(\hat{\sigma}^*\right)}, \text{ where } \hat{\sigma}^* = \mathrm{Var}_{F_n}\left(\hat{\theta}^*\right). \tag{2.6}$$

Here, the sample $\mathbf{X}_n^{**} = \left(X_1^{**}, X_2^{**}, \ldots, X_n^{**}\right)$ is the second phase bootstrap sample which is sampled with replacement from the first phase bootstrap sample $\mathbf{X}_n^* = \left(X_1^*, X_2^*, \ldots, X_n^*\right)$ and $\sigma_{F_n^*}$ denotes the standard error with respect to the double bootstrap sample $\mathbf{X}_n^{**} = \left(X_1^{**}, X_2^{**}, \ldots, X_n^{**}\right)$.

The algorithm for approximating the standard error of the bootstrap standard error is as follows:

A. Generate the first bootstrap sample $\mathbf{X}_n^*(1) = \left(X_{11}^*, X_{12}^*, \ldots, X_{1n}^*\right)$ from F_n.

i. Generate the first double bootstrap sample $\mathbf{X}_n^{**}(1)=\left(X_{11}^{**},X_{12}^{**},\ldots,X_{1n}^{**}\right)$ from F_n^*, where F_n^* is similar to, (2.1), but defined in terms of $\mathbf{X}_n^*(1)=\left(X_{11}^*,X_{12}^*,\ldots,X_{1n}^*\right)$, which was defined in terms of $\mathbf{X}_n$.

ii. Using the first double bootstrap sample $\mathbf{X}_n^{**}(1)=\left(X_{11}^{**},X_{12}^{**},\ldots,X_{1n}^{**}\right)$ calculate the statistic $\hat{\theta}^{**}(1)=\hat{\theta}\left(\mathbf{X}_n^{**}(1)\right)$.

iii. Independently repeat above two steps B_1 times to obtain the bootstrap replications $\hat{\theta}^{**}(1),\hat{\theta}^{**}(2),\ldots,\hat{\theta}^{**}(B_1)$ and then calculate

$$\hat{\sigma}^*(1)=\sqrt{\frac{1}{B_1-1}\sum_{b=1}^{B_1}\left(\hat{\theta}^{**}(b)-\hat{\theta}^{**}(\cdot)\right)^2}\text{, where } \hat{\theta}^{**}(\cdot)=\frac{1}{B_1}\sum_{b=1}^{B_1}\hat{\theta}^{**}(b).$$

B. Independently repeat step A B_2 times to obtain B_2 bootstrap replications of the standard error $\hat{\sigma}^*(1),\hat{\sigma}^*(2),\ldots,\hat{\sigma}^*(B_2)$.

C. Finally calculate the Monte Carlo approximation to $\hat{\hat{\sigma}}$, i.e.,

$$\hat{\hat{\sigma}}_{B_2}=\sqrt{\frac{1}{B_2-1}\sum_{b=1}^{B_2}\left(\hat{\sigma}^*(b)-\hat{\sigma}^*(\cdot)\right)^2}\text{, where } \hat{\sigma}^*(\cdot)=\frac{1}{B_2}\sum_{b=1}^{B_2}\hat{\sigma}^*(b).$$

Efron (1979) illustrated the double bootstrap with an example applied to linear discriminant analysis (Frangos, 1984:101–104).

2.3 Two sample test for equality of means

In this section we discuss the two sample test for equality of means by using the bootstrap and the algorithm to perform the test. The main focus of Chapter 1 was on two sample tests, usually under the assumption of normality ($N(\mu,\sigma^2)$) of the population distribution, where the test statistic under the null hypothesis consequently has a known distribution such as the *t*-distribution. Without the normality assumption, the bootstrap method can be applied to approximate the distribution of such test statistics. Other non-parametric methods also exist. Furthermore, large sample approximations of the distribution of the test statistics exist. The following brief discussion of executing the bootstrap hypothesis test for equality of means, will illustrate the applicability of the bootstrap method in two-sample comparison tests.

2.3.1 Bootstrap hypothesis test for equality of means

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a random sample from an unknown distribution F with mean μ_X and let $\mathbf{Y}_m = (Y_1, Y_2, \ldots, Y_m)$ be a random sample from an unknown distribution G with mean μ_Y.

Then the objective is to test the null hypothesis denoted by:

$$H_0 : \mu_X = \mu_Y . \tag{2.7}$$

The hypothesis concerns only the means and doesn't involve the variance or any other distributional parameter. Efron and Tibshirani (1993:224) suggested the following algorithm for computation of the bootstrap test statistic for testing equality of means as follows:

A. Calculate the means of the samples data $\mathbf{X}_n$ and $\mathbf{Y}_m$, and refer to it as $\bar{X}$ and $\bar{Y}$.

B. Calculate the test statistic $t_{obs} = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$, (2.8)

where $S_X^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \Big/ (n-1)$ and $S_Y^2 = \sum_{j=1}^{m} (Y_j - \bar{Y})^2 \Big/ (m-1)$.

C. Calculate the mean $\bar{W}$ of the pooled data $\mathbf{W}_N = (X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)$ where $N = n + m$.

D. Transform the two data sets to conform to the null hypothesis, i.e., $\tilde{X}_i = X_i - \bar{X} + \bar{W}$ and $\tilde{Y}_j = Y_j - \bar{Y} + \bar{W}$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

E. Let the EDF of $\tilde{\mathbf{X}}_n$ be denoted by F_n and let the EDF of $\tilde{\mathbf{Y}}_m$ be denoted by G_m.

F. Create the first bootstrap samples $\tilde{X}_i^*(1)$ and $\tilde{Y}_j^*(1)$ by sampling with replacement from $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{Y}}_m$ respectively, as was described before. From the first bootstrap samples $\tilde{\mathbf{X}}_n^*$ and $\tilde{\mathbf{Y}}_m^*$ calculate the means $\bar{\tilde{X}}^*(1)$ and $\bar{\tilde{Y}}^*(1)$, and the variances $S_{\tilde{X}^*}^2(1)$ and $S_{\tilde{Y}^*}^2(1)$.

G. Then calculate the first bootstrap statistic as: $t^*(1) = \dfrac{\bar{\tilde{X}}^*(1) - \bar{\tilde{Y}}^*(1)}{\sqrt{\dfrac{S_{\tilde{X}^*}^2(1)}{n} + \dfrac{S_{\tilde{Y}^*}^2(1)}{m}}}$, (2.9)

where $S_{\tilde{X}^*}^2$ and $S_{\tilde{Y}^*}^2$ are the sample variances based on the bootstrap samples.

H. Repeat the two latter steps B times to obtain a vector of test statistics: $t^*(1), t^*(2), \dots, t^*(B)$.

I. Calculate the Monte Carlo approximation of the p-value of the test, i.e.,

$$p_{boot} = \frac{1}{B} \sum_{b=1}^{B} I\left(t^*(b) \geq t_{obs}\right), \tag{2.10}$$

where t_{obs} is defined in (2.8).

The estimator p_{boot} has no interpretation as an exact probability, but like all bootstrap estimates, it is only guaranteed to be accurate if the sample size goes to infinity. If this is the case, p_{boot} can be interpreted as follows: if the value of p_{boot} is small, then the evidence against H_0 becomes strong. In traditional two sample tests, the equal variance assumption is attractive for the *t*-test because it simplifies the form of the resulting distribution of the test statistic. But, usually there is no compelling reason to assume equal variances and in such cases, we consider the bootstrap hypothesis test for comparing the two means.

2.4 Two sample test for equality of distributions

Two sample tests for equality of two distributions are now considered in this section, and both the permutation test and the bootstrap test will be discussed.

2.4.1 Permutation test for equality of distributions

The permutation test is a non-parametric test that is based on the available data, rather than the distribution. If the two independent random samples $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y}_m = (Y_1, Y_2, \dots, Y_m)$ are observed, and drawn from possibly different probability distributions F and G, respectively, then the goal of the statistician is to test the null hypothesis of no difference between F and G, denoted by:

$$H_0 : F = G. \tag{2.11}$$

If the null hypothesis is true, there is no difference between the probabilistic behaviour of the two random samples. A hypothesis test, of which the permutation test is an example, is a formal way of deciding whether or not the data support the decision to reject the null hypothesis.

The permutation test involves the following:

Combine the two samples $\mathbf{X}_n$ and $\mathbf{Y}_m$ in one vector, and order the elements from small to large. Also, in a separate vector, keep track of the origin of these ordered values by recording if the element was an X-value or a Y-value. Denote the ordered vector of values by $\mathbf{W}_N$ and let $\mathbf{g}_N$ denote the vector that indicates which group each ordered observation belongs to. Also, let N equal the combined sample size $n+m$. The vector $\mathbf{g}_N$ consists therefore of n X's and m Y's, and there are $\binom{N}{n} = \frac{N!}{n!m!}$ possible $\mathbf{g}$ vectors. The following result is of importance for the permutation test:

Permutation Lemma: Under $H_0 : F = G$, the vector $\mathbf{g}$ has probability $1 \Big/ \binom{N}{n}$ of equalling any one of its possible arrangements (Efron and Tibshirani, 1993:207).

Thus, any permutation of X's and Y's is equally likely if the null hypothesis is true. The test statistic is therefore expressed as follows:

$$\hat{\theta} = \bar{X} - \bar{Y} = \frac{1}{n} \sum_{g_i = X} W_i - \frac{1}{m} \sum_{g_i = Y} W_i , \tag{2.12}$$

where $\sum_{g_i = X} W_i$ indicates the sum of the W_i over values of $i = 1,2,\ldots,N$ having $g_i = X$.

Equation (2.12) can be seen as a function of $\mathbf{g}_N$ and $\mathbf{W}_N$, i.e. $\hat{\theta}$ can be expressed as $\hat{\theta} = S(\mathbf{g}_N, \mathbf{W}_N)$. Let $\mathbf{g}^p$ denote any one of the $\binom{N}{n}$ possible vectors of n X's and m Y's, and define the permutation replication of $\hat{\theta}$ as:

$$\hat{\theta}^p = S(\mathbf{g}^p, \mathbf{W}_N) = \bar{X}^p - \bar{Y}^p = \frac{1}{n} \sum_{g_i^p = X} W_i - \frac{1}{m} \sum_{g_i^p = Y} W_i , \text{ for } i = 1,2,\ldots,N. \tag{2.13}$$

There are $\binom{N}{n}$ permutation replications $\hat{\theta}^p$ and the distribution that places probability $1 \Big/ \binom{N}{n}$ on each one of $\hat{\theta}^p$ is called the permutation distribution of $\hat{\theta}^p$.

The following is the Monte-Carlo algorithm for the computation of the two-sample permutation test statistic:

A. Calculate all $\binom{N}{n}$ permutations of possible ways that the n X's and m Y's can be arranged, i.e. find $\mathbf{g}_N(1), \mathbf{g}_N(2), \ldots, \mathbf{g}_N\binom{N}{n}$ and their corresponding $\mathbf{W}_N$ vectors.

B. Choose B independent vectors $\mathbf{g}^p(1), \mathbf{g}^p(2), \ldots, \mathbf{g}^p(B)$ as well as their corresponding $\mathbf{W}_N$ vectors, each consisting of n X's and m Y's and each being randomly selected from the set of all $\binom{N}{n}$ possible such vectors. (B will usually be at least 1000; see **Table 2.1**).

C. Calculate the first test statistic: $\hat{\theta}^p(1) = \bar{X}^p - \bar{Y}^p$.

D. Repeat the latter two steps B times to obtain the distribution of the permutation statistics: $\hat{\theta}^p(1), \hat{\theta}^p(2), \ldots, \hat{\theta}^p(B)$.

E. Approximate the permutation p-value with the following expression:

$$p_{perm} = \frac{1}{B}\sum_{b=1}^{B} I\left(\hat{\theta}^p(b) \geq \hat{\theta}\right), \tag{2.14}$$

where $\hat{\theta}$ is the statistic calculated from the original data set and $I(\cdot)$ is the indicator function.

Table 2.1 below, obtained from Efron and Tibshirani (1993: 211), illustrates the influence of the number of permutation replications on the achieved significance level (ASL):

Table 2.1: Influence of the number of permutations on the ASL

ASL_{perm}:	0.500	0.250	0.100	0.050	0.025
B:	100	299	900	1901	3894

The sample sizes of the original samples $\mathbf{X}_n$ and $\mathbf{Y}_m$, also have an influence, and it is known that the number of permutations can be very large even in cases with moderate sample sizes. It is for this reason that the Monte Carlo approach is adopted instead, whereby only a small fraction of ASL, but still a large number B, of the permutations are used.

2.4.2 Bootstrap test for equality of distributions

As before, let $\mathbf{X}_n = (X_1, X_2, ..., X_n)$ be an i.i.d. sample from the probability distribution F and $\mathbf{Y}_m = (Y_1, Y_2, ..., Y_m)$ be an i.i.d. sample from possibly different probability distribution G. Then the null hypothesis below is of interest:

$$H_0 : F = G, \tag{2.15}$$

which implies that, under H_0, there is no difference between the probability structure that generated $\mathbf{X}_n$ and $\mathbf{Y}_m$. When both populations are approximately normal, the two-sample t procedures (discussed in Chapter 1) compare the two population means satisfactorily. The bootstrap, however, can be used to compare the two population distributions, without the normality assumption and without the restriction to comparison of means only.

The important idea is that the bootstrap should mimic the conditions surrounding the separate samples that produced the original data. The bootstrap hypothesis test is also based on the test statistic:

$$t(\mathbf{X}_n, \mathbf{Y}_m) = t = \bar{X} - \bar{Y}, \tag{2.16}$$

but the distribution of the test statistic is unknown.

According to Efron and Tibshirani (1993, 220–222), the algorithm for computation of the bootstrap test for testing equality of probability distributions can be carried out as follows:

A. Calculate the means of the two original datasets and let $t(\mathbf{X}_n, \mathbf{Y}_m) = t = \bar{X} - \bar{Y}$.

   Denote the combined sample of $\mathbf{X}_n$ and $\mathbf{Y}_m$ by

   $\mathbf{V}_N = (X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_m)$ and let $\hat{F}_0$ denote the empirical distribution belonging to $\mathbf{V}_N$, putting probability $\frac{1}{n+m}$ on each of the elements of $\mathbf{V}_N$.

B. Draw $n+m$ observations by sampling with replacement from $\mathbf{V}_N$. Call the first n observations $\mathbf{X}_n^*(1) = (X_1^*, X_2^*, ..., X_n^*)$ and the remaining m observations $\mathbf{Y}_m^*(1) = (Y_1^*, Y_2^*, ..., Y_m^*)$.

C. Calculate the means $\bar{X}^*$ and $\bar{Y}^*$ of the bootstrap samples $\mathbf{X}_n^*(1)$ and $\mathbf{Y}_m^*(1)$.

D. Then calculate the first bootstrap test statistic

$t\left(\mathbf{X}_n^*(1), \mathbf{Y}_m^*(1)\right) = t^* = \bar{X}^* - \bar{Y}^* = t^*(1)$.

E. Repeat the two latter steps B times to obtain the statistics $t^*(1), t^*(2), \ldots, t^*(B)$.

F. Approximate the p-value by calculating the expression:

$$p_{boot} = \frac{1}{B}\sum_{b=1}^{B} I\left(t^*(b) \geq t\right), \tag{2.17}$$

where, $I(\cdot)$ denotes an indicator function and t is defined in A above, i.e. the observed difference in the means of the original data.

2.4.3 Remarks

It should be noted that permutations should be drawn in a way that is consistent with the null hypothesis and with the study design. In the two-sample problems discussed, focus has been pointed to cases where the null hypothesis implies two identical population distributions. We may also wish to compare population means, proportions, standard deviations, or other statistics using the permutation test. Resampling randomly reassigns observations to the two groups. The permutation test can be used to approximate the accuracy of the test statistics obtained from the matched pairs design discussed in Chapter 1, on which it randomly permutes the two observations within each pair separately. A variety of comparisons are again possible, which can be studied in the literature (see Welch and Gutierrez, 1988).

To test the hypothesis of no relationship between two variables, the correlation is the common measure of association, but not the only one.

The t-test statistic is based on standardizing the difference of means in a way to get a statistic that has a t-distribution when H_0 is true. The permutation test works directly with the difference of means (or some other statistic) and estimates the sampling distribution by applying resampling steps. The t-test gives accurate p-values if the sampling distribution of the difference of means is at least approximately normal. The permutation test gives accurate p-values even when the sampling distribution is not close to normal.

The permutation p-value is said to be exact, while the bootstrap p-value is said to be an approximation, but in practice the two methods often give similar results.

The bootstrap histograms center near the value of the statistic $\hat{\theta}$, while that of permutation histograms are centered near zero (Efron and Tibshirani, 1993:216–218). These different ways should be interpreted correctly.

The standard deviation of the permutation distribution is not a dependable estimate of standard error for $\hat{\theta}$, while the bootstrap standard deviation is a trustworthy estimator.

2.5 Bootstrapping time series data

A brief discussion of Time Series models will be given in Chapter 3. We now only consider the application of the bootstrap method in Time Series Analysis.

According to Wei (1990), a time series consists of serially correlated data, and while the data may be identically distributed, the data of a time series are not independent. In nonparametric situations the dependence may be complex, or our knowledge of it may be limited, so that neither of the approaches discussed above may be feasible.

Let $Z_1, Z_2, \ldots, Z_n$ be observed data from a strictly stationary time series $\{Z_t, t \in \mathbb{Z}\}$, for example. The assumption of stationarity implies that the joint probability law of $(Z_t, Z_{t+1}, \ldots, Z_{t+k})$ does not depend on t for any $k \geq 0$. Assume also that the time series is weakly dependent, i.e., the collection of random variables $\{Z_t, t \leq 0\}$ is approximately independent of $\{Z_t, t \geq k\}$ when k is large enough. An example of a weak dependence structure is given by n-dependence under which $\{Z_t, t \leq 0\}$ is (exactly) independent of $\{Z_t, t \geq k\}$ if $k > n$. Due to the dependence between the observations, even the most basic methods involved in applied statistical work become challenging. The bootstrap method that will be discussed for time series data is based on so-called bootstrap residuals and is referred to as a model based bootstrap method. This is the method explaining how to deal with stationary dependent data by using the bootstrap.

2.5.1 Bootstrapping residuals

Statisticians usually fit suitable models to data when analysing it, and various time series models are usually possible. Once a model is fitted satisfactorily the model based bootstrap demands that residuals are determined from the fitted model. From these residuals a new series is generated by incorporating random samples taken from the residuals, into the fitted model. The residuals are typically recentred to have the same mean as the innovations of the model. According to Lahiri (2003: 24), when the sequence of originally observed variables $Z_1, Z_2, \ldots, Z_n$ is stationary, it has been shown that, under suitable regularity conditions, a version of the model based bootstrap approximation to the standardized least squares estimators of the parameters involved in the model, are more accurate than the normal approximations. This is not true for nonstationary sequences. We will briefly illustrate some of the concepts involved below. The following steps are guidelines for deriving bootstrap estimators of the standard errors of the estimators of the parameters of the autoregressive model of order p.

A. Given some stationary time series data $Z_1, Z_2, \ldots, Z_n$, estimate the parameters of the model, i.e., by using the method of least squares. Then we wish to estimate the standard deviation of the estimators of the autoregressive coefficients $\hat{\phi} = \left(\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p\right)$ of the $\mathrm{AR}(p)$ model. The $\mathrm{AR}(p)$ model is defined in Section 3.6, i.e. $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t$.

B. Use the formulation of the $\mathrm{AR}(p)$ model: $Z_t = \hat{\phi}_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + \ldots + \hat{\phi}_p Z_{t-p} + e_t$, for $t = p+1, p+2, \ldots, n$. Then obtain the residuals:

$$e_t = Z_t - \hat{\phi}_1 Z_{t-1} - \hat{\phi}_2 Z_{t-2} - \ldots - \hat{\phi}_p Z_{t-p}, t = p+1, p+2, \ldots, n. \qquad (2.18)$$

C. Center the residuals $\tilde{e}_t = e_t - \dfrac{1}{n-p} \sum_{t=p+1}^{n} e_t$, for $t = p+1, p+2, \ldots, n$. (2.19)

D. Sample with replacement from the centred residuals $\tilde{e}_{p+1}, \tilde{e}_{p+2}, \ldots, \tilde{e}_n$ to obtain bootstrap residuals $\tilde{e}^*_{p+1}, \tilde{e}^*_{p+2}, \ldots, \tilde{e}^*_n$.

E. Use the Bootstrap residuals to recursively generate new samples. This is not an easy task, especially not for the $\mathrm{AR}(p)$ model with large p. The task is much simpler for the $\mathrm{AR}(1)$ model where:

$$Z_t = \phi_1 Z_{t-1} + e_t, \ t = 2, 3, \ldots, n. \qquad (2.20)$$

The least squares estimator for ϕ_1, is denoted by $\hat{\phi}_1$. If (2.19) has been calculated, bootstrap residuals can be determined from the centered residuals, i.e., $\tilde{e}_2^*, \tilde{e}_3^*, \ldots, \tilde{e}_n^*$. By taking $Z_1^* = Z_1$, and by using $\hat{\phi}_1$ in place of ϕ_1, in (2.20) above, a new bootstrap sample can recursively be determined as follows:

$$\begin{aligned} Z_2^* &= \hat{\phi}_1 Z_1^* + \tilde{e}_2^* \\ Z_3^* &= \hat{\phi}_1 Z_2^* + \tilde{e}_3^* \\ &\vdots \\ Z_n^* &= \hat{\phi}_1 Z_{n-1}^* + \tilde{e}_n^*. \end{aligned}$$

For the $\text{AR}(p)$ this is much more complicated, involving all the estimators $\hat{\phi} = \left(\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p\right)$, and by estimating, choosing or calculating initial values for $Z_{t-1}, \ldots, Z_{t-p}$, whereafter Z_t^* can be determined recursively, for $t = p+1, \ldots, n$. A "burn in" period for the process usually is demanded. Several methods exist in literature, of which some demand distributional assumptions.

F. For the $\text{AR}(1)$ model, we can now use the first bootstrap sample $Z^*(1) = \left(Z_1^*, Z_2^*, \ldots, Z_n^*\right)$ to estimate $\hat{\phi}_1^*(1)$. By repeating the previous steps B times, the bootstrap estimate of the standard deviation of the $\text{AR}(1)$ parameter is:

$$\hat{\sigma}_{1,B} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\phi}_1^*(b) - \hat{\phi}_1^*(.)\right)^2}, \text{ where } \hat{\phi}_i^*(.) = \frac{1}{B}\sum_{b=1}^{B}\hat{\phi}_i^*(b), \quad i = 1, 2, \ldots, p.$$

G. For the $\text{AR}(p)$ model, Efron and Tibshirani (1993) recommend the following procedure:

i. Use the bootstrap residual in (D) to get Z^*'s from:

$$\begin{aligned} Z_{p+1}^* &= \hat{\phi}_1 Z_p^* + \hat{\phi}_2 Z_{p-1}^* + \ldots + \hat{\phi}_p Z_1^* + \tilde{e}_{p+1}^* \\ Z_{p+2}^* &= \hat{\phi}_1 Z_{p+1}^* + \hat{\phi}_2 Z_p^* + \ldots + \hat{\phi}_p Z_2^* + \tilde{e}_{p+2}^* \\ &\vdots \\ Z_n^* &= \hat{\phi}_1 Z_{n-1}^* + \hat{\phi}_2 Z_{n-2}^* + \ldots + \hat{\phi}_p Z_{n-p}^* + \tilde{e}_n^* \end{aligned}$$

and set $Z_i^* = Z_i$, $i = 1, 2, \ldots, p$. (You can use one of the other methods, but that will change the algorithm slightly).

ii. We now have $Z_1^*, Z_2^*, \ldots, Z_n^*$. We can use these to repeat the procedure in step (A) and get $\hat{\phi}^* = \left(\phi_1^*, \phi_2^*, \ldots, \phi_p^*\right)$, the bootstrap estimates of the autoregression parameters.

iii. Repeat steps (i) and (ii) B times to obtain $\hat{\phi}^*(1), \hat{\phi}^*(2), \ldots, \hat{\phi}^*(B)$.

iv. We can now estimate the bootstrap estimate of standard deviation of the autoregression parameters as:

$$\hat{\sigma}_{i,B} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\phi}_i^*(b) - \hat{\phi}_i^*(.)\right)^2}, \; i = 1,2,\ldots,p,$$

where

$$\hat{\phi}_i^*(.) = \frac{1}{B}\sum_{b=1}^{B}\hat{\phi}_i^*(b).$$

The techniques discussed will be used in Chapter 4.

CHAPTER 3

TIME SERIES ANALYSIS

In Chapter 4, new tests involving the bootstrap method applied to time series data, will be developed. Therefore, it is necessary to present a brief summary of general time series theory. In Section 3.1 below a brief discussion of time series and related basic definitions are provided. In Section 3.2 the autocovariance and the autocorrelation functions are defined and discussed. In Section 3.3 the partial autocorrelation function (PACF) is discussed briefly. In section 3.4 the sample autocovariance and the sample autocorrelation are briefly discussed and in Section 3.5 a discussion of the sample PACF is given. In Section 3.6 the autoregressive (AR) process is highlighted with an example of the order one AR process. Section 3.7 concludes the chapter with the short introduction to the moving average (MA) process, and the definition of some time series frequently encountered in statistical practice.

3.1 Time series definitions

According to Wei (1990:1–2), a time series is a set of data which is collected chronologically, usually in terms of equally spaced time intervals. The intrinsic nature of a time series is that the observations are correlated and the time orders of observations are important. An important part of time series analysis is the selection of a suitable probability model for the data. Having chosen a model, the next steps are to estimate the parameters of interest, to perform goodness-of-fit tests and to use the fitted model to enhance the understanding of the mechanism generating the series.

The following four steps are important in time series analysis:

- Graphical examination of the series, checking in particular if there is a trend, seasonal components, any apparent sharp changes in behaviour and outlying observations;
- Remove the trend and seasonal components to obtain stationary residuals. These processes involve transformation (which is briefly discussed in Chapter 1, Section 1.4) and differencing the data;
- Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function which will be discussed later in this chapter;

- Do forecasting with the adequate model that has been achieved.

According to Wei (1990:6–10) the following definitions capture some of the key-stone basic concepts in time series analysis:

- A stochastic process is a family of time indexed random variables $X(\omega,t)$, where ω belongs to a sample space and t belongs to an index set.
- Consider a finite set of random variables $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ from a stochastic process $\{X(\omega,t): t = 0, \pm 1, \pm 2, \ldots\}$. Then the n-dimensional distribution function is defined by:

$$F\left(x_{t_1}, x_{t_2}, \ldots, x_{t_n}\right) = p\left(\omega : x(\omega,t_1) \le x_{t_1}, x(\omega,t_2) \le x_{t_2}, \ldots, x(\omega,t_n) \le x_{t_n}\right) \qquad (3.1)$$

- A process is said to be n^{th} – order stationary in distribution if its n-dimensional distribution function F is time invariant, i.e. if

$$F\left(x_{t_1}, x_{t_2}, \ldots, x_{t_n}\right) = F\left(x_{t_1+k}, x_{t_2+k}, \ldots, x_{t_n+k}\right), \qquad (3.2)$$

  for any integer $t_1, t_2, \ldots, t_n$ and k.
- A process is said to be strictly stationary if (3.2) is true for any n. In other words, shifting the time origin by an amount k has no effect on the joint distribution, which depends only on the intervals between $t_1, t_2, \ldots, t_n$.
- A process is said to be n^{th} order weakly stationary if all its joint moments up to order n exists and are time invariant.
- For a given real-valued process $\{X_t : t = 0, \pm 1, \pm 2, \ldots\}$, the mean function of the process is defined as:

$$\mu_t = E(X_t). \qquad (3.3)$$

- The variance function of the process is defined as:

$$\sigma_t^2 = E\left[(X_t - \mu_t)(X_t - \mu_t)\right]. \qquad (3.4)$$

- The covariance function between X_{t_1} and X_{t_2} is defined as:

$$\gamma(t_1,t_2) = E\left[\left(X_{t_1} - \mu_{t_1}\right)\left(X_{t_2} - \mu_{t_2}\right)\right]. \qquad (3.5)$$

- The correlation function between X_{t_1} and X_{t_2} is defined as:

$$\rho(t_1,t_2) = \frac{\gamma(t_1,t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}. \qquad (3.6)$$

In the following section we concentrate on the terms 'autocovariance and autocorrelation functions'.

3.2 The autocovariance and autocorrelation functions

The following brief introduction of these two important concepts, i.e., the autocovariance and autocorrelation functions, is based on the content of Wei (1990:10–11) and Fuller (1996:7–12).

For a stationary process $\{X_t\}$, $E(X_t)=\mu$ denotes the mean and $\operatorname{Var}(X_t)=E(X_t-\mu)^2=\sigma^2$ denotes the variance, which are both constants. If two variables X_{t_1} and X_{t_2} are considered, and if $t_1=s$ and $t_2=t$, then the covariances between two variables X_s and X_t denoted by $\operatorname{Cov}(X_s,X_t)$, are functions only of the time difference $|t-s|$, $s<t$. The joint distribution of X_s and X_t depends only on $|t-s|$, $s<t$, which is referred to as the lag. Thus the autocovariance function $\gamma(t_1,t_2)$ also depends on $|t-s|$, $s<t$. If $t_1=t$ and $t_2=t+k$, we denote the autocovariance function $\gamma(t_1,t_2)=\gamma(t,t+k)$ by:

$$\begin{aligned}\gamma_k &= E\left[(X_t-\mu)(X_{t+k}-\mu)\right]\\ &=\operatorname{Cov}(X_t,X_{t+k}).\end{aligned} \tag{3.7}$$

γ_k is referred to as the autocovariance coefficient at lag *k*. The size of an autocovariance coefficient depends on the unit in which X_t is measured.

It is useful to standardize the autocovariance function to produce a function called the autocorrelation function (ACF), which measures the correlation between X_t and X_{t+k}. If $t_1=t$ and $t_2=t+k$, then the ACF is denoted by:

$$\rho_k=\frac{\operatorname{Cov}(X_t,X_{t+k})}{\sqrt{\operatorname{Var}(X_t)}\sqrt{\operatorname{Var}(X_{t+k})}}=\frac{\gamma_k}{\gamma_0}, \tag{3.8}$$

where γ_k is the autocovariance defined in (3.7) above and $\gamma_0=\operatorname{Var}(X_t)=$ $\operatorname{Var}(X_{t+k})=\sigma^2$.

The properties of the autocovariance and ACF for a stationary process are as follows:

(a) $\gamma_0 = \mathrm{Var}(X_t) \Rightarrow \rho_0 = 1$.

(b) $|\gamma_k| \le \gamma_0 \Rightarrow |\rho_k| \le 1$.

(c) $\gamma_k = \gamma_{-k} \Rightarrow \rho_k = \rho_{-k}$, $\forall k$, i.e. γ_k and ρ_k are even functions and hence symmetric about the time origin, $k = 0$. This follows from the fact that the time difference between X_t and X_{t+k}, and X_t and X_{t-k} are the same. Therefore, the ACF is often plotted only for nonnegative lags, and the plot is referred to as a correlogram.

The lag k of the autocovariance γ_k and autocorrelation ρ_k is discrete if the time series is discrete and continuous if the time series is continuous.

In the following section the general concept of partial autocorrelation function is introduced. A necessary condition for a function to be the autocovariance or autocorrelation function of some process is that it should be positive semidefinite. It is now possible to define a useful concept, i.e. White noise processes.

A process $\{a_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $E(a_t) = \mu_a$, usually assumed to be 0, constant variance $\mathrm{Var}(a_t) = \sigma_a^2$ and $\gamma_k = \mathrm{Cov}(a_t, a_{t+k}) = 0$ for all $k \ne 0$. By definition, it immediately follows that a white noise process $\{a_t\}$ is stationary with the autocovariance function

$$\gamma_k = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \ne 0, \end{cases}$$

the autocorrelation function

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \ne 0, \end{cases}$$

and the partial autocorrelation function (also see the discussion in the following paragraph)

$$\phi_{kk} = \begin{cases} 1 & k = 0 \\ 0 & k \ne 0. \end{cases}$$

Since by definition $\rho_0 = \phi_{00} = 1$ for any process, when we talk about the autocorrelation and partial autocorrelations, we refer only to ρ_k and ϕ_{kk} for $k \neq 0$. The basic phenomenon of the white noise process is that the ACF and PACF are identically equal to zero. Although this process hardly ever occurs in applied time series, it plays an important role as basic building block in the construction of time series models.

A white noise process is Gaussian if its joint distribution is normal. In the following discussions, unless mentioned otherwise, $\{a_t\}$ is always referred to as a zero mean Gaussian white noise process.

3.3 The partial autocorrelation function (PACF)

According to Wei (1990:12–16), the partial autocorrelation is the correlation between X_t and X_{t+k} after the mutual linear dependency on the intervening variables $X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}$ has been removed. The PACF is defined by:

$$\operatorname{Corr}\left(X_t, X_{t+k} \middle| X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}\right) = \phi_{kk} = \frac{\operatorname{Cov}\left[\left(X_t - \hat{X}_t\right)\left(X_{t+k} - \hat{X}_{t+k}\right)\right]}{\sqrt{\operatorname{Var}\left(X_t - \hat{X}_t\right)}\sqrt{\operatorname{Var}\left(X_{t+k} - \hat{X}_{t+k}\right)}}, \quad (3.9)$$

where ϕ_{kk} denotes the PACF between X_t and X_{t+k} at lag *k*. $\hat{X}_t$ and $\hat{X}_{t+k}$ are the linear estimates given by $\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \ldots + \beta_{k-1} X_{t+k-1}$ and $\hat{X}_{t+k} = \alpha_1 X_{t+k-1} + \alpha_2 X_{t+k-2} + \ldots + \alpha_{k-1} X_{t+1}$, where β_i and α_i, $1 \leq i \leq k-1$ are the mean square linear regression coefficients obtained from minimizing $E\left(X_t - \hat{X}_t\right)^2$ and $E\left(X_{t+k} - \hat{X}_{t+k}\right)^2$ respectively.

In the next section, the estimators for the parameters of the autocovariance and the autocorrelation functions are derived, because a stationary time series is characterized by its mean μ, variance σ^2, autocorrelation ρ_k and partial autocorrelation ϕ_{kk}.

The exact values of these parameters can be calculated if the ensemble of all possible realizations is known. Otherwise, they can be estimated if multiple independent realizations are available. However, in most applications, it is difficult or impossible to obtain multiple realizations. Most available time series constitute only a single realization. This makes it impossible to calculate the ensemble average. However, for a stationary process we have a natural alternative of replacing the ensemble average by the time average. In the following, we examine conditions under which we can estimate the mean and autocovariance and hence the autocorrelations by using time average.

3.4 Estimation of the autocovariance and the autocorrelation

Unfortunately the ACF does not unequally specify a time series. It is possible to find other normal and non-normal processes with the same ACF (or very close to the same ACF), which makes interpretation of the sample behaviour difficult. However, the concept remains to be an important guide to the properties of a time series as well as the sample autocorrelation coefficients of an observed time series, which measure the correlation between observations at different distances apart. These coefficients often provide insight into the probability model which generated the data. To assess the degree of dependence in the data and to select a model for the data we use the sample ACF. If there is a belief that the data are realized values of a stationary time series $\{X_t\}$, then the sample ACF will be useful in estimating the ACF of $\{X_t\}$. This estimate suggests a way to decide on which of the other stationary time series models are suitable candidates to represent the dependence in the data.

Wei (1990:19–22) defines the sample autocovariance of realization $\{X_t : t = 1, 2, ..., n\}$ of a time series by:

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}\left[(X_t - \bar{X})(X_{t+k} - \bar{X})\right], \; k = 0, 1, 2, ..., n-1, \tag{3.10}$$

where $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$ denotes the sample mean of the time series.

The autocorrelations of the time series $\{X_t\}$ are defined in (3.8) and are estimated from an observed series $\{X_t : t = 1, 2, ..., n\}$ by the sample autocorrelations defined by:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}\left[\left(X_t - \bar{X}\right)\left(X_{t+k} - \bar{X}\right)\right]}{\sum_{t=1}^{n}\left(X_t - \bar{X}\right)^2}, \qquad k = 1,2,... \qquad (3.11)$$

$$= \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \hat{\rho}_{-k}.$$

Therefore, the sample ACF is symmetric about the origin $k = 0$.

In the following section the estimation of the PACF is discussed. This estimator is referred to as the sample PACF.

3.5 Sample PACF

The sample PACF of lag k is the autocorrelation between X_t and X_{t-k} that is not accounted for by lags 1 through $k-1$. Sample PACF of a realization $\{X_t : t = 1,2,...,n\}$ of a time series is defined by:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{i=1}^{k}\hat{\phi}_{ki}\hat{\rho}_{k+1-i}}{1 - \sum_{i=1}^{k}\hat{\phi}_{ki}\hat{\rho}_i}, \qquad (3.12)$$

where $\hat{\phi}_{k+1,i} = \hat{\phi}_{ki} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-i}$, $i = 1,2,...,k$. The sample PACFs $\hat{\phi}_{k+1,k+1}$ are consistent estimators of the theoretical partial autocorrelations.

3.6 Autoregressive (AR) process

Two useful ways of expressing a time series process exist. One representation implies writing a process X_t as a linear combination of a sequence of uncorrelated random variables, and the other involves writing the process in a way in which the value of X at time t is regressed on its own past values plus a random shock. We refer to the representations as the moving average and the autoregressive representations respectively.

In paragraph 2.5.1 the $\text{AR}(p)$ and $\text{AR}(1)$ processes were used in a bootstrap application.

In this section the general idea of AR processes is introduced and the first order AR process is defined. According to Fuller (1996:39–41) the autoregressive process of order p, also known as $\text{AR}(p)$ process, is given by:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t .$$

By introducing the backshift operator B^j by $B^j X_t = X_{t-j}$, it follows that $X_t = \phi_1 B^1 X_t + \phi_2 B^2 X_t + \ldots + \phi_p B^p X_t + \varepsilon_t$. Equivalently,

$$\phi_p(B) X_t = \varepsilon_t , \tag{3.13}$$

where $\phi_p(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \ldots - \phi_p B^p$. Also B is known as a lag operator.

Wei (1990:32) states that if $\sum_{k=1}^{p} |\phi_k| < \infty$ then the process is said to be invertible. For a process to be stationary the roots of $\phi_p(B) = 0$ must lie outside the unit circle. It is usually assumed that the error term $\{\varepsilon_t\}$ of (3.13) is normally distributed with mean zero and unit variance. This is usually done by standardizing the time series to unit variance.

The representation of the AR process plays an important role in understanding the mechanism of forecasting. AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock. An example is given below, where the order of the AR process is chosen to be equal to one.

3.6.1 First-order Gaussian AR process (AR(1)-process)

Let $\{X_t\}$ be a stationary series satisfying the following process:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad \text{for } t = 0, \pm 1, \pm 2, \ldots , \tag{3.14}$$

where $\{\varepsilon_t\} \sim N\left(0, \sigma_\varepsilon^2\right)$ independently, $\phi < 1$, X_s and ε_t are uncorrelated for $s < t$.

By taking the expectation on each side of (3.14), with $E(\varepsilon_t) = 0$, it follows that $E(X_t) = 0$. The autocovariance of the $\text{AR}(1)$ process is given by:

Since $\gamma_0 = \text{Var}(X_t)$, and $\text{Var}(X_t) = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2$, and $\gamma_1 = \phi_1 \gamma_0$,

$$\begin{aligned}\gamma_k &= \text{Cov}(X_t, X_{t-k}) \\ &= \text{Cov}(\phi_1 X_{t-1} + \varepsilon_t, X_{t-k}) \\ &= \phi_1 \text{Cov}(X_{t-1}, X_{t-k}) + \text{Cov}(\varepsilon_t, X_{t-k}) \\ &= \phi_1 \gamma_{k-1} \\ &\vdots \\ &= \phi_1^{|k|} \gamma_0\end{aligned}$$

Since $\gamma_0 = \text{Var}(X_t)$, and $\text{Var}(X_t) = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2$, and $\gamma_1 = \phi_1 \gamma_0$, we can conclude that

$$\gamma_k = \begin{cases} \phi_1 \gamma_1 + \sigma_\varepsilon^2, & k = 0 \\ \phi_1^{|k|} \gamma_0, & k = 1, 2, \ldots \end{cases} \tag{3.15}$$

The ACF can be found using the definition and the observation.

$$\gamma_k = \gamma_{-k},$$

which leads to the following:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1^{|k|} \gamma_0}{\gamma_0} = \phi_1^{|k|}, \quad k = 0, \pm 1, \ldots \tag{3.16}$$

From the linearity of the covariance function in each of its arguments and the fact that ε_t is uncorrelated with X_{t-1} it follows that:

$$\begin{aligned}\gamma_0 &= \text{Cov}(X_t, X_t) \\ &= \text{Cov}(\phi_1 X_{t-1} + \varepsilon_t, \phi_1 X_{t-1} + \varepsilon_t) \\ &= \phi_1^2 \text{Cov}(X_{t-1}, X_{t-1}) + \text{Cov}(\varepsilon_t, \varepsilon_t) \\ &= \phi_1^2 \gamma_0 + \sigma_\varepsilon^2, \text{ which complies with (3.15).}\end{aligned}$$

Hence, it follows that the variance of the first order AR process is:

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} = \sigma_X^2 \tag{3.17}$$

AR processes of higher order are more complicated regarding their analysis, and we will rather continue with the next paragraph, which contains a brief definition of the general concepts regarding the moving average process. Also, the problem of determining whether the process is stationary or nonstationary will be addressed.

3.7 Moving average (MA) process

The MA process of order q is denoted by $\mathrm{MA}(q)$ and described by:

$$\begin{aligned} X_t &= \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \\ &= \varepsilon_t - \left(\theta_1 B^1\varepsilon_t + \theta_2 B^2\varepsilon_t + \ldots + \theta_q B^q\varepsilon_t\right) \\ &= \left(1 - \left(\theta_1 B^1 + \theta_2 B^2 + \ldots + \theta_q B^q\right)\right)\varepsilon_t \\ &= \theta_q(B)\varepsilon_t \end{aligned} \tag{3.18}$$

where $\{\varepsilon_t\}$ is a purely random process with mean zero and variance σ_ε^2, $\{\theta_t\}$ are constants and B is the backshift operator.

Because $1+\theta_1^2+\ldots+\theta_q^2<\infty$, a finite moving average process is always stationary, according to Wei (1990: 46). This moving average process is invertible if the roots of $\theta_q(B)=0$ lie outside the unit circle. Box and Jenkins (1976) call a process invertible if it can be written in an autoregressive (AR) form. Wei (1990: 26) also explains the concept of invertibility. Wei (1990:47–53) proved that a stationary process that contains no deterministic component that can be forecast or predicted exactly from its own past can always be expressed in the form of (3.18). These processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time.

From Wei (1990: 52), for (3.18), the variance is given by:

$$\begin{aligned} \gamma_0 &= \mathrm{Cov}(X_t, X_t) \\ &= \sigma_\varepsilon^2 \sum_{t=0}^{q} \theta_t^2 \text{, where } \theta_0 = 1. \end{aligned} \tag{3.19}$$

The other autocovariances at lag k are given by:

$$\begin{aligned} \gamma_k &= \mathrm{Cov}(X_t, X_{t-k}) \\ &= \begin{cases} \sigma_\varepsilon^2\left(-\theta_k + \theta_1\theta_{k-1} + \cdots + \theta_{q-k}\theta_q\right) & k = 1,2,\ldots,q \\ 0 & k > q \end{cases}. \end{aligned} \tag{3.20}$$

As γ_k does not depend on the time t, and the mean is constant, the $\mathrm{MA}(q)$ process is second-order stationary for all values $\{\theta_t\}$.

The ACF of the $\mathrm{MA}(q)$ process at lag k is given by the following expression:

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k-1} + \cdots + \theta_{q-k}\theta_q}{1+\theta_1^2 + \cdots + \theta_q^2} & k = 1,2,\ldots,q \\ 0 & k > q \end{cases}. \tag{3.21}$$

A special property of the MA process is that the ACF *cuts off* at lag q.

3.8 Remarks

More useful time series representations exist. It is true that a stationary and invertible process can be represented either in a moving average form or in an autoregressive form. However, a problem with either representation is that it may contain too many parameters. This is true even for a finite order moving average and a finite order autoregressive model as it often takes a high order model for good approximation. In general, a large number of parameters reduce efficiency in estimation. Thus, in model building, it may be necessary to include both autoregressive and moving average terms in a model. This leads to the following useful mixed autoregressive moving average (ARMA) process:

$$\phi_p(B)X_t = \theta_q(B)\varepsilon_t, \tag{3.22}$$

where $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$.

We refer to this process as an $\mathrm{ARMA}(p,q)$ process, in which p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively. The stationary and invertible ARMA process can be written in an autoregressive representation as:

$$\pi(B)X_t = \varepsilon_t,$$

where $\pi(B) = \dfrac{\phi_p(B)}{\theta_q(B)} = \left(1 - \pi_1 B - \pi_2 B^2 - \cdots\right)$. This process can also be written as a moving average representation:

$$X_t = \psi(B)\varepsilon_t,$$

where $\psi(B) = \dfrac{\theta_q(B)}{\phi_p(B)} = \left(1 + \psi_1 B + \psi_2 B^2 - \cdots\right)$.

The stationarity conditions depend on the autoregressive part, i.e., the roots of $\phi_p(B)=0$ should be larger than one. The invertibility conditions only depend on the moving average part, i.e., the roots of $\theta_q(B)=0$ should also be larger than one. Many useful series exist in this family, for example, the $\text{ARMA}(1,1)$ process.

In model building, it may be necessary to include both autoregressive and moving average terms in a model, which leads to a mixed autoregressive moving average (ARMA) process: $\phi_p(B)X_t=\theta_q(B)a_t$ where $\phi_p(B)=1-(\phi_1 B+\ldots+\phi_p B^p)$ and $\theta_q(B)=1-(\theta_1 B+\ldots+\theta_q B^q)$. Properties of this model are discussed in Wei (1990:56–64).

Several nonstationary models exist, such as the ARIMA models, which are defined by:

$$\phi_p(B)(1-B)^d X_t=\theta_q(B)\varepsilon_t, \tag{3.23}$$

where the stationary AR operator $\phi_p(B)=1-\phi_1 B-\cdots-\phi_p B^p$ and the invertible MA operator $\theta_q(B)=1-\theta_1 B-\cdots-\theta_q B^q$ share no common factors.

$(1-B)^d X_t$ is a "differenced series". To understand the meaning of the term "differencing", we consider the following: Although many time series are nonstationary, due to some equilibrium forces, different parts of these series behave very much alike except for their difference in the local mean levels. Box and Jenkins (1976: 85) refer to this kind of nonstationary behavior as homogeneous nonstationary. In terms of the ARMA models, the process is nonstationary if some roots of its AR polynomial do not lie outside the unit circle. However, by the nature of homogeneity, the local behavior of this kind of homogeneous nonstationary series is independent of its level. Hence, by letting $\Psi(B)$ be the autoregressive operator describing the behavior, we have

$$\Psi(B)(X_t+C)=\Psi(B)X_t$$

for any constant C. This implies that $\Psi(B)$ must be of the form

$$\Psi(B)=\phi(B)(1-B)^d,$$

for some $d>0$ where $\phi(B)$ is a stationary autoregressive operator. Thus, a homogeneous nonstationary series can be reduced to a stationary series by taking a suitable difference of the general series. In other words, the series $\{X_t\}$ is nonstationary but its *d*-th differenced series, $\left\{(1-B)^d X_t\right\}$ for some integer $d \geq 1$, is stationary. For example, if the *d*-th differenced series follows a white noise phenomenon, we have

$$(1-B)^d X_t = \varepsilon_t.$$

The parameter θ_0 plays very different roles for $d=0$ and $d>0$. When $d=0$, the original process is stationary. However, when $d \geq 1$, θ_0 is called the deterministic trend term and is often omitted from the model unless it is really needed. The resulting homogeneous nonstationary model in (3.23) has been referred to as the autoregressive integrated moving average model of order (p,d,q) and is denoted as the ARIMA(p,d,q) model. When $p=0$, the ARIMA(p,d,q) model is also called the integrated moving average model or order (d,q) and is denoted as the IMA(d,q) model. The Random Walk Model, the ARIMA$(0,1,1)$ also known as the IMA$(1,1)$ model belongs to this class of models.

All or some of these models may be applicable in Chapters 4 and 5.

CHAPTER 4

NEW TESTS INVOLVING THE BOOTSTRAP METHOD

4.1 Introduction

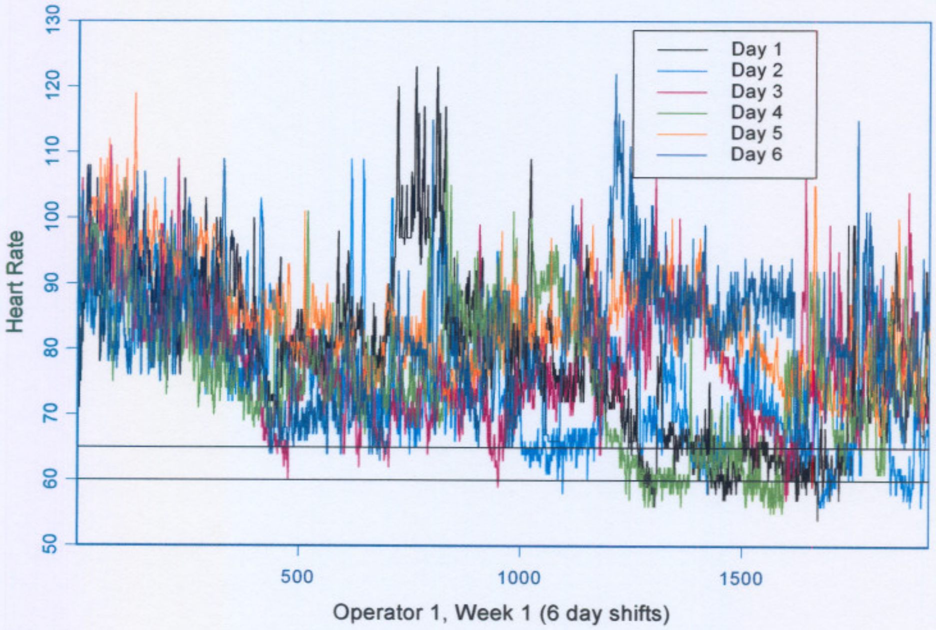
The purpose of this chapter is to determine tests, parametric and non-parametric, for equality of trends against any suitable alternative hypothesis, in two or more time series, i.e. when the usual i.i.d. assumptions and distribution assumptions usually associated with two sample comparisons, are relaxed. The new proposed tests will involve the bootstrap method, and we also wish to determine the power of the new tests.

The motivation for developing tests of this nature was founded when data, obtained from operators of expensive equipment working in a quarry, were analysed. In order to understand the applicability of these types of tests, a brief discussion of the practical situation of concern, now follows. Similar situations are abundant in various analytical fields.

Data collection was designed to gain information on critical factors which would be reflective of operator fatigue, since the objective of the analysis was to develop a fatigue management program for the industry concerned. Possible causal factors of fatigue had to be identified and investigated in order to establish possible trends and recognition of fatigue predictors so that effective interventions could be designed. Data collection comprised of various sub components, including physiological functioning. For example, *heart rates* were measured using polar watches, as well as other variables, including *dietary habits, sleeping patterns, environmental factors* such as *noise*, *light* and *temperature* in the work place and at home, *psychological factors* such as concentration, reaction time and information processing speed. The necessary design ethics and risk factors were in place and sampling was done randomly. Since the client was interested in possible ways to minimise accidents due to decreasing alertness of the operators, data related to "alertness" were studied. The factor "Work Shifts" (i.e. the duration, schedule, etc.) were suspected of influencing non-alertness and accident risk. Various complicated computer-based tests of alertness failed in the circumstances because of unforeseen practical matters out of control, but the heart beats of the operators were available and useable.

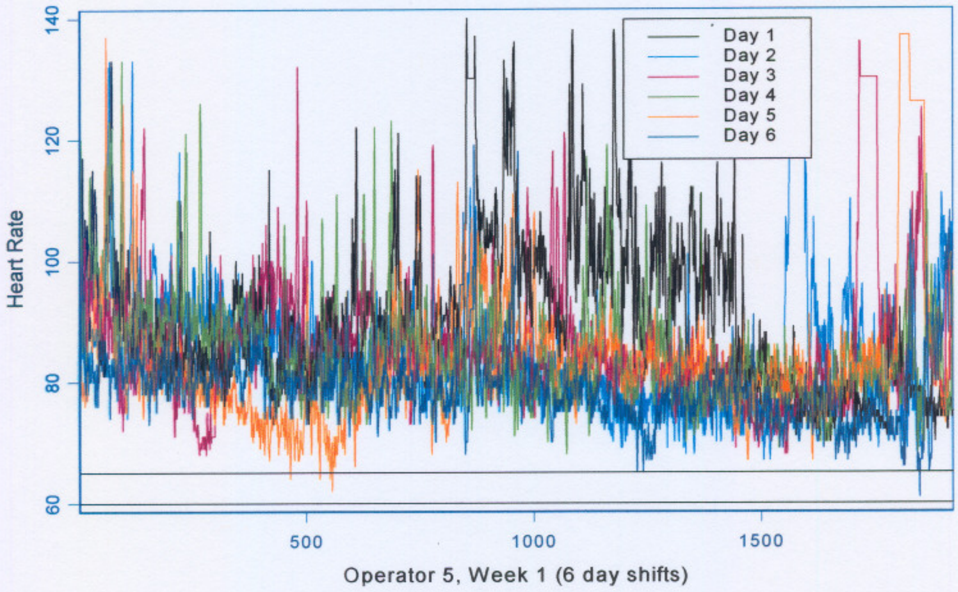
It is assumed in the health-related literature that for normal, healthy workers in the situation of interest, heart beat values below 65 are associated with increasing lack of alertness and potential accidents. Operator heart beats were measured every 15 seconds while on duty, for three weeks, during which the workers all worked on several day shifts, night shifts and morning shifts of 8 hours each. Several shift combinations were put to test. To indicate the influence of shifts (day, night and morning shifts) on the fatigue-patterns and behaviour of the operators during working hours, first glance graphical indications such as the examples below, indicated that operators may indeed not be alert and capable of handling the equipment in many situations. It is also true that trends are not particularly useful since accidents occur at individual points when heart beats drop (i.e. when an operator falls asleep) which depend on various factors such as 'time into shift', 'different alertness tolerance of operators', and many other factors mentioned above regarding off shift behaviour and circumstances. The following graphs are examples of typical behaviour of the heart beats of operators and shift related patterns.

Fig. 1: Time series representing the day shift heart beats of Operator1, say for week1.



Operator 5 below seems to be by day a 'safer' worker than the first, since he stays alert, except for an occasional dip into unalertness (which may result in an accident – only one accident is necessary to establish loss).

Fig. 2: Time series representing the day shift heart beats of Operator 5, for week 1.



The night shifts (22:00-06:00) for each operator, presented by graphs similar to Fig. 1 and 2 above, revealed similar differences between the behaviour of the operators, but also appears to be different from the day-shift data as is evident from Fig. 3 and 4.

Fig. 3: Time series representing the night shift heart beats of Operator 1, for week 2.

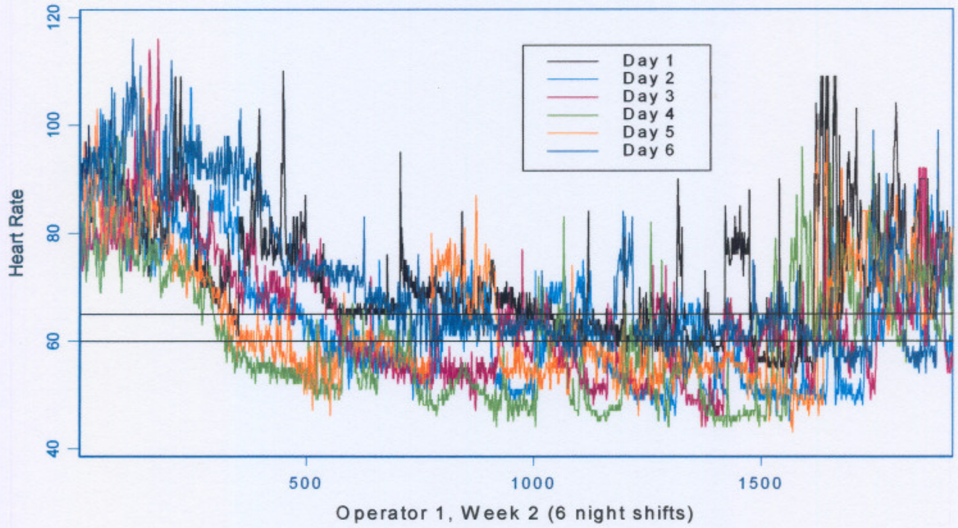
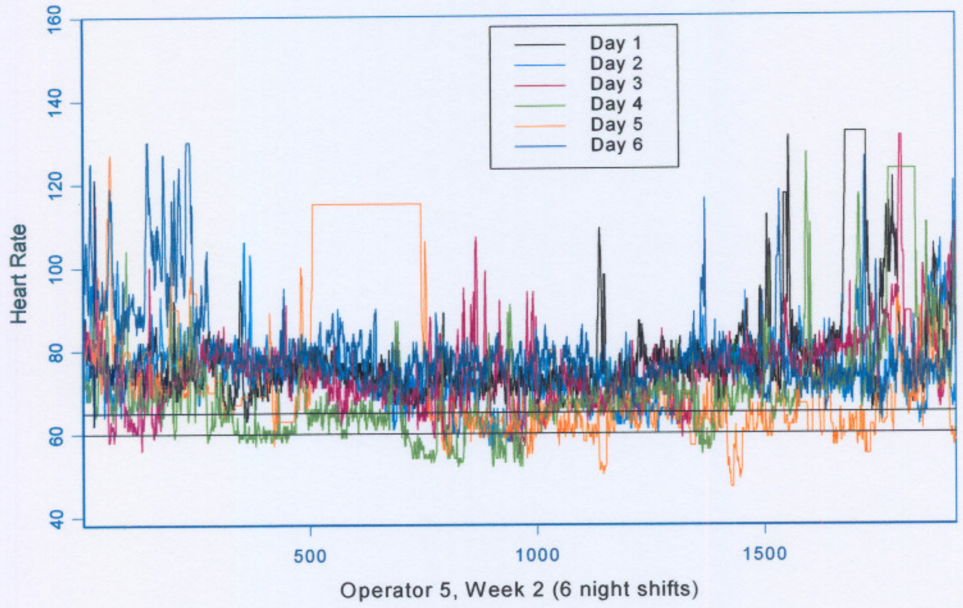


Fig. 4: Time series representing the night shift heart beats of Operator 5, for week 2.



- The graphs above reveal interesting potential hypotheses to be tested, such as: Accidents are more likely to occur during night shifts than during day shifts in general or for a specific operator. Or: An operator is equally alert during morning, day and night shifts. Or: Operators are more alert during morning shift than during night shifts and many more.
- The heart rates of some operators (operator 1 for example) are low very early in the shifts. This tendency lasts throughout the week. This tendency differs from operator to operator.

From similar comparative graphs the way in which the individual operators react to day shifts, night shifts and morning shift from days 1 – 6 of the different shifts as the weeks progressed, forms a picture. For example, the graphs show that throughout the study, the night shifts present problem zones, for all the operators. Furthermore it seems that the day shifts become more risk prone towards midweek. Some of the operators experienced problems at the beginning of the morning shifts as well, but in general it seems that the morning shifts were less risky as far as fatigue related accidents are concerned.

It is clear, that most of the hypotheses of concern can be formulated in the form:

$H_0 : \mu_1 = \mu_2$, where μ_i represents the mean of the i^{th} population of interest.

Also, multiple comparisons such as $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = ...$ are of interest, which is not part of the present study.

More specific hypotheses can be formulated, such as:

- $H_0 : \mu_1^{OP1} = \mu_2^{OP1}$, where $\mu_1^{OP1} \equiv$ mean heart beat of operator 1 during situation 1, and μ_2^{OP1} is the mean heart beat of Operator 1 in situation 2. Also
- $H_0 : \mu_1^{OP1} = \mu_2^{OP2}$, where $\mu_1^{OP1} \equiv$ mean heart beat of operator 1 during situation 1, and μ_2^{OP2} is the mean heart beat of Operator 2 in situation 2.

It appeared that serial correlation was present within each data set as well as other dependence structures in the two data sets of concern, whichever hypothesis was tested, were indeed present in all cases. By ignoring serial correlation and dependence structures, wrong conclusions may result, if the usual parametric or non-parametric two sample tests are applied.

The following parametric tests are not applicable on the data because the two sets of data are not independent in some cases, and because of the autocorrelation structures in each time series dataset: The *z*-test, *t*-test, Mann-Whitney *U* test, the *t*-test for dependent samples and Wilcoxon matched-pair sign-rank test for dependent samples.

4.2 Relating methods from literature

Several testing procedures have been explored in the literature recently, utilizing different statistical viewpoints in comparing populations or characteristics of populations, in situations were the usual tests mentioned above, are not applicable.

(i). For example: McGregor and Babb (1989) studied the behaviour of *F* test-statistics based on the differences of two time series, having nonnegative trends. These trends are taken to be proportional, and the observations were observed at the same time points where the differences are computed. The object of the test is to determine the equality of trends against inequality of trends provided the paired differences are independently and identically distributed normal variables. In more detail, the two time series have nonnegative trends which are proportional, as is seen from the following formulations.

Let $t_1 < \ldots < t_n$ be a fixed set of equispaced time points at which both series are observed. Considering two time series given by:

$$Y_i = \mu_i + \varepsilon_i \text{ (treated as a Control series) and}$$

$$Z_i = \pi_i \mu_i + \eta_i \text{ (treated as an Experimental series),}$$

where $\mu_i \equiv \mu(t_i)$ is a non-random nonnegative function of t_i, $\pi_i = \pi(t_i)$ is an unknown nonnegative parameter, and both $\varepsilon_i = \varepsilon(t_i)$ and $\eta_i = \eta(t_i)$ are random error terms with zero mean for all $t_i, i = 1, 2, \ldots, n$.

If the proportionality constant $\pi_i < 1$, then relative to the series Y_i the series Z_i is reduced, otherwise if proportionality constant $\pi_i > 1$ then Z_i is increased, $i = 1, 2, \ldots, n$.

By denoting the difference between the two time series, by X_i, we have:

$$\begin{aligned} X_i &= Y_i - Z_i \\ &= (\mu_i + \varepsilon_i) - (\pi_i \mu_i + \eta_i) \\ &= \mu_i + \varepsilon_i - \pi_i \mu_i - \eta_i \\ &= \mu_i (1 - \pi_i) + \varepsilon_i - \eta_i. \end{aligned}$$

Let $\delta_i = \varepsilon_i - \eta_i$, $i = 1, 2, \ldots, n$. Then

$$X_i = (1 - \pi_i)\mu_i + \delta_i, \; i = 1, 2, \ldots, n.$$

Ali (1973) found the exact distribution of:

$$T = \frac{\overline{X}}{S_X / \sqrt{n}},$$

if $\{X_i\}$, $i = 1, 2, \ldots, n$ follow a multivariate normal distribution with mean zero and the covariance matrix diagonal elements being unity and off diagonal elements being the correlation ρ. Furthermore, he showed that the tails of the distribution of T are thicker than those of the standard $(\rho = 0)$ t-distribution, for $(\rho > 0)$, and they are thinner for $(\rho < 0)$.

If δ_i are independent $N\left(0,\sigma^2\right)$ and if $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ denotes the sample mean and $S_X^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ denotes the sample variance, then McGregor and Babb (1989) showed that:

$$E\left(n\overline{X}_n^2\right) = \sigma^2 + n^{-1}(1-\pi)^2\left(\sum \mu_i^2 + 2\sum_{i<j}\mu_i\mu_j\right)$$

$$E\left(S_X^2\right) = \sigma^2 + n^{-1}(1-\pi)^2\left(\sum \mu_i^2 + 2(n-1)^{-1}\sum_{i<j}\mu_i\mu_j\right).$$

If $\pi = 1$, then both $n\overline{X}_n^2$ and S_X^2 are unbiased estimators of σ^2 and therefore

$$F = \frac{nX_n^2}{S_X^2} \sim F(1, n-1).$$

Since the population means $\{\mu_i\} \geq 0$, if $\pi \neq 1$, $n\overline{X}_n^2$ will be larger than $E\left(S_X^2\right)$, which is clearly a motivation of using the F-test to test the null hypothesis $\pi = 1$ against the alternative hypothesis $\pi \neq 1$.

Suppose the errors for the two given time series $\{\varepsilon_i\}$ and $\{\eta_i\}$ are governed by first order autoregressive (AR(1)) processes $\varepsilon_i = \rho\varepsilon_{i-1} + f_i$ and $\eta_i = \rho\eta_{i-1} + g_i$ respectively, with the same autoregressive parameter ρ, where $0 < |\rho| < 1$ and the $\{f_i\}$ and $\{g_i\}$ are sets of independent normal random variables with zero means. Then, the difference of the two AR(1) processes given by:

$$\delta_i = \rho\delta_{i-1} + e_i,$$

where $\{e_i\}$ are independent normal random variables with zero mean and assume without loss of generality that they have unit variances.

Under the conditions that the processes are governed by an AR(1) process, a method similar to that of Daniels (1956) is used for deriving the approximated large-sample distribution of F when $\pi = 1$, and it is found that after transformation and some mathematics, McGregor and Babb (1989) found that the approximate probability density function of F is:

$$d(f)=\sqrt{\left\{\frac{1-\rho^2}{(n-1)\pi}\right\}}\left[\Gamma\left(\frac{n-1}{2}\right)\right]^{-1}\Gamma\left(\frac{n}{2}\right)\sqrt{f^{-1}u^n}\left(\frac{1-\rho u}{1-\rho^2 u}\right)\left\{1+O\left(n^{-1}\right)\right\},$$

where $0<\mathrm{u}<1,\ 0<|\rho|<1$ and

$$2(n-1)\rho^2 u=(n-1)\left(1-\rho^2\right)+f(1-\rho)^2-\left[\left\{(n-1)\left(1-\rho^2\right)+f(1-\rho)^2\right\}^2-\left\{2(n-1)\rho\right\}^2\right]^{\frac{1}{2}}.$$

Using the L'Hopital's Rule, McGregor and Babb (1989) showed that $\lim_{\rho\to 0} d(f)=g(x)$, where $g(x)$ is the exact probability density function of F, with 1 and $(n-1)$ degrees of freedom, i.e.,

$$g(x)=\Gamma\left(\frac{n}{2}\right)\left[\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)\right]^{-1}\sqrt{\frac{1}{(n-1)x}}\left(\frac{n-1}{n-1-x}\right)^{\frac{n}{2}} \text{ with}$$

$$\Gamma(n)=\int_0^{\infty} x^{n-1}e^{-x}dx.$$

(ii). In another interesting paper, Aminzadeh (1999) combines two correlated time series $\mathbf{Z}$ and $\mathbf{Y}$ in the sense of estimating $P(\mathbf{Z}<\mathbf{Y})$, especially with the aim of future prediction. The main purpose of the paper is to estimate the probability that, given the values of the two time series in the past, $\mathbf{Z}<\mathbf{Y}$ at a specific time in the future. Aminzadeh (1999) states the following:

Let $\mathbf{Z}=Z_1,Z_2,\ldots,Z_n$ and $\mathbf{Y}=Y_1,Y_2,\ldots,Y_m$ represent observations of two correlated time series $\mathbf{Z}$ and $\mathbf{Y}$, respectively. Let Z_{n+L} and Y_{m+k} denote values of $\mathbf{Z}$ and $\mathbf{Y}$ at the future time $n+L$ and $m+k$ respectively, for $L,k=1,2,3,\ldots$. Often in practice people face the problem of estimating $Q=P\left(Z_{n+L}<Y_{m+k}\,|\,\mathbf{Z},\mathbf{Y}\right)$, where $n+L=m+k$. In this paper Aminzadeh assumed that values of the random variables $\mathbf{Z}$ and $\mathbf{Y}$ changes over time.

Three models of stationary time series are used to represent behavior of the time series $\mathbf{Z}$ and $\mathbf{Y}$, namely AR, MA and ARMA processes. Furthermore, it is assumed that the white noise process associated with the series $\mathbf{Z}$ and $\mathbf{Y}$, has a bivariate normal distribution. The behaviour of the $\mathbf{Z}$ and $\mathbf{Y}$ series is represented by the same model, and estimation of Q is of interest.

(a) Autoregressive models

Let the series $\mathbf{Z}$ and $\mathbf{Y}$ be governed by $\mathrm{AR}(p)$ processes given by $Z_t-\mu=\sum_{i=1}^{p}\phi_i(Z_{t-i}-\mu)+a_t$ and $Y_t-\xi=\sum_{i=1}^{p}\theta_i(Y_{t-i}-\xi)+b_t$ respectively, where, ϕ_i, θ_i for $i=1,2,...,p$ are the AR parameters and a_t, b_t are white-noise processes with $E(a_t)=E(b_t)=0$, $\mathrm{Var}(a_t)=\sigma_Z^2$, $\mathrm{Var}(b_t)=\sigma_Y^2$ and $\mathrm{Cov}(a_t,a_{t\pm s})=\mathrm{Cov}(b_t,b_{t\pm r})=0$ for $r,s\neq 0$.

Since correlation exists between $\{a_t\}$ and $\{b_t\}$, it is assumed that the joint probability distribution of a_t and b_t is a bivariate normal distribution, with

$$\rho=\frac{\mathrm{Cov}(a_{t\pm s},b_{t\pm r})}{\sigma_Z\sigma_Y},\ \forall\ r \text{ and } s.$$

Let $Z_n(L)=E(Z_{n+L}|\mathbf{Z})$ and $Y_m(k)=E(Y_{m+k}|\mathbf{Y})$. Using the $\mathrm{AR}(p)$ processes, it is found that:

$$Z_n(L)=\mu+\sum_{i=1}^{p}\phi_i(Z_n(L-i)-\mu) \text{ and } Y_m(k)=\xi+\sum_{i=1}^{p}\theta_i(Y_m(k-i)-\xi).$$

Furthermore, let $e_n(L)$ and $\varepsilon_m(k)$ represent the differences between actual and expected values of $\mathbf{Z}$ and $\mathbf{Y}$ at times $n+L$ and $m+k$ respectively, denoted by:

$$e_n(L)=Z_{n+L}-Z_n(L) \text{ and } \varepsilon_m(k)=Y_{m+k}-Y_m(k),$$

which represent the residuals of the series at future times and it is shown that $e_n(L)=\sum_{v=0}^{L-1}R_v a_{n+L-v}$ and $\varepsilon_m(k)=\sum_{w=0}^{k-1}\delta_w b_{m+k-w}$, where

$$R_0=1, R_1=\phi_1,\ ...,R_v=\sum_{i=1}^{p}\phi_i R_{v-i} \text{ and } \delta_0=1,\delta_1=\theta_1,\ ...,\delta_w=\sum_{i=1}^{p}\theta_i\delta_{w-i}.$$

Using the assumption for the white noise processes, a_t and b_t, the following results are derived:

$\operatorname{Var}(e_n(L)) = \operatorname{Var}(Z_{n+L}) = \sigma_Z^2 \sum_{v=0}^{L-1} R_v^2$, $\quad \operatorname{Var}(\varepsilon_m(k)) = \operatorname{Var}(Y_{m+k}) = \sigma_Y^2 \sum_{w=0}^{k-1} \delta_w^2$

and $\operatorname{Cov}(e_n(L), \varepsilon_m(k)) = \rho\sigma_Z\sigma_Y \sum_{v=0}^{L-1} \sum_{w=0}^{k-1} R_v \delta_w$.

Since $e_n(L)$ and $\varepsilon_m(k)$ are both linear combinations of normal variables, their marginal distributions are normal. Therefore, Z_{n+L} and Y_{m+k} are normally distributed. Hence, $Q = P(Z_{n+L} < Y_{m+k} | \mathbf{Z}, \mathbf{Y}) = \Phi(\delta)$, where

$\delta = \dfrac{Y_m(k) - Z_n(L)}{\sqrt{\operatorname{Var}(e_n(L)) + \operatorname{Var}(\varepsilon_m(k)) - 2\operatorname{Cov}(e_n(L), \varepsilon_m(k))}}$, with Φ the cumulative distribution function of the standard normal distribution.

Note that the estimation of Q requires estimators of the parameters involved in δ. For all the models that follow, the same procedure is used to determine the quantity δ.

(b) Moving average processes (MA)

The moving average processes of order q for the series $\mathbf{Z}$ and $\mathbf{Y}$ are defined by Aminzadeh (1999) as: $Z_t - \mu = a_t - \sum_{j=1}^{q} \gamma_j a_{t-j}$ and $Y_t - \xi = b_t - \sum_{j=1}^{q} \varphi_j b_{t-j}$, where γ_j, φ_j for $j = 1, 2, ..., q$ are MA parameters. For these models it was shown that $Q = P(Z_{n+L} < Y_{m+k} | \mathbf{Z}, \mathbf{Y}) = \Phi(\delta)$, where

$$\begin{aligned}\delta &= \frac{Y_m(k) - Z_n(L)}{\sqrt{\operatorname{Var}(Z_{n+L}) + \operatorname{Var}(Y_{m+k}) - 2\operatorname{Cov}(Z_{n+L}, Y_{m+k})}} \\ &= \frac{(\xi - \mu)}{\sqrt{\sigma_Z^2\left(1 + \sum_{j=1}^{q} \gamma_j^2\right) + \sigma_Y^2\left(1 + \sum_{j=1}^{q} \varphi_j^2\right) - 2\rho\sigma_Z\sigma_Y \sum_{j=0}^{q} \sum_{j=0}^{q} \gamma_j \varphi_j}}.\end{aligned}$$

(c) Autoregression moving average processes (ARMA)

The autoregressive moving average processes of order p and q for the $\mathbf{Z}$ and $\mathbf{Y}$ series are:

$$Z_t - \mu = \sum_{i=1}^{p} \phi_i \left(Z_{t-i} - \mu \right) - \sum_{j=1}^{q} \gamma_j a_{t-j} + a_t \text{ and } Y_t - \xi = \sum_{i=1}^{p} \theta_i \left(Y_{t-i} - \xi \right) - \sum_{j=1}^{q} \varphi_j b_{t-j} + b_t,$$

where $\phi_i, \theta_i, \gamma_j, \varphi_j$ for $i = 1, 2, ..., p; j = 1, 2, ..., q$ are ARMA parameters. Hence, $Q = P\left(Z_{n+L} < Y_{m+k} \middle| \mathbf{Z}, \mathbf{Y} \right) = \Phi(\delta)$, where

$$\delta = \frac{Y_m(k) - Z_n(L)}{\sqrt{\mathrm{Var}\left(Z_{n+L}\right) + \mathrm{Var}\left(Y_{m+k}\right) - 2\mathrm{Cov}\left(Z_{n+L}, Y_{m+k}\right)}}$$

$$= \frac{(\xi - \mu) + \sum_{i=1}^{p} \theta_i \left(Y_m(k-i) - \xi \right) - \sum_{i=1}^{p} \phi_i \left(Z_n(L-i) - \mu \right)}{\sqrt{\mathrm{Var}\left(e_n(L)\right) + \mathrm{Var}\left(\varepsilon_m(k)\right) - 2\rho\sigma_Z\sigma_Y \mathrm{Cov}\left(e_n(L), \varepsilon_m(k)\right)}}.$$

Similar expressions are derived for other time series models. Often in practice the condition of the stationary time series is not valid. In this case, the difference transformations are performed on each series separately to move closer to stationary conditions.

(iii). Ducharme and Ledwina (2003) developed an efficient and adaptive nonparametric test for the two-sample problem when two independent samples are available:

$$X_i \sim \begin{cases} F(\cdot), & i = 1, \ldots, m \\ G(\cdot), & i = m+1, \ldots, N \end{cases},$$

where F and G are unknown continuous distribution functions. The testing problem is $H_0 : F = G$ vs. $H_1 : F \neq G$.

(iv). Alonso and Maharaj (2006) propose a procedure based on subsampling for testing the equality of the generating processes of two stationary time series that may or may not be assumed to be independently generated. Basawa et al. (1984), Maharaj (1996), Guo (1999) and Maharaj (2000) also proposed test procedures which involve the estimation of models. Coates and Diggle (1986), Swanepoel and VanWyk (1986) and Timmer et al. (1999) developed test procedures based on spectral estimation. With the exception of Maharaj (2000), the above-mentioned methods are only applicable to series that are assumed to be independently generated from each other. We will not discuss these methods in more detail.

4.3 Developing the new tests

Keeping the previous two paragraphs in mind, i.e. knowing which tests are not applicable, and after studying and searching for test procedures to test for equality of the means, new tests involving non-parametric techniques are developed below. The new methods can be clarified under the category "model-based resampling methods", discussed in Chapter 2, paragraph 2.5. Resampling is done on residuals and is based on specific time series models, as described in Davison and Hinkley (1997, 389–391).

The criteria set for the new methods involve satisfactory accurate significance levels and high power.

In Section 4.4 below, a brief discussion on the term "serial correlation" will be given. In Section 4.5, a test involving two independent samples $\mathbf{X}$ and $\mathbf{Y}$ will be discussed, where both $\mathbf{X}$ and $\mathbf{Y}$ are serially correlated internally. Tests for two dependent samples $\mathbf{X}$ and $\mathbf{Y}$, where both $\mathbf{X}$ and $\mathbf{Y}$ are serially correlated sets of data, will be discussed in Section 4.6.

4.4 Serial Correlation

Usually, in regression models, it is assumed that the error terms are independent of each other. In most applications involving time series data, for example in macro-economic and financial variables, this assumption is no longer valid, due to many reasons such as the "cyclical" nature of the data.

According to standard literature, when errors associated with observations of different time periods are related to each other or when the observations themselves are associated with observations of different time periods, we refer to the phenomenon as serial correlation. The ACF defined in (3.8) at lag k, expresses numerically the value of the serial correlation.

Most often it is the case that errors associated with adjacent observations are correlated, and errors for observations which are "far apart" are not. Thus serial correlation is only defined for data sets exhibiting a natural ordering of some sort. One example of some type of serial correlation is referred to as first-order serial correlation. This means that errors in one time period are directly correlated with errors in the next time period (or the previous time period). In most applications, this type of serial correlation will be positive and that should lead to interesting discussions. Usually discussions of serial correlation will be parallel to discussions of heteroscedasticity.

Regarding serial correlation, answers are usually demanded for the following three questions:

1. What are the consequences of serial correlation in data analysis, especially regarding estimation procedures, hypothesis testing, etc?
2. How can corrections be made to the estimation procedure to allow for serial correlation?
3. Are there tests available to test for the presence of serial correlation in the data set?

According to Kutner et al. (2005: 490 – 498), the effects of Serial Correlation on Least Squares Estimates are serious. If serial correlation is present in the data, least squares estimators will still be unbiased, but no longer Best Linear Unbiased Estimators (BLUE). Moreover, in the case of positive serial correlation, the estimates of standard errors will be lower than they should be (i.e. they will be biased downward). This will have consequences for confidence intervals which will be too short, and the null hypothesis will be rejected too often. Also R^2 values will be higher than they should be, and the estimator of the error variance will be smaller than it should be.

The following correction procedure for serially correlated data can be considered and applied:

Consider the following multiple regression model which allows for the order serial correlation:

$y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t$, where $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, $0 \le |\rho| < 1$, and it is assumed that $v_t \sim N\left(0, \sigma_v^2\right)$, $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$ and that v_t is independent of ε_t but ε_t is not independent of other error terms, such as the ε_{t-1}. It is clear that the value of the error term in the current period is determined by reducing the value of the error in the previous period and adding an effect of a different random variable with mean equal to zero. This type of error behaviour is referred to as a first order autoregressive (AR(1)) process, as we have seen in Chapter 3. The questions to be answered are:

- How does the value of the error in the current time period affect the error value in other periods?
- Can serial correlation be removed from the data?

To answer these questions, note that it has been proved that $\operatorname{Var}(\varepsilon_t) = \dfrac{\operatorname{Var}(v_t)}{1-\rho^2}$ (see Neter el al, 2005: 487) and the correlation between error terms fades as the time periods move further apart, i.e.

$\operatorname{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho \times \operatorname{Var}(\varepsilon_t)$, $\operatorname{Cov}(\varepsilon_t, \varepsilon_{t-2}) = \rho^2 \times \operatorname{Var}(\varepsilon_t)$ and

$\operatorname{Cov}(\varepsilon_t, \varepsilon_{t-3}) = \rho^3 \times \operatorname{Var}(\varepsilon_t)$, and so on.

Furthermore, suppose ρ were known. There is a simple procedure for transforming the data so that serial correlation is no longer present.

Consider an observation from the model: $y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t$ in the previous time period, i.e., $y_{t-1} = \beta_1 + \beta_2 x_{2t-1} + \cdots + \beta_k x_{kt-1} + \varepsilon_{t-1}$. Multiply this equation by ρ to obtain $\rho y_{t-1} = \beta_1 \rho + \beta_2 \rho x_{2t-1} + \cdots + \beta_k \rho x_{kt-1} + \rho \varepsilon_{t-1}$. Subtract this equation from the original model equation observed in time period t. The following equation is obtained:

$$y_t^* = \beta_1 (1-\rho) + \beta_2 x_{2t}^* + \cdots + \beta_k x_{kt}^* + v_t,$$

where $y_t^* = y_t - \rho y_{t-1}$, $x_{jt}^* = x_{jt} - \rho x_{jt-1}$, $j = 2,3,\ldots,k$ and $v_t = \varepsilon_t - \rho \varepsilon_{t-1}$.

For the transformed model, the error term v_t is serially uncorrelated, so that the assumptions of the Gauss Markov theorem are once again satisfied. Thus if we run least squares on the transformed model, estimates of the β_i, $i = 1,2,\ldots,k$ will be BLUE.

Note that the transformation $y_t^* = y_t - \rho y_{t-1}$ usually can not be applied in practice, since the value of ρ is unknown. However, the following iterative practical procedure can be employed (see Neter et al., 2005: 492):

A. Run ordinary least squares, and obtain residual values $\hat{\varepsilon}_t$, $t = 0,1,2,\ldots,n$.

B. Regress the following model:

$$\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + v_t ,$$

and obtain the least squares estimate $\hat{\rho}$.

C. Using the value of $\hat{\rho}$, transform the data the way discussed above and run least squares on the transformed data. The Durbin-Watson test is then employed to test whether the error terms for the transformed model are uncorrelated. If the test indicates that they are uncorrelated, the procedure terminates. Otherwise, go back to B.

If the sample size is large, the estimates obtained from this procedure should be close to the estimates obtained if ρ were known.

A very popular test for serial correlation is the Durbin-Watson test. It is based on the intuitive idea that if there is first order serial correlation present, the sample correlation between the residual in period t and period $t-1$ should be far away from 0.

The Durbin Watson test statistic is defined as:

$$DW = 2(1-\hat{\rho}),$$

where $\hat{\rho}$ is defined as before, i.e. the regression coefficient when $\hat{\varepsilon}_t$ is regressed on $\hat{\varepsilon}_{t-1}$. To apply the test, critical values are tabulated. Among the sources that describe the Durbin-Watson test are Chou (1989), Montgomery and Peck (1992), and Neter et al. (1983), which also describe other alternative approaches to deal with autoregression. More sources discussing autocorrelation are Schmidt and Taylor (1970), Banks and Carson (1984) and Bowerman *et al.* (2005).

4.5 Two independent samples, where both X and Y are serially correlated

To understand the relationships that are active in the data in this case, it is necessary to consider the structure between the data sets, as well as structured within each data set.

4.5.1 Relationships in the data

Let $\{X_i,\ i=1,2,...,n\}$ and $\{Y_i,\ i=1,2,...,n\}$ be two independent samples, then $\{X_i\}$ does not correspond to $\{Y_i\}$, but X_i depends on X_{i-1} as well as on previous members of the X-series, and Y_i depends on Y_{i-1} as well as on previous members of the Y-series. Several ways exist for modelling these time series, and the discussion that follows will be based on a general formula. However, illustrations of the application of the bootstrap method (to determine bootstrap critical values) for this study, will be done by using the simple structure of the AR (1) model. More complicated structures will follow in follow-up research.

Suppose $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are two independent data sets, each correlated within. The following simple model can be used to describe the behaviour of the X-values around its mean μ_X:

$$X_i = \mu_X + \varepsilon_i, \tag{4.1 a}$$

where the ε_i's are dependent on previous ε-values in some way. A general relationship between the ε_i's is described by using any general dependence model, for example the nonstationary $\text{ARIMA}(p,d,q)$ model: $\phi(B)(1-B)^d \varepsilon_i = \theta_q(B)u_i$, where the u_i, $i=1,2,\cdots,n$ are i.i.d. random variables. Here $E(u_i)=0$, $\text{Var}(u_i)=\sigma_u^2$, $\phi(B)=1-\phi_1 B-\cdots-\phi_p B^p$ is the stationary AR operator, and $\theta_q(B)=1-\theta_1 B-\cdots-\theta_q B^q$ is an invertible MA operator sharing no common factors between them. If $d=q=0$ and $p=1$, the AR(1) process is defined, which has a simple structure. Similarly, $Y_1, Y_2, \ldots, Y_m$ can be defined around its mean μ_Y by the model:

$$Y_j = \mu_Y + \vartheta_j, \tag{4.1 b}$$

where the $\vartheta_j's$ depend on previous ϑ-values according to a general structure: $\varphi(B)(1-B)^d \vartheta_j = \delta_q(B)w_j$, where the w_j are i.i.d. random variables, with $E(w_j)=0, \text{Var}(w_j)=\sigma_w^2$, $j=1,2,\ldots,m$. $\varphi(B)=1-\varphi_1 B-\cdots-\varphi_p B^p$ is a stationary AR operator and $\delta_q(B)=(1-\delta_1 B-\cdots-\delta_q B^q)$ is an invertible MA operator sharing no common factors between then. For $d=q=0$ and $p=1$, the AR(1) process is obtained, which will be persued further.

Consider the AR(1) case: For equations (4.1) above, the ε_i's and ϑ_j's are both defined then as:

$$\varepsilon_i=\rho_X\varepsilon_{i-1}+u_{iX} \text{ and } \vartheta_j=\rho_Y\vartheta_{j-1}+w_{jY},$$

where u_{iX} are i.i.d. random variables and $E(u_{iX})=0$, $\operatorname{Var}(u_{iX})=\sigma_u^2$, $i=1,2,\ldots,n$. Also w_{iY} are i.i.d. random variables $E(w_{jY})=0$, $\operatorname{Var}(w_{jY})=\sigma_w^2$, $j=1,2,\ldots,m$.

Then for $i=1,2,\ldots,n$,

$$X_i=\mu_X+\rho_X\varepsilon_{i-1}+u_{iX}.$$

This is equivalent to

$$X_i-\mu_X=\rho_X(X_{i-1}-\mu_X)+u_{iX}, \tag{4.2}$$

so that

$$u_{iX}=(X_i-\mu_X)-\rho_X(X_{i-1}-\mu_X).$$

Similarly, for $j=1,2,\ldots,m$,

$$Y_j=\mu_Y+\rho_Y\vartheta_{j-1}+w_{jY},$$

which is the same as

$$Y_j-\mu_Y=\rho_Y(Y_{j-1}-\mu_Y)+w_{jY}, \tag{4.3}$$

so that

$$w_{jY}=(Y_j-\mu_Y)-\rho_Y(Y_{j-1}-\mu_Y).$$

4.5.2 Testing procedure

The hypotheses to be tested are:

$H_0:\mu_X=\mu_Y$, vs. $H_1:\mu_X>\mu_Y$ (or any other alternative hypotheses).

If **X** and **Y** were independent and uncorrelated within each vector of data, the traditional test statistic to be applied would be:

$$t=\frac{\bar{X}_n-\bar{Y}_m}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}},$$

where $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{m+n-2}$. From Chapter 1 we know that t follows a t_{n+m-1} distribution under certain conditions, which is not valid in the present situation, due to the correlation structures within each data set. Since the distribution of t is not known in the present conditions, the decision rule for the new situation will not involve the t_{n+m-1}-distribution.

H_0 will be rejected if $t \geq C_{n,m}(\alpha)$ for the one-sided alternative hypothesis above, where $C_{n,m}(\alpha)$ is the true critical value defined by:

$$P_{H_0}(t \geq C_{n,m}(\alpha)) \approx \alpha. \tag{4.4}$$

In the traditional parametric, i.i.d. situations, $C_{n,m}(\alpha)$ would typically be $z(\alpha)$ or $t_{n+m-1}(\alpha)$ depending on whether σ^2 is known/unknown and on the sample sizes. But if the parametric and i.i.d. assumptions are not true, the distribution of t is unknown, as well as $C_{n,m}(\alpha)$. Therefore, the bootstrap approximation of the distribution of t will be used to determine bootstrap approximations for the exact critical values $C_{n,m}(\alpha)$. Resampling residuals seems to be a handy tool in the effort of obtaining bootstrap data sets.

4.5.3 Bootstrap critical values

From equations (4.2) and (4.3), $\mathbf{X}$ and $\mathbf{Y}$ can be expressed in terms of residuals $\hat{\mu}_{iX}$ and $\hat{w}_{jY}$ respectively, for $i = 2,3,\ldots,n$ and $j = 2,3,\ldots,m$:

$$\hat{u}_{iX} = (X_i - \bar{X}) - \hat{\rho}_X(X_{i-1} - \bar{X}) \text{ and } \hat{w}_{jY} = (Y_j - \bar{Y}) - \hat{\rho}_Y(Y_{j-1} - \bar{Y}).$$

Keep in mind that the residuals $\hat{u}_{iX}$ and $\hat{w}_{jY}$ cannot be computed unless appropriate estimators for ρ_X and ρ_Y are formulated. Standard estimators for the autocorrelation function in these cases are defined by:

$$\hat{\rho}_X = \frac{\sum_{i=2}^{n}(X_i - \bar{X}_n)(X_{i-1} - \bar{X}_n)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} \text{ and } \hat{\rho}_Y = \frac{\sum_{j=2}^{m}(Y_j - \bar{Y}_m)(Y_{j-1} - \bar{Y}_m)}{\sum_{j=1}^{m}(Y_j - \bar{Y}_m)^2}.$$

The bootstrap samples must adhere to the following imitation of the real data (see (4.2) and (4.3):

$$\begin{aligned} X_i^* - \bar{X}_n &= \hat{\rho}_X (X_{i-1}^* - \bar{X}_n) + u_{iX}^*, (i = 2,\dots,n), \\ Y_j^* - \bar{Y}_m &= \hat{\rho}_Y (Y_{j-1}^* - \bar{Y}_m) + w_{jY}^*, (j = 2,\dots,m), \end{aligned} \tag{4.5}$$

where the u_{iX}^* and w_{jY}^* are i.i.d. variables respectively, drawn from the centered residuals $\hat{u}_{iX}^c$ and $\hat{w}_{iX}^c$ defined below, with probability $1/(n-1)$ and probability $1/(m-1)$ respectively. Efron and Tibshirani (1993) suggested using the initial values $X_1^* = X_1$ and $Y_1^* = Y_1$ for each bootstrap sample, but other possibilities exist, for example, the medians of the two samples. For the present study, the values $X_1^* = X_1 - \bar{X}_n$ and $Y_1^* = Y_1 - \bar{Y}_m$ are chosen as fixed initial values when generating the respective bootstrap samples from the centered residuals, because the two values represent the location of the two datasets:

$$\hat{u}_{iX}^c = \left(X_i - \bar{X}_n\right) - \hat{\rho}_X\left(X_{i-1} - \bar{X}_n\right) - (n-1)^{-1} \sum_{i=2}^{n} \left\{\left(X_i - \bar{X}_n\right) - \hat{\rho}_X\left(X_{i-1} - \bar{X}_n\right)\right\}, \tag{4.6}$$

for $i = 2,\dots,n$ and

$$\hat{w}_{jX}^c = \left(Y_j - \bar{Y}_m\right) - \hat{\rho}_Y\left(Y_{j-1} - \bar{Y}_m\right) - (m-1)^{-1} \sum_{j=2}^{m} \left\{\left(Y_j - \bar{Y}_m\right) - \hat{\rho}_Y\left(Y_{j-1} - \bar{Y}_m\right)\right\},\ j = 2,\dots,m-1.$$

Instead of (4.2) and (4.3), the model assumptions $E\left(u_{iX}\right) = 0,\ i = 2,\dots,n$ and $E\left(w_{jY}\right) = 0,\ j = 1,2,\dots,m$ are imitated better.

To summarize, the first bootstrap sample $\mathbf{X}^* = \left(X_1^*, X_2^*, \dots, X_n^*\right)$ is generated as follows:

A. Draw with replacement from $u_{1X}^c, u_{2X}^c, \dots, u_{(n-1)X}^c$ with probability $1/n-1$ to find u_{1X}^{c*}.

B. Determine $\hat{\rho}_X$ and $\bar{X}_n$ from the initial sample.

C. Take $X_1^* = X_1 - \bar{X}_n$ and determine X_2^* from (4.5).

D. Repeat step A above to obtain u_{2X}^{c*} and use the values of X_2^*, $\bar{X}_n$ and $\hat{\rho}_X$ to determine X_3^*, and repeat the process to find $X_4^*, X_5^*, \dots, X_n^*$.

E. In the same way, find $\mathbf{Y}^* = \left(Y_1^*, Y_2^*, \dots, Y_m^*\right)$.

The bootstrap samples can now be used to obtain bootstrap approximations of $C_n(\alpha)$.

Bootstrap estimated critical values are obtained from the bootstrap imitation of (4.4):

$$P^*_{H_0^*}(t^* \geq C_{n,m}(\alpha)) \approx \alpha, \tag{4.7}$$

where

$$t^* = \frac{\left(\bar{X}_n^* - \bar{Y}_m^*\right) - \left(\bar{X}_n - \bar{Y}_m\right)}{S_p^* \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$= \frac{\left(\bar{X}_n^* - \bar{X}_n\right) - \left(\bar{Y}_m^* - \bar{Y}_m\right)}{S_p^* \sqrt{\frac{1}{n} + \frac{1}{m}}}, \text{ and } S_p^{*2} = \frac{(n-1)S_{X^*}^2 + (m-1)S_{Y^*}^2}{n+m-2}. \tag{4.8}$$

The formulation of t^* accommodates the null-hypothesis.

These quantities are calculated by using the bootstrap samples $\mathbf{X}^*$ and $\mathbf{Y}^*$. The following procedure should be used to determine bootstrap approximated critical values:

A. Generate the first bootstrap samples $\mathbf{X}^*(1) = \left(X_1^*, X_2^*, \ldots, X_n^*\right)$ and $\mathbf{Y}^*(1) = \left(Y_1^*, Y_2^*, \ldots, Y_m^*\right)$ as described above and calculate the first t^* value according to (4.7). This will be referred to as t_1^*.

B. Generate the second bootstrap samples $\mathbf{X}^*(2) = \left(X_1^*, X_2^*, \ldots, X_n^*\right)$ and $\mathbf{Y}^*(2) = \left(Y_1^*, Y_2^*, \ldots, Y_m^*\right)$ as above and determine t_2^*.

C. Repeat the process B times to produce test-statistics values $t_1^*, t_2^*, \ldots, t_B^*$.

D. Sort the t^* values from smallest to largest to yield $t_{(1)}^*, t_{(2)}^*, \ldots, t_{(B)}^*$.

E. Then the bootstrap approximation of $C_{n,m}(\alpha)$ is $\hat{C}_{n,m}(\alpha) = t^*_{[B(1-\alpha)]}$.

F. H_0 is rejected if $t \geq \hat{C}_{n,m}(\alpha)$.

4.6 Two dependent samples X and Y, both are serially correlated

We again consider the structural relationships between as well as within the data sets.

4.6.1 Relationships in the data

The data sets in this paragraph are two dependent data sets $\mathbf{X}$ and $\mathbf{Y}$, and both $\mathbf{X}$ and $\mathbf{Y}$ are serially correlated within. The new test is based on data pairs, as was done when the *t*-test for dependent data were discussed previously in Chapter 1. This means that $\{X_i,\ i=1,2,...,n\}$ and $\{Y_i,\ i=1,2,...,n\}$ are the two dependent sets and that $\{X_i\}$ corresponds to $\{Y_i\}$, but since both sets of data are serially correlated as well, X_i depends on X_{i-1} as well as on previous members of the X-series, and Y_i depends on Y_{i-1} as well as on previous members of the Y-series, as was mentioned in Section 4.5 above.

Let $\mathbf{X}$ and $\mathbf{Y}$ be dependent data sets and assume that the sample sizes are equal, i.e. of size *n*, with both $\mathbf{X}$ and $\mathbf{Y}$ serially correlated. Pair the data and let $\{(X_i, Y_i),\ i=1,2,...,n\}$ represent the pairs of subjects. Then $\{D_i = X_i - Y_i,\ i=1,2,...,n\}$ will represent the correlated differences $D_1, D_2, \dots, D_n$. For theoretical purposes, these pairs must be modelled, the D_i's can be modelled around the mean difference $\mu_D = \mu_X - \mu_Y$ as follows: $D_i = \mu_D + a_i \quad (i=1,2,\dots,n)$ where the a_i's are dependent on previous a-values. This dependence can be expressed in a general form such as the ARIMA formulation discussed in Section 4.5 above. In practical situations, the time series resulting from the pairs of data D_i, will first be processed to determine the specific type of relationship that exists, i.e., the time series applicable must be identified. To our knowledge there is no rule that governs the form of the dependence structures in the differences (*D*) with those in the samples $\mathbf{X}$ and $\mathbf{Y}$. Therefore we express the dependence structure in *D* in a general way and model fitting must be done first. If the differences (*D*) forms an AR(1) process, we will proceed in the way explained below, but the procedures developed provide for many other time series. The illustration of the bootstrap application will be in terms of the simpler AR(1) form and imitate the previous techniques.

As before, the following condition has to be satisfied which is analogous to (4.2):

$$D_i - (\mu_X - \mu_Y) = \phi_D(D_{i-1} - (\mu_X - \mu_Y)) + u_{iD}, \tag{4.9}$$

where the u_{iD} are i.i.d. random variables, and therefore

$D_i - \mu_D = \phi_D(D_{i-1} - \mu_D) + u_{iD}$.

4.6.2 Testing procedure

The hypotheses of no difference between the series $\mathbf{X}$ and $\mathbf{Y}$ are formulated in terms of the differences, given by:

$H_0 : \mu_D = 0$ vs. $H_1 : \mu_D > 0$ or any other alternative hypothesis.

This hypothesis is tested by using the test statistic:

$$t_{dep} = \frac{\bar{D}}{\sqrt{\hat{\sigma}^2(\bar{D})}}, \tag{4.10}$$

where $\hat{\sigma}^2(\bar{D})$ denotes the estimated variance of $\bar{D}$, and *dep* denotes the dependence of the two original datasets $\mathbf{X}$ and $\mathbf{Y}$.

For the estimation of $\hat{\sigma}^2(\bar{D})$, standard time series results revised in Chapter 3 and repeated below, entails the following procedure:

For a stationary process $\{Z_t\}$, the mean is denoted by $E(Z_t) = \mu$ and variance is denoted by $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$, the covariances between Z_t and Z_{t+k} are given by $\gamma_k = Cov(Z_t - \mu)(Z_{t+k} - \mu)$, and the autocorrelation at lag k between Z_t and Z_{t+k} is $\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}$, where $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$, under stationarity of the process Z_t. In paired data it means that $\phi_D = \frac{\gamma_1}{\gamma_0}$, and therefore

$$\hat{\phi}_D = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}, \tag{4.11}$$

where $\hat{\gamma}_0 = \frac{1}{n}\sum_{i=1}^{n}(D_i - \bar{D})^2$ and $\hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^{n-1}(D_i - \bar{D})(D_{i+1} - \bar{D})$ (Wei, 1990: 10).

An estimator for $\hat{\sigma}^2(\bar{D})$ is (see Anderson, 1976: 55):

$$\hat{\sigma}^2(\bar{D}) = \frac{\hat{\gamma}_0(1 + \hat{\phi}_D)}{n(1 - \hat{\phi}_D)}, \tag{4.12}$$

with all expressions defined above.

As before, in (4.10) above, the distribution of t_{dep} is not known under the pair-wise assumptions, and as before, the bootstrap critical values associated with the test statistic must be determined, in the same manner it was done in Section 4.5 above.

4.6.3 Bootstrap critical values

Analogous to (4.6) and (4.9), it follows that the centred residuals are:

$$\hat{u}^c_{iD} = \left(D_i - \bar{D}\right) - \hat{\phi}_D\left(D_{i-1} - \bar{D}\right) - (n-1)^{-n} \sum_{i=2}^{n} \left\{\left(D_i - \bar{D}\right) - \hat{\phi}_D\left(D_{i-1} - \bar{D}\right)\right\}, \quad i = 2, \dots, n, \quad (4.13)$$

where $\bar{D} = \bar{X} - \bar{Y}$ $\left(\text{or } \bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i\right)$, and

$$\hat{\phi}_D = \frac{\sum_{i=2}^{n}(D_i - \bar{D})(D_{i-1} - \bar{D})}{\sum_{i=1}^{n}(D_i - \bar{D})^2}. \quad (4.14)$$

The bootstrap sample must adhere to the following bootstrap imitation of (4.9):

$$D_i^* - \bar{D} = \hat{\phi}_D(D_{i-1}^* - \bar{D}) + u_{iD}^{c*}, (i = 2, \cdots, n), \quad (4.15)$$

where u_{iD}^{c*} are drawn independently from the centred residuals $\hat{u}^c_{iD}$ with probability $1/(n-1)$, $i = 2, \dots, n$.

Define $D_1^* = D_1 - \bar{D}$ as a fixed initial value for the purpose of generating the respective bootstrap samples. The bootstrap time series of differences are determined recursively as follows as was explained before:

A. Draw with replacement from $u_{1D}^c, u_{2D}^c, \dots, u_{(n-1)D}^c$ with probability $1/(n-1)$ to find u_{1D}^{c*}.

B. Determine $\hat{\phi}_D$ and $\bar{D}$ from the data.

C. Take $D_1^* = D_1 - \bar{D}$ and determine D_2^* from (4.15).

D. Repeat the previous steps to obtain D_3^*. By repeating the process, $\mathbf{D}^* = \left(D_1^*, D_2^*, \dots, D_n^*\right)$ can be obtained.

The bootstrap critical value $\hat{C}_n(\alpha)$ is defined by the bootstrap imitation of (4.4):

$$P^*_{H_0^*}(t^*_{dep} \geq \hat{C}_n(\alpha)) \approx \alpha, \tag{4.16}$$

where

$$t^*_{dep} = \frac{\bar{D}^* - \bar{D}}{\sqrt{\dfrac{\hat{\gamma}^*_0(1+\hat{\phi}^*_D)}{n(1-\hat{\phi}^*_D)}}}, \quad \hat{\gamma}^*_0 = \frac{1}{n}\sum_{i=1}^{n}(D^*_i - \bar{D}^*)^2$$

and

$$\hat{\phi}^*_D = \left[\frac{1}{n}\sum_{i=1}^{n-1}(D^*_i - \bar{D}^*)(D^*_{i+1} - \bar{D}^*)\right] \Big/ \hat{\gamma}^*_0 .$$

By formulating the test statistic in this way, it is "under the null-hypothesis".
Note that all the above quantities are defined in terms of the bootstrap sample, and can be calculated. As before, this process is repeated B times to produce bootstrap test-statistics values $t^*_{dep1}, t^*_{dep2}, \cdots, t^*_{depB}$ whereafter these values are sorted, and presented by $t^*_{dep(1)}, t^*_{dep(2)}, \cdots, t^*_{dep(B)}$. The bootstrap approximation for $C_n(\alpha)$ is then $\hat{C}_n(\alpha) = t^*_{dep[B(1-\alpha)]}$. H_0 is rejected if $t_{dep} \geq \hat{C}_n(\alpha)$.

The properties of the new tests will now be evaluated in Chapter 5.

CHAPTER 5

MONTE CARLO SIMULATION STUDIES

5.1 Introduction

The purpose of the procedures developed in Chapter 4 above is to develop tests with desirable properties, to compare characteristics of two populations, if the samples drawn from the populations (or if two sets of observations are available from two situations) are presented in time series form, i.e. if serially correlated data are available. Test procedures are developed below for situations when the two samples (or sets of observations) are independent or dependent. This purpose will be met if the new test procedures, which are based on bootstrap critical values, perform well in terms of trustworthy probabilities of Type I errors, i.e. if the probabilities achieved by using these tests of falsely rejecting the null hypothesis, are close to the prescribed significance levels. The new bootstrap-based tests should also be powerful, i.e. the probability of rejecting the null hypothesis if the alternative is true, should be high.

In this chapter, results of Monte-Carlo studies will be presented regarding the objectives mentioned above. These studies were conducted to

(i) illustrate the validity of the bootstrap approximation of the exact critical values, when two sample tests were performed to compare the means of two time series

(ii) determine the power of the newly proposed testing procedures.

All simulations were carried out on a 400MHZ Pentium 4 Personal Computer using Microsoft FORTRAN Power Station 6.6 and the MSIMSL library. The normal random numbers were generated by using the function called *drnnor*.

The tests were performed for the independent and the dependent cases described in Chapter 4 above and were applied to data satisfying the AR(1) model formulation. We report only for equal sample sizes n and m in the independent case although the sample sizes may differ as is clear from Chapter 4. We display results for limited sample sizes, although the study was performed for many sample sizes. Throughout $B = 1000$ bootstrap replications were used according to the algorithms given in Chapter 4, and 2000 independent Monte Carlo trials ($MC = 2000$, where MC denotes the number of Monte Carlo trials) were performed. The level of accuracy of the tests can be improved by using larger numbers of bootstrap replications and more Monte Carlo trials.

The first focus of the Monte Carlo studies is to examine the reliability of the bootstrap critical values of the tests concerned, with regard to the achieved probabilities of Type I errors, i.e. $P_{H_0 \text{ is true}}(Reject\ H_0)$, where the following hypotheses will be considered throughout this study:

$H_0: \mu_X = \mu_Y$ against the one sided alternative hypothesis $H_1: \mu_X > \mu_Y$ in the case of independent data sets. These formulations are equivalent to the following: $H_0: \mu_X - \mu_Y = 0$ vs. $H_1: \mu_X - \mu_Y > 0$.

For the dependent case, the hypotheses of interest are $H_0: \mu_D = 0$ against $H_1: \mu_D > 0$, with μ_D defined in paragraph 4.6. In both cases the null hypotheses are rejected if the test statistics concerned exceed the obtained bootstrap critical values, and the null hypotheses are not rejected otherwise.

The second focus of the Monte Carlo studies is to determine the estimated power of the tests and to investigate possible influential factors on the size of the power of the tests. Included in the Monte Carlo studies, the performance of the bootstrap critical values is therefore also compared with the behaviour of standard two sample tests described in Chapter 1, for choices of the parameters ideal for the usual tests (when ρ_X and ρ_Y are both equal to 0, for example, which imply uncorrelated data, i.e. independent data with $N(0,1)$ distributions). In this way, not only the estimated probability of Type I errors by using bootstrap critical values is evaluated, but it is also compared to the existing 'best' tests used in practical statistics.

Since the Monte Carlo studies were designed to evaluate the aims of the study, more specific detail will follow in the paragraphs below.

5.2 Monte Carlo approximation of the probability of type I error

To determine the estimated probability of a Type I error, i.e. $P_{H_0 \text{ is true}}(Reject\ H_0)$, data were generated under the null hypotheses for both the independent case (with the autocorrelation coefficients denoted by (ρ_X, ρ_Y)) and the dependent case (with the autocorrelation coefficient denoted by (ϕ_D)). For both cases, sample sizes

$n = \{30, 50, 100, 150\}$ are displayed although smaller and larger values for n and m were applied to determine small and large sample behaviour of the tests.

Also note the following:

- Samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be generated to fluctuate around pre-selected means μ_X and μ_X for the independent case. Various values of μ_X and μ_Y have been considered but we report in this study only results for values of $\Delta\mu = \mu_X - \mu_Y$ equal to 0, since this is the situation associated with the null hypothesis. For example, if $\mu_X = \mu_Y = 1$, then $\Delta\mu = 0$ and if $\mu_X = \mu_Y = 0.5$, then $\Delta\mu = 0$.
- Various values of μ_D have been considered in the overall Monte Carlo study, but for determining critical values, only the case where $\mu_D = 0$ is important, since this is associated with the null hypothesis.
- Various values of ρ_X and ρ_Y are considered. ρ_X and ρ_Y values used in the studies include $\rho_X = \rho_Y$ values {0.0, 0.1, 0.3, 0.5, 0.7, 0.9}, and also values of $\rho_X \neq \rho_Y$. From these results the influence of the correlation on the credibility of the bootstrap critical values can be determined. We report only on a selection of the results. Similar behaviour was found for other parameter values.
- When the data were generated, u_{iX} and w_{jY} ($i = j = 1, ..., n$) defined in (4.2) and (4.3) respectively were assumed to originate from normal distributions having means equal to 0 and variances equal to 1. This is also assumed for the dependent data giving rise to the differences and u_{iD} defined in (4.9), $i = 1, ..., n$.
- $C_{n.m}(\alpha)$ and $C_n(\alpha)$ defined in (4.7) and (4.16) respectively are estimated using the procedures described in paragraphs 4.5 and 4.6 respectively. The following levels of significance were used: $\alpha = 0.01, 0.05, 0.10$.
- The processes were repeated MC (Monte Carlo) times, with $MC = 2000$. For each Monte Carlo trial the t-statistic defined in paragraph 4.5.2. (denoted in terms of MC as $t(mc)$), as well as t_{dep} defined in (4.10) (denoted by $t_{dep}(mc)$) were evaluated. Also, $\hat{C}_{n.m}(\alpha)$ and $\hat{C}_n(\alpha)$ are determined for each Monte Carlo trial (denoted by $\hat{C}_{n.m,mc}(\alpha)$ and $\hat{C}_{n,mc}(\alpha)$).

- The Monte Carlo estimation of probability of Type I error for the independent and dependent cases respectively are then determined by:

$$\frac{1}{MC}\sum_{mc=1}^{MC} I\left(t(mc) > \hat{C}_{m,n,mc}(\alpha)\right) \text{ and } \frac{1}{MC}\sum_{mc=1}^{MC} I\left(t_{dep}(mc) > \hat{C}_{n,mc}(\alpha)\right).$$

If the new procedures are successful, it is expected that equation (4.7) for the independent case and equation (4.16) for the dependent case should be satisfied if the number of bootstrap iterations and Monte Carlo trials increase. Although extensive studies were performed, only an adequate selection of results will be presented to prove the point of interest, which in this paragraph, is the good behaviour of the bootstrap critical values. The results will be displayed in Tables 5.1 below.

5.3 Monte Carlo approximation of the power of the new tests

To determine the estimated powers of the tests, i.e. $P_{H_1 \text{ is true}}(Reject\ H_0)$, data were generated under the alternative hypotheses for both the independent case (with the autocorrelation coefficients denoted by (ρ_X, ρ_Y)) and the dependent case (with the autocorrelation coefficient denoted by (ϕ_D)) as before. For both cases, sample sizes $n = \{30, 50, 100, 150\}$ are displayed although smaller and larger sample sizes were used as well.

Also note the following:

- Various values of μ_X and μ_Y have been considered and we report in this study results for values of $\Delta\mu = \mu_X - \mu_Y$ equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, since these are the situations associated with the alternative hypothesis. We report in the independent and dependent cases only for $\Delta\mu$ = 0.5, since similar behaviour was found for other values of $\Delta\mu$.
- Various values of μ_D have been considered in the overall Monte Carlo study, and we report estimated powers for cases where μ_D = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, since these values are associated with the alternative hypothesis.
- Various values of ρ_X and ρ_Y are considered. ρ_X and ρ_Y values used in the studies include ρ_X (= ρ_Y) values {0.1, 0.3, 0.5, 0.7, 0.9}, and also values of $\rho_X \neq \rho_Y$. From these results the influence of the correlation on the power of the

tests can be determined. The same values were used for ϕ_D in the dependent case. We report only on a selection of these.

- When the data were generated, u_{iX} and w_{jY} $(i = j = 1, ..., n)$ defined in (4.2) and (4.3) respectively were assumed to originate from normal distributions having means equal to 0 and variances equal to 1. For the dependent case, dependence was induced by generating samples **X** and **Y** from a bivariate Normal distributions, i.e., from a $N\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{bmatrix}\sigma_X^2 & \rho_\varepsilon\sigma_X\sigma_Y \\ \rho_\varepsilon\sigma_X\sigma_Y & \sigma_Y^2\end{bmatrix}\right)$ - distribution.

  Various values of σ_X^2, σ_Y^2 and ρ_ε were used but we report only for $\rho_\varepsilon = 0.5$, $\sigma_X^2 = 1$ and $\sigma_Y^2 = 2$. Similar results were found for other values of the parameters.
- Again, $C_{n.m}(\alpha)$ and $C_n(\alpha)$ defined in (4.7) and (4.16) respectively are estimated using the procedures described in paragraphs 4.5 and 4.6 respectively. The following levels of significance are used: $\alpha = 0.01, 0.05, 0.10$.
- The processes were repeated MC (Monte Carlo) times, with $MC = 2000$. For each Monte Carlo trial the t-statistic defined in paragraph 4.5.2. (denoted in terms of MC by $t(mc)$), as well as t_{dep} defined in (4.10) (denoted by $t_{dep}(mc)$) were evaluated. Also, $\hat{C}_{n.m}(\alpha)$ and $\hat{C}_n(\alpha)$ is determined for each Monte Carlo trial (denoted by $\hat{C}_{n.m,mc}(\alpha)$ and $\hat{C}_{n,mc}(\alpha)$).
- The Monte Carlo estimation of power of the tests for the independent and dependent cases respectively are then determined under the alternative hypotheses by:

$$\frac{1}{MC}\sum_{mc=1}^{MC} I\left(t(mc) > \hat{C}_{m,n,mc}(\alpha)\right) \text{ and } \frac{1}{MC}\sum_{mc=1}^{MC} I\left(t_{dep}(mc) > \hat{C}_{n,mc}(\alpha)\right).$$

If the new procedures are successful, it is expected that the Monte Carlo estimations of the power of the tests, i.e. the estimated values of $P_{H_1 \text{ is true}}(Reject\ H_0)$, should become higher for non-local alternatives, i.e. for parameter values not supporting the null hypothesis. Again, although extensive studies were performed, only an adequate selection of results will be presented to prove the points of interest, which is in this case, that the new tests have high power even in highly correlated situations. The results will be displayed in Tables 5.2 below.

5.4 Results of Monte Carlo studies

5.4.1 Results regarding the validity of the bootstrap critical values:

The results regarding the validity of the bootstrap critical values are displayed in Tables 5.1a to Table 5.1d in Appendix A. Table 5.1a (for two independent sets of data) and Table 5.1b (for two dependent sets of data) display the estimated Type I errors obtained for the new procedures in comparison with the estimated Type I errors obtained by applying the traditional t – test, for example. This is done to verify the validity or not of the bootstrap procedure by comparing it to a test that is trustworthy in the null conditions, i.e. the t – test.

5.4.2 Results regarding the powers of the new tests:

The results regarding the powers of the new tests are displayed in Tables 5.2a, Tables 5.2b and Tables 5.3 in Appendix A.

Table 5.2a (for two independent sets of data) and Table 5.2b (for two dependent sets of data) display the estimated powers of the new tests, if autocorrelation is present in the time series. Various values were assigned to (ρ_X, ρ_Y), with $\rho_X = \rho_Y$ as well as with $\rho_X \neq \rho_Y$, to determine the influence of the autocorrelation on the powers of the test. Tables 5.2a and Tables 5.2b only report on cases where $\rho_X = \rho_Y$, since similar behaviour were found for cases where $\rho_X \neq \rho_Y$.

Tables 5.3 report on cases where $\rho_X = \rho_Y = 0.0$, i.e. in the absence of autocorrelation, where data are generated under the alternative hypothesis. Comparisons regarding estimated power achieved are therefore made between the new tests and the traditional t – test in conditions favourable to the t – test.

The table numeration includes information in brackets of the applicable samples sizes and the value of $\Delta\mu$ respectively. In these tables, the estimated powers obtained when the traditional t – test would have been applied (wrongly) in the presence of autocorrelation, are displayed to point out the influence of the correlation structures on both the new tests as well as on the traditional testing procedures.

5.5 Conclusions

- By studying Table 5.1a and Table 5.1b, it is clear that with regard to the validity of the bootstrap critical values in comparison with t-percentiles, in the absence of serial correlation, the new tests, for the independent case, provide approximated estimated probabilities of Type I error close to the prescribed levels, even for small sample sizes of (30 and 50). This is in a slightly lesser way true for the dependent case, for $n = m = 30$. For larger samples sizes, it is evident that the new tests achieve satisfactory approximations of the probabilities of Type I error, which indicates that the test procedure making use of bootstrap critical values, perform well in conditions suitable for the traditional t-tests. The bootstrap procedure, however, is developed to better situations where serial correlation is present, and where normality of the data is not required. Then the t-test is not applicable. Non-parametric tests, such as the Mann-Whitney test for the independent case, or the Signed Rank test in the paired sample case, must be used. But because of the autocorrelation present in the data, these tests are not applicable. Tables 5.1a and b merely reveal that the bootstrap critical values can be used in null-conditions successfully when serial correlation is not present.
- By studying Table 5.1c and 5.1d, it is clear that with regard to the validity of the bootstrap critical values, the new tests for the independent and dependent cases, provide estimated powers larger than the t-test in the absence of serial correlation. Considering this, together with the observations contained in the previous remark, we can conclude that the bootstrap-based test procedure is at least as powerful as the traditional t-test.
- By studying Tables 5.2a (30,0.0), 5.2a (50,0.0), 5.2a (100,0.0) and 5.2a (150,0.0) it is clear that, in the presence of serial correlation, the test procedure based on bootstrap critical values, produces approximated Type I errors much closer to the prescribed levels than the t-test especially if serial correlation increases, although these bootstrap-based approximated values are slightly inaccurate in case of heavier serial correlation. It is clear that this procedure should be preferred to the traditional t-test, which performs untrustworthy when mild to heavy correlation is present. As the sample sizes increase, the situation improves for the bootstrap test procedure, but the traditional t-test becomes unreliable and does not attain the prescribed levels, but produces liberal significance levels far above the prescribed levels.

To summarize, these tables reveal that, when $\Delta\mu$=0.0, i.e. under null hypothesis conditions, in the presence of small autocorrelations, the type I errors achieved by the new tests are much closer to the prescribed levels than those achieved by the t - test. This indicates that the procedures based on bootstrap critical values, although not completely accurate for small samples, are more reliable than wrongly applied t - tests, especially if serial correlation is large. If very small autocorrelation is present in the data sets, the t - test behaves well but in all cases when $\Delta\mu$ = 0.0, the new tests provide more reliable results, contrary to the behaviour of the traditional tests.

- Tables 5.2a(30,0.5), 5.2a(50,0.5), 5.2a(100,0.5) and 5.2a(150,0.5) reveal that, for all values of $\Delta\mu$ > 0.0, i.e. under alternative hypothesis conditions, for all sample sizes, and for all values of $(\rho_X,\rho_Y)>(0.0,0.0)$, the new methods provide estimated power less than the t - test. However, the t - test proved to be unreliable and uncontrolled in the presence of serial correlation, i.e. rejecting too easily, and therefore the powers achieved by the t - test are not to be believed at the significance level prescribed.
- By studying Tables 5.2b(30,0.0), 5.2b(50,0.0), 5.2b(100,0.0) and 5.2b(150,0.0) it is clear that, for all cases, the two procedures behave almost identically if serial correlation is present in null conditions. Both procedures are slightly disturbed by the correlation structures. For the dependent case, there is little to choose between the two methods.
- By studying Tables 5.2b(30,0.5), 5.2b(50,0.5), 5.2b(100,0.5) and 5.2b(150,0.5), the achieved estimated powers seem to be identical for the two procedures. One explanation for this phenomenon is that all samples were generated from bivariate normal errors, which favours the t-distribution. Limited studies involving other distributions indicated that in the dependent case, the resulting t - statistic will suffer from serial correlation when errors are not normally distributed. The bootstrap procedure copes well with non-normality. Follow-up studies are on the way to persue the case where errors are not normally distributed.
- Overall, it is clear that the estimated powers of all the tests decrease in the presence of heavy autocorrelated structures, but we can claim that the new tests are remarkably useful and stable for large sample sizes in comparison with the t - test, when small and moderate correlation is present in the data. This is

reassuring, since no other tests are available in the literature for such complex situations.

- We finally conclude that the bootstrap method can be applied successfully in two-sample comparative situations if serial correlation is present, for both independent and dependent samples. The proposed new testing procedures produce Type I error approximations close to the prescribed significance levels and the achieved power estimations are more reliable than when the t – test is applied wrongly, especially in the independent case. In cases where heavy correlation structures are present, the power decreases in all sample sizes, especially for small significance levels, as can be expected.

5.6 Remarks and recommendations

We conclude this study with the following remarks:

To analyse the data regarding the operators working in the quarry, it is necessary to characterise the time series involved. If AR(1) models are identified, the methods described can be implemented. If other models are identified, similar procedures will be followed, after the parameters of the models have been estimated. These methods will form part of another study.

We have succeeded in developing new bootstrap-based tests that fulfilled the aims of this study. Satisfactory estimated Type I errors were obtained as well as high powers for the new tests.

ANNEXURE A

1. Results regarding the validity of the bootstrap critical values:

Table 5.1a: Estimated Type I errors determined with bootstrap critical values for the independent case, in comparison with $t-$test critical values for the traditional two-sample $t-$test when $\mu_X = \mu_Y = 0.0$ and (ρ_X, ρ_Y)=(0.0, 0.0). All standard errors are less than 0.01.

Estimated Type I errors		$n=m$=30	$n=m$=50	$n=m$=100	$n=m$=150
	α				
Bootstrap critical values	0.01	0.015	0.012	0.017	0.015
	0.05	0.058	0.057	0.064	0.053
	0.10	0.106	0.103	0.107	0.108
Traditional $t-$test critical values	0.01	0.009	0.006	0.014	0.009
	0.05	0.047	0.049	0.060	0.049
	0.10	0.098	0.101	0.106	0.102

Table 5.1b: Estimated Type I errors determined with bootstrap critical values for the dependent case, in comparison with $t-$test critical values for the traditional one-sample $t-$test when $\mu_X = \mu_Y = 0.0$ and (ρ_X, ρ_Y)=(0.0, 0.0). All standard errors are less than 0.01.

Estimated Type I errors		$n=m$=30	$n=m$=50	$n=m$=100	$n=m$=150
	α				
Bootstrap critical values	0.01	0.017	0.013	0.009	0.009
	0.05	0.070	0.048	0.053	0.046
	0.10	0.117	0.097	0.102	0.102
Traditional $t-$test critical values	0.01	0.015	0.013	0.007	0.010
	0.05	0.064	0.051	0.051	0.042
	0.10	0.112	0.096	0.099	0.100

Table 5.1c: Estimated powers for the new tests for the independent case, with $\mu_X = 0.5$, $\mu_Y = 0.0$ in the absence of serial correlation. All standard errors are less than 0.01. Estimated powers when traditional t – tests are applied, are included.

Estimated Type I errors		$n = m = 30$	$n = m = 50$	$n = m = 100$	$n = m = 150$
	α				
Bootstrap critical values	0.01	0.446	0.623	0.899	0.977
	0.05	0.650	0.824	0.969	0.997
	0.10	0.759	0.902	0.986	0.999
Traditional t – test critical values	0.01	0.329	0.551	0.881	0.973
	0.05	0.584	0.797	0.968	0.996
	0.10	0.715	0.886	0.985	0.999

Table 5.1d: Estimated powers for the new tests for the dependent case, with $\mu_X = 0.5$, $\mu_Y = 0.0$ in the absence of serial correlation. All standard errors are less than 0.01. Estimated powers when traditional t – tests are applied, are included.

Estimated Type I errors		$n = m = 30$	$n = m = 50$	$n = m = 100$	$n = m = 150$
	α				
Bootstrap critical values	0.01	0.259	0.376	0.701	0.875
	0.05	0.502	0.640	0.884	0.968
	0.10	0.640	0.762	0.938	0.987
Traditional t – test critical values	0.01	0.258	0.384	0.701	0.885
	0.05	0.499	0.637	0.880	0.972
	0.10	0.635	0.762	0.938	0.987

2. Results regarding the type I errors and the powers of the new tests, for independent samples

Tables 5.2a (for two independent sets of data) below display the estimated type I errors and the estimated powers of the new tests. The table numeration includes information in brackets of the applicable samples sizes and the value of $\Delta\mu$ respectively.

Table 5.2a (30, 0.0): Estimated type I error rates of the new test for the independent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated type I error rates when traditional t – test are applied, are included.

n = 30, ($\Delta\mu$ =0.0)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.018	0.031	0.033
	0.05	0.061	0.076	0.083
	0.10	0.108	0.121	0.138
t – test	0.01	0.019	0.046	0.090
	0.05	0.066	0.114	0.177
	0.10	0.112	0.173	0.236

Table 5.2a (30, 0.5): Estimated powers of the new test for the independent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated powers when traditional t – test are applied, are included.

n = 30, ($\Delta\mu$ =0.5)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.404	0.288	0.255
	0.05	0.596	0.474	0.383
	0.10	0.723	0.597	0.479
t – test	0.01	0.341	0.325	0.349
	0.05	0.579	0.540	0.502
	0.10	0.716	0.654	0.587

Table 5.2a (50, 0.0): Estimated type I error rates of the new test for the independent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated type I error rates when traditional $t-$test are applied, are included.

n = 50, ($\Delta\mu$ =0.0)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.017	0.018	0.027
	0.05	0.067	0.064	0.069
	0.10	0.119	0.108	0.118
$t-$test	0.01	0.019	0.044	0.089
	0.05	0.077	0.112	0.175
	0.10	0.129	0.176	0.229

Table 5.2a (50, 0.5): Estimated powers of the new test for the independent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated powers when traditional $t-$test are applied, are included.

n = 50, ($\Delta\mu$ =0.5)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.555	0.374	0.244
	0.05	0.760	0.627	0.438
	0.10	0.858	0.743	0.559
$t-$test	0.01	0.548	0.509	0.455
	0.05	0.771	0.727	0.625
	0.10	0.863	0.810	0.697

Table 5.2a (100, 0.0): Estimated type I error rates of the new test for the independent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated type I error rates when traditional $t-$test are applied, are included.

n = 100, ($\Delta\mu$ =0.0)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.018	0.024	0.021
	0.05	0.055	0.063	0.065
	0.10	0.111	0.117	0.115
$t-$test	0.01	0.019	0.049	0.094
	0.05	0.071	0.121	0.177
	0.10	0.126	0.179	0.239

Table 5.2a (100, 0.5): Estimated powers of the new test for the independent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated powers when traditional $t-$test are applied, are included.

n = 100, ($\Delta\mu$ =0.5)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.822	0.619	0.365
	0.05	0.940	0.829	0.582
	0.10	0.969	0.893	0.700
$t-$test	0.01	0.848	0.786	0.648
	0.05	0.951	0.903	0.786
	0.10	0.972	0.936	0.837

Table 5.2a (150, 0.0): Estimated type I error rates of the new test for the independent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated type I error rates when traditional t − test are applied, are included.

n = 150, ($\Delta\mu$=0.0)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.7,0.7)
	α			
New test	0.01	0.016	0.012	0.021
	0.05	0.058	0.052	0.061
	0.10	0.112	0.099	0.112
t − test	0.01	0.022	0.039	0.095
	0.05	0.076	0.109	0.185
	0.10	0.135	0.173	0.252

Table 5.2a (150, 0.5): Estimated powers of the new test for the independent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated powers when traditional t − test are applied, are included.

n = 150, ($\Delta\mu$=0.5)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.945	0.784	0.475
	0.05	0.991	0.928	0.716
	0.10	0.999	0.969	0.832
t − test	0.01	0.962	0.912	0.797
	0.05	0.996	0.969	0.897
	0.10	0.999	0.986	0.935

3. Results regarding the type I errors and the powers of the new tests, for dependent samples

Table 5.2b (30, 0.0): Estimated type I error rates of the new test for the dependent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated type I error rates when traditional $t-$test are applied, are included.

n = 30, ($\Delta\mu$=0.0)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.021	0.021	0.035
	0.05	0.080	0.070	0.090
	0.10	0.125	0.127	0.145
$t-$test	0.01	0.017	0.024	0.033
	0.05	0.074	0.069	0.086
	0.10	0.119	0.123	0.141

Table 5.2b (30, 0.5): Estimated powers of the new test for the dependent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated powers when traditional $t-$test are applied, are included.

n = 30, ($\Delta\mu$=0.5)				
(ρ_X, ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.207	0.169	0.113
	0.05	0.429	0.354	0.255
	0.10	0.557	0.482	0.363
$t-$test	0.01	0.207	0.163	0.104
	0.05	0.422	0.347	0.257
	0.10	0.543	0.477	0.359

Table 5.2b (50, 0.0): Estimated type I error rates of the new test for the dependent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated type I error rates when traditional $t-$test are applied, are included.

n = 50, ($\Delta\mu$=0.0)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.013	0.017	0.030
	0.05	0.057	0.066	0.074
	0.10	0.108	0.114	0.130
$t-$test	0.01	0.015	0.018	0.033
	0.05	0.056	0.068	0.074
	0.10	0.103	0.112	0.127

Table 5.2b (50, 0.5): Estimated powers of the new test for the dependent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with ρ_X = ρ_Y. All standard errors are less than 0.02. Estimated powers when traditional $t-$test are applied, are included.

n = 50, ($\Delta\mu$=0.5)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.321	0.205	0.133
	0.05	0.585	0.429	0.307
	0.10	0.715	0.576	0.427
$t-$test	0.01	0.330	0.211	0.141
	0.05	0.585	0.436	0.307
	0.10	0.716	0.578	0.437

Table 5.2b (100, 0.0): Estimated type I error rates of the new test for the dependent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated type I error rates when traditional t − test are applied, are included.

n = 100, ($\Delta\mu$ =0.0)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.011	0.013	0.017
	0.05	0.049	0.064	0.049
	0.10	0.109	0.118	0.111
t − test	0.01	0.011	0.015	0.016
	0.05	0.046	0.066	0.053
	0.10	0.108	0.118	0.112

Table 5.2b (100, 0.5): Estimated powers of the new test for the dependent case with $\mu_X = 0.5$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated powers when traditional t − test are applied, are included.

n = 100, ($\Delta\mu$ =0.5)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.597	0.383	0.194
	0.05	0.834	0.658	0.427
	0.10	0.916	0.782	0.561
t − test	0.01	0.607	0.396	0.211
	0.05	0.839	0.661	0.439
	0.10	0.915	0.777	0.565

Table 5.2b (150, 0.0): Estimated type I error rates of the new test for the dependent case with $\mu_X = 0.0$, $\mu_Y = 0.0$ and various values of (ρ_X, ρ_Y), with $\rho_X = \rho_Y$. All standard errors are less than 0.02. Estimated type I error rates when traditional $t-$test are applied, are included.

n = 150, ($\Delta\mu$=0.0)				
(ρ_X,ρ_Y)		(0.1,0.1)	(0.3,0.3)	(0.5,0.5)
	α			
New test	0.01	0.009	0.011	0.014
	0.05	0.060	0.055	0.064
	0.10	0.111	0.109	0.117
$t-$test	0.01	0.010	0.013	0.015
	0.05	0.059	0.057	0.070
	0.10	0.110	0.105	0.117

ANNEXURE B

Source Code

```
!==========================================================================!
! MONTE CARLO METHOD FOR IDENTIFYING THE POWER OF THE TWO
! TEST STATISTICS
!==========================================================================!


        program Powertest_value
        use msimsl
        implicit none

!       Declaration of variables
integer*4 b, n, i, brep, k, j, l, m, mcarlo, mc, ii, jj, ll, bj, bk, bl
real alpha1, alpha2, alpha3, rhoaax, rhoaay, rhoaaxdep, rhoaaydep
real mux, muy, muxdep, muydep, rhoe

parameter (n = 30, m = 30,brep=1000, alpha1=0.1, alpha2=0.05, alpha3=0.01 )
parameter (mcarlo = 2000, mux = 0.0, muy = 0.0, muxdep = 0.0, muydep = 0.0)
real*8 t01pers,t05pers,t10pers,dt01pers,dt05pers,dt10pers, c, dc

! vir n=30
  parameter (t01pers = 2.392377,t05pers = 1.671553, t10pers = 1.296319)
  parameter (dt01pers = 2.462021,dt05pers = 1.699127, dt10pers = 1.311434)
! vir n=50
!  parameter (t01pers = 2.365002,t05pers = 1.660551, t10pers = 1.29025)
!  parameter (dt01pers = 2.404892,dt05pers = 1.676551, dt10pers = 1.299069)
! vir n=100
!  parameter (t01pers = 2.345328,t05pers = 1.652586, t10pers = 1.285842)
!  parameter (dt01pers = 2.364606,dt05pers = 1.660391, dt10pers = 1.290161)
! vir n=150
!  parameter (t01pers = 2.338926,t05pers = 1.64995, t10pers = 1.284399)
!  parameter (dt01pers = 2.351635,dt05pers = 1.655145, dt10pers = 1.287259)
! vir n=200
!  parameter (t01pers = 2.3452,t05pers = 1.6526, t10pers = 1.2858)
!  parameter (dt01pers = 2.3452,dt05pers = 1.6526, dt10pers = 1.2858)

! the following line contains normal quantiles when the rhos are 0, to be applied for
! all the various sample sizes.
! parameter (t01pers = 2.3263,t05pers = 1.6448, t10pers = 1.2816)


integer*4 rx(n-1), ry(n-1), rd(n-1),index1, index2, index3
real*8 x(n), y(n), d(n), mean_x, mean_y, mean_d, adj_x(n), adj_y(n),adj_d(n)
real*8 xind(n-1),xpri(n-1),yind(n-1),ypri(n-1), dind(n-1),dpri(n-1)
real*8 rhox,rhoy,rhod,varx,vary,varpooled,vard,statistic(mcarlo),dstatistic(mcarlo)
real*8 meand,denom, sigey, sigex
real*8 errx(n-1), erry(n-1), aresx(n),aresy(n),bx(n),by(n),gamma0, gamma1,
stdpooled,brhod
```

```
real*8 errd(n-1), aresd(n),bd(n), bdind(n-1), bdpri(n-1), meanbd, bvard, bgamma0,
bgamma1
real*8 meanbx, meanby, bvarx, bvary, bvarpooled, bstdpooled, bmeand
real*8 c_val1(mcarlo),c_val2(mcarlo),c_val3(mcarlo)
real*8 dc_val1(mcarlo),dc_val2(mcarlo),dc_val3(mcarlo), xdep(n), ydep(n), ex(n),
ey(n)
real*8 normx(n), normy(m), exnorm(n), eynorm(n), resx(n), resy(n), resxdep(n),
resydep(n)
real*8 probxy1,probxy2,probxy3, probd1,probd2,probd3,
stderrxy1,stderrxy2,stderrxy3
real*8 stderrd1,stderrd2,stderrd3, vardnum, vardden,  bvardnum, bvardden
real*8 rmcrhox(mcarlo),rmcrhoy(mcarlo),rmcrhod(mcarlo),rhoxstderr,rhoystderr,
rhodstderr
integer countxy1, countxy2, countxy3, countd1, countd2, countd3
integer countxy1t, countxy2t, countxy3t, countd1t, countd2t, countd3t
real*8 bstatistic(brep), rhoxmean, rhoymean, rhodmean
real*8 dbstatistic(brep),rest_rhox,rest_rhoy,rest_rhod
real*8 probxy1t,probxy2t,probxy3t, probd1t,probd2t,probd3t
real*8 stderrxy1t,stderrxy2t,stderrxy3t
real*8 stderrd1t,stderrd2t,stderrd3t

!         Opening the writable output external files
          open(3 , file ="mu00\30power05_01.doc")

rhoe = 0.5d0

rhoaaxdep = 0.1d0
rhoaaydep = 0.1d0

sigey = 2.0d0

rhoaax = 0.1d0
rhoaay = 0.1d0

countxy1 = 0.0d0
countxy2 = 0.0d0
countxy3 = 0.0d0
countd1 = 0.0d0
countd2 = 0.0d0
countd3 = 0.0d0

countxy1t = 0.0d0
countxy2t = 0.0d0
countxy3t = 0.0d0
countd1t = 0.0d0
countd2t = 0.0d0
countd3t = 0.0d0

c = ((1-rhoe**2)/(rhoe**2))**0.5
dc = (1 + c**2)**0.5
```

```
call rnset(0)

! Monte Carlo loop begin here
do mc = 1, mcarlo

! generate the errors from standard normal distribution function
call drnnor(n, normx)
call drnnor(m, normy)
call drnnor(n, exnorm)
call drnnor(n, eynorm)

! Generate the samples to fit an AR(1) models
do i = 1, n
        ex(i) = exnorm(i)
        ey(i) = sigey * ((exnorm(i) + c*eynorm(i))/dc)
         if (i .eq. 1) then
                resx(i) = 0.0d0
                resxdep(i) = 0.0d0
                resydep(i) = 0.0d0
        else
                resx(i) = mux + rhoaax*(resx(i-1) - mux) + normx(i)
                resxdep(i) = muxdep + rhoaaxdep*(resxdep(i-1) - muxdep) + ex(i)
                resydep(i) = muydep + rhoaaydep*(resydep(i-1) - muydep) + ey(i)
        end if
        x(i) = resx(i)
        xdep(i) = resxdep(i)
        ydep(i) = resydep(i)
enddo

l=0
k=0
do l = 1, m
        if (l .eq. 1) then
                resy(l) = 0.0d0
        else
                resy(l) = muy + rhoaay*(resy(l-1) - muy) + normy(l)
        end if
y(l) = resy(l)
end do

d = xdep - ydep

!       Calculating means and centering the series
        mean_x = sum(x)/size(x)
        mean_y = sum(y)/size(y)
        mean_d = sum(d)/size(d)
        adj_x = x - mean_x
        adj_y = y - mean_y
        adj_d = d - mean_d
```

```
! Calculating the test statistic for independent between and dependent within the
variables
varx = sum(adj_x**2)/(dfloat(n-1)) !variance for x's
vary = sum(adj_y**2)/(dfloat(m-1)) !variance for y's

!pooled variance
varpooled = ((dfloat(n-1)*varx) + (dfloat(m-1)*vary))/(dfloat(m+n-2))
stdpooled = varpooled**0.5 !pooled standard diviation
meand = mean_x - mean_y ! means difference
denom = (dfloat(n+m)/dfloat(n*m))**0.5
statistic(mc) = (meand)/(stdpooled*denom)

!       Calculating the first order autoregressive coefficients
        do ii = 1, (n-1)
          xind(ii) = adj_x(ii)
          xpri(ii) = adj_x(ii+1)
        end do

        do jj = 1, (m-1)
          yind(jj) = adj_y(jj)
          ypri(jj) = adj_y(jj+1)
        end do

        do ll = 1, (n-1)
          dind(ll) = adj_d(ll)
          dpri(ll) = adj_d(ll+1)
        end do

        gamma0 = (sum(adj_d**2))/(dfloat(n))
        gamma1 = (sum(dpri*dind))/(dfloat(n))

        rhox = (sum(xpri*xind))/(sum(adj_x**2))
        rhoy = (sum(ypri*yind))/(sum(adj_y**2))
        rhod = gamma1/gamma0

        rmcrhox(mc) = rhox
        rmcrhoy(mc) = rhoy
        rmcrhod(mc) = rhod

!       Calculating the test statistic for dependence between and within variable
        vardnum = gamma0*(1.0 + rhod)
        vardden = n*(1.0 - rhod)
        vard = (vardnum/vardden)
        dstatistic(mc) = (mean_d)/(vard)**0.5

!       Calculating the residuals
        errx = (xpri - rhox*xind) - ((sum(xpri - rhox*xind))/(n-1))
        erry = (ypri - rhoy*yind) - ((sum(ypri - rhoy*yind))/(m-1))
        errd = (dpri - rhod*dind) - ((sum(dpri - rhod*dind))/(n-1))
```

```
!       Bootstrap loop begins
do b = 1, brep

! Generate the random numbers, calculate the bootstrap samples using the bootstrap
! residual that are sampled with replacement from the residuals with prob. 1/(n-1).
call rnund(n-1, n-1, rx)
call rnund(m-1, m-1, ry)
call rnund(n-1, n-1, rd)
bk=0
bl=0
do bk = 1, n
        if (bk .eq. 1) then
                aresx(bk) = x(1)-mean_x
                aresd(bk) = d(1)-mean_d
        elseif (bk .ge. 2) then
                aresx(bk) = mean_x + rhox*(aresx(bk-1) - mean_x) + errx(rx(bk-1))
                aresd(bk) = mean_d + rhod*(aresd(bk-1) - mean_d) + errd(rd(bk-1))
        end if
end do

do bl = 1, m
        if (bl .eq. 1) then
                aresy(bl) = y(1)- mean_y
        elseif (bl .ge. 2) then
                aresy(bl) = mean_y + rhoy*(aresy(bl-1) - mean_y) + erry(ry(bl-1))
        end if
end do

!       Obtain the bootstrap means, together with bootstrap centered samples
meanbx = sum(aresx)/size(aresx)
meanby = sum(aresy)/size(aresy)
meanbd = sum(aresd)/size(aresd)
bx = aresx - mean_x
by = aresy - mean_y
bd = aresd - meanbd

! Calculating the bootstrap test statistics for independence between and dependent
!  within variables
bvarx = sum(bx**2)/(dfloat(n-1))
bvary = sum(by**2)/(dfloat(m-1))
bvarpooled = ((dfloat(n-1)*bvarx) + (dfloat(m-1)*bvary))/(dfloat(m+n-2))
bstdpooled = bvarpooled**0.5
bmeand = (meanbx - mean_x) - (meanby - mean_y)
bstatistic(b) = bmeand/(bstdpooled*denom)

! Calculating the first order autoregressive varince and covarince of the difference
! between variables
        do bj = 1, (n-1)
                bdind(bj) = bd(bj)
                bdpri(bj) = bd(bj+1)
        end do
```

```
bgamma0 = sum(bd**2)/dfloat(n)
bgamma1 = sum(bdpri*bdind)/dfloat(n)
brhod = bgamma1/bgamma0

! Calculating the bootstrap test statistics for dependence between and within variables
bvardnum = bgamma0*(1.0 + brhod)
bvardden = n*(1.0 - brhod)
bvard = bvardnum/bvardden
dbstatistic(b) = ((meanbd-mean_d)/bvard**0.5)

end do !Bootstrap loop ends

!  Sort the bootstrap statistics, and find the indeces for the percentile values
call dsvrgn(brep, bstatistic, bstatistic)
call dsvrgn(brep, dbstatistic, dbstatistic)
index1 = nint(brep*(1.0 - alpha1))
index2 = nint(brep*(1.0 - alpha2))
index3 = nint(brep*(1.0 - alpha3))

! Determine the critical values for level:  1%, 5% and 10%
c_val1(mc) = bstatistic(index1)
c_val2(mc) = bstatistic(index2)
c_val3(mc) = bstatistic(index3)
dc_val1(mc) = dbstatistic(index1)
dc_val2(mc) = dbstatistic(index2)
dc_val3(mc) = dbstatistic(index3)

if (statistic(mc) .ge. c_val1(mc)) then
        countxy1 = countxy1 + 1.0d0
end if

if (statistic(mc) .ge. c_val2(mc)) then
        countxy2 = countxy2 + 1.0d0
end if

if (statistic(mc) .ge. c_val3(mc)) then
        countxy3 = countxy3 + 1.0d0
end if

if (dstatistic(mc) .ge. dc_val1(mc)) then
        countd1 = countd1 + 1.0d0
endif

if (dstatistic(mc) .ge. dc_val2(mc)) then
        countd2 = countd2 + 1.0d0
endif

if (dstatistic(mc) .ge. dc_val3(mc)) then
        countd3 = countd3 + 1.0d0
endif
```

```
if (statistic(mc) .ge. t10pers) then
        countxy1t = countxy1t + 1.0d0
end if

if (statistic(mc) .ge. t05pers) then
        countxy2t = countxy2t + 1.0d0
end if

if (statistic(mc) .ge. t01pers) then
        countxy3t = countxy3t + 1.0d0
end if

if (dstatistic(mc) .ge. dt10pers) then
        countd1t = countd1t + 1.0d0
endif

if (dstatistic(mc) .ge. dt05pers) then
        countd2t = countd2t + 1.0d0
endif
if (dstatistic(mc) .ge. dt01pers) then
        countd3t = countd3t + 1.0d0
endif

enddo

rest_rhox = sum(rmcrhox)/mcarlo
rest_rhoy = sum(rmcrhoy)/mcarlo
rest_rhod = sum(rmcrhod)/mcarlo

probxy1 = dfloat(countxy1)/mcarlo
probxy2 = dfloat(countxy2)/mcarlo
probxy3 = dfloat(countxy3)/mcarlo
probd1 = dfloat(countd1)/mcarlo
probd2 = dfloat(countd2)/mcarlo
probd3 = dfloat(countd3)/mcarlo

stderrxy1 = ((probxy1 * (1 - probxy1))/mcarlo)**0.5
stderrxy2 = ((probxy2 * (1 - probxy2))/mcarlo)**0.5
stderrxy3 = ((probxy3 * (1 - probxy3))/mcarlo)**0.5

stderrd1 = ((probd1 * (1 - probd1))/mcarlo)**0.5
stderrd2 = ((probd2 * (1 - probd2))/mcarlo)**0.5
stderrd3 = ((probd3 * (1 - probd3))/mcarlo)**0.5

probxy1t = dfloat(countxy1t)/mcarlo
probxy2t = dfloat(countxy2t)/mcarlo
probxy3t = dfloat(countxy3t)/mcarlo


probd1t = dfloat(countd1t)/mcarlo
```

```
probd2t = dfloat(countd2t)/mcarlo
probd3t = dfloat(countd3t)/mcarlo

stderrxy1t = ((probxy1t * (1 - probxy1t))/mcarlo)**0.5
stderrxy2t = ((probxy2t * (1 - probxy2t))/mcarlo)**0.5
stderrxy3t = ((probxy3t * (1 - probxy3t))/mcarlo)**0.5

stderrd1t = ((probd1t * (1 - probd1t))/mcarlo)**0.5
stderrd2t = ((probd2t * (1 - probd2t))/mcarlo)**0.5
stderrd3t = ((probd3t * (1 - probd3t))/mcarlo)**0.5

rhoxmean = sum(rmcrhox)/mcarlo
rhoymean = sum(rmcrhoy)/mcarlo
rhodmean = sum(rmcrhod)/mcarlo
rhoxstderr = ((sum((rmcrhox - rhoxmean)**2))/(mcarlo-1))**0.5
rhoystderr = ((sum((rmcrhoy - rhoymean)**2))/(mcarlo-1))**0.5
rhodstderr = ((sum((rmcrhod - rhodmean)**2))/(mcarlo-1))**0.5

! Print the results to external file
Write(3,*) "@@@@@@@@@@@@Results@@@@@@@@@@@@@@@@"
write(3,*) "n = ",n," m = ", m," B = ", brep, " MC = ", mcarlo
print*, "============================================================"
print*, "The estimated coefficient parameters and their bootstrap standard errors"
print*, "============================================================"
print*, "rhoaax = ",rhoaax, " rhoaay = ",rhoaay, " rhoaaxdep = ",rhoaaxdep, "
rhoaaydep = ",rhoaaydep
print*, ""
write(3,"(A12,F8.4,A12,F8.4,A12,F8.4)") "rhoaax = ",rhoaax, " rhoaay = ",rhoaay, "
rhoaaxdep = ",rhoaaxdep, " rhoaaydep = ",rhoaaydep
write(3,*) ""
print*, ""
print*, "The values for mux and muy are", mux, muy
print*, ""
write(3,"(A30,F8.4,F8.4)") "The values for mux and muy are", mux, muy
write(3,*) ""
print*, "The values for muxdep and muydep are", muxdep, muydep
print*, ""
write(3,"(A30,F8.4,F8.4)") "The value for muxdep and muydep are", muxdep,
muydep
write(3,*) ""
print*, "The values for est_rhox is ", rest_rhox," and for est_rhoy is ", rest_rhoy
print*, ""
write(3,"(A30,F8.4,A20,F8.4)") "Est_rhox is ", rest_rhox," and Est_rhoy is ",
rest_rhoy
write(3,*) ""
print*, "The std error for rhox is ", rhoxstderr," and for rhoy is ", rhoystderr
print*, ""
write(3,"(A30,F8.4,A20,F8.4)") "Est_rhox std error is ", rhoxstderr," and for est_rhoy
is ", rhoystderr
write(3,*) ""
print*, "The value for rhod is", rest_rhod
```

```
print*, ""
write(3,"(A30,F8.4)") "The value for rhod is", rest_rhod
write(3,*) ""
print*, "The std error for rhod is ", rhodstderr
print*, ""
write(3,"(A30,F8.4)") "The std error for rhod is ", rhodstderr
write(3,*) ""

print*,"=================================================="
print*,"Two ind samples, where both X and Y are serially correlated for 99TH
PERCENTILE"
print*,"=================================================="
print*, ""
print*, "The powers are", probxy3, probxy3t
print*, ""
print*. "The bootstrap estimated standard error for the power is", stderrxy3, stderrxy3
print*, ""
write(3,*)"================================================="
write(3,*)"Two ind samples, where both X and Y are serially correlated for 99TH
PERCENTILE"
write(3,*)"================================================="
write(3,"(A12,F8.4)") "The power is", probxy3
write(3,*) ""
write(3,"(A12,F8.4)") "The wrong power is", probxy3t
write(3,*) ""

write(3,"(A56,F8.4)") "The bootstrap estimated standard error for the power is",
stderrxy3
write(3,*) ""
write(3,"(A65,F8.4)") "The bootstrap estimated standard error for the wrong power
is",stderrxy3t
write(3,*) ""
print*,"=================================================="
print*,"Two dep samples, where both X and Y are serially correlated for 99TH
PERCENTILE"
print*,"=================================================="
print*, "The powers are", probd3, probd3t
print*, ""
print*, "The bootstrap estimated standard error for the power is", stderrd3, stderrd3
print*, ""

write(3,*)"================================================="
write(3,*)"Two dep samples, where both X and Y are serially correlated for 99TH
PERCENTILE"
write(3,*)"================================================="
write(3,"(A12,F8.3)") "The power is", probd3
write(3,*) ""
write(3,"(A12,F8.3)") "The wrong power is", probd3t
write(3,*) ""
write(3,"(A56,F8.3)") "The bootstrap estimated standard error for the power is",
stderrd3
```

```
write(3,*) ""
write(3,"(A65,F8.3)") "The bootstrap estimated standard error for the wrong power is", stderrd3t
write(3,*) ""

print*,"============================================================"
print*,"Two ind samples, where both X and Y are serially correlated for 95TH PERCENTILE"
print*,"============================================================"
print*, "The powers are", probxy2, probxy2t
print*, "The bootstrap estimated standard error for the power is", stderrxy2, stderrxy2t
print*, ""

write(3,*)"============================================================"
write(3,*)"Two ind samples, where both X and Y are serially correlated for 95TH PERCENTILE"
write(3,*)"============================================================"
write(3,"(A12,F8.4)") "The power is", probxy2
write(3,*) ""
write(3,"(A12,F8.4)") "The wrong power is", probxy2t
write(3,*) ""
write(3,"(A56,F8.4)") "The bootstrap estimated standard error for the power is", stderrxy2
write(3,*) ""
write(3,"(A65,F8.4)") "The bootstrap estimated standard error for the wrong power is", stderrxy2t
write(3,*) ""

print*,"============================================================"
print*,"Two dep samples, where both X and Y are serially correlated for 95TH PERCENTILE"
print*,"============================================================"
print*, "The powers are", probd2, probd2t
print*, "The bootstrap estimated standard error for the power is", stderrd2, stderrd2t
print*, ""

write(3,*)"============================================================"
write(3,*)"Two dep samples, where both X and Y are serially correlated for 95TH PERCENTILE"
write(3,*)"============================================================"
write(3,"(A12,F8.3)") "The power is", probd2
write(3,*) ""
write(3,"(A12,F8.3)") "The wrong power is", probd2t
write(3,*) ""
write(3,"(A56,F8.3)") "The bootstrap estimated standard error for the power is", stderrd2
write(3,*) ""
write(3,"(A65,F8.3)") "The bootstrap estimated standard error for the wrong power is", stderrd2t
write(3,*) ""
```

```
print*,"============================================================"
print*,"Two ind samples, where both X and Y are serially correlated for 90TH
PERCENTILE"
print*,"============================================================"
print*, "The powers are", probxy1, probxy1t
print*, "The bootstrap estimated standard error for the power is", stderrxy1, stderrxy1t
        print*, ""

write(3,*)"============================================================"
write(3,*)"Two ind samples, where both X and Y are serially correlated for 90TH
PERCENTILE"
write(3,*)"============================================================"
write(3,"(A12,F8.4)") "The power is", probxy1
write(3,*) ""
write(3,"(A12,F8.4)") "The wrong power is", probxy1t
write(3,*) ""
write(3,"(A56,F8.4)") "The bootstrap estimated standard error for the power is",
stderrxy1
write(3,*) ""
write(3,"(A65,F8.4)") "The bootstrap estimated standard error for the wrong power
is", stderrxy1t
write(3,*) ""

print*,"============================================================"
print*,"Two dep samples, where both X and Y are serially correlated for 90TH
PERCENTILE"
print*,"============================================================"
print*, "The probability of rejecting null hypothesis when it is false is", probd1,
probd1t
print*, "The bootstrap estimated standard error for the power is", stderrd1, stderrd1t
        print*, ""

write(3,*)"============================================================"
write(3,*)"Two dep samples, where both X and Y are serially correlated for 90TH
PERCENTILE"
write(3,*)"============================================================"
write(3,"(A12,F8.3)") "The power is", probd1
write(3,*) ""
write(3,"(A12,F8.3)") "The wrong power is", probd1t
write(3,*) ""
write(3,"(A56,F8.3)") "The bootstrap estimated standard error for the power is",
stderrd1
write(3,*) ""
write(3,"(A65,F8.3)") "The bootstrap estimated standard error for the wrong power
is", stderrd1t
write(3,*) ""

Write(3,*) "@@@@@@@@@@@@End @@@@@@@@@@@@@@@@@"
  close(Unit=3)
        end program Powertest_value
```

BIBLIOGRAPHY

AMINZADEH, M.S. 1999. Estimation of $P(Z < Y)$ for correlated stochastic time series models. *Journal of Applied Mathematics and Computation*, **104:**179–189.

ALI, M.M. 1973. Distribution of Student's ratio for correlated observations. *Statistiche Hefte*, **14:**350–356.

ALONSO, A.M. and MAHARAJ, E.A. 2006. On the comparison of time series using subsampling. *Working Paper* **#05–07** (2), Universidad Carlos III de Madrid.

ANDERSON, O.D. 1976. *Time Series Analysis and Forecasting, the Box-Jenkins approach.* Buterworths, London and Boston.

BANKS, J. and CARSON, J.S. 1984. *Discrete-event system simulation.* Englewood Cliffs, NJ: Prentice Hall, Inc.

BARNETT, V. and LEWIS, T. 1994. *Outliers in statistical data*, 3rd ed. Chichester: John Wiley and Sons.

BASAWA, I., BILLARD, L., and SRINIVASAN, R. 1984. Large-sample test of homogeneity for time series. *Biometrika*, **71**(1):203–206.

BOWERMAN, B.L., O'CONNELL, R.T. and KOEHLER, A.B. 2005. *Forecasting, Time Series and Regression An applied approach*, 4th ed. United State of America: Thomson Brooks/Cole.

BOX, G. E. P. and JENKINS, G. M. 1976. *Time Series Analysis: Forecasting and Control*, 2nd ed. Holden Day, San Francisco.

COATES, D.S. and DIGGLE, P.J. 1986. Tests for comparing two estimated spectral densities. *Journals of Time Series Analysis*, **7:** 7–20.

CHOU, Y. 1989. *Statistical analysis for business and economics.* New York: Elsevier.

COHEN, J. 1977. *Statistical Power Analysis for the Behavioral Science.* New York: Academic Press.

COHEN, J. 1988. *Statistical Power Analysis for the Behavioral Science*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Association, Publishers.

DANIELS, H.E. 1956. The approximate distribution of serial correlation coefficients. *Biometrics*, **43:**169–185.

DANIEL, W. 1978. *Applied Nonparametric Statistics.* Geogia State University: Houghton Mifflin Company.

DANIEL, W. 1990. *Applied Nonparametric Statistics*, 2nd ed. Boston: PWS-Kent Publishing Company.

DAVISON, A. and HINKLEY, D. 1997. *Bootstrap methods and their application*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press.

DIGGLE, P.J. and FISHER, N.I. 1991. Nonparametric comparison of cumulative periodograms. *Applied Statistics*, **40**(3): 423–434.

DUCHARME, G.R. and LEDWINA, T. 2003. Efficient and nonparametric test for the two-sample problem. *The Annals of Statistics*, **31**(6):2036–2058.

EFRON, B. 1979. Bootstrap methods: another look at the jackknife. *The annals of statistics*, **7**:1–26.

EFRON, B. 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, **68**:589–599.

EFRON, B. and TIBSHIRANI, R. 1993. *An introduction to the bootstrap.* New York: Chapman & Hall.

FRANGOS, C.C. 1984. *The jackknife, the bootstrap and other nonparametric methods in statistics.* University of the Witwatersrand, Johannesburg.

FRANSES, P.H. 1998. *Time series models for business and economic forecasting.* UK: Cambridge University Press.

FULLER, W.A. 1996. *Introduction to statistical time series*, 2nd ed. NY: Wiley interscience.

GUO, J.H. 1999. A nonparametric test for the parallelism of two first-order autoregressive processes. Australia: NZ *Journal of Statistics*, **41**(1):59–65.

HARTLEY, H.O. 1940. Testing the homogeneity of a set of variances. *Biometrika*, **31**:249–255.

HARTLEY, H.O. 1950. The maximum F-ratio as a shortcut test for heterogeneity of variance. *Biometrika*, **37**:308–312.

HOLLANDER, M. and WOLFE, D.A. 1999. *Nonparametric statistical methods.* New York: John Wiley and Sons.

HOWELL, D.C. 1992. *Statistical methods for psychology*, 3rd ed. Boston: PWS–Kent Publishing Company.

KIRK, R. 1995. *Experimental design: Procedures for the behavioral science*, 2nd ed. Belmont, CA: Brooks/Cole Publishing Company.

LAHIRI, S. 2003. *Resampling methods for dependent data*, Springer series in statistics, Springer.

MAHARAJ, E.A. 1996. A significance test for classifying ARMA models. *Journal of Statistical Computing. Simulation*, **54**:305–331.

MAHARA, J.E.A. 2000. Clusters of time series. *Journal Classification*, **17**(2):297–314.

McGREGOR, J.R. and BABB, J.C. 1989. Serially correlated differences in the paired comparison of time series. *Biometrics*, **76**(4):735–739.

MONTGOMERY, D.C. and PECK, E.A. 1992. *Introduction to linear regression analysis*, 2nd ed. New York: John Wiley and Sons, Inc.

NETER, J., KUTNER, M.H., NACHTSHEIM, C.J., and LI, W. 2005. *Applied linear regression models*, 5th ed. NY: McGraw-Hill/Irwin, Inc..

NETER, J., WASSERMAN, W. and KUTNER, M.H. 1983. *Applied linear regression models*, 3rd ed. Homewood, IL: Richard D. Irwin, Inc.

RICE, J.A. 1995, *Mathematical statistics and analysis*, 2nd ed, University of California, Berkeley: Duxbury Press.

SANTANA, L. 2004. *Data dependent choice of the tuning parameter for the modified bootstrap*. North West University: Potchefstroom Campus.

SCHMIDT, J.W. and TAYLOR, R.E. 1970. *Simulation and analysis of industrial systems*. Homewood, IL: Richard D. Irwin, Inc.

SHESKIN, D.J. 2000. *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd ed. Western Connecticut State University: Chapman & Hall.

SNEDECOR, G. W. and COCHRAN, W. G. 1980. *Statistical Methods*, 7th ed. Iowa State University Press.

STEVENS, J. 1996. *Applied multivariate statistics for the social sciences*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

SWANEPOEL, J.W.H. and VAN WYK, J.W.J. 1986. The comparison of two spectral density functions using the bootstrap. *Journals of Statistical Computation, Simulation*, **24**:271–282.

SWANEPOEL, C.J. 1992. *New results on the jackknife and bootstrap methodologies with application to biostatistics*. University of the Witwatersrand.

TIMMER, J., LAUK, M., VACH, W. and LUCKING, C.H. 1999. A test for a difference between spectral peak frequencies. *Computational, Statistics & Data Analysis*, **30**(1):45–50.

WEI, W. 1990. *Time series analysis: Univariate and multivariate methods*. Addison Wesley publishing company.

WELCH, W.J. and GUTIERREZ, L.G. 1988. Robust permutation tests for matched-pairs designs. *Journal of the American Statistical Association,* **83**(402):450–455.

WILCOXON, F. 1945. Individual comparisons by ranking methods. *Biometrics,* **1**:80-83.