CHAPTER 3 DATA AND METHODOLOGY

Chapter 3

DATA AND METHODOLOGY

3.1 Data

The data set consists of daily closing prices of trading/services stocks listed on the Kuala Lumpur Stock Exchange (KLSE). The period of study is from 3 January 1995 to 2 November 2000. Forty stocks were selected from ninety stocks listed under the trading/services sector as at 2 November 2000 for the analysis. The daily closing prices were obtained from the newspaper, i.e., The Star and the New Straits Times, and the KLSE monthly Investors Digest. Stocks included in the sample are those whereby data on the daily closing prices are available for every trading day in the period used for this study. This excludes stocks that had been suspended or not traded for any day in the period of analysis. Also, only stocks that remained listed in trading/services sector in the entire period of study are included.

These daily stock prices were corrected for capital adjustments, i.e., bonus issue, rights issue, bonus issue and rights issue, stock split, and consolidation. The detail for calculating the return after the capital adjustment is as follows:

(i) *Bonus issue*

Example: Company A proposed a bonus issue of 1 for 4 and Ex-date: 15.3.1998.

Price of Stock A on 14.3.1998 (t-1) : P_{t-1}

Price of Stock A on 15.3.1998 (t) : P_t

Period	No. of Shares	Price	Investment value
t −1	4	P_{t-1}	$4P_{t-1}$
t	5	P_t	$5P_t$

$$\therefore \text{Return}_t = \ln\left(\frac{5P_t}{4P_{t-1}}\right)$$

(ii) *Rights issue*

Example: Company B proposed a rights issue of 2 for 5 @ RM1.10 and Ex-date: 22.7.1999.

Price of Stock B on 21.7.1999 (t-1) : P_{t-1}

Price of Stock B on 22.7.1999 (t) : P_t

Period	No. of Shares	Price	Investment value
t −1	5	P_{t-1}	$5P_{t-1}$ + 2(RM1.10)
t	7	P_t	$7P_t$

$$\therefore \text{Return}_t = \ln\left(\frac{7P_t}{5P_{t-1} + \text{RM}2.20}\right)$$

(iii) *Bonus issue and rights issue*

Example: Company C proposed a bonus issue of 1 for 3 and a rights issue of 1 for 3 @ RM1.50 and Ex- date: 20.1.2000.

Assume: the bonus issue is not entitled to the rights issue and vice versa.

Price of Stock C on 19.1.2000 (t-1) : P_{t-1}

Price of Stock C on 20.1.2000 (t) : P_t

Period	No. of Shares	Price	Investment value
$t-1$	3	P_{t-1}	$3P_{t-1}$ + 1(RM1.50)
t	5	P_t	$5P_t$

$$\therefore \text{Return}_t = \ln\left(\frac{5P_t}{3P_{t-1} + \text{RM}1.50}\right)$$

(iv) *Stock split*

Example: Share with par value of RM1.80 is split into 2 shares, i.e., 90 cents per share. This adjustment procedure is the same as what we do for bonus issue of 1 for 1.

Period	No. of Shares	Price	Investment value
$t-1$	1	P_{t-1}	P_{t-1}
t	2	P_t	$2P_t$

$$\therefore \text{Return}_t = \ln\left(\frac{2P_t}{P_{t-1}}\right)$$

(v) *Consolidation*

Example: 3 shares of RM1.20 each become 1 share of RM1.20. This adjustment procedure is the opposite of what we do for bonus issue.

Period	No. of Shares	Price	Investment value
$t-1$	3	P_{t-1}	$3P_{t-1}$
t	1	P_t	P_t

$$\therefore \text{Return}_t = \ln\left(\frac{P_t}{3P_{t-1}}\right)$$

The stocks used in this study were divided into two categories, i.e., small companies with market capitalization below RM700 million and large companies with market capitalization above RM700 million as at 30 October 2000. Therefore, twenty stocks are in the small company category and the large company category, respectively. A list of these stocks with their respective market capitalization is presented in *Appendix A*.

Futhermore, to examine the persistency and consistency of the calendar anomalies over time, the whole period of study, i.e., 3 January 1995 to 2 November 2000 is divided into two subperiods. Subperiod 1 is from 3 January 1995 to 30 June 1997 and subperiod 2 is from 1 July 1997 to 2 November 2000. This marks two subperiods before and after the start of the financial crisis.

3.2 Methodology

The daily stock returns R_t are computed as follows:

$$R_t = \ln P_t - \ln P_{t-1}$$

where R_t is the daily stock return at time t,

P_t is the daily closing price at time t, and

P_{t-1} is the daily closing price at time (t-1).

For the purpose of this study, the main null hypotheses (H_0) of calendar anomalies to be tested on forty stocks of the trading/services sector are:

(1) The pre-holiday effect - the difference in mean returns for the trading days before holidays (pre-holidays) and the other trading days equals zero.

(2) The half-monthly effect – the difference in mean returns for the first half of a trading month and second half of a trading month equals zero.

(3) The time-of-the-month effect – the mean returns for the first third of a month, second third of a month and last third of a month are equal.

The first hypothesis aims to test whether there is any statistically significant difference between the mean returns for the trading days before holidays (pre-holidays) and the other trading days. Therefore, the trading days are divided into two groups. The first group is the trading days before the holidays, i.e., pre-holidays and the second group is the trading days after excluding the day before the holidays, i.e., the other trading days. The mean returns for the first group are then compared with the mean returns for the second group. The null and alternative hypotheses (H_0 and H_a, respectively) for testing the pre-holiday effect are as follows:

$$H_0(1) : \mu_{HOL} = \mu_O$$

$$H_a(1) : \mu_{HOL} \neq \mu_O$$

where μ_{HOL} is the mean return of a stock for the trading days before holidays and μ_O is the mean return for the other trading days. If the null hypothesis is rejected, we can conclude that there is a statistically significant difference between the mean returns for the pre-holiday trading days and the other trading days. In this study, we consider tests at the five per cent significance level.

When the stock market is closed because of a holiday, the returns for the first trading day after the holiday are omitted from the data set. If Thursday is a holiday, then the return for Friday is omitted in this study. The holidays in this study include all the public holidays observed by the Federal Territory where KLSE is located. These amount to thirteen different public holidays, i.e., New Year's Day, Federal Territory Day, Chinese New Year, Hari Raya Puasa, Hari Raya Haji, Labour Day, Awal Muharram, Wesak Day, Birthday of DYMM SPB Yang Dipertuan Agong, Birthday of Prophet Muhammad, National Day, Deepavali and Christmas Day. These dates for the period of study, from 3 January 1995 to 2 November 2000, are obtained from the diary.

The second hypothesis compares the mean returns of a stock for the first half of a trading month with the mean returns for the second half of a trading month. It is testing for the half-monthly effect on each stock, i.e., testing whether there is any statistically significant difference between the mean returns for the first half and the second half of a trading month. Therefore, the trading month is split into two halves. Following Ariel (1987), the first half consists of the last trading day of the previous month and the first eight trading days of the month. The second half consists of the nine days prior to the last trading day of the calendar month. The null and alternative hypotheses (H_0 and H_a, respectively) for testing the half-monthly effect are as follows:

$H_0(2): \mu_{FH} = \mu_{SH}$

$H_a(2): \mu_{FH} \neq \mu_{SH}$

where μ_{FH} is the mean return of a stock for the first half of a trading month and μ_{SH} is the mean return for the second half of a trading month. If the null hypothesis is rejected, this means that there is a statistically significant difference between the mean returns for the first half and the second half of a trading month.

The third hypothesis tests the equality of mean returns for each one third of a month. It is based on the inclusion of mean returns of all trading days of a month. It aims to test whether there is any statistically significant difference between the mean returns for the first third of a month, second third of a month, and last third of a month. Therefore, the trading days will be divided into three groups. Following Kohers and Patel (1999), the first third of a month consists of the 28th day of the previous month through the 7th day of the month. The second third of a month extends from the 8th day of the month through the 17th day. The last third of a month consists of the 18th day of the month through the 27th day. The null and alternative hypotheses (H_0 and H_a, respectively) for testing the time-of-the-month effect are as follows:

$H_0(3): \mu_{FT} = \mu_{ST} = \mu_{LT}$

$H_a(3)$: At least two μ's are not equal

where μ_{FT} is the mean return of a stock for the first third of a month, μ_{ST} is the mean return for the second third of a month, and μ_{LT} is the mean return for the last third of a month. If the null hypothesis is rejected, we can conclude that there is a statistically significant difference between the mean returns for the first third, second third, and last third. For the significant cases, i.e., rejection of the null hypothesis of equality in mean returns, we conduct a further analysis using multiple comparisons test to look for pairwise mean differences that lead to the rejection.

3.2.1 Test for Normality

Many prior researches regarding financial ratios have shown that financial ratio distributions are not normal. They suggest that one of the reasons for this non-normality is because of the presence of outliers. Some of them suggest that the data should be transformed to square root or natural log if it is found that the distribution is not normally distributed. Many tests of hypothesis assume normality. Therefore, it is approriate to test for normality before proceeding further.

The normality test can be performed by using the one-sample Kolgomorov-Smirnov test. This test is useful when the mean and variance of the population are unknown as it uses the estimates from the sample itself. The Kolgomorov-Smirnov test is a test of goodness-of-fit. This test is concerned with the degree of agreement between the cumulative distributions of observed (sample) relative frequencies and expected (theoretical) relative frequencies. The test determines whether the distribution of the cumulative relative frequencies of the sample compares reasonably with the distribution of the cumulative relative frequencies expected under the null hypothesis. The null and alternative hypotheses (H_0 and H_a, respectively) are as follows:

$H_0(4) : F_S(x) = F_T(x)$

$H_a(4) : F_S(x) \neq F_T(x)$

where x represents the return,

$F_S(x)$ is the observed cumulative frequency distribution of a random sample of n observations, and

$F_T(x)$ is the theoretical cumulative frequency of a normal distribution.

The Kolgomorov-Smirnov test concerns itself with the absolute value of the maximum deviation (difference), of the expression $F_S(x)$ - $F_T(x)$ and is designated by the symbol D. Then, the test statistic in this test is given by

$$D = \text{maximum} \mid F_S(x) - F_T(x) \mid$$

where the expression $\mid F_S(x) - F_T(x) \mid$ indicates the absolute of the difference. The decision rule is to accept H_0 if the D value is smaller or equal than the critical value of $D_{n,\alpha}$ at the α per cent level of significance for sample size n. We will reject H_0 if the D value is greater than the critical value of $D_{n,\alpha}$ at the α per cent level of significance for sample size n. If we reject the null hypothesis, we can conclude that the data is not normally distributed.

3.2.2 Test for Equality of Variances

Some parametric tests of equality of mean returns used in this study assume homogeneity of variances. The Levene test will be used for this assumption of equal population variances. The null and alternative hypotheses (H_0 and H_a, respectively) for the Levene test are as follows:

$H_0(5) : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$

$H_a(5)$: At least two variances are different

where t is the number of groups of interest (or treatments).

The test statistic for this test is stated as below:-

$$F_0 = \frac{\left[\sum_{i=1}^{t} n_i\left(\overline{w_{i.}} - \overline{w_{..}}\right)^2\right] / [t-1]}{\left[\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(w_{ij} - \overline{w_{i.}}\right)^2\right] / [N-t]} \sim F_{t-1,N-t} \quad \text{under } H_0(5)$$

where $w_{ij} = | x_{ij} - \overline{x_{i.}} |$ is the absolute difference between the j^{th} observation of the unit receiving i^{th} treatment and the sample mean of the i^{th} treatment,

$\overline{w_{i.}} = \dfrac{\sum_{j=1}^{n_i} w_{ij}}{n_i}$ is the mean of the absolute differences for the i^{th} treatment,

$\overline{w_{..}} = \dfrac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} w_{ij}}{N}$ is the overall mean common to all the absolute differences,

N is total number of observations, and

n_i is sample size for the i^{th} treatment.

The appropriate reference distribution for F_0 is the F distribution with (t – 1) and (N – t) degrees of freedom. The null hypothesis would be rejected if F_0 is greater than the critical value of $F_{\alpha,\ t-1,\ N-t}$ at the α per cent level of significance. If we reject the null hypothesis, the assumption of equal population variances is violated, i.e., variances are unequal.

3.2.3 Hypothesis Tests

Two types of statistical tests are used in this study, i.e., parametric tests and nonparametric tests.

3.2.3.1 Parametric Tests

The assumption of normality must be satisfied when a parametric test is applied. The parametric tests that are carried out in this study include the following:

(i) Two Independent Samples t Test

This test is appropriate to use for comparing two treatment means in this study. The null hypotheses to be tested are $H_0(1)$ and $H_0(2)$. If the assumption of equal population variances is justified, the sample variances computed from the two samples will each be an estimate of the common variance, σ^2. We obtain a single estimate of σ^2 by pooling the two sample estimates. This pooled estimate can be obtained by computing the weighted average of the two sample variances, where the weight are the degrees of freedom. Then, the pooled estimate of the common σ^2 is given by

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

The standard error of the estimator, $\overline{x_1}$ - $\overline{x_2}$ is

$$\sqrt{\left(\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}\right)} = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Therefore, the test statistic to be used is as below:

$$t_0 = \frac{\left(\overline{x_1}-\overline{x_2}\right)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2} \text{ under } H_0(1) \text{ and } H_0(2)$$

where $\overline{x_1}$ is the mean return for the tradings before holidays (i.e., pre-holidays) for $H_0(1)$ or the first half of a trading month for $H_0(2)$,

$\overline{x_2}$ is the mean return for the trading days after excluding the day before the holidays (i.e., the other trading days) for $H_0(1)$ or second half of a trading month for $H_0(2)$,

s_1^2 is the sample variance for the tradings before holidays (i.e., pre-holidays) for $H_0(1)$ or the first half of a trading month for $H_0(2)$,

s_2^2 is the sample variance for the trading days after excluding the day before the holidays (i.e., the other trading days) for $H_0(1)$ or second half of a trading month for $H_0(2)$,

n_1 is the sample size for the tradings before holidays (i.e., pre-holidays) for $H_0(1)$ or the first half of a trading month for $H_0(2)$, and

n_2 is the sample size for the trading days after excluding the day before the holidays (i.e., the other trading days) for $H_0(1)$ or second half of a trading month for $H_0(2)$.

t_0 follows the t distribution with (n_1+n_2-2) degrees of freedom under the null hypothesis. We will reject H_0 if the absolute value of t_0 is greater than the critical value of $t_{\alpha/2, n_1+n_2-2}$ at the α per cent level of significance. The rejection of a null hypothesis shows that there is a statistically significant difference between the mean returns for the trading days before holidays (pre-holidays) and other trading days for $H_0(1)$ or the mean returns for the first half and the second half of a trading month for $H_0(2)$.

When the population variances are unequal, even though the populations are normally distributed, the t distribution as mentioned above are not suitable for comparing two treatment means in this study.

Then, the appropriate test statistic to be used is given by

$$t' = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df'} \quad \text{under } H_0(1) \text{ and } H_0(2)$$

When the population variances are not equal, the degrees of freedom is not (n_1+n_2-2). We have to use a modified value of the degrees of freedom (df') as below:

$$df' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

(ii) *One-way Analysis of Variance (ANOVA)*

One-way ANOVA is used to test for a significant difference among several means. It is an extension of the t-test for the difference between two means that was discussed above. The assumptions for this test are that the data for the treatments must come from normally distributed populations with equal variances and samples are selected randomly and independently. The term of treatment refers to any factor that the experimenter controls. This test is called the one-way or single factor analysis of variance because only one factor is investigated. We use a fixed effects model to compare the effects of treatment on a dependent variable. The fixed effects model is as below:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2,\ldots,t \\ j = 1,2,\ldots,n_i \end{cases}$$

where X_{ij} is the $(ij)^{th}$ observation,

μ is a parameter common to all treatments called the overall mean,

τ_i is a parameter unique to the i^{th} treatment called i^{th} treatment effect, and

ε_{ij} is a random error term such that is normally and independently distributed with mean zero and a constant variance, σ^2 $[\varepsilon_{ij} \sim NID(0, \sigma^2)]$.

We are interested in testing the equality of the t treatment means, that is,

$H_0(6)$: $\mu_1 = \mu_2 = \ldots = \mu_t$

$H_a(6)$: At least two μ's are not equal

The one-way ANOVA is used to test $H_0(3)$, i.e., to test whether there is any statistically significant difference between the mean returns for the first third of a month, second third of a month, and last third of a month.

The test statistic for this test is given by

$$F_0 = \frac{SS_{Treatment}/(t-1)}{SS_E/(N-t)} = \frac{MS_{Treatment}}{MS_E} \sim F_{t-1,\,N-t} \text{ under } H_0(6)$$

where $SS_{Treatment}$ is the sum of squares for treatments and measures the variation explained by the differences between the treatment means,

SS_E is the sum of squares for error and it is a measure of the unexplained variability, obtained by calculating a pooled measure of the variability within the t samples,

$MS_{Treatment}$ is the mean square for treatments, and

MS_E is the mean square for error.

The ratio of the variability between-groups to the variability within-groups, F_0 follows a F distribution with $(t-1)$ and $(N-t)$ degrees of freedom under the null hypothesis. If F_0 is greater than $F_{\alpha,\,t-1,\,N-t}$ at the α per cent level of significance, it will lead us to reject the null hypothesis. This rejection means at least two of the mean returns for each one third of a month are different. Thus, the time-of-the-month effect exists.

(iii) Two-way Analysis of Variance (ANOVA) with unequal cell frequencies

Two-way ANOVA with unequal cell frequencies is adopted to test for the equality of row treatment effects and column treatment effects. It also determines whether there is a significant interaction between the row and column treatments. The two-way ANOVA is applied to all the hypotheses of $H_0(1)$, $H_0(2)$, and $H_0(3)$. This is because besides the factor considered in each hypothesis, another factor which is month-of-the-year effect may exist. Say the treatment factor is column factor and the month-of-the-year effect is row factor. This test requires each cell to contain at least one observation ($n_{ij} \geq 1$ for all i, j). Let X_{ijk} be the observed returns when the row factor is at the i^{th} level (i = 1,2,...,12) and the column factor is at the j^{th} level (j = 1,2,...,t) for the k^{th} replicate (k = 1,2,...,n_{ij}).The observations may be described by the linear statistical model as below:

$$X_{ijk} = \quad \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\ldots,12 \\ j = 1,2,\ldots,t \\ k = 1,2,\ldots,n_{ij} \end{cases}$$

where X_{ijk} is the $(ijk)^{th}$ observation,

μ is overall mean effect,

α_i is the effect of the i^{th} level of the row factor,

β_j is the effect of the j^{th} level of column factor,

$(\alpha\beta)_{ij}$ is the effect of the interaction between α_i and β_j, and

ε_{ijk} is a random error term such that is normally and independently distributed with mean zero and a constant variance, σ^2 [$\varepsilon_{ijk} \sim NID(0, \sigma^2)$].

Both row and column factors, are of equal interest. Specifically, we are interested in testing hypothesis about the equality of row treatment effects, that is

$H_0(7a) : \alpha_1 = \alpha_2 = \ldots = \alpha_{12} = 0$

$H_a(7a) :$ At least one $\alpha_i \neq 0$

The test statistic is given by

$$F_{\text{Row Treatment}} = \frac{SS_{\text{RowTreatment}}/11}{SS_E/(N-12t)}$$

$$= \frac{MS_{\text{RowTreatment}}}{MS_E} \sim F_{11,\,N-12t} \text{ under } H_0(7a)$$

where $SS_{\text{Row Treatment}}$ is the sum of squares due to the row factor,

SS_E is the sum of squares due to error,

$MS_{\text{Row Treatment}}$ is the mean square due to the row factor, and

MS_E is the mean square due to error.

Besides that, we are also interested in testing the hypothesis about the equality of column treatment effects, that is

$H_0(7b) : \beta_1 = \beta_2 = \ldots = \beta_t = 0$

$H_a(7b) :$ At least one $\beta_j \neq 0$

The test statistic is given by

$$F_{\text{Column Treatment}} = \frac{SS_{\text{ColumnTreatment}}/(t-1)}{SS_E/(N-12t)}$$

$$= \frac{MS_{\text{ColumnTreatment}}}{MS_E} \sim F_{t-1,\,N-12t} \text{ under } H_0(7b)$$

where $SS_{\text{Column Treatment}}$ is the sum of squares due to the column factor,

SS_E is the sum of squares due to error,

$MS_{\text{Column Treatment}}$ is the mean square due to the column factor, and

MS_E is the mean square due to error.

In addition, we are also interested in determining whether the row and column factors interact. Thus, we wish to test

$H_0(7c)$: $(\alpha\beta)_{ij} = 0$

$H_a(7c)$: At least one $(\alpha\beta)_{ij} \neq 0$ for all i, j

The test statistic is given by

$$F_{Interaction} = \frac{SS_{Interaction} / 11(t-1)}{SS_E / (N-12t)}$$

$$= \frac{MS_{Interaction}}{MS_E} \sim F_{11(t-1),\, N-12t} \text{ under } H_0(7c)$$

where $SS_{Interaction}$ is the sum of squares due to interaction between the row and column factors,

SS_E is the sum of squares due to error,

$MS_{Interaction}$ is the mean square due to interaction between the row and column factors, and

MS_E is the mean square due to error.

If there are differences between the row treatment effects, then the $MS_{Row\ Treatment}$ will be significantly larger than MS_E. Similarly, if there are column treatment effects or interaction is present, then the corresponding mean squares ($MS_{Column\ Treatment}$ or $MS_{Interaction}$, respectively) will be significantly larger than MS_E. If we assume that the model is adequate and that the error term ε_{ijk} is normally and independently distributed with mean zero and a constant variance σ^2, then $F_{Row\ Treament}$, $F_{Column\ Treatment}$, and $F_{Interaction}$ are distributed as F distribution with (11), (t-1), and 11(t-1) degrees of freedom in the numerator, respectively, and (N – 12t) degrees of freedom in the denominator, under H_0(7a), H_0(7b), and H_0(7c), respectively. If we reject the null hypothesis H_0(7a), we conclude that the month-of-the-year effect is present after controlling for the column treatment. If H_0(7b) is rejected, then the pre-holiday, half-monthly or time-of-the-month effect (column treatment) exists in this study, depending on which of H_0(1), H_0(2) or H_0(3) is tested, after controlling for the month-of-the-year effect (row treatment). In addition, if H_0(7c) is rejected, we conclude that there is a significant interaction between the column treatment and month-of-the-year effect.

(iv) Duncan's Multiple Range Test

Suppose that the null hypothesis is rejected in conducting an analysis of variance involving more than two treatments. This is the case for H_0(3). Thus, we conclude that there are differences between the treatment means, but exactly where the differences occur may not be clear. Therefore, further comparisons among groups of treatment means may be useful. The procedures for making these comparisons are usually known as multiple comparison methods. In order to identify specific segments of a month with significantly different returns in the third hypothesis H_0(3), the Duncan's multiple range test is utilized. One of the advantages of the procedure, which has been widely used, is its simplicity. To apply this test, the t treatment averages are arranged in ascending order. The standard error of each average is determined as below:

$$s_{\bar{x}_i} = \sqrt{\frac{MS_E}{n_h}}$$

where

$$n_h = \frac{t}{\sum_{i=1}^{t}\left(\frac{1}{n_i}\right)}$$

A pair of means is only significantly different at the α per cent level of significance if the observed difference is greater than the critical value, R_p. The critical value is given as below:

$$R_p = D_{\alpha;\,p,\,\text{error df}}\sqrt{\frac{MS_E}{t}\left\{\sum_{i=1}^{t}\left(\frac{1}{n_i}\right)\right\}}$$

where α is the significance level,

p is the number of means encompassed in the range being tested (p=2,3),

error df is the number of degrees of freedom for error, and

MS_E is mean square due to error and is based on (N-t) degrees of freedom for the one-way ANOVA or (N-12t) degrees of freedom for the two-way ANOVA.

3.2.3.2 Nonparametric Tests

If the assumption of normality does not hold, alternative nonparametric or distribution-free tests are used. These tests assume no knowledge about the underlying populations except maybe that the variables of interest are continuous. The primary disadvantage for the tests is that they do not utilize all the information provided by the sample and thus are inefficient. However, the nonparametric tests are more efficient than the parametric tests with serious departures from normality. The nonparametric tests that are used in this study are as follows:

(i) Mann-Whitney Test

A nonparametric test alternative to the two independent samples t test is the Mann-Whitney test. We are interested in testing the null hypothesis $H_0(1)$ and $H_0(2)$ that the two samples come from identical populations against the alternative hypothesis that the two populations have unequal means. For each sample, the returns are arranged in ascending order. A rank of 1 is assigned to the smallest observation, 2 assigned to the second smallest, and so on. Let W_1 be the sum of the ranks of the observations in the first sample. Let W_2 be the sum of the ranks of the observations in the second sample. We use the related statistics as below:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

$$U = \min(U_1, U_2)$$

where n_1 is the number of observations in the first sample, and

n_2 is the number of observations in the second sample.

The test statistic U is used for a two-tailed test by comparing this value to the critical value of U_0. The null hypothesis would be rejected if the calculated U value is smaller or equal than the critical value of U_0 at the α per cent level of significance. If we reject the null hypothesis, we concude that there is a statistically significant difference between the mean returns for the trading days before holidays (pre-holidays) and other trading days for $H_0(1)$ or the mean returns for the first half and the second half of a trading month for $H_0(2)$.

(ii) Kruskal-Wallis Test

In many situations a researcher may feel that one or more of the assumptions underlying the one-way ANOVA are not met. Therefore, the H-test, or the Kruskal-Wallis test may provide a suitable alternative test, which makes use of ranks rather than original observations. It tests the null hypothesis that all t populations have the same probability distribution against the alternative hypothesis that at least two of the t populations have distributions that differ in location. To perform a Kruskal-Wallis test, first the observations are replaced by ranks in ascending order from 1(for the smallest observations) to N (for the largest observations) in the combined set of data. In the case of ties, each of the tied observations (observations having the same value) is replaced by the average of the ranks. The test statistic for this test is as below:

$$H = \frac{12}{N(N+1)}\left(\sum_{i=1}^{t}\frac{R_{i.}^2}{n_i}\right) - 3(N+1) \quad \sim \chi^2_{(t-1)} \text{ under } H_0(3)$$

where $R_{i.}$ is the sum of the ranks in the i^{th} treatment.

If there are k number of groups with ties, the H statistic is adjusted by dividing by

$$1 - \frac{\sum_{j=1}^{k}\left(m_j^3 - m_j\right)}{N^3 - N}$$

where m_j is the number of tied observations in the j^{th} group of tied observations.

Then, the adjusted test statistic is given by

$$H = \frac{\frac{12}{N(N+1)}\left(\sum_{i=1}^{t}\frac{R_{i.}^2}{n_i}\right) - 3(N+1)}{1 - \frac{\sum_{j=1}^{k}\left(m_j^3 - m_j\right)}{N^3 - N}} \sim \chi^2_{(t-1)} \text{ under } H_0(3)$$

The computed value of H is compared to the critical value of $\chi^2_{\alpha,2}$ for $H_0(3)$. If the calculated H value is greater than the critical value of $\chi^2_{\alpha,2}$, we reject the null hypothesis and conclude that there is a difference in location among the distributions of mean returns for each one third of a month. Thus, the time-of-the-month effect exists in this study.

3.3 Statistical Programs

Statistical software programs are used in this study for analyzing the data. This includes SPSS (Statistical Package for Social Science) and Microsoft Excel.