A TEST FOR NON-GAUSSIAN DISTRIBUTIONS ON THE JOHANNESBURG STOCK EXCHANGE AND ITS IMPLICATIONS ON FORECASTING MODELS BASED ON HISTORICAL GROWTH RATES.

By

Lloyd A Corker

A mini-thesis submitted in partial fulfilment of the requirements for the degree of Magister Commercii in the Faculty of Economic and Management Sciences, University of The Western Cape.

December 2002

KEYWORDS

Stable distribution
Heavy-tail
Capital Market Theory
Gaussian
Symmetry
Infinite variance
Continuous time paths
Diffusion processes
Central Limit Theorem
Arbitrage

ABSTRACT

If share price fluctuations follow a simple random walk then it implies that forecasting models based on historical growth rates have little ability to forecast acceptable share price movements over a certain period. The simple random walk description of share price dynamics is obtained when a large number of investors have equal probability to buy or sell based on their own opinion. This simple random walk description of the stock market is in essence the Efficient Market Hypothesis, EMT.

EMT is the central concept around which financial modelling is based which includes the Black-Scholes model and other important theoretical underpinnings of capital market theory like mean-variance portfolio selection, arbitrage pricing theory (APT), security market line and capital asset pricing model (CAPM). These theories, which postulates that risk can be reduced to zero sets the foundation for option pricing and is a key component in financial software packages used for pricing and forecasting in the financial industry.

The model used by Black and Scholes and other models mentioned above are Gaussian, i.e. they exhibit a random nature. This Gaussian property and the existence of expected returns and continuous time paths (also Gaussian properties) allow the use of stochastic calculus to solve complex Black-Scholes models. However, if the markets are not Gaussian then the idea that risk can be reduced to zero can lead to a misleading and potentially disastrous sense of security on the financial markets.

This study project test the null hypothesis – share prices on the JSE follow a random walk – by means of graphical techniques such as symmetry plots and Quantile-Quantile plots to analyse the test distributions. In both graphical techniques evidence for the rejection of normality was found. Evidence leading to the rejection of the hypothesis was also found through nonparametric or distribution free methods at a 1% level of significance for Anderson-Darling and Runs test.

The study project further aims to fit a Levy-Stable distribution to the data as suggested by Mandelbrot (1963) and then by Fama (1965) to describe the distributions of speculative prices. Stable parameters were calculated and further evidence obtained for the rejection of the null hypothesis. The distributions obtained indicated skewed heavy-tailed distributions becoming Gaussian with increasing differences in prices in accordance with the Central Limit Theorem.

The study project concludes that the nature of the Leptokurtic distributions obtained for the change in price for the first few weeks are indeed non-Gaussian and therefore holds serious consequences for the application of classical capital market theory.

DECLARATION

I declare that *A test for Non-Gaussian distributions on the Johannesburg Stock Exchange and its implications on forecasting models based on historical growth* rates is my own work, that all sources I have used have been accurately reported and acknowledged, and that this document has not previously in its entirety or in part been submitted at any university in order to obtain an academic qualification.

Lloyd A Corker

December 2002

ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr S Huda, my supervisor, for his continual support and advice on economic matters throughout my commerce studies and for his believe and confidence in me. Special thanks to Dr Theunis JvW Kotze for his counsel and valued insight on statistical matters during this study project. Thanks to all the departmental staff especially Meagan Langeveldt, departmental secretary, for all her patience, diligence and true professionalism.

Where I have succeeded it is only through the motivation and unfailing encouragement of my friends everywhere –

Jerome, Dylan, Samantha, Cheryl, Luzelle, Marie-Anne, Delia, Cedric, Tania, Roderick, Leona, Yolanda, Marcelle, Ruth, Natasha, Richard, Jerry, Ntobeko, John, Stephanie, Prudence, Piet, Ivan, Lionel, Egroy, Bruce, Justice, Dominic, Shabier, Mary-Anne, Venetia.........

Deidré thank you for making me smile again.

Amigos para siempre

Table of contents

LIST OF TABLES

List of Figures

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE

Chapter 1 of this study project introduces the reader to the research topic in terms of the following:

- The relative importance of the research topic and its impact on modelling and forecasting speculative prices.
- Defining the objectives of the study project.
- Defining the scope of the study project.
- Outlining the approach that is followed in the study project.
- Data used for the analysis.
- Describing the methodology that was followed in researching the topic.

1.2 BACKGROUND

Classical economic theory has had considerable success in advancing the understanding of financial markets in the last fifty years. As a result, research theories such as the Capital Asset Pricing Model (CAPM), Efficient Market Hypothesis (EMH), Arbitrage Pricing Theory (APT) and the Black-Scholes model have played an important role in the success of financial risk management during this time. Physicists started in the recent time applying their concepts and methods to study economic problems. For this work the name Econophysics has been established. Much work in the field of Econophysics is focused on understanding the statistical properties of financial time series.

In the last years it has been recognised in the physical sciences that unpredictable time series and stochastic processes are not synonymous.

Specifically, chaos theory has shown that unpredictable time series can arise from deterministic non-linear systems. Also, empirical research has brought more clearly into focus some of the deficiencies of classical economic theory; such as share price dynamics having a non-Gaussian (non-normal) distribution i.e. share price changes do not follow a random walk.

The study is of importance to all who still work on the premise that share price dynamics exhibit random walk properties such as in the fields of quantitative finance, econometrics, optimisation, multivariate analysis and many other fields. If the markets are not Gaussian (normal) then the false expectation of finite variance, expected returns and continuous time paths - implying risk can be reduced to zero in capital market theory - can lead to a false sense of security on the financial markets.

1.3 OBJECTIVE OF THE STUDY PROJECT

The objective of the study is to subject a popular assumption about the behaviour of stock market prices following a simple random walk to empirical test. It further aims to model stock market price dynamics on the Johannesburg Stock Exchange using the Levy-Stable process.

1.4 SCOPE OF THE STUDY PROJECT

The focus of the project is to test for non-Gaussian behaviour of historical stock market price fluctuations using Johannesburg Stock Exchange All-Share Index determined by the closing prices from 04 January 1988 to 31 December 2001. It also attempts to estimate the distribution function using Stable laws as suggested by Mandelbrot (1963) and Fama (1965). It restricts comments to the impact this study project will have on important theoretical underpinnings of capital market theory like mean-variance portfolio selection, APT, security market line and CAPM, which are based on the concept of speculative price changes exhibiting Gaussian behaviour.

1.5 APPROACH

This study project is based on a similar study done by Kunst et al (1991) – 'Testing for randomness and Stability on the Austrian stock market' and by J. Huston McCulloch (1978) – 'Continuous Time Processes with Stable Increments'. This study also places emphasis on graphical analyses in testing for non-Gaussian distributions. The non-parametric test is done using DataPlot (version 7/2002) and the maximum likelihood estimation of Stable parameters is calculated using STABLE 3.04 (for Windows 95/98/NT/2000, 2002, 6 February).

1.6 DATA USED

The object of the tests performed in this study is logarithmic weekly changes of the Johannesburg Stock Exchange All-Share Index determined by the closing prices from 04 January 1988 to 31 December 2001.

The study is disadvantaged as the All-Share Index presents problems over that of individual stocks in that a survival bias is present due to companies being floated and de-listed on The Johannesburg Stock Exchange for the test period. Also, changes between averages tend to bring in biases against the random walk model by temporal aggregation (Kunst et al, 1991).

The natural logarithmic form of the random walk model is used which states that the logarithm of the price at time, $t + 1$, is the sum of the logarithm of the price at time, t, and a residual at time, $t + 1$, i.e.

$$R_{mt} = \text{Ln}P_{(t + \Delta t)} - \text{Ln}P_t \qquad (1)$$

Where,

R_{mt}: market return, in period t;

P_t: price index at period t;

$P_{(t + \Delta t)}$: price index at period $t + \Delta t$

Ln: natural log.

The taking of natural logarithms induces stationarity in the time-series to some extent and is justified both theoretically and empirically. Theoretically, logarithmic returns are analytically more tractable when linking together sub-period returns to form returns over longer intervals (Poshakwale, 1996). Empirically logarithmic returns are more likely to be normally distributed which is a prior condition of standard statistical techniques (Strong, 1992). The taking of logarithms is also a natural assumption from keeping the price from ever going negative.

1.7 RESEARCH METHODOLOGY APPLIED

The research methodology consists of three parts. The first part uses graphical techniques to analyse distributional symmetry through the graphical analyses of symmetry plots. In addition, Quantile-Quantile plots are analysed for randomness of the time-series data.

The second part uses distribution free (non-parametric) test for non-Gaussian distributions, namely the Runs test and the Anderson-Darling test.

The third part estimates the Stable parameters using Levy-Stable laws introduced by Paul Levy during his investigations of the behaviour of sums of independent random variables in the early 1920's. The method of maximum likelihood estimation is used to estimate the Stable parameters, which is widely used in many software packages such as S-Plus and STABLE.

CHAPTER 2

LITERATURE SURVEY

2.1 OBJECTIVE

Chapter 2 attempts to make explicit and clear to the reader the central issues that shapes and bounds the research question. It provides the context within which the study project is located as well as the gaps and contradictions existing in the literature.

2.2 LITERATURE SURVEY

The first approach to solving financial problems with the help of scientific methods came from Louis Bachelier in 1900. He modelled price dynamics as an ordinary random walk where prices can go up and down due to a variety of many independent random causes. Despite his theory, anticipating Einstein's theory of Brownian motion published in 1905, Bachelier's model of the stock market was, however, too simple and failed to capture many of the crucial aspects of price fluctuations, for example the possibility of extreme events on the financial markets such as crashes and market booms.

The Gaussian distribution used to model the stock market is ubiquitous in all branches of natural and social sciences and this is basically due to the Central Limit Theorem: the sum of independent, or weakly dependent, random disturbances, all of them with finite variance, results in a Gaussian random variable (Masoliver, 1999).

Knowing the probability distribution of speculative prices is one of the most important problems in mathematical finance yet in spite of its importance for both theoretical and practical applications the problem is yet unsolved. Gaussian models are widely used in finance although, as Kendall (1953) first noticed, the Gaussian distribution does not fit financial data well especially at the tails of the distribution. It is at the tails where price distributions are crucial in the analysis of financial risk. Therefore, obtaining a reliable distribution that captures the areas in the tails of the distributions has profound consequences for estimating and modelling risk.

One of the first attempts to explain the appearance of heavy-tails in financial data was made by Mandelbrot (1963) who based on Pareto Levy-Stable laws – also called Levy-Stable or Stable laws - obtained a Leptokurtic distribution. Although most papers reject the Gaussian distribution in favour of the Stable distribution such as Stevenson and Bear (1970) and Dusak (1973); Hall, Brorsen and Irwin (1989) and Hudson, Leuthold and Sarassoro (1987) claim that futures returns are not Stable.

Assuming the financial world to be Gaussian would imply the existence of expected returns and continuous time paths both properties essential to important theoretical underpinnings of capital market theory like mean-variance portfolio selection, Capital Asset Pricing Model (CAPM), Efficient Market Hypothesis (EMH), Arbitrage Pricing Theory (APT) and security market line. A Gaussian model of the financial world would also simplify complex stochastic models needed to price options and other derivative instruments. Should the distributions be non-Gaussian Leptokurtic distributions (heavy-tailed) then expected returns would fail to exist and market processes would not be continuous pure diffusion processes, essential to financial modelling theory, but rather an infinite amount of discontinuities (Bamberg, 2001).

However, assuming the financial world to be best described by a Stable process would imply frightening mathematical consequences to the classical capital market theorist. Stable distributions have infinite sample variance and therefore expected returns do not exist. Also there will exist an infinite number of discontinuities in any finite time interval such as our time series data (McCulloch, 1978). It is these discontinuities restricting the use of normal increment diffusion processes, which has discouraged the use of Stable distributions. However, as early as 1976 Robert C. Merton (an advocate of normal increment diffusion processes) pointed out that the continuity of a pure diffusion process (Gaussian process) is actually one of its drawbacks.

J. Huston McCulloch (1978) showed that discontinuities are not statistically unmanageable or economically unreasonable and in fact are actually appealing in their own right for a price series.

CHAPTER 3

APPLICATION OF RESEARCH METHODOLOGY

3.1 OBJECTIVE

The aim of chapter 3 is to elaborate and apply the research methodology outlined in chapter 1 in order to establish some common ground with the reader for the sake of understanding the analyses and results in chapter 4.

3.2 RESEARCH METHODOLOGY

As outlined in chapter 1 the natural logarithmic form of the random walk model is used:

$$R_{mt} = \text{Ln}P_{(t+\Delta t)} - \text{Ln}P_t \,;\ t = 1,2,3,\ldots,n.$$

Where n is the total number of daily closing prices on the All-Share Index from 04 January 1988 to 31 December 2001, i.e. 3487 recorded observations.

The model is tested taking daily, weekly and monthly price differences of the All-Share Index. Weekly price differences is defined as the difference in the closing price on consecutive Mondays while monthly differences is defined as Δt = 20 days i.e.

Daily:	$\Delta t = 1, 2, 3, \ldots\ldots\ldots\ldots, n - 1.$
Weekly:	$\Delta t = 5, 10, 15, \ldots\ldots\ldots, n - 5.$
Monthly:	$\Delta t = 20, 40, 60, \ldots\ldots., n - 20.$

Day1 represents the price change of one day i.e. $\Delta t = 1$ and Day25 represents a price change of 25 days i.e. $\Delta t = 25$ and so on.
Week1 represents the price change of one week i.e. $\Delta t = 5$ (5 trading days in a week) and Week25 represents a price change of 25 weeks i.e.
$\Delta t = 25(5) = 125$ and so on.

Month1 represents the price change of one month i.e. $\Delta t = 20$ (20 trading days in a month) and Month25 represents a price change of 25 months i.e.
$\Delta t = 25(20) = 500$ and so on.

Thus Week25 should have the same time-series data and therefore the same distribution as Day125 (25×5 trading days) and Month10 should equal Day200 (10×20 trading days).

3.3 GRAPHICAL ANALYSES

This part of the study project uses graphical techniques to analyse the distributions of weekly price changes of the All-Share Index to draw conclusions about the distributions being Gaussian or non-Gaussian by analysing

- Symmetry plots and
- Quantile-Quantile plots

3.3.1 SYMMETRY PLOTS

Symmetry Plots are useful for assessing the symmetry of the data and hence whether the data fulfil one of the characteristics of a Gaussian distribution. A Gaussian distribution is symmetrical about its mean (or median) value.

The construction of a symmetry plot first entails sorting the data values from smallest to largest. Then all the values to the left and to the right of the median are collected and plotted a respective distance from the median. This process is repeated for the pair of points that are second closest to the median, then third closest, and so on.

If the distribution of the price changes of the All-Share Index is symmetric, the points will lie close to the diagonal line. If the distribution is positively skewed, the points will deviate above the line and if the distribution is negatively skewed, the points will deviate below the line i.e. a non-Gaussian distribution.

3.3.2 QUANTILE-QUANTILE PLOTS

For this study project Quantile-Quantile plots are used to compare the test distribution with a Gaussian distribution. The construction of a Quantile-Quantile plot entails sorting the *n* observed data points into ascending order, so that:

$$x_1 \leq x_2 \leq \ldots\ldots \leq x_n$$

These observed values are then plotted against one axis of the graph while on the other axis the plot will show:

$$F^{-1}((i-r_{adj}) / (n+n_{adj}))\ ;$$

where i is the rank of the respective observation, r_{adj} and n_{adj} are adjustment factors (≤ 0.5) which ensure that the p-value for the inverse probability integral will fall between 0 and 1, and F^{-1} denotes the inverse of the probability integral for the respective standardized distribution.

For detrended Gaussian plots the plots are the differences between the observed and expected values. If the sample is from a Gaussian distribution the plots should cluster in a horizontal band around zero with no distinguishable pattern.

3.4 NON-PARAMETRIC TEST

Non-parametric or distribution free methods are used to process data if the researcher knows nothing about the parameters of the variable of interest in the population (such as the mean or the standard deviation). To test the null hypothesis - the data follows a Gaussian distribution - we make use of two well-known test for normality, Anderson-Darling and Runs test.

3.4.1 Anderson-Darling test

The Anderson-Darling procedure is a general test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. It is a modification to the much used Kolmogorov-Smirnov (K-S) test and gives much more weight to the tails than does the K-S test. The Anderson-Darling test makes use of the Gaussian distribution (for this study project) in calculating critical values which is much more sensitive than the distribution-free K-S test.

We define the Anderson-Darling test as:

H_0: The data follows a Gaussian distribution.

H_a: The data do not follow a Gaussian distribution.

Test Statistic: The Anderson-Darling test statistic is defined as

$A^2 = -N - S$ where

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N} [\ln F(y_i) + \ln(1 - F(Y_{N+1-i}))]$$

and F is the cumulative Gaussian distribution with y_i the ordered data pair (Anderson-Darling and Shapiro-Wilk tests, 2003, March 17).

The test is a one-sided test and the hypothesis that the distribution is of a specific form (Gaussian) is rejected if the test statistic, A, is greater than the critical value at a 1% level of significance.

3.4.2 Runs Test

The Runs test is used to decide if a data set is from a random process. This is done by using the laws of probability to estimate the number of runs one would expect by chance. There are several alternative formulations of the Runs test in literature but a run is generally defined as a series of increasing or decreasing values. The number of increasing, or decreasing, values is the length of the run. In a random data set, the probability that the $(l + 1)^{th}$ value is larger or smaller then the l^{th} value follows a Binomial distribution, which forms the basis of the Runs test.

Our Runs test comprises the following:

- Runs of length exactly l for l = 1, 2,……., 10
- Number of Runs of length l
- Expected number of Runs of length l
- Standard deviation of the number of Runs of length l
- A z-score where the z-score is defined to be

$$Z_i = \frac{Y_i - \overline{Y}}{S}$$

where $\overline{Y}$ is the sample mean and s is the sample standard deviation (Runs Test, 2003, March 17).

H_0: The data follows a Gaussian distribution.

H_a: The data do not follow a Gaussian distribution.

A 5 % level of significance is used with an absolute z-score greater than 1.96 (two-tailed test) indicating non-randomness.

3.5 ESTIMATING THE DISTIRBUTION FUNCTION

After testing for normality, the study project attempts to find an appropriate distribution to describe the data set. As mentioned in chapter 1 it was first suggested by Mandelbrot (1963) and then by Fama (1965) to use Stable laws to describe the distributions of speculative prices.

Stable distributions are a rich class of probability distributions (including the Gaussian distribution) that allow for skewness and heavy-tails and have many intriguing mathematical properties, Weron (2001).

Definition 3.1. A random variable X is Stable (in the broad sense) if for X_1 and X_2 independent copies of X and any positive constants a and b,

$$aX_1 + bX_2 \stackrel{D}{=} cX + d,$$

for some positive c and some $d \in \mathbb{R}$. The random variable is strictly Stable (or Stable in the narrow sense) if $d = 0$ for all choices for a and b (Goovaerts et al,).

Stable distributions have been proposed as models for many complex physical and economic systems. There are several reasons for using a non-Gaussian Stable distribution to describe a system. Firstly, there are strong theoretical reasons for expecting a non-Gaussian Stable model, e.g. when expecting extreme events giving rise to a Cauchy distribution (Feller, 1971).

The second reason is the Generalised Central Limit Theorem, which says that when a sample size is large the sample mean possesses an approximately Gaussian distribution if the random sample is taken from any distribution with a finite mean, and a finite variance (Nolan, 1999). Fund and risk managers who base their forecasting models on the idea that risk can be reduced to zero – a property of a Gaussian distribution - particularly appreciate this property of Stable distributions.

The third argument for modelling with Stable distributions is empirical: many large data sets exhibit heavy-tails and skewness and therefore expected returns do not exist.

The fourth argument is these distributions have the convenient and appealing property that when two random variables drawn from Stable distributions having the same characteristic exponent, α, (see section 3.6) are added together, the resulting sum will have the same shaped distribution (McCulloch, 1978). Thus if the markets are Stable then the distribution for say daily price change will look the same for weekly price changes and monthly price changes and so on solving the problem of which statistics to use over different time frames.

Examples of Stable distributions being used in finance and economics are given in Mandelbrot (1963), Fama (1965), Embrechts, Klüppelberg, and Mikosch (1997), Cheng and Rachev (1995), McCulloch (1996) and Belkacem, Véhel and Walter (1996).

3.6 STABLE DISTRIBUTIONS

Paul Levy introduced Stable laws during his investigations of the behaviour of sums of independent random variables in the early 1920's (Weron, 2001). However, despite having the solution for this family of distributions the general form of the Stable distributions is not available. The expensive need for computational power and their frightening properties in continuous time context have caused this family of distributions to be largely ignored by practitioners. But with the increase of availability of reliable computer software programs to estimate these distributions Stable models are becoming more attractive to practitioners. This is not surprising because unlike the 2-parameter Gaussian distribution (mean and variance) the Stable model has 4 parameters and can therefore fit a data set much better (Nolan, 1999).

Stable distributions are peculiar in that they include heavy-tailed distributions as well as the Gaussian distribution. But they go further than the Gaussian distribution in that they can be Leptokurtic (heavy-tailed) but also have the same functional form with possibly different parameters, a property of the Gaussian distribution. A Gaussian distribution is thus a special type of a Stable distribution. Based on the above property, definition 3.1. can be restated as: those distributions for which the following identity in distribution holds for any number $n \geq 2$:

$$\sum_{i=1}^{n} X_i \overset{D}{=} C_n X + D_n$$

where X_i are identical independent copies of X and C_n, D_n are constants (Focardi, 2001).

Though we lack closed form formulae for the probability density functions what we have are their moment generating function. This in turn can take on different kinds of parameterisations (Nolan, 1999). But for numerical purposes it is useful to use the moment generating function,

$$\operatorname{Ln} \varphi_0(x) = \begin{cases} -\sigma^{\alpha}|x|^{\alpha}\{1 + i\beta \sin(x) \tan\frac{\pi\alpha}{2}[(\sigma|x|)^{1-\alpha}-1]\} + i\mu_0 x, & \alpha \neq 1 \quad (2) \\ -\sigma|x|\{1 + i\beta \sin(x)\frac{2}{\pi}\operatorname{Ln}(\sigma|x|)\} + i\mu_0 x, & \alpha = 1 \end{cases}$$

The four parameters (α, β, σ and μ) of the Stable model are:

- $\alpha \in (0,2]$, an index of stability or peakness in the tails. It describes the rate at which the tails of the distribution taper off. If $\alpha = 2$ then the distribution is Gaussian with finite variance but if $\alpha < 2$ then the variance is infinite.
- $\beta \in [-1,-1]$, an index of skewness in the distribution. If $\beta = 0$, the Stable densities are symmetric if $\beta < 0$ the distribution is skewed to the left if $\beta > 0$ the distribution is skewed to the right.
- $\sigma > 0$, is a scale parameter, which describes the width of the peak in the distribution.
- $\mu \in \mathbb{R}$, is a location parameter, which describes the shift of the peak of the distribution.

STABLE is used to estimate the parameters of the Stable distribution.

CHAPTER 4

RESULTS OF ANALYSES

4.1 OBJECTIVE

Chapter 4 provides a discussion of the results of the study project. The discussion follows the same structure as the research methodology outlined in chapter 3.

4.2 RESULTS

Analyses was done for price changes of the following time series:

- Day1, Day10, Day25 and Day50;
- Week1, Week10, Week25 and Week50;
- Month1, Month10, Month25 and Month50.

4.3 GRAPHICAL ANALYSES

This part of the study project uses graphical techniques to analyse the distributions of weekly price changes of the All-Share Index to draw conclusions about the distributions being Gaussian or non-Gaussian by analysing

- Symmetry plots and
- Quantile-Quantile plots

4.3.1 SYMMETRY PLOTS

The symmetry plots obtained clearly show positively skewed distributions from Day1 up to Week10. This indicates more probability in the right tail areas, contrary to most speculative price time series data, which are often negatively skewed. These distributions are not Gaussian. Increasing the price change (Δt) causes symmetry to become negatively skewed. Creating a motion picture of symmetry plots from Day1 to Day2000 (Month100) shows symmetry to move back and forth across the median as Δt increases from positively skewed to negatively skewed. However as Δt increases the deviations in the tail areas gets closer to the median (see Figure 4.1a – i). Analyses of these symmetry plots, shows the variation of probability in the tail areas with changing skewness. The skewness of the distributions is of importance to the choice of statistics one would use for analyses of time series data and has deep implications for risk management. For example assuming a Gaussian distribution would seriously underestimate an extreme event like a market boom, if the distribution for that market were positively skewed.

SYMMETRY PLOTS

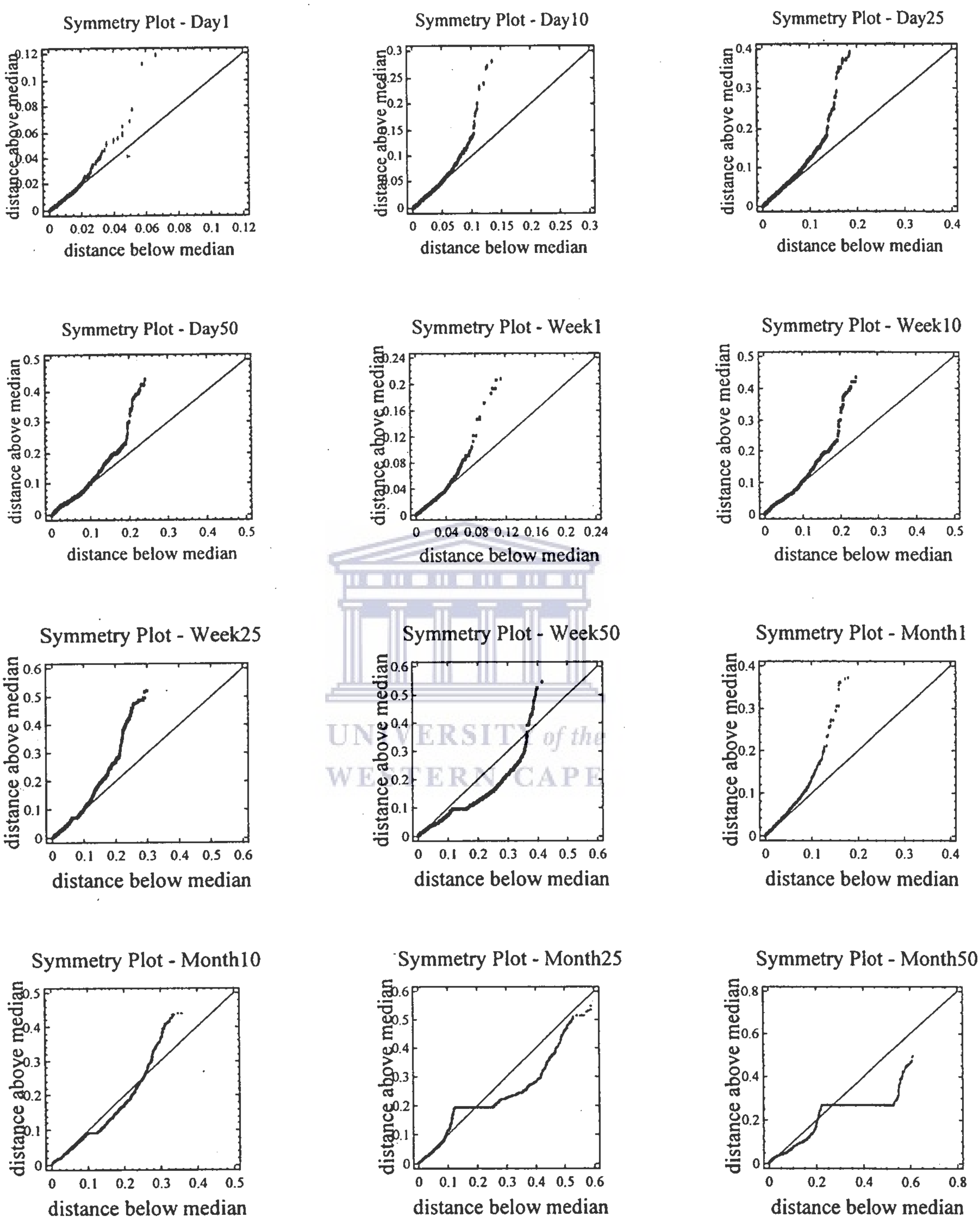

Fig. 4.1a - l. Symmetry plots of daily, weekly and monthly logged price differences of the All-Share Index

4.3.2 QUANTILE-QUANTILE PLOTS

Much the same can be said for Quantile-Quantile plots. The plot for Day1 clearly shows strong deviation from the normal in the positive tail (see Figure 4.2a – i). If the sample is from a Gaussian distribution the plots should cluster in a horizontal band around zero with no distinguishable pattern. The pattern becomes less distinguishable as Δt increases implying a tendency to normality.

QUANTILE-QUANTILE PLOTS

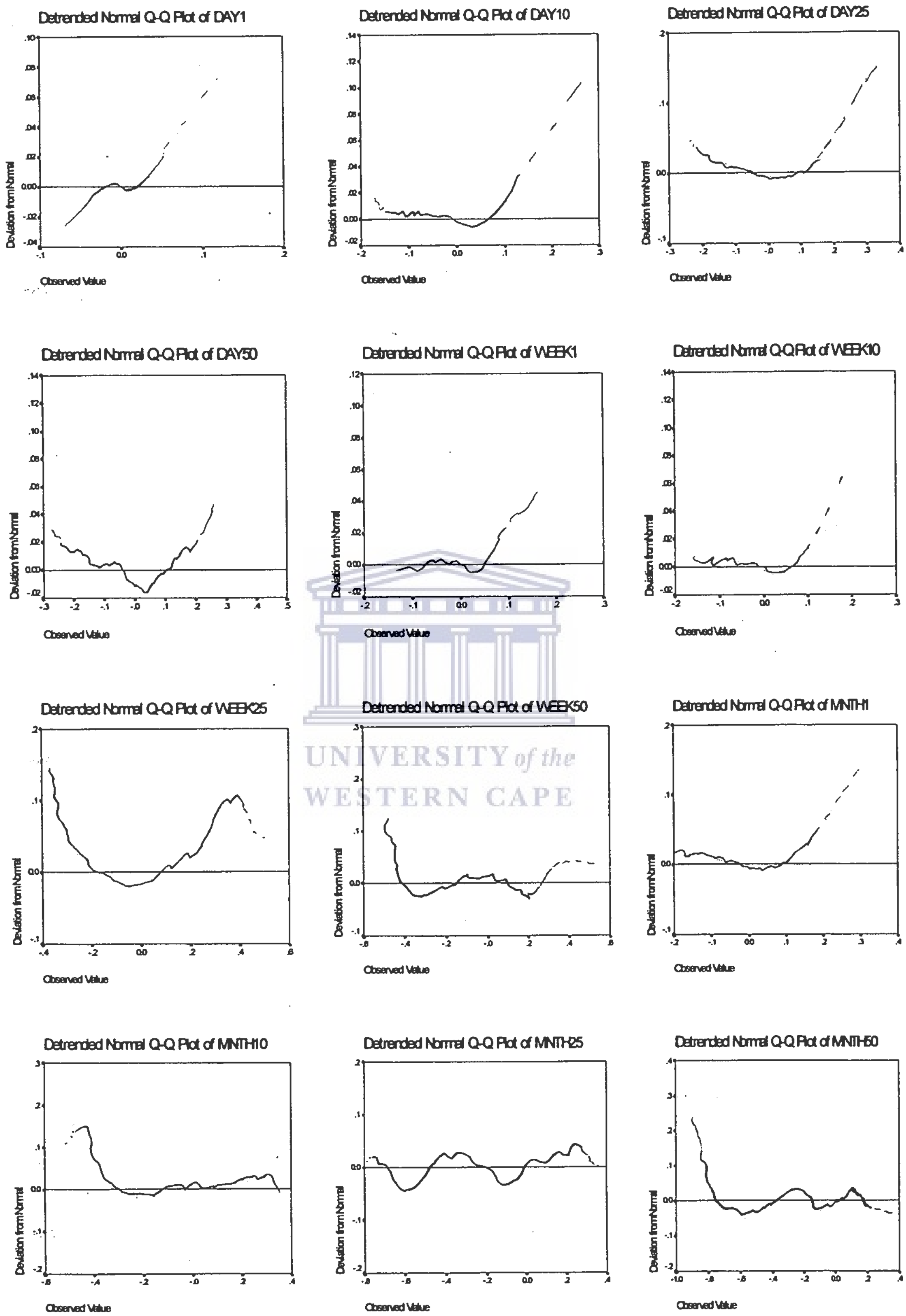

Fig. 2a - l. Quantile-Quantile plots of daily, weekly and monthly logged price differences of the All-Share Index

4.4 NON-PARAMETRIC TEST

4.4.1 Anderson-Darling Test

We defined the Anderson-Darling test as:

H_0: The data follows a Gaussian distribution.

H_a: The data do not follow a Gaussian distribution.

In all cases the null hypothesis was rejected at 1% level of significance. These distributions, what ever they are, are not Gaussian.

Table 4.2.2.1 Anderson-Darling Test

	Mean	Std. Dev.	Anderson-Darling Test Statistic Value	Adjusted Test Statistic Value	H_o: Data Normally Distributed At 1% L.o.S.
Day 1	-0.50×10^{-3}	0.108×10^{-1}	40.75703	40.80371	Rejected
10	-0.50×10^{-2}	0.408×10^{-1}	12.06139	12.07520	Rejected
25	-0.122×10^{-1}	0.673×10^{-1}	12.14745	12.16136	Rejected
50	-0.241×10^{-1}	0.941×10^{-1}	13.82702	13.84286	Rejected
Week 1	-0.247×10^{-2}	0.208×10^{-1}	21.01522	21.03932	Rejected
10	-0.497×10^{-2}	0.418×10^{-1}	11.95635	11.97006	Rejected
25	-0.602×10^{-1}	0.136721	17.42562	17.44560	Rejected
50	-0.120731	0.171819	9.624641	9.635678	Rejected
Month 1	-0.101×10^{-1}	0.596×10^{-1}	15.12489	15.14231	Rejected
10	-0.954×10^{-1}	0.1600515	8.358234	8.367860	Rejected
25	-0.225619	0.1963332	11.54947	11.56277	Rejected
50	-0.388609	0.2349333	23.35362	23.38052	Rejected

4.4.2 Runs Test

In all cases for both Runs above and below the median as well as Runs up and Runs down, the null hypothesis was rejected.

Table 4.2.2.2 Runs test Above and Below

Runs Above and Below Median					
	Number of Runs above and below (median)	Expected number of Runs	Large sample test statistic	P-value	H_o: Data Random at 1% L.o.S.
Day 1	1507 (-0.00052)	1744.0	-8.01235	1.112×10^{-15}	Reject
10	377 (-0.00613)	1744.0	-46.2954	0.0	Reject
25	225 (-0.017252)	1744.0	-51.445	0.0	Reject
50	195 (-0.029231)	1744.0	-52.4614	0.0	Reject
Week 1	577 (-0.002753)	1744.0	-39.5197	0.0	Reject
10	195 (-0.02923)	1744.0	-52.4614	0.0	Reject
25	101 (-0.07100)	1744.0	-55.646	0.0	Reject
50	72 (-0.0937154)	1744.0	-56.6285	0.0	Reject
Month 1	247 (-0.014058)	1744.0	-50.6997	0.0	Reject
10	73 (-0.0886833)	1744.0	-56.5946	0.0	Reject
25	82 (-0.190867)	1744.0	-56.2897	0.0	Reject
50	40 (-0.268782)	1744.0	-57.7126	0.0	Reject

Table 4.2.2.3 Runs test Up and Down

Runs Up and Down					
	Number of Runs up and down median	Expected number of Runs	Large sample test statistic	P-value	H_o: Data Random at 1% L.o.S.
Day 1	2190	2324.33	-5.37666	7.6096×10^{-8}	Reject
10	1511	2324.33	-32.655	0.0	Reject
25	1509	2324.33	-32.7353	0.0	Reject
50	1495	2324.33	-33.2978	0.0	Reject
Week 1	1495	2324.33	-33.2978	0.0	Reject
10	1495	2324.33	-33.2978	0.0	Reject
25	1448	2324.33	-35.186	0.0	Reject
50	1412	2324.33	-36.6323	0.0	Reject
Month 1	1510	2324.33	-32.6952	0.0	Reject
10	1411	2324.33	-36.6724	0.0	Reject
25	1270	2324.33	-42.337	0.0	Reject
50	1074	2324.33	-50.2112	0.0	Reject

4.5 MAXIMUM LIKELIHOOD ESTIMATION OF STABLE PARAMETERS

The quantile estimate of SO parameters of the Stable model for the time-series under consideration is given in Table 4.3.1. Stable curves were fitted for data sets with outliers and data sets without outliers using maximum likelihood estimation procedure in STABLE.

The table indicates a Leptokurtic distribution for the first 25 weeks and a Gaussian distribution for the 50th week. The distribution obtained for time series week1 has an α = 1.6192 (with outliers) this is in accordance with Fama's (1965) results on US stocks and Kunst et al (1991) on Austrian stocks. Thus one can conclude that a weekly price change for the All-Share Index has a non-Gaussian distribution. Without outliers α = 1.6348 indicating that non-normality is not induced by outliers in the time series for week1. Outliers in the time-series were identified and deleted using StatGraphics.

A comparison of the results show the outliers in the data only slightly affecting the skewness of the distributions with outliers making the distributions more positively skewed. The location and width of the peaks of the distributions are seemingly unaffected by outliers.

What is very noticeable is the decreasing size of the amount of distribution found in the tails as Δt increases. There is a gradual transition from a strong Leptokurtic to a Gaussian distribution i.e. a decrease in the area in the tails (α = 2 and β = 0) as expected by the Central Limit Theorem (CLT).

The Table below illustrates that the Gaussian distribution is a poor descriptor of the time-series data for the first few weeks but with a tendency to a Gaussian process as Δt increases.

Table 4.3.1 Stable Model with Maximum Likely Hood Estimators

Stable Model with Maximum Likely Hood Estimator								
Quantile Estimate of SO Parameters	α outliers	α	β outliers	β	σ outliers	σ	μ outliers	μ
Day 1	1.567241	1.748783	-0.04687	-0.20948	0.005635	0.00530	-0.00044	-0.00034
10	1.680056	1.704422	0.16932	0.113521	0.025804	0.024272	-0.00790	-0.00677
25	1.788537	1.829951	0.54540	0.760619	0.042537	0.041637	-0.02048	-0.20747
50	1.572887	1.637102	0.25753	0.201643	0.053564	0.53483	-0.03300	-0.3199
Week 1	1.619213	1.634876	0.02993	-0.00765	0.015952	0.01591	-0.00290	-0.00278
10	1.541894	1.637102	0.24786	0.201643	0.054178	0.053483	-0.03540	-0.03199
25	1.787482	1.742298	0.82927	0.798098	0.087850	0.08554	-0.08335	-0.81510
50	2.000000	2.000000	0.00000	0.000000	0.128394	0.128512	-0.09470	-0.09370
Mnth 1	1.786871	1.785238	0.56809	0.513148	0.03625	0.035771	-0.01676	-0.01714
10	2.0000	2.00000	0.00000	0.000000	0.122445	0.118920	-0.09202	-0.08870
25	2.0000	2.00000	0.0000	0.000000	0.176787	0.165348	-0.2171	-0.19090
50	2.0000	0.00000	0.0000	0.000000	0.297882	0.299258	-0.2859	-0.26880

4.6 ECONOMIC AND STATISTICAL IMPLICATIONS

The Stable hypothesis makes two assumptions: (1) the population variance of the distribution of the first differences is infinite, and (2) empirical distributions conform best to a Stable distribution (Fama, 1965).

The infinite variance assumption (variance grows with the size of the sample) has important implications from a statistical and economic standpoint. From a purely statistical standpoint the sample variance will be a meaningless measure of dispersion while its median and other fractiles will all exist and be reasonable. Statistical tools based on a finite variance assumption such as least squares regression will give misleading answers (Mandelbrot, 1963).

From an economic perspective infinite variance will produce ever growing fluctuations in economic phenomena but perhaps more seriously if infinite variance exist then expected share prices and expected (discrete) returns do not exist and therefore the important theoretical underpinnings of capital market theory like mean-variance portfolio selection, Capital Asset Pricing Model (CAPM), Efficient Market Hypothesis (EMH), Arbitrage Pricing Theory (APT) and security market line are no longer valid (Bamberg, 2001).

Bamburg and Dorfleitner, 2001, showed that any assumption of discrete returns as normally distributed clearly ignores that the Gaussian distribution is not closed under multiplication:

From equation (1) we have

$$R_{mt} = \mathrm{Ln}\left(\frac{P_t}{P_{t-1}}\right) \qquad (3)$$

i.e.

$$e^{R_{mt}} \left(\frac{P_t}{P_{t-1}}\right) \qquad (4)$$

i.e.

$$P_t = (P_{t-1})\ e^{R_{mt}} \qquad (5)$$

from which we can see that R_{mt} is the continuously compounded rate of return.

From the definition of R_{mt}

$$P_t = P_0 e^{R_1 + R_2 + \dots + R_t} = P_0 e^{R_{0t}} \qquad (6)$$

where the log return R_{0t} corresponds to the time interval [0,*t*].

From equation (6) we get the additivity property

$$R_1 + R_2 + \dots\dots R_t = R_{0t} \qquad (7)$$

The log return R_t^P of a portfolio is related to the log returns $R_t^{(i)}$ of the single stocks by the non-linear formula

$$R_t^P = \ln\left(\sum_{i=1}^{n} a_i e^{R_t^{(i)}}\right), \qquad (8)$$

where $a_1, a_2, \ldots, a_n$ are the fractions of the invested capital. But because of the non-linearity of equation (8) traditional capital market theory prefers to work with percentage price changes, i.e.

$$DR_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1 = e^{R_t} - 1 \qquad (9)$$

We now have a linear relation between the returns of single stocks and the portfolio return, i.e.

$$DR_t^P = \sum_{i=1}^{n} a_i DR_t^{(i)} \qquad (10)$$

where $R_t \in \mathbb{R}$ and $DR_t \in [-1,\infty)$. Thus any assumption of discrete returns as normally distributed clearly contradicts property (10) and also ignores that the Gaussian distribution is not closed under multiplication.

To show that heavy-tailed log return distributions of any kind leads to non-existing expected prices and discrete returns Bamberg and Dorfleitner used probability theory:

If a random variable $X \geq 0$ has finite expectation E(X), then the upper tail behavior must be of the following type:

$$\lim_{x\to\infty} x[P(X > x)] = 0 \qquad (11)$$

but the negation of (11) is

$$\text{as } x \to \infty \; x[P(X > x)] \not\to 0 \Rightarrow E(X) = \infty. \qquad (12)$$

With $X = e^R$ and from (11)

$$\lim_{x\to\infty} xP(e^R > x) = \lim_{x\to\infty} xP(R > \ln x) = \lim_{x\to\infty} e^z P(R > z) = \lim_{z\to\infty} e^z \frac{c}{z^\alpha} = \infty \quad (13)$$

But this violates (11) as $\lim_{x \to \infty} x[P(X > x)] \neq 0$ and with (12) we have

$$E(X) = E(e^R) = \infty$$

Thus any heavy-tailed log returns distribution or Leptokurtic distribution of any kind leads to non-existing expected prices and non-existing discrete returns with far-reaching consequences for traditional capital market theory. The origin of the observed Leptokurtosis is still being debated. There are several models trying to explain these heavy-tail distributions, Stable model - Mandelbrot (1963), power law exponent - Lux (1996), Non-Gaussian because of uneven activity during trading hours – Clark (1973), quasi-Stable stochastic process with finite variance – Mantegna, Stanley (1994) and diffusion process superimposed with a Poisson-driven process – Merton (1976).

Also it is well known that continuous time stochastic processes - Gaussian processes such as mean-variance portfolio selection, arbitrage pricing theory (APT), security market line and CAPM – has serially independent Stable increments implying an $\alpha = 2$ i.e. Gaussian. However from equation (2) we see that when $\alpha < 2$ there will exist an infinite number of discontinuities in any finite time interval such as our time-series data. It is these discontinuities restricting the use of normal increment diffusion processes, which has discouraged the use of Stable distributions. However, as early as 1976 Robert C. Merton (an advocate of normal increment diffusion processes) pointed out that the continuity of a pure diffusion process (Gaussian process) is actually one of its drawbacks. He suggested that financial models introduce discontinuities by adding to a background diffusion process a Poisson-driven process that provides occasional discontinuities at irregular intervals.

J. Huston McCulloch (1978) showed that discontinuities ($\alpha < 2$) are not statistically unmanageable or economically unreasonable and in fact are actually appealing in their own right for a price series. His reasoning is as follows: transaction prices are not defined in continuous time since only a number of transactions take place in any time interval but buying and selling offers do exist in continuous time which implies that 'continuous time price series' must mean a bid or an asked or a bid-ask-mean price offer series, rather than an actual transaction price series. But these offers are discontinuous. When an extremal offer is taken up, withdrawn, or supplanted by a better offer, the price of the security changes discontinuously the instant the alteration occurs. McCulloch found most of these discontinuities to be very small just like almost all the discontinuities in a Stable process except when occasionally some important news arrives, causing a larger discontinuity. Mandlebrot (1963) interprets these discontinuities as spontaneous changes in price that leave supply and demand behind and cites them as justification for closing markets when they move by more than a certain amount.

As mentioned in chapter 3.3 some of the frightening properties of Stable distributions (discontinuities and infinite variances) have stifled the use of this family of distributions because of its contradiction with traditional capital market theory. Stable markets would also rule out the practicality of arbitrage because it is impossible to buy and sell simultaneously in two markets. Arbitrage lends itself to a Gaussian process because to make the operation successful a continuous sample path is required. This comes about when an arbitrageur ascertains the price in one market, makes a transaction in a second market, and then quickly returns to the first market to close out his position. A continuous sample path is required to ensure that he is able to close his position as quickly as possible so that the actual closing price in the first market is 'virtually the same' as the price he originally ascertained (McCulloch, 1978).

But the reality of arbitrage is that no arbitrageur is able to act with infinite speed to take advantage of the continuous path of a Gaussian market. Even though a Gaussian market is everywhere continuous there is no limit to how far the price can move in a few seconds. McCulloch (1978) found that because of the greater cohesion in Stable markets, the speed necessary is actually less in a Stable market than a Gaussian market.

CHAPTER 5

SUMMARY

5.1 OBJECTIVE

The objective of chapter 5 is to summarise the results obtained and to conclude with some brief remarks and recommendations.

5.2 SUMMARY OF RESULTS

In the presence of survival bias due to the use of an index and diminishing time-series, the results overwhelmingly support the alternative hypothesis and rejection of randomly changing prices on JSE being Gaussian.

The results are confirmed by symmetry and Quantile-Quantile plots which highlights the problem of which statistics to assume to characterise the changing nature of the distributions over time.

The Anderson-Darling and Runs test all reject the Gaussian description at the 1% level of significance of the JSE All-Share Index over time.

The Stable parameters calculated show that the distributions are strongly Leptokurtic in the first few weeks with a gradual tendency to a Gaussian distribution, in keeping with classical mathematics of finance theory.

It was also shown that any heavy-tailed log return distribution or Leptokurtic distribution of any kind leads to non-existing expected prices and non-existing discrete returns with far-reaching consequences for traditional capital market theory.

5.3 CONCLUDING REMARKS

While ample evidence exist to reject the Gaussian distribution for a non-Gaussian heavy-tailed distribution, the true distribution and statistics of price changes remains to be settled. The use of the economically appealing Stable distribution should be based on the following considerations:

- The ability to correctly fit the tails of a model to empirical data;
- The assumption of finite variance;
- The assumption of finite mean;
- Their evaluation has to conform to the entire theoretical formulation of economics.

5.4 RECOMMENDATIONS

More research needs to be done in the areas of finance theory, which deals with infinite expected future prices, and expected rates of return e.g., Samuelson (1976) and also Robert C. Merton's (1976) Poisson-driven process that provides occasional discontinuities at irregular intervals. Hopefully this research will reconcile trading behaviour with the speculative price distributions obtained.

LIST OF REFERENCES:

"Anderson-Darling and Shapiro-Wilk tests", (2003, March14 – last update), (Information Technology Laboratory at the National Institute of Standards and Technology), Available:
http://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm
(Accessed: 2003, March 17)

Bamberg, G. and Dorfleitner, G. (2001). "Heavy Tails and Traditional Capital Market Theory", Available:
http://www.wiso.unaugsburg.de/ibo/Downloads/Heft177.pdf
(Accessed: 2002, October 21)

Belkacem, L., Véhel, J. and Walter, C. (1996), "CAPM, Risk and Portfolio Selection in 'Stable' Markets", *INRIA*, no. 2776

Cheng, B.N. and Rachev, S.T. (1995), "Multivariate Stable future prices", *Mathematical Finance*, vol. 5, pp.133-53

Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, vol. 41, pp.135-155

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), "Modelling extremal events for insurance and Finance", *Springer Verlag, Heidelberg.*

Fama, J. (1963), "Mandelbrot and the Stable Paretian Hypothesis", *Journal of Business*, vol. 35, pp. 420-429

Focardi, S.M. (2001), "Heavy tails, scaling and Stable laws: A critical look at modeling extremal events in economic and financial phenomena", (The Intertek Group), Available:
http://www.theintertekgroup.com/scaling.pdf
(Accessed: 2002, September 15)

Goovaerts, M., De Schepper, A., Dhaene, J. and Kaas, R. “Stable laws and the distribution of cash-flows”, (2003, March13 – last update), (Katholieke Universiteit Leuven), Available: http://www.econ.kuleuven.ac.be/tew/academic/actuawet/pdfs/Stable.pdf (Accessed: 2003, March 17)

Kunst, R.M., Reschenhofer, E. and Rodler, K. (1991), “Analysis of Austrian Stocks: Testing for Stability and Randomness”. *Journal of Empirical Economics*, vol. 16, pp. 465–477

Lux, T. (1996), "The Stable Paretian Hypothesis and the Frequency of Large Returns: An Examination of Major German Stocks", *Applied Financial Economics*, vol. 6, pp. 463 - 475

Mandelbrot, B. (1963), “Variation of certain speculative prices”, *Journal of business*, vol. 35, pp. 394 – 419

Mantegna, R.N. and Stanley, H.E. (1994), “Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight”, *Physics Rev. Lett.*, 73, pp. 2946-2949

Masoliver, J., Montero, M. and Porrà, J.M. (1999), A dynamical model describing stock market price distributions. *Physica A*, vol. 283, pp. 559–567

McCulloch, J.H. (1978), “Continuous Time Processes with Stable Increments”, *Journal of Business*, vol. 51, no. 44, pp. 601–620

McCulloch, J.H. (1996), "Financial applications of Stable distributions", *Statistical Methods in Finance*, *Handbook of Statistics*. vol. 14.

Merton, R.C. (1976), “Option Pricing when Underlying Stock Returns Are Discontinuous”, *Journal of Financial Economics*, no. 3, pp. 125-144

Nolan, JP. (1997), "STABLE 3.04", computer program available to download at http://academic2.american.edu/~jpnolan/Stable/Stable.html (Accessed: 2002, September 15)

Nolan, J.P. (1999), "Fitting Data and Assessing Goodness-of-fit with Stable Distributions". *Preprint, American University*, Available: http://academic2.american.edu/~jpnolan/Stable/Stable.html (Accessed: 2002, September 15)

Nolan, J.P. (2001), "Maximum likelihood estimation of Stable parameters", *Preprint, American University*, Available: http://academic2.american.edu/~jpnolan/Stable/Stable.html (Accessed: 2002, September 15)

Poshakwale, S. (1996), "Evidence on the Weak-form efficiency and the day of the week effect in the Indian Stock Market", *Finance India*, vol. 10(3), pp. 605 – 616

Quantile-Quantile Plots. Statsoft Electronic Textbook.(StatSoft), (1984-2003 – copyright date)Available: http://www.statsoftinc.com/textbook/glosq.html (Accessed: 2002, September 15)

"Runs test", (2003, March14 – last update), (Information Technology Laboratory at the National Institute of Standards and Technology), Available: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35d.htm (Accessed: 2003, March 17)

Strong, N. (1992), "Modelling abnormal returns: A review article", *Journal of Business Finance and Accounting*, Vol. 19, pp. 533-553

Weron, R. (2001), "Levy-Stable Distributions Revisited: Tail Index > 2 Does Not Exclude The Levy-Stable Regime", *International Journal of Modern Physics* C, Vol. 12, no. 2, pp. 209-223