Helsinki University of Technology Laboratory of Computational Engineering Publications
Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja
Espoo 2002 Report B 33

ESSAYS ON ASSET RETURN DISTRIBUTIONS

Juuso Töyli

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Laboratory of Computational Engineering Publications
Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja
Espoo 2002 Report B 33

ESSAYS ON ASSET RETURN DISTRIBUTIONS

Juuso Töyli

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Electrical Engineering, for public examination and debate in Auditorium S4 at Helsinki University of Technology (Espoo, Finland) on the 20th of December, 2002, at 12 noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Computational Engineering

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Laskennallisen tekniikan laboratorio

Distribution:
Helsinki University of Technology
Laboratory of Computational Engineering
P. O. Box 9203
FIN-02015 HUT
FINLAND
Tel. +358 9 451 4826
Fax. +358 9 451 4830

Available from http://lib.hut.fi/Diss/2002/isbn9512262541/

E-mail: juuso.toyli@hut.fi

ISBN 951-22-6253-3 (printed)
ISBN 951-22-6254-1 (PDF)
ISSN 1455-0474
Picaset Oy
Helsinki , Finland 2002

Abstract

In this thesis we discuss the asset returns. Our work was initially motivated by Mantegna's and Stanley's results (1995) that put forward the stable distribution as a model of asset returns and demonstrated the scaling property that seemed to be present in the data. Nevertheless, that work raised several questions both theoretically interesting and practically challenging such as: what is the effect of measurement quantity on the inference concerning asset returns, which are the proper quantities to look at, does the scaling exists, and if so what are its limits, are there characteristic times on asset returns, how the possible time-dependent variance affects the inference?

When exploring these issues, we became concerned about the possible variability and time-dependency of the shape of the asset return distribution in addition to the time-dependent variance. Thus, we speculated that the possible variability of the shape could have been one reason behind the contradictory results concerning the best fitting model of asset returns. Furthermore, the anomalies related to the mean returns and standard deviations led us to raise the question whether the shape of the asset return distribution shows similar kinds of anomalies. Finally, since we noticed that much debate has been had about the time-independent and time-dependent models but there has been relative few studies where these various models have been compared using the same datasets, especially high frequency data, this has been done is thesis and quite surprising results were obtained.

In order to address these questions we studied Standard & Poor's 500 daily index data of the New York Stock Exchange from more than 32 years. In addition, we used a high frequency data recorded on about 20 seconds time-interval over three years time period. For comparison reasons we also studied a small market, namely the Helsinki Stock Exchange all shares return index (HEX) over seven year period. Moreover, we used an artificial data to demonstrate some effects of measurement quantities.

Our results show that the proper variable to look at is the logarithmic return. Initially, for short time horizon or holding periods, the truncated Lévy distribution was found to fit the data quite well. Since this is not a stable distribution, the scaling behaviour observed for short times should break down for longer times. Thus, we demonstrated that the characteristic time of the breakdown of scaling is of the order of few days. Furthermore, the analysis of convergence of the kurtosis showed that it takes place within few months.

When we investigated the various time-independent models of asset returns being simple normal distribution, Student t-distribution, Lévy, truncated Lévy, general stable distribution, mixed diffusion jump, power exponential distribution, and compound normal distribution, the results indicated that all models, excluding the simple normal distribution, are, at least, quite reasonable descriptions of the data. Surprisingly, however all other time-independent models except the normal distribution usually outperform the time-dependent GARCH(1,1) model for time horizons shorter than about four hours although the fine grained data evidently includes time dependencies. However, the GARCH-model is on average the best model for daily returns, and especially for periods of time when the return generating process cannot be assumed normal.

In the case of the variability of the shape of asset return distribution, our results showed that the shape of the distribution does not vary from one weekday to another. However, substantial deviations over time were observed while there are also temporal periods when it is reasonable to assume the return generating process as normal. The known time-dependencies were found inadequate in explaining these deviations. Furthermore, the results indicate that the return distribution approaches normal when the time interval used to calculate returns is increased.

Finally, our findings led us to raise three questions for future research to address. First, we speculated that there seem to be periods of "business as usual" when the return generating process is well described by the normal distribution. However, for some reason – for example, external shock, bubble formation – every now and then also periods of ferment emerge. These periods are characterised by higher volatility and increased time-dependencies. Second, the poor performance of GARCH(1,1) model on high frequencies lead us to question whether the assumption of GARCH that returns are normally distributed with time-dependent parameters is reasonable and whether it should be substituted with some other model where also the shape is allowed to vary over time. Such a model could, at least in theory, capture the business as usual periods and periods of ferment. Third, although we were surprised by the poor performance of GARCH(1,1) on high frequencies, we were reluctant to generalise this finding before a more detailed analysis. However, if this behaviour is typical for financial data, it could also be a source for further insight to the return generating process.

Acknowledgements

In writing this thesis I have become greatly indebted the combined efforts and contributions of many people and institutions: my gratitude to all of them.

First and foremost, I am greatly indebted to Academy Professor Kimmo Kaski. This thesis would not have realized without his guidance, support, and encouragement. I acknowledge Professor Antti Kanto for his valuable support and advice. My special appreciation goes to Mr. Jussi Liesiö who initially brought me and Kimmo Kaski in contact. Thanks are also due to Professor Janos Kertesz, Mr. Laszlo Gillemot, Mr. Laszlo Kullman, and Mr. Marko Sysi-Aho with whom I have had a privilege to collaborate.

I am especially grateful to Professor Jukka Perttunen and Professor Seppo Pynnönen whose comments and suggestions improved the thesis at the manuscript stage.

Professor Jorma Taina and Professor Sten-Olof Hansén deserve special thanks for their support. I am also thankful to Professor Tomi Seppälä for inspiring me to do research.

The financial support of Graduate School in Computational Methods of Information Technology, Oskar Öflunds Stiftelse, Osuuspankkiryhmän Tutkimussäätiö, Suomen Arvopaperimarkkinoiden Edistämissäätiö, and Säästöpankkien Tutkimussäätiö is gratefully acknowledged.

Last but not least, I acknowledge the steady encouragement of my parents, my sister, and my grandparents, all of whom expressed a continual faith in this project.

Helsinki, November 2002

Juuso Töyli

List of Publications

This thesis consists of the following papers:

I. Kullman, L. – Töyli, J. – Kertesz, J. – Kanto, A. – Kaski, K (1999) **Characteristic times in stock market indices**. Physica A: Statistical Mechanics and its Applications. Vol. 269:1, 98-110.

II. Kullman, L. – Töyli, J. – Kertesz, J. – Kanto, A. – Kaski, K (2000) **Break-Down of Scaling and Convergence to Gaussian**. International Journal of Theoretical and Applied Finance. Vol. 3:3, 371-373.

III. Gillemot, L. – Töyli, J. – Kertesz, J. – Kaski, K. (2000) **Time independent models of asset returns revisited**. Physica A: Statistical Mechanics and its Applications. Vol. 282:1-2, 304-324.

IV. Töyli, J. – Kaski, K. – Kanto, A. (2002) **Variability of Hex Return Distribution**. The Finnish Journal of Business Economics. Vol. 51:1, 64-89.

V. Töyli, J. – Kaski, K. – Kanto, A. (2002) **On the shape of asset return distribution**. Communications in Statistics – Simulation and Computation. Vol. 31:4, 489-521.

VI. Töyli, J. – Sysi-Aho, M. – Kaski, K. (2002) **Models of Asset Returns: Changes of Pattern from Tick by Tick to 30 Days Holding Period**. Submitted for publication in Quantitative Finance.

Author's Contributions

The initial idea of the research presented in the included papers owes in many ways to Academy Professor Kimmo Kaski who built the multi-cultural research team where the author was one of the initial members. In case of Paper 1, the author has a major role in checking the background and he took intensively part into preliminary analysis. All the reported Lévy fittings were made by Laszlo Kullman (The C code used was the same as Mantegna and Stanley used). The author has also an important role in manuscript writing. In the short Paper 2, the author's role was more of a supportive co. author and the manuscript was prepared by Laszlo Kullman although reviewed and commented by the author.

The idea of the Paper 3 originates from the author such that visiting student Laszlo Gillemot did most of the analysis and coding under author's supervision. The author had the leading role in the manuscript writing and result interpretation.

In the Papers 4 and 5, the author did all the coding, the analyses, and most of the manuscript writing. The original idea of Paper 4 owes to Professor Antti Kanto who also mentored the author during the initial steps (this paper was prepared parallel to the Paper 1 although published much later). In general, the author prepared the two manuscripts and the professors commented them.

The nature of the paper 6 was to summarize, clarify and extended the analyses and results of our multi-year project in exploring asset returns. Here M.Sc. Marko Sysi-Aho made few preliminary analyses, prepared the high frequency data, and did some coding. Nevertheless, the author made all the reported analyses and wrote the manuscript.

Contents

List of Tables

List of Figures

1. Introduction

This Chapter will review the background for the research and summarize the purposes of the Papers included in this thesis. Abbreviations and conceptual definitions used are also shortly listed.

1.1 Background of research

The shape of the asset[1] return distribution has been intensively studied since Louis Bachelier's (1900) doctoral dissertation "The Theory of Speculation". Bachelier sketched a model where the changes in the logarithms of bond prices were normally distributed by invoking the central limit theorem. He further assumed that the price changes are independent and identically distributed (*iid*). Although Bachelier clearly anticipated the rediscovery of Brownian motion by Einstein, it was Osborne (1959) who formally specified the asset return generating process as Brownian motion. The classical theory of finance is mainly based on this assumed normality. Assumed because the early scholars did not formally test their assumptions – they merely assumed that if the market were as assumed then the suggested model would follow. However, Mandelbrot (1963) and Fama (1965a) first reported fundamental differences from the normality: empirical return distributions are fat-tailed and peaked when compared to normal distribution (i.e., they are *leptokurtic)*.

Thereafter it has been commonly accepted that daily empirical return distributions are fat-tailed and peaked. However, the asset return series also seem to posses some additional empirical features that are counter to *iid* and normality hypotheses. For example, a few studies suggest that the return distributions might be skewed (Kon 1984; Fielitz – Rozell 1983). Asset returns are also not usually found to be autocorrelated like the squared returns, variances, and volumes are (Schwert 1989). Furthermore, the standard deviation of Monday's returns is found to be higher than that of other days (Fama 1965a). A statistically significant difference in the mean return of Mondays compared to that of other days has been reported (French 1980; Gibbons – Hess 1981) and non-stationary mean-excess returns have been discovered during the first trading

[1] The concept of asset is a generic name for a financial instrument that can be bought or sold, like stocks, currencies, gold, bonds, etc.

week of January (Keim 1983). In addition, the turn of the month effect – i.e., the mean return for stocks is positive only for days immediately prior to and during the first half of a calendar month and indistinguishable from zero for days during the last half of the month – has been found out (Ariel 1987). When the returns of individual stocks are studied – bimodality, i.e., the distribution has two peaks – is occasionally observed (Kanto et. al. 1998). These findings suggest that the return generating process is likely to be *complex*, *non-linear* and that returns might be *dependent*.

Although the empirical characteristics of asset returns are well-know, the research has not resulted in a conclusive view about the best model for asset returns. The models of asset returns can be divided to time-independent and time-dependent categories. The well-known time-independent models, which can capture some of the non-normalities present in empirical financial data series, include stable distribution (Mandelbrot 1963; Fama 1965a), Student t-distribution (Blattberg – Gonedes 1974; Hagerman 1978), Mixed diffusion-jump (Merton 1976), and Compound normal distribution[2] (Kon 1984). The question which one of these models is best fitting remains open but in general the research shows that, in case of daily returns, normal distribution comes out as the worst fitting model and that the Student t, the mixed diffusion jump, and the compound normal distributions outperform the stable distribution (Blattberg – Gonedes 1974; Hagerman 1978; Officer 1972: Hsu et. al. 1974). In contrast, the ranking order between the Student t-distribution, the mixed diffusion jump, and the compound normal distribution seems not to be uniquely clear (Kon 1984, Akgiray et. al. 1987, Tucker 1992, Tucker et. al 1988). On the other hand, monthly returns are generally regarded normally distributed (Ariel 1987).

The question whether the variance converges to a finite limiting value is related to the time-dependency. The stable distribution results in a model that assumes infinite variance. However, recent research has found out that the variance of the asset return distribution is finite but time-dependent in a complex non-linear manner (Perry 1983; see also Tucker 1992). This time-dependency could explain the volatility clustering[3] observed in financial time series and the leptokurtic unconditional return distributions. The models that assume the time-dependency include, for example, (generalised) autoregressive heteroscedastic ([G]ARCH) models (Engle 1982; Bollerslev 1986; 1987), stochastic volatility (SV) models (see Taylor 1985; Taylor 1994), and models based on chaos theory or on fractals leading to complex dynamics (Hsied 1991, Mandelbrot 1999). In general, these models lead to martingale differ-

[2] Also called mixture normal

[3] Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes (Mandelbrot 1963).

ences, which implies that the realisations of a stochastic process are uncorrelated but not necessarily independent. These dependencies are then modelled in the conditional variance equation but the point forecasts of returns cannot be improved. This possible time-dependency does not necessarily imply market inefficiency since it is enough for the efficiency to hold that the return-generating process can be represented as a martingale (Fama 1970).

Despite of the evidence against the stable distribution, Mantegna and Stanley (1995) observed scaling property and put forward again the stable distribution as a model of asset returns but they also pointed out an approximately exponential decay in the tails of the distribution. However, this result may partially be due to the use of differences instead of logarithmic returns and perhaps more specially it may be due to the use of high frequency transaction level data in the analysis. Nevertheless, Mantegna and Stanley's work raises at least the following important questions. What is the effect of measurement quantity on the inference concerning asset returns? Which are the proper quantities to look at? If the scaling exists, what are its limits? How the possible time-dependent variance affects the inference? Are there characteristic times on asset returns? In addition to that, the linear and non-linear dependencies, especially that of the second moment, have been widely elaborated in the literature but it seems that limited attention has been paid to the possible variability and time-dependence of the shape of the asset return distribution. The possible variability of shape could have been one reason behind the contradictory results concerning the best fitting model of asset returns. It is also both theoretically interesting and practically challenging to know whether the shape of asset return distribution has similar kind of anomalies than mean returns and standard deviations. Much debate has also been made around the time-independent and time-dependent models but there are relative few studies where these models are compared together with the same datasets. This thesis will explore these issues.

1.2 Purpose of research

The purpose of the first two papers – "*Characteristic times in stock market indices*" and "*Break-Down of Scaling and Convergence to Gaussian*" - was to find out where are the limits of scaling[4] in the stock market data and to examine the effect of measurement quantity (logarithmic return versus simple difference) on the inference. In these two papers, we applied the biased view that

[4] Scaling means that the shape of return distribution remains the same regardless of time scale. Thus, for example, if one minute returns are Lévy distributed with a given characteristics exponent, then also daily and monthly returns should be Lévy distributed with the same characteristics exponent.

the (truncated) Lévy distribution is an adequate description of stock return distribution for short times and studied, based on this assumption, the crossover to the Gaussian behaviour. Along with that, we examined the effects of non-linear dependencies by filtering them out with ARCH, GARCH and IGARCH methods.

In the third paper – "*Time Independent Models of Asset Returns Revisited*" – we re-examined the well-known time-independent models of asset returns (simple normal distribution, Student *t*-distribution, Lévy, truncated Lévy, general stable distribution, mixed diffusion jump, and compound normal distribution) and studied further the effects of the use of the simple differences and logarithmic returns on the inference. In addition, we showed that the likelihood ratio test can be used to discriminate between mixed diffusion jump and compound normal model.

The purpose of the paper "*Variability of Hex Return Distribution*" was to explore the possible variability in the shape of the unconditional distribution. This problem was studied with the help of the following questions: does the shape of distribution vary from on weekday to another, is the shape of distribution constant over time, and does the shape of the distribution vary along the time horizon used to calculate the returns? In the next paper "*On the shape of asset return distribution*", we continued on the same path. Here we extended the data to cover a major market (S&P 500 daily index) and took into account the known linear and non-linear time dependencies. Thus, a forth question "how the known linear and non-linear time dependencies affect the inference concerning the shape of the distribution" was also studied. With these questions, the nature of these two papers was descriptive and they seek *not* to give any detailed quantitative or modelling based *explanation* of the possible time variation of the shape of the asset return distribution.

The final paper of this bundled thesis "*Models of Asset Returns: Changes of Pattern from Tick by Tick to 30 Days Holding Period*" summarized, clarified and extended the analyses and results of our multi-year project to explore asset returns. This paper focused to study the effects of (i) different time periods, (ii) different holding periods and (iii) non-linear dependencies on the conclusions concerning the best fitting time-independent model of asset return distribution (iv) and to compare the time-independent models to a simple time-dependent model. A special interest was paid to the evolution of the properties of asset returns when very short holding periods (high frequency data) are being analysed and to the question whether the models of asset return distribution found good on daily data also provide good description of data on shorter time-intervals.

1.3 Abbreviations and conceptual definitions

Anomaly is used in the financial literature to describe peculiarities present in financial data series.

Asset is a generic name for a financial instrument that can be bought or sold, like stocks, currencies, gold, bonds, etc.

Brownian motion. See *Wiener process*.

Martingale is a stochastic process where the realisations are uncorrelated but not necessarily independent.

Lévy distribution is used to refer a special case of general stable distributions where the location and skewness parameters are assumed zero.

Stylized empirical facts. Econophysicists have introduced the concept *stylized empirical facts* to refer to the systematic empirical facts that support neither the *iid* nor the normality hypotheses.

Volatility clustering means that large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.

A *Wiener process* (also referred to as *Brownian motion*) is a particular type of Markov stochastic process. It has been used in physics to describe the motion of particle that is subject to a large number of small molecular shocks.

Table 1: Abbreviations

Abbreviation	**Description**
ARCH	Autoregressive heteroscedastic
CND	Compound normal distribution
FFT	Fast Fourier Transform
GARCH	Generalised autoregressive heteroscedastic
GED	Generalised error distribution
GSD	General stable distribution
iid	Independent and identically distributed
KDE	Kernel density estimate
MDJ	Mixed diffusion-jump
ML	Maximum likelihood
pdf	Probability density function
PED	Power exponential distribution
STU	Student t-distribution
SV	Stochastic volatility
TLF	Truncated Lévy distribution

2. Characteristics of Asset Returns

This Chapter discusses the terms of measurement (logarithmic return, simple price change, and percentage return) used when analysing asset returns. Their differences are highlighted and the conclusion that logarithmic return is theoretically the most justifiable measure to use is presented. In addition, the empirical facts that support neither the iid nor the normality hypotheses (stylized empirical facts) and anomalies present in financial data series are reviewed.

2.1 Terms of measurement

The asset returns can be measured in terms of logarithmic return, simple price change, and percentage return. In order to define these precisely, let P_t be the asset price at time t. The simple price difference is then denoted by D_t and reads as follows:

$$D_t = P_t - P_{t-1} \tag{2.1}$$

The variability of a simple price difference is an increasing function of the price level and this might bias the inference at least when the price level increases significantly during the analysis period. Fortunately, the use of logarithmic return (r_t) neutralises most of this effect (Fama 1965a, 45). Logarithmic return reads as follows:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \tag{2.2}$$

In addition, the change in logarithmic price is the yield, with continuous compounding, from holding a security for the period in questions. Proof is as follows (Fama 1965a, 45):

$$\begin{aligned} \frac{P_{t+1}}{P_t} &= \exp\left(\ln\left(\frac{P_{t+1}}{P_t}\right)\right) \\ P_{t+1} &= P_t \exp\left(\ln(P_{t+1}) - \ln(P_t)\right) \end{aligned} \tag{2.3}$$

When applying logarithmic return, continuous time generalisations of discrete time results are easier and returns over more than one day are simple functions of single day returns (Taylor 1986, 13). The third way to define asset return is the percentage return (R_t) that is numerically very close to logarithmic return for small changes.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.4}$$

In addition to these points, the return aggregation is of importance in most financial applications. Table 2 summarises the difference between percentage returns and logarithmic returns. w_i denotes the weight of asset i, t denotes the time, and p portfolio. The data in Table 2 indicates that when the aggregation is done across time it is more convenient to work with logarithmic returns and, in case of aggregation across assets, percentage return results in a simpler expression.

Table 2 : Return aggregation

Aggregation	**Temporal**	**Cross-section**
Percent return	$R_{it} = \prod_t (1+R_{it}) - 1$	$R_{pt} = \sum_i w_i R_{it}$
Logarithmic return	$r_{it} = \sum_t r_{it}$	$r_{pt} = \ln\left(\sum_i w_i e^{r_{it}}\right)$

Source: RiskMetrics[tm] – Technical Document 1996, 49

However, since percentage return and logarithmic return are very close for small changes, it is common to approximate a portfolio return in case of logarithmic return as:

$$r_{pt} \cong \sum_{i=1}^{N} w_i r_{it} \tag{2.5}$$

This leads to a situation where the one-day model computed with r_t extends easily to returns of greater than one day. In general, the use of logarithmic return is commonly accepted among the financial researchers. Despite of that fact, in some early econophysics studies differences were favoured (see Mantegna – Stanley 1995). This is likely to lead to different interpretations. Figure 8 illustrates the behaviour of the three measures when applied to S&P 500 daily index. Especially, the effect of price level on price variability, in case of the simple price differences, is clearly visible form it. Hereafter the concept return is used to refer to logarithmic return for concise presentation if not otherwise indicated.

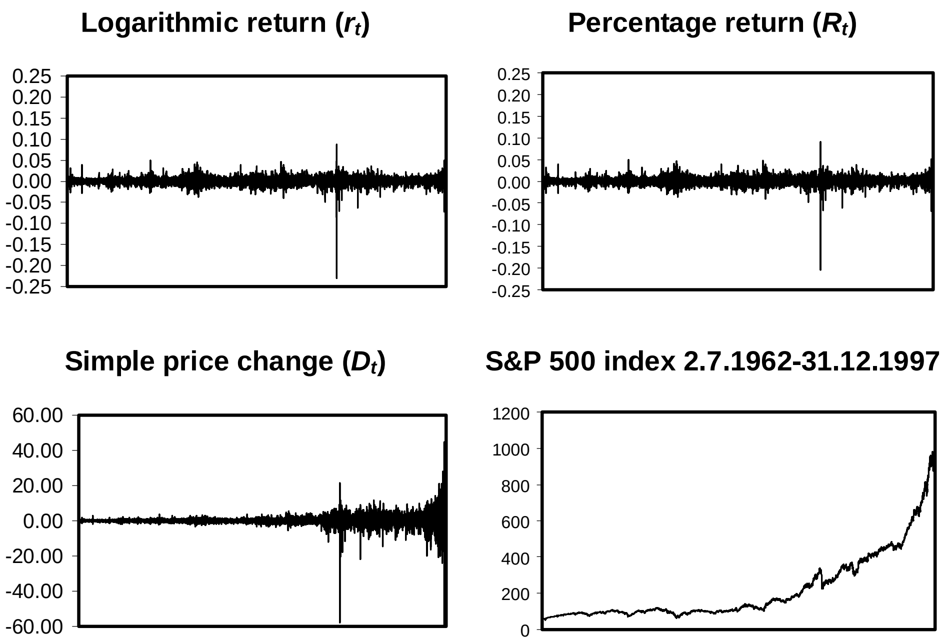


Figure 1: Terms of measurement

2.2 Anomalies and stylized empirical facts

The classical theory of finance has its roots on Bachelier's work (1900) and relies on the assumption that asset returns are identically and independently distributed (IID). The normality assumption is usually then added that leads to the Brownian motion behaviour (see Osborne 1959). However, the empirical research has found several systematic empirical facts that support neither the *iid* nor the normality hypotheses. Econophysicists have introduced the concept *stylized empirical facts* (see, e.g., Cont 2001) to refer to these. In turn, *"anomaly"* is the concept usually used in the financial literature to describe peculiarities present in financial data series. In this thesis, these concepts are used interchangeably.

According to Cont (2001, 224) asset returns typically demonstrate the following properties. (1) Return series are usually not autocorrelated. (2) Return distributions have heavy tails when compared to those of normal distribution. (3) Gain/loss asymmetry that, however, is not observed in exchange rates. (4) Aggregational Gaussianity – i.e., the shape of return distribution is not the same at different time scales and, when time interval increases, the shape becomes more and more like normal distribution. (5) Intermittency – i.e., returns display, at any time scale, high degree of variability. (6) Return series show volatility clustering and (7) conditional heavy tails – i.e., after correcting the returns for volatility clustering, the return series still has heavy tails. (8) Slow decay of autocorrelation in absolute returns. (9) Leverage effect – i.e., most measures of volatility are negatively correlated with returns of that asset. (10) Volume/volatility correlation. (11) Asymmetry in time scales (coarse-grained measures [long time scale] of volatility predict fine-scale [short time scale] volatility better than the other way round).

The research has also found various *anomalies* related to weekdays and calendar months. These include the findings that the standard deviation of Monday's returns is higher than that of other days (Fama 1965a). A statistically significant difference in the mean return of Mondays compared to that of other days (French 1980; Gibbons – Hess, 1981). Non-stationary mean-excess returns in the first trading week of January (Keim 1983). Along with that, Ariel reported the turn of the month effect, i.e., the mean return for stocks is positive only for days immediately prior to and during the first half of calendar months and indistinguishable from zero for days during the last half of the month (Ariel 1987). In contrast to large international markets, somewhat more contradictory results concerning small markets have been reported. For example, the anomalies related to weekdays on Finnish market are not generally regarded strong (Booth et. al. 1992). In addition, when the returns of individual stocks are being analysed, bimodality - i.e., the distribution has two peaks - is occasionally observed (Kanto et. al. 1998). In contrast to individual stocks, bimodality has not been reported when stock portfolios or indices have been studied.

The above-discussed findings concerning the characteristics of asset returns seem to suggest that the return generating process might be *non-linear* and that returns might be mutually *dependent*. This possible dependency does not necessarily imply market inefficiency since it is enough for efficiency that the return-generating process can be represented as a martingale[5] (Fama 1970). Although the classical assumption that asset returns are identically independently distributed (iid) does not necessitate any statements about the shape of the return distribution it does necessitate the independence. The possible time-dependency can be divided in linear and non-linear categories. The linear dependency is usually measured by autocorrelation. Autocorrelation is defined as follows:

[5] Martingale means that the realisations of a stochastic process are uncorrelated but not necessarily independent.

$$C(\tau) = corr(r(t, \Delta t), r(t + \tau, \Delta t)) \tag{2.6}$$

where *corr* denotes the sample correlation. A typical finding is that the asset returns for daily and weekly holding periods are modestly autocorrelated for one day or one week lag (typically less than 0.2) and slightly stronger for longer time periods (see Fama et. al. 1988) but, for longer than one lags, the autocorrelation tends to vanish completely. There is also convincing evidence that suggests that it is safe to assume autocorrelation zero for all practical purposes for longer than 15 minutes time lags (Cont et. al. 1997, see Cont 2001). In contrast to the negligible return's autocorrelation, squared returns, variances, and volumes are autocorrelated (Schwert 1989). This finding suggests that returns might have non-linear dependencies. Linear and non-linear dependencies could result in increased probability mass in the tails of the observed return distribution, which in turn is likely to lead to biased inference about the shape of the distribution and its variability over time.

An undisputable exception from the classical asset returns' normality assumption is that the empirical return distributions indicate substantial excess kurtosis. A large positive value for kurtosis indicates that the tails of the distribution are longer than those of a normal distribution are; a negative value for kurtosis indicates shorter tails (becoming like those of a box-shaped uniform distribution). It was Mandelbrot and Fama who first reported this fundamental deviation from the normality (Mandelbrot 1963, Fama 1965a). The kurtosis is defined as follows:

$$k' = \frac{E(x_t - \mu)^4}{\sigma^4} - 3 = k - 3 \tag{2.7}$$

where E is the expectation operator, μ is the mean of random return variable x_t, and σ is the standard deviation. For normal distribution, the value of k is three. It is typical practise to deduct three from the k (as in Eq. 2.7) so that the value of kurtosis coefficient k' for the normal distribution is zero. However, it needs to be pointed out that the existence of above discussed dependencies in the data is likely to increase kurtosis. Furthermore, it is known that the kurtosis and skewness behave very differently as the time interval is increased when the different terms of measurement are used. According to Lau and Wingender (1989), ignoring this phenomenon might have been one reason for the contradictory results concerning skewness and kurtosis.

The following issue is the possible skewness, which is more questionable property than kurtosis. However, since the possible skewness is fundamental to portfolio investment decisions, to most mainstream financial models, and to many statistical testing procedures relating to asset returns, it needs to be discussed. The skewness is defined as follows:

$$\gamma = \frac{E(x_t - \mu)^3}{\sigma^3} \tag{2.8}$$

where E is the expectation operator, μ is the mean of random return variable x_t, and σ is the standard deviation. For normal distribution $\gamma=0$, and for non-symmetric distributions γ is non-zero. Skewness is positive when right hand tail is heavier and negative when left hand tail is heavier.

There are relatively few studies considering the skewness in stock market data and the results are contradictory. Negative, positive, and non-existent skewness has been reported. For example, a few studies indicate that the return distributions are skewed (Kon 1984, Fielitz – Rozell 1983). Schwert found out that the skewness is varying from positive to negative from one sample period to another when he analysed U.S. stock prices from 1802 to 1987 (Schwert 1987). Alles and Kling (1994) reached a conclusion when studying a wide class of assets (NYSE, AMEX and NASDAQ stock indices, US government long-term and medium-term bond indices, and mortgage, corporate and government agency bond indices) that smaller capitalised stock indices are more negatively skewed than larger stock indices and that the skewness of stock indices follows a business-cycle-related variation over time so that skewness tend to be more negative during economic upturns and less negative, even positive, during downturns.

Alles and Kling's findings are in line with Schwert's results but partly contradictory to Badrinath and Chatterjee's (1991). This contradictory might be due to the measurement errors caused by market frictions such as non-synchronous trading and bid-ask bounce (Alles – Kling 1994). It has also been suggested that the slight skewness observed in empirical returns might be due to the dependencies in the data that are not properly taken into account when a significant skewness is observed (see Alles – Kling 1994; Lau – Wingender 1989). Thus, the current research has not resulted in a conclusive view and there is an ongoing debate around this issue but, in general, it has not been regarded as a problem for practical purposes.

There are at least four different alternative approaches to test the significance of skewness. First alternative is to assume that returns are iid normally distributed and then apply the standard test for skewness (see Alles-Kling 1994). Second alternative is to adjust the significance value to take into account the observed autocorrelation (see Alles-Kling 1994). Third way is to filter the auto-correlation out from the data and then apply the standard tests for skewness. Fourth alternative is to test wide range of distributions and consider the kurtosis and skewness together (see Badrinath – Chatterjee 1988, 1991). In contrast to linear-dependency, the possible effect of non-linear dependencies on the skewness estimates has not been extensively assessed by the research.

An additional crucial assumption is whether the second moment is finite or infinite. This issue has been intensively discussed since Mandelbrot's (1963) and Fama's (1965a) seminal works that suggested the stable distribution, which does not necessarily have a finite second moment, as a model of security returns. The later research has provided substantial evidence that does *not* support the stable hypothesis. (Blattberg – Gonedes 1974, Hagerman 1978, Officer 1972, Hsu et. al. 1974, Perry 1983, Tucker 1992). In contrast, Mantegna and Stanley recently reached the opposite result again, but this result may partially be due to the use of differences instead of logarithmic returns and perhaps more specially it may be due to the use of high frequency transaction level data in the analysis (Mantegna – Stanley 1995, see Kullman et. al. 1999). In general, it is usually accepted that the variance is finite although time-dependent in a complex manner when daily data is being analysed (Perry 1983, Tucker 1992, see also Fama – French 1988).

One strong evidence that supports finite second moment is that the he shape of return distribution of the same asset calculated over various different intervals (for example, daily, weekly, monthly, or annual holding periods) seems to become more and more like normal distribution when time interval increases. In general, the research generally indicates that monthly returns are well described by stationary normal distribution (Hagerman 1978, Akgiray – Booth 1987, Akgiray 1989). This finding is in line with the central limit theorem (CLT). In addition, it seems that logarithmic returns do not have long term properties but the squared returns and absolute returns do have. However, these findings may be due to the aggregation and non-stationary (Lobato – Savin 1998).

3. Asset Return Generating Processes and Distributions

This Chapter reviews the different aims and approaches to the modelling of asset returns. After that, various statistical analysis based models of asset returns are discussed. These include kernel density estimate (KDE), normal distribution, general stable distribution (GSD), truncated Lévy flight distribution (TLF), Student t-distribution (STU), power exponential distribution (PED), mixed diffusion jump (MDJ), compound normal distribution (CND), RiskMetric^{tm} model, (G)ARCH type models, and stochastic volatility (SV) models. The previous research concerning these models is reviewed and the statistical comparison of the models is discussed.

It needs to be pointed out that the asset *pricing* models such as capital asset pricing model (Sharpe 1964, Lintner 1965) (CAPM) and arbitrage pricing model (Ross 1976) (APM) are beyond the scope of this thesis. The models considered here focus on the modelling the return generating process and/or the resulting distribution of returns not the theoretically correct asset price or market equilibration process. Nevertheless, the understanding of the return generating process and the resulting distribution of asset returns inevitably constitute the basis for the asset pricing models.

3.1 Modelling approaches

There are three distinct approaches to model the unconditional distribution of asset returns. The first approach is to model the assumed return generating stochastic process resulting in a distribution of returns. The second approach is to seek a distribution that empirically fits with the observed data and the third approach, which is probably the most challenging, is to model the behaviour of individual agents who act on the market. It needs to be pointed out that the aims of modelling are also different. Some studies seek to *understand* the return generating process while others are merely *descriptive*. Different kinds of modelling approaches are needed for such broad aims.

The modelling of the stochastic return generating process seems to be appealing since it is based on the understanding of the return generating process. However, the processes often become so complex that the resulting distribution cannot be analytically derived (see Epps – Epps 1976; Oldfield et.

al. 1977; Tauchen – Pitts 1983) although some of the models in this category lead to rather simple formulations with reasonably good explanatory power. For example, the models based on Brownian motion lead to normally distributed returns (Osborne 1959) while the mixed diffusion process (Merton 1976) results in easily tractable probability density function.

The second approach, i.e., to seek a distribution that empirically fits with the observed data, was first applied by Fama and Mandelbrot (Mandelbrot 1963; Fama 1965a). Their works were facilitated by the fact that the empirical distributions did not seem to be normal. However, it should be pointed out that the stable distribution also has an intuitively appealing underlying process but this process was not the starting point. The studies in this category are often more descriptive than focused on the understanding of the return generating process and some of the later suggested mixture models lack completely an underlying stochastic process (see, for example, Bookstaber– McDonald 1987; Rachev – SenGupta 1993). In general, most of the models in this category are mixture distributions that include the models based on stochastic process as special cases. Therefore, these models are necessarily at least as well fitting as their stochastic process based special cases. However, the cost is an increased complexity and lack of theoretical justification. Some of the well-known generalised time-independent mixtures are discussed by Rachev and SenGupta (1993) as well as Bookstaber and McDonald (1987).

In these two approaches, the modelling problem can be further divided in two questions. The first question is whether the parameters of the distribution (and/or generating process) are stationary and the second question asks what the functional shape of the distribution is. The past research concentrated initially on the shape of the distribution and only recently on the time-dependency of the parameters, especially that of variance.

The third modelling approach is probably the most challenging one since it seeks to model individual agents. Such an approach could, at least theoretically, produce ultimate understanding about the generating process. A few preliminary studies indicate that artificial markets are able to produce data that behaves like the real market data. Hence, crashes, jumps, time dependent variance, and other complex time-dependencies can be produced. However, the return generating process tends to become too complex to be represented analytically and this approach is hardly suitable for modelling empirical distributions. This is not a problem because these markets seek in the first place understanding how the actions of individual agents and their interactions affect the market as a whole. (See, for example, LeBaron 2000; Arthur et. al. 1997; LeBaron et. al. 1998; Lux – Marchesi 1999; for generic discussion about agent based models see Bonabeau 2002)

3.2 Time independent models

3.2.1 Kernel density estimate

Asset returns are typically modelled with parametric models. However, these models always necessitate tight assumptions about the return generating process. For visualisation purposes and, at least, for preliminary analysis non-parametric approaches might be more suitable. Nevertheless, such a model is necessarily a descriptive in its nature since it lacks the stochastic process based explanation why returns should be distributed according to it. A particularly simple, although flexible, non-parametric model is the kernel density estimate (KDE). The kernel estimator with kernel K is defined as follows:

$$\hat{f}(x) = \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{x - r_t}{h}\right) \tag{3.1}$$

where r_t is the asset return at time t, the parameter n is the number of data points, and h is the smoothing parameter. This kernel estimator can be considered to be a sum of bumps placed at the observations where the kernel function K defines the shape of the bumps and the smoothing parameter h defines their width. If the kernel function K is a probability density function (pdf), it follows from the definition that $\hat{f}$ is also a pdf (Silverman 1986).

In general, the inference is not sensitive to the selection of kernel function, and under Gaussian kernel the minimum approximate mean integrated square error is achieved when the smoothing parameter is (Silverman 1986):

$$h_{opt} = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \tag{3.2}$$

This choice works well when the underlying population is normally distributed but it tends to over-smooth if the population is multimodal. A more robust estimator can be achieved with an adaptive estimate of the spread of distribution instead of σ in Eq. (3.2). The adaptive estimate is defined as:

$$\text{A} = \min\,(\text{standard deviation, interquartile range}/1.34) \tag{3.3}$$

In the case of normal distribution, the two arguments are theoretically equal. Silverman (1986, 48) has argued that the results could be improved further if the factor (4/3)^(1/5)=1.0592 in Eq. (3.2) is reduced slightly, for example, to 0.9.

The estimation of kernel density is straightforward and evident from the definition. However, there is also a Fast Fourier Transform (FFT) based algorithm that speeds up the computation remarkably (for the algorithm see Silverman 1986, 63-65). The next issue to consider is how to simulate data from the kernel density estimate $\hat{f}$. Let us assume that $\hat{f}$ is constructed by the Kernel method with non-negative kernel function K and window width h. Under these assumptions univariate realisation from $\hat{f}$ can be generated as follows (Silverman 1986, 141-143):

1. Choose I uniformly with replacement from $\{1,\ldots,n\}$
2. Generate ε to have probability function K (3.4)
3. Set $Y=X_I+h\varepsilon$

If the realisations of Y are required to have first and second moment properties as those observed in the sample $\{X_1,\ldots,X_n\}$, then the step three in Eq.(3.4) should be replaced with

$$Y = \overline{X} + \frac{(X - \overline{X} + h\varepsilon)}{\sqrt{1 + h^2 \dfrac{\sigma_K^2}{\sigma_X^2}}} \tag{3.5}$$

where $\overline{X}$ and σ_X^2 are the sample mean and variance of $\{X_i\}$ and σ_K^2 is the variance of the Kernel K. (Silverman 1986, 143-144)

3.2.2 Normal distribution

The classical theory of finance is mainly based on the assumed normality having its roots in Bachelier's work at the beginning of this century (see Bachelier 1900). Thus, it is common to assume that returns behave as random walkers, which implies that the returns are identically and independently distributed (*iid*) having a zero expectation value and constant variance. The normality assumption is usually then added that leads to the Brownian motion (see Osborne 1959). In general, the assumed normality is probably the most essential assumption of classical theory of finance. For example, the capital asset pricing model and the mean-variance portfolio theory rely strongly on the (multivariate) normality of returns.

One elegant way to characterise this is to assume that returns are generated by a Brownian motion. Let S be the price whose absolute change is then dS and the relative change is then dS/S that is assumed to follow arithmetic Brownian motion as follows:

$$\frac{dS}{S} = \mu dt + \sigma dZ \tag{3.6}$$

where dZ is the standard Wiener process[6] (see Osborne 1959 for discussion). The above equation can also be written as:

$$dS = \mu S dt + \sigma S dZ \tag{3.7}$$

where μS stands for instantaneous expected drift rate and σS for instantaneous standard deviation at time t. This can further be expressed in discrete form:

$$\Delta S = \mu S \Delta t + \sigma S \Delta Z \tag{3.8}$$

where ΔZ is $\varepsilon\sqrt{\Delta t}$ and $\varepsilon \sim N(0,1)$. By applying Ito's lemma (see Hull 1997, 255-256 for derivation):

$$\Delta G_t \sim N((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2 \Delta t) \tag{3.9}$$

where ΔG is defined as ΔlnS_t. Thus, the logarithmic return follows a normal distribution.

The sampling from normal distribution is straightforward. There are numerous simulation algorithms but the Box-Muller method is probably the most well-known (see Box – Muller 1958). The algorithm reads as follows:

1. Generate U_1, U_2 ~ Uniform(0, 1)
2. Compute $$\begin{aligned} X_1 &= \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ X_2 &= \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{aligned} \tag{3.10}$$

Here X_1 are X_2 independent pseudorandom standard normal observations.

3.2.3 General stable distribution

It was Mandelbrot (1963) and Fama (1965a) who first reported fundamental differences of asset returns from the normality: empirical asset return distributions turned out to be leptokurtic and have longer tails than normal distribution. These deviations have thereafter been observed by many others. Facilitated by the fact that empirical return distributions seemed to show more kurtosis than predicted by the normal distribution Mandelbrot (1963) introduced

---

[6] A *Wiener process* is a particular type of Markov stochastic process. It has been used in physics to describe the motion of particle that is subject to a large number of small molecular shocks. It is also sometimes referred to as *Brownian motion.*

the General stable distribution[7] (GSD) as a model of asset returns. His findings were strongly supported by Fama's (1965a) seminal work. The probability density function of general stable distribution cannot be written in closed form but the characteristics function reads as follows:

$$F_{gsd}(q) = \exp\left\{-\Delta t \gamma\left(i\delta + |q|^{\alpha}\left(1 + i\beta \tan\left(\frac{\alpha\pi}{2}\right)\frac{z}{|z|}\right)\right)\right\} \tag{3.11}$$

where Δt is the time difference between two successive asset prices values, β is the skewness parameter, δ the location parameter, $\gamma > 0$ is the scaling parameter, and $0 < \alpha \leq 2$ is the characteristics exponent. The probability density function can be derived by the inverse Fourier transformation:

$$f_{gsd}(x) = F^{-1}\{F_{gsd}(q)\}(x) \tag{3.12}$$

General stable distribution has the following properties. (1) It is the Cauchy distribution if $\alpha=1$ and $\beta=0$. (2) It is the normal distribution if $\alpha=2$ and $\beta=0$. (3) It is Lévy distribution when $\beta=\delta=0$. (4) All moments of order $r<\alpha$ are finite. (5) Except when $\alpha=2$, in which case moments of all orders are finite. (6) If a sum of independent identically distributed random variables has a limiting distribution, it must be a stable distribution. Thus, the non-normal stable distributions generalise the central limit theorem to cases where the second moments of the summed random variables are infinite. (7) A sum of independent stable random variables will be stable with characteristic exponent α' if each summed random variable is distributed with α'.

In order to derive the probability density functions of the Lévy, Truncated Lévy Flight (see Section 3.2.4) and General Stable distributions from their characteristic function the following approximate numerical algorithm can be applied (Gillemot et. al. 2000):

1. Make choice of points (q_j) to determine values of the characteristic function $F(q)$, i.e., $F_j = F(q_j)$.
2. With the Inverse Fast Fourier Transformation (IFFT) compute values of the corresponding probability density function, $f(x_j)$.
3. Use spline interpolation to estimate the values of the probability density function for arbitrary intermediate points.

Although the use of the spline technique leads to minor numerical deviation from the exact results, it can be approached arbitrarily close by choosing q_j's more densely. Along with this, it might be of interest to generate α-stable ran-

[7] In this research the concept Lévy distribution is used to refer a special case of General Stable Distributions where the location and skewness parameters are assumed zero.

dom variables. One algorithm to achieve this is as follows (Janicki – Weron 1994, 109-126):

3. Generate V ~ Uniform(-π/2, π/2)
4. Generate W ~ Exp(1) (3.13)
5. Compute $X = \frac{\sin(\alpha V)}{\{\cos(V)\}^{\alpha^{-1}}} \left\{ \frac{\cos(V - \alpha V)}{W} \right\}^{\alpha^{-1}(1-\alpha)}$

The above algorithm contains the familiar Box-Müller method for the generation of normally distributed random variables as special case (α=2).

Figure 2 illustrates the shape of the Lévy distribution (β=δ=0) with α=0.7, 1.0 (Cauchy), 1.3, 1.7, 2.0 (normal). The smaller α, the sharper the "body" of the distribution, and fatter the tails, as shown in the right panel.

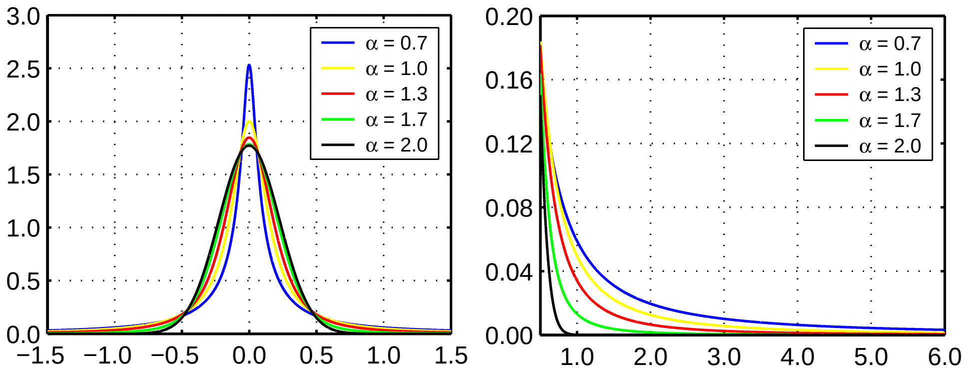


Figure 2: Shape of Lévy distribution

It needs also to be pointed out that according to Blattberg and Gonedes (1974) the symmetric stable distribution (β=0) is obtained from a normal distribution whose variance is drawn from a strictly positive stable distribution. Thus, this model actually implies that the return generating process is a stochastic process where the variance is a *iid* random variable (Hsieh 1991).

3.2.4 Truncated Lévy flight

For the GSD the variance (except for α=2) is infinite, which for analysis purposes especially due to limited statistics in the tails of the distributions, is inconvenient. Moreover, a few major stock markets, e.g. New York Stock Exchange, have introduced "circuit breakers", to limit large amplitude variation in stock prices. Mantegna and Stanley (1995) suggested that when main part of the distribution fits with Lévy distribution, but the tails seem to decay exponentially. In order to capture these properties of the return generating process, some scheme of truncating the distribution seemed plausible and was first proposed by Mantegna and Stanley (1994). This can be realized simply by assuming a finite variance for the asset return distribution, and the model is then called Truncated Lévy Flight model. Following this idea Bouchaud and Potters (1997) developed a smooth Truncated Lévy Flight model, in which the distribution was assumed to be Lévy type (*β=δ=0*) in the middle, but towards the tails an exponential cut-off was made to become more dominant. Only the characteristic function of the distribution can be written in closed form:

$$F_{tlf}(q) = \exp\left\{ -\Delta t \gamma \cdot \frac{\left(q^2 + l^2\right)^{\alpha/2} \cos\left(\alpha \cdot \arctan\left(\frac{|q|}{l} \right) \right) - l^\alpha}{\cos\left(\frac{\pi \alpha}{2} \right)} \right\}$$

where *l≥0* is the cut-off parameter and the definitions of the other quantities are the same as for the Generalised stable distribution (see Section 3.2.3). The probability density function of this three parameter distribution is derived by using the inverse Fourier transformation $f_{gsd}(x) = F^{-1}\{F_{gsd}(q)\}(x)$ and the same approximate numerical method as in the case of General stable distribution can be applied (see Section 3.2.3). It should be noted that when the cut-off parameter *l≠0*, the distribution is no longer stable, and all the moments exist.

3.2.5 Student t-distribution

Blattberg and Gonedes (1974) introduced the Student t-distribution (STU) as a model of asset returns. Since then, this model has become a standard benchmark for new and more complex models (see, for example, Kon 1984). The Student t-distribution seems to be rather attractive model since it is able to capture the excess kurtosis observed in financial time-series, it has finite second moment in contrast to stable distributions, it is easy and fast to estimate, and its mathematical properties are well known.

As in case of the symmetric stable distribution (*β=0*), the Student t-distribution can be obtained from a normal distribution whose variance is random variable. In contrast to the symmetric stable distribution, the variance is now drawn from inverted gamma distribution (Hsieh 1991; Blattberg – Gonedes 1974). This specification is in line with the observation that the variances of financial time series might be non-stationary. Nevertheless, the Student t-distribution is mainly a descriptive model that seems to fit financial data well for which reason it has been widely applied in finance. For example, it has been used to achieve more feasible asset allocations in a downside-risk framework (Lucas - Klaassen 1998; see also Harlow 1991).

The Student t density function with location parameter *m*, scale parameter *H>0*, and degrees of freedom parameter *v>0* reads as follows:

$$f(x \mid m, H, v) = \frac{v^{\frac{1}{2}v}}{B\left(\frac{1}{2}, \frac{1}{2}v\right)} \left[v + H(x-m)^2\right]^{-\frac{1}{2}(v+1)} \sqrt{H} \tag{3.14}$$

where *B(· , ·)* is the Beta function. The Student t-distribution has the following properties: $E(x)=m$, for *v>1* and $Var(x)=H^{-1}v/(v-2)$, for *v>2*. In general, all moments of order *r>v* are finite. Furthermore, when *v=1* the Student density function is the Cauchy density function and when $v \to \infty$ the Student t-distribution converges to the normal distribution. If a Student random variable *x* with *v>2* is standardised by dividing $x-E(x)$ by $\sqrt{Var(x)}$, the resulting variable has (1) fatter tails than the density function of standardised normal random variable and (2) it has larger values in the neighbourhood of mean (3) and it is symmetric. A few such Student t densities with unit variances, zero means, and varying values of shape parameter (*v*) as well as a normal distribution ($v \to \infty$) are plotted in Figure 3.

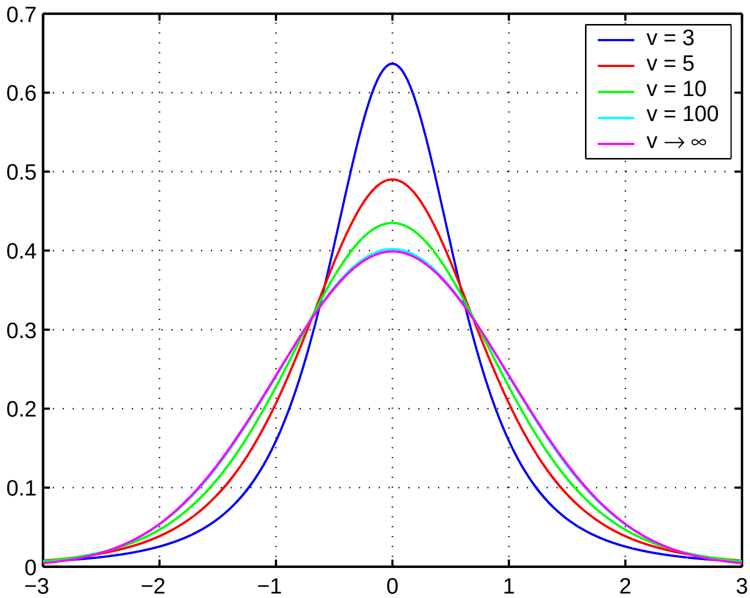


Figure 3: Student t density functions for five values of *v*

The maximum likelihood estimation of the Student t parameters is relative simple with numerical methods. However, the random sampling algorithm is quite lengthy and complicated. For this algorithm, the reader is referred to Devroye 1986.

3.2.6 Power exponential distribution

In the Paper 4 of this bundled thesis, the power exponential distribution was introduced as a time-independent model of asset returns. However, with a different parameterisation, the power exponential distribution (PED) is called generalised error distribution (GED), which is a model considered by Taylor 1994 and Hsu 1982.

The choice of power exponential was argued for by the facts that (1) the selected model family should include the traditional model – i.e., normal distribution - as a special case, (2) the model should be able to capture the observed excess kurtosis and long tails in empirical security return distributions, (3) the model should at least allow the addition of skewness parameter at a later stage of the research if needed, and (4) the model should have a finite second moment. (Töyli et. al. 2002a) Nevertheless, to our knowledge there is no systematic process that would generate power exponentially distributed returns and, thus, strictly speaking, the return series cannot be power exponentially distributed. However, as a descriptive model, the power exponential distribution is a promising model. Power exponential density function reads as follows:

$$p(y\mid\theta,\sigma,\beta)=\frac{\left\{\Gamma\left[\frac{3(1+\beta)}{2}\right]\right\}^{\frac{1}{2}}}{(1+\beta)\left\{\Gamma\left[\frac{(1+\beta)}{2}\right]\right\}^{\frac{3}{2}}\sigma}\exp\left[-\left\{\frac{\Gamma\left[\frac{3(1+\beta)}{2}\right]}{\Gamma\left[\frac{(1+\beta)}{2}\right]}\right\}^{\frac{1}{(1+\beta)}}\left|\frac{y-\theta}{\sigma}\right|^{\frac{2}{(1+\beta)}}\right] \quad (3.15)$$

where the parameters are $-\infty < y < \infty$, $\beta > -1$, $-\infty < \theta < \infty$, and $\sigma > 0$. Power exponential distribution is *normal* when $\beta = 0$, *Laplace* (*=double exponential)* when $\beta = 1$, and approaches *uniform* when $\beta \to -1$.

In literature, it seems that a common practice is to limit β further to be at the maximum unity (see Welsh 1996, Box – Tiao 1973). However, it needs to be pointed out that this choice is not feasible in studies that model financial data since such limitation would mean that the maximum excess kurtosis is three. This is inadequate since larger kurtosis and even heavier tails than in double exponential distribution ($\beta = 1$) have been observed in financial data. Fortunately, there is no mathematical constrains to limit β to be unity, although this would simplify the integration in Bayesian analyses. Relaxing this assumption also leads to a situation, in which the maximum likelihood estimates for location and scale parameters do not necessarily exist although the shape parameter is known (see Agró 1995). In Figure 4 a few density functions with varying values of the shape parameter β are shown.

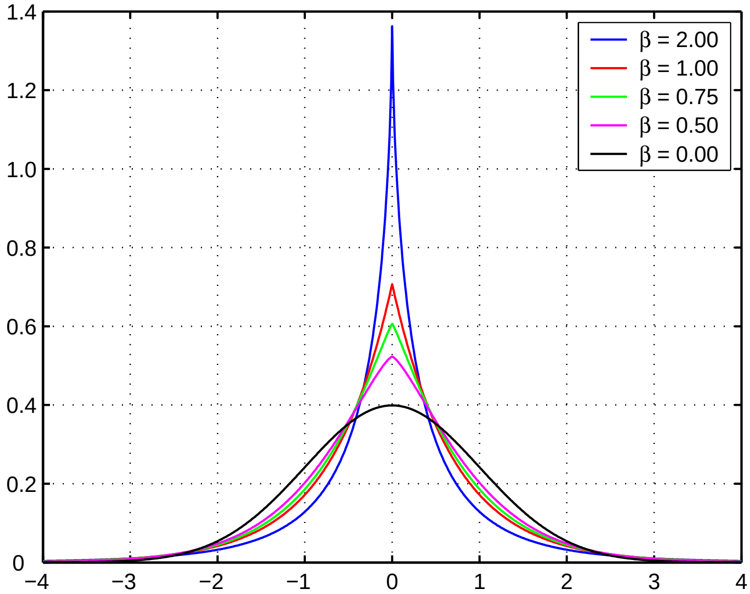


Figure 4: Power exponential density functions for five values of β.

The estimation of PED's parameters, especially the shape parameter β, is non-trivial since, as for the maximum likelihood estimation of parameters, it is noted that these estimates cannot be extracted analytically and they are not necessarily either efficient or unbiased when β is greater than zero (Agrò 1995). There are two alternative solutions to overcome this inconvenience. *First*, ML estimates can be solved numerically and their feasibility, when β is larger than zero, can be investigated with the help of a simulation study. *Second*, the Bayesian paradigm (Bayes 1763; Laplace 1774; 1812), which is based on the notion that after the data is collected it is known for sure, can be applied. Thus, the inference is based on the conditional distributions of the parameters given the data.

In case of ML estimates, we (Töyli et. al. 2002b) tested their feasibility by constructing a simulation study where a sample of 1 000 power exponentially distributed random numbers were drawn 5 000 times for each value of β, incremented by 0.1 from zero to two. The parameters of power exponential distribution were estimated from each sample by numerically minimising the negative logarithmic likelihood function with the algorithm from MATLAB's optimisation toolbox. In Table 3 the results for the mean of each parameter estimate, their standard errors (columns labelled with s.e.), and the true value of β are shown. The data in the table suggest that the maximum likelihood estimates approximate very well the true parameters. Consequently, it seems that this method can be used to estimate the parameters. (Töyli et. al. 2002b)

Table 3: Means of maximum likelihood estimates (Töyli et. al. 2002b)

True β*	β	s.e.	θ	s.e.	σ	s.e.
0.0	-0.0019	1.01E-03	0.0003	4.51E-04	0.9992	3.15E-04
0.1	0.0972	1.06E-03	0.0006	4.40E-04	0.9993	3.31E-04
0.2	0.1982	1.18E-03	0.0002	4.49E-04	0.9993	3.50E-04
0.3	0.2952	1.21E-03	0.0006	4.31E-04	0.9994	3.64E-04
0.4	0.3996	1.27E-03	-0.0007	4.26E-04	0.9994	3.83E-04
0.5	0.4966	1.36E-03	-0.0005	4.04E-04	0.9990	4.00E-04
0.6	0.5973	1.39E-03	0.0001	3.94E-04	0.9993	4.21E-04
0.7	0.6971	1.48E-03	0.0000	3.74E-04	0.9987	4.38E-04
0.8	0.7980	1.54E-03	0.0007	3.58E-04	0.9993	4.58E-04
0.9	0.8992	1.63E-03	0.0004	3.37E-04	0.9994	4.77E-04
1.0	0.9968	1.71E-03	-0.0002	3.26E-04	0.9986	4.93E-04
1.1	1.1013	1.75E-03	0.0000	3.09E-04	0.9999	5.12E-04
1.2	1.2039	1.80E-03	-0.0001	2.89E-04	1.0003	5.40E-04
1.3	1.2995	1.89E-03	0.0000	2.69E-04	0.9988	5.49E-04
1.4	1.4029	1.94E-03	-0.0002	2.55E-04	1.0001	5.79E-04
1.5	1.5058	1.99E-03	0.0002	2.41E-04	1.0003	5.91E-04
1.6	1.6064	2.08E-03	0.0001	2.29E-04	1.0007	6.16E-04
1.7	1.7064	2.09E-03	-0.0002	2.14E-04	1.0006	6.37E-04
1.8	1.8073	2.23E-03	0.0001	2.03E-04	1.0009	6.61E-04
1.9	1.9056	2.28E-03	0.0003	1.87E-04	0.9999	6.76E-04
2.0	2.0075	2.31E-03	-0.0001	1.81E-04	1.0015	6.86E-04

*For each value of beta 5000 simulation with a sample of 1 000 were generated and the parameter were estimated from these simulations

In the Bayesian approach, the inference is based on the conditional distributions of the parameters given the data. This conditional distribution is defined by

$$p(\theta \mid y) \propto L(\theta \mid y) p(\theta) \tag{3.16}$$

where $\theta' = (\theta_1, \ldots, \theta_k)$ is a vector of k parameters and $y' = (y_1, \ldots, y_n)$ is a vector of n observations. The likelihood function $L(\theta \mid y)$ plays a significant role since it is the function through which the data y modifies prior knowledge of θ included in its prior distribution $p(\theta)$. This formulation leads as such to an adaptive model that is capable of learning from experience.

In our work (Töyli et. al. 2002a, Töyli et. al. 2002b) the parameter estimation procedure was based on the work by Box and Tiao (1973). The selection of the prior distributions for θ and σ followed Jeffreys' rule and resulted in non-informative priors. We also assumed the location parameter θ to be independent of σ and β *a priori* and log(σ) locally uniform and independent of β *a priori*. For the non-normality parameter β we used uniform prior distribution (see Töyli et. al. 2002a for discussion). The form of $P_u(\beta \mid y)$ for the numerical calculation of the marginal β was chosen as follows (Töyli et. al. 2002a):

$$P_u(\beta \mid y) \propto \int_{-\infty}^{\infty} e^{\ln\left\{\Gamma\left(1+\frac{1}{2}n(1+\beta)\right)\right\} - n\ln\left\{\Gamma\left(1+\frac{1}{2}(1+\beta)\right)\right\} - \frac{n(1+\beta)}{2}\ln\left\{\left[\sum_{i=1}^{n}|y_i-\theta|^{\frac{2}{1+\beta}}\right]\right\} \pm \ln(c)} d\theta \tag{3.17}$$

in which $-1<\beta<\infty$, $-\infty<\theta<\infty$, and n=1,2,3…, and c is some arbitrary positive constant used to avoid overflows in computation. By following the same logic, the $P_u(\theta \mid y)$ to be used in computation can be written as follows (Töyli et. al. 2002a):

$$p_u(\theta \mid y) \propto \int_{-1}^{\infty} e^{-\frac{n(1+\beta)}{2}\ln\left(\left[\sum_{i}|y_i-\theta|^{\frac{2}{1+\beta}}\right]\right) + \ln\left(\Gamma\left(\frac{2+n(1+\beta)}{2}\right)\right) - n\ln\left(\Gamma\left(\frac{3+\beta}{2}\right)\right) \pm \ln(c)} d\beta \tag{3.18}$$

An additional issue is how to sample power exponentially distributed random numbers. Johnson and Tadikamalla have proposed several alternatives and compared their computational efficiency (Johnson 1979; Tadikamalla 1980). The simplest, but accurate, way is based on Gamma-distribution. It is noted that this approach is not computationally most efficient. Thus, let us define $\alpha = 2/(1+\beta)$. Now, if X is power exponentially distributed with unit variance, zero mean, and shape parameter α, then $|X|^{\alpha}$ has a gamma distribution with shape parameter α^{-1} and scale parameter 1. Based on that, the simulation algorithm is straightforwardly (Johnson 1979; Tadikamalla 1980):

1. Generate $u \sim \text{Gamma}(\alpha^{-1},1)$
2. Let $x = |u|^{\alpha^{-1}}$ (3.19)
3. Attach a random sign to x

In order to obtain a specific mean μ and variance σ^2, let x be[8]:

4. $$x = \mu + x\frac{\sigma\sqrt{\Gamma(\alpha^{-1})}}{\sqrt{\Gamma(3\alpha^{-1})}} \tag{3.20}$$

3.2.7 Mixed diffusion-jump

Merton (1976) proposed mixed diffusion jump process (MDJ) as a model of asset returns. This model is based on the assumption that the returns are gener-

[8] There is a typo in this equation in the paper "On the shape of asset return distribution".

ated by a geometric Brownian motion and an independent compound Poisson process with normally distributed jump amplitudes. The resulting unconditional distributions can be skewed and leptokurtic. Consequently, this model is able to capture the main deviations from normality observed in empirical asset return distributions. Merton has also shown how this model can be used in option pricing. This results in an increased realism when compared to the basic Black and Scholes model (Merton 1976).

The idea of this model is that the asset price dynamics constitute from normal and abnormal vibrations. (1) Normal vibrations in price are, for example, due to the temporary imbalance between supply and demand, changes in capitalisation rates, changes in economic outlook, or other new information that causes marginal changes in the asset value. This part of the variation is modelled with the standard geometric Brownian motion with a constant variance per unit time. (2) Abnormal vibrations in price are then due to the arrival of new important information about the asset that has more than a marginal effect on price. Since it is reasonable to expect that there are quiet and active times in the stock market, this process is in its nature discrete. Merton in turn decided to model this component with a Poisson process. (Merton 1976)

In generic form mixed diffusion jump process can be described by the following stochastic differential equation (conditional on $S(t)=S)$:

$$\frac{dS}{S} = (\alpha - \lambda k)dt + \sigma dZ + dq \tag{3.21}$$

where α is the instantaneous expected return; σ^2 is the instantaneous variance of the return, conditional on no arrivals of new information (i.e., Poisson event does not occur); dZ is a standard Wiener process; $q(t)$ is an independent Poisson process; dZ and dq are assumed to be independent; λ is the mean of arrivals per unit time; $k \equiv E(Y-1)$ where $(Y-1)$ is the random variable percentage change in the asset price if the Poisson event occurs; and E is the expectation operator over the random variable Y. (Merton 1976, 128-129) The evolution of logarithmic price change can be described as follows:

$$\ln \frac{S(t+\Delta t)}{S(t)} = \mu \Delta t + \sigma B_t(\Delta t) + \sum_{n=1}^{P_t(\Delta t)} J_{t,n} \tag{3.22}$$

where t is the absolute time, Δt is time difference between two successive asset prices, $S(t)$ is the asset price at time t, $B_t(\Delta t)$ is the standard Brownian motion, $P_t(\Delta t)$ is a Poisson counting process with parameter λ, $J_{t,n}$ is a normal random variable with mean μ_j and standard deviation σ_j, and $\mu = \alpha - 2^{-1}\sigma^2$ where α the instantaneous conditional expected rate of return per unit time for the

Brownian motion part of the process, and σ^2 is the instantaneous conditional variance of this rate. (Gillemot et. al. 2000)

Assuming that λ, μ, σ, μ_j, and σ_j are constants, the probability density function of mixed diffusion jump reads as follows (Gillemot et. al. 2000):

$$f(x\,|\,\Theta)=\sum_{n=0}^{\infty}\left(\frac{e^{-\lambda\Delta t}(\lambda\Delta t)^n}{n!}\right)N\left(X\,|\,\mu\Delta t+n\mu_j,\sqrt{(\sigma\Delta t)^2+n\sigma_j^2}\right) \tag{3.23}$$

where $\Theta=(\lambda, \mu, \sigma, \mu_j, \sigma_j)$; λ is the parameter of Poisson process, μ_j and σ_j are the mean and standard deviation of a normally distributed jump. Along with these, μ and σ are the respective mean and standard deviation resulting from the Brownian motion. The distribution is leptokurtic when $\lambda>0$ and skewed for $\mu_j\neq 0$ and moments of all order exist. (Akgiray – Booth 1987, 271; Gillemot et. al. 2000). The mean and variance are:

$$\begin{aligned} E(X) &= \mu+\lambda\mu_j \\ Var(X) &= \sigma^2+\lambda\left(\sigma_j^2+\mu_j^2\right) \end{aligned} \tag{3.24}$$

Figure 5 shows a few mixed diffusion jump density functions that are scaled to have unit variance and zero mean.

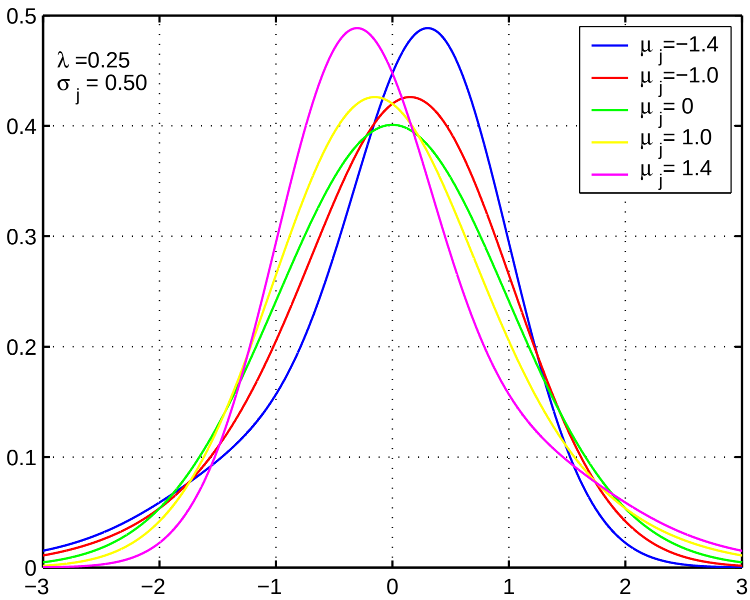


Figure 5: A few mixed diffusion-jump density functions

The estimation of the parameters of this model is again straightforward since the maximum likelihood function is well behaving and relatively simple. The only difficulty is related into the infinite summation that needs to be truncated in numerical estimation algorithms. For example, Gillemot et. al. (2000) apply n=20 as the truncation point.

3.2.8 Compound normal distribution

The latest suggestion of time-independent models, which can be seen as a mainstream model, is the compound normal model (CND) proposed by Kon (Kon 1984). The idea of compound normal distribution (also called mixture normal distribution) is that the returns are drawn from several distinct normal distributions. Evident by this is a special case of the most traditional normal distribution model.

CND leads to an intuitively appealing interpretation since it can be argued that the asset prices are driven by information coming from several information sources. For example, there might be a non-information distribution (i.e. the usual noise), a firm-specific distribution, an industry-specific distribution, and a market-wide distribution (see Kon 1984, 149). These information sources are then modelled with normal distributions that are weighted according to the importance of the particular information source. The observed distribution is then this weighted distribution.

Mathematically, the compound normal probability density function is simply a weighted sum of individual normal distribution and reads as follows:

$$f(x \mid \phi) = \sum_{i=1}^{K} \alpha_i P(x \mid \mu_i, \sigma_i^2) \tag{3.25}$$

where $\phi=(\alpha_1,\ldots,\alpha_K,\ \mu_1,\ldots\ \mu_K,\sigma_1,\ldots,\sigma_K)$ is the parameter vector. In general, moments of all order exist and the mean and variance are given by (Argiray – Bouth 1987, 272):

$$\begin{aligned} E(x) &= \sum_{i=1}^{K} \alpha_i \mu_i \\ Var(x) &= \sum \alpha_i \left[(\mu_i - m)^2 + \sigma_i^2\right] \end{aligned} \tag{3.26}$$

The parameters of the compound normal model are usually estimated with maximum likelihood estimation. This method is convenient since the likelihood function is simple but there is minor difficulty in this since the number of normal distributions to be included in the model is usually *not* known *a priori*.

Thus, also this issue needs to be estimated. Recall that $L(x \mid \phi, K) \geq L(x \mid \phi, K-1)$ but at a certain value of K this difference becomes insignificant. This can be determined with the standard likelihood ratio test. It can be shown (Kon 1984) that the test statistic, given by $-2(\log(L(x \mid \phi, K-1))) - \log(L(x \mid \phi, K))$, is approximately chi-square distributed with three degrees of freedom.

The likelihood ratio test has generally been used to estimate the number of normal distributions included into the model (Kon 1984, Akgiray-Booth 1987, Tucker 1992). However, if the number of normal distributions is decided on the basis of likelihood ratio test and the comparison of competitive models is based on Schwarz criteria (like also has been a common practise), this might lead to a situation where compound normal model is unfairly discriminated. In these cases, the number of normal distributions included into the model can be decided with the help of Schwarz criteria (Schwarz 1978, Akgiray-Booth 1987) that is as follows:

1. Calculate $SC = \log L(x \mid \phi) - d \log T^{0.5}$ (where $L(x \mid \phi)$ is the maximum likelihood function value, *d* the number of independent parameters, and *T* is the sample size) for each model.
2. Select the model with largest *SC*. (3.27)

In general, this model is able to capture the excess kurtosis and skewness of the data. It includes the traditional normal distribution as a special case, has finite moments, and has an intuitive interpretation that is well in line with other findings and theoretical reasoning. In addition, according to Tucker (1992) the mixed diffusion-jump model (see Section 3.2.7) can be arbitrary closely approximated by a finite mixture of normal densities (i.e., compound normal model).

The simulation from compound normal distribution is quite trivial. This can be done by (1) determining the normal distribution – i.e., information source and (2) then drawing from this information source; for example, with the familiar Box-Muller algorithm. On step 1, the weights of distributions define the corresponding probabilities of information sources.

3.3 Time dependent models

3.3.1 General model

The time-independent models of asset returns discussed in Section 3.2 are able to capture the heavy tails of empirical distributions when compared to those of normal distribution. General stable distribution, mixed diffusion-jump, and compound normal models are also able to replicate the possible skewness. However, none of the time-independent models can model the volatility clustering or other time-dependencies evident in financial data series. A more generic time-dependent model can be achieved by dropping out iid assumption and defining that the returns are generated by the following stochastic process:

$$r_t = \mu + \sigma_t \varepsilon_t \tag{3.28}$$

where r_t is the return at time t, σ_t is the time-dependent standard deviation, μ is the expected value, and ε_t is some probability distribution. The differences in the model specifications include the distribution of ε_t and the equation for the time-dependent standard σ_t. It needs to be pointed out that this possible dependency does not necessarily imply market inefficiency since it is enough for efficiency that the return-generating process can be represented as a martingale (Fama 1970). The expected value (μ) is usually not modelled as time-dependent (note also the lack of t subscript) since return series are usually not autocorrelated (Töyli et. al. 2002a; Töyli et. al. 2002b; Cont et. al. 1997; Schwert 1989; see Cont 2001; see Fama 1988). However, such addition would be simple to implement and, after that, the model could produce autocorrelated time series.

It is also worth to contrast the model in Eq. (3.28) in more detail with the time-independent models since these have similarities. Blattberg and Gonedes (1974) pointed out that the symmetric stable distribution is obtained from a normal distribution whose variance is drawn from a strictly positive stable distribution, that the Student t-distribution is obtained from a normal distribution whose variance is drawn from an inverted gamma distribution, and that Clark's (1973) model[9] is a normal distribution whose variance is drawn from a log normal distribution. Thus, all these mixture models can be written in the form: $r_t = \sigma_t z_t$, where z_t is iid standard normal, and σ_t is another ***iid*** random variable. (Hsieh 1991).

The models assuming time-dependency can be divided in two categories based on the fact whether the time-dependency of variance is modelled with a stochastic process or via some deterministic equation. The former category includes stochastic volatility (SV) models, and models based on chaos theory leading to complex dynamics. The latter category includes autoregressive heteroscedastic (ARCH) type models, and models based on different kinds of exponentially smoothing schemes.

[9] Clark tested a hypothesis that the distribution of price change is subordinate to a normal distribution. Thus, the price series evolves at different rates during identical intervals of time. The empirical test were done with a specification where independent increments of process *X(t)* were normally distributed, directed by a process *T(t)*, whose independent increments were lognormally distributed.

3.3.2 Variance as exponentially weighted moving average

Probably the simplest widely used specification for time-dependent model is the RiskMetric^{tm} model created by J.P. Morgan (RiskMetrics^{tm} – Technical Document 1996). J.P. Morgan's well-known and widely applied value-at-risk[10] framework, which is a standard benchmark for new risk management models, is based on this model where the time-dependency is defined with the help of exponential smoothing (RiskMetrics^{tm} – Technical Document 1996, 51) as follows:

$$\begin{aligned} r_t &= \sigma_t z_t \\ \sigma_t &= \sqrt{\lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2} \end{aligned} \tag{3.29}$$

where r_t stands for logarithm return, $z_t \sim N(0,1)$, and the parameter λ ($0<\lambda<1$) is the decay factor. RiskMetric^{tm} assumes λ=0.94 for 1-day volatility and λ=0.97 for 1-month volatility (RiskMetrics^{tm} – Technical Document 1996, 100). This approach is simple, fast to calculate, and it is close to IGARCH models. However, it has been criticised by the academic community and improvements have been suggested (see, for example, Ahlstedt 1998).

3.3.3 Autoregressive conditionally heteroscedastic models

It was Engle (1982) who suggested an autoregressive heteroscedastic (ARCH) type model with an application to inflation time series. Bollerslev (1986, 1987) specified GARCH type model that is a generalised ARCH. These models have thereafter been further developed and different variations have been suggested, for example, the surveys by Bollerslev et. al. (1992), Bera and Higgins (1993) and Bollerslev et. al. (1994) report several hundred papers.

In the seminal paper by Engle (1982) a discrete time stochastic process (ε_t) is defined as an ARCH model of the form

$$\begin{aligned} &\varepsilon_t = z_t h_t^{1/2} \\ &z_t \; i.i.d., E(z_t) = 0, Var(z_t) = 1 \end{aligned} \tag{3.30}$$

where h_t is a time-varying, positive, and measurable function of the information set at time t-1. The conditional variance of ε_t equals to h_t and may change

[10] During the past few years, the value-at-risk techniques have enjoyed substantial success among the academic as well as commercial communities. The value-at-risk is based on the idea that the downside risk is measured with a single figure given a probability level (usually 95%). Thus, the value-at-risk describes the maximum amount of possible loss with 95 % probability. For summary, see Simons 1996.

over time. The iid assumption results in serially uncorrelated ε_t with zero mean. The following time-dependent parameterisation for h_t was suggested by Engle (1982):

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad \text{, where } \alpha_0 > 0, \alpha_i > 0 \text{ for all } i \tag{3.31}$$

The stochastic process of mean in Eq. (3.31) is a martingale that in turn is enough for market efficiency to hold (Fama 1970). The time-dependent equation for the conditional variance is also able to capture the volatility clustering observed in financial time series. However, the estimation of ARCH with large values of q is rather difficult and often leads to violation of the non-negativity constraints on α_i*'s*.

Bollerslev (1986) suggested the GARCH model as an alternative that also allows dependency in the past values of $h_{t.}$ The GARCH model of order (p, q) is given by:

$$h_h = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \quad \text{, where } \alpha_0 > 0, \alpha_i > 0, \beta > 0 \text{ for all } i \tag{3.32}$$

It can be shown (see Bollerslev 1986) that GARCH model is an infinite-order ARCH model with exponentially decaying weights for large lags. A high order ARCH can therefore be substituted by a low-order GARCH model. The GARCH model can be clarified further by noting that the conditional variance h_t in Eq. (3.32) depends linearly on the past behaviour of the squared values in an autoregressive AR(q) process and on the conditional variance itself in a moving average MA(p) process.

In order to capture the common empirical finding that shocks in variance tend to be persistent when using high frequency data, i.e., the sum of parameters α_i and β_i is close to one. In order to be able to capture this empirical behaviour, Bollerslev and Engle specified the family of IGARCH models in which the parameters sum to unity as being integrated in variance. In the IGARCH models, shocks to the conditional variance are persistent in the sense that they remain important for forecasts of all horizons. However, there is the problem that unconditional variance for simple IGARCH(1,1) does not exists. (see for further discussion Bollerslev – Engle 1993) Other variations and improvements also exist (see Bollerslev et. al. 1992, Bera and Higgins (1993) and Bollerslev et. al. (1994) for collections of references)

In general, (G)ARCH type models are able to capture the volatility clustering observed in financial-time series, they can explain the excess kurtosis, and do not violate the market efficiency. Figure 6 shows the time-dependent standard deviation extracted from HEX index (returns were multiplied by 100).

The blue line indicates the time dependent standard deviation calculated based on GARCH model and red line that derived from RiskMetric[tm] model. The two models seem to produce very similar behaviour. This is not a surprise since the functional forms of these two models are close to each other.

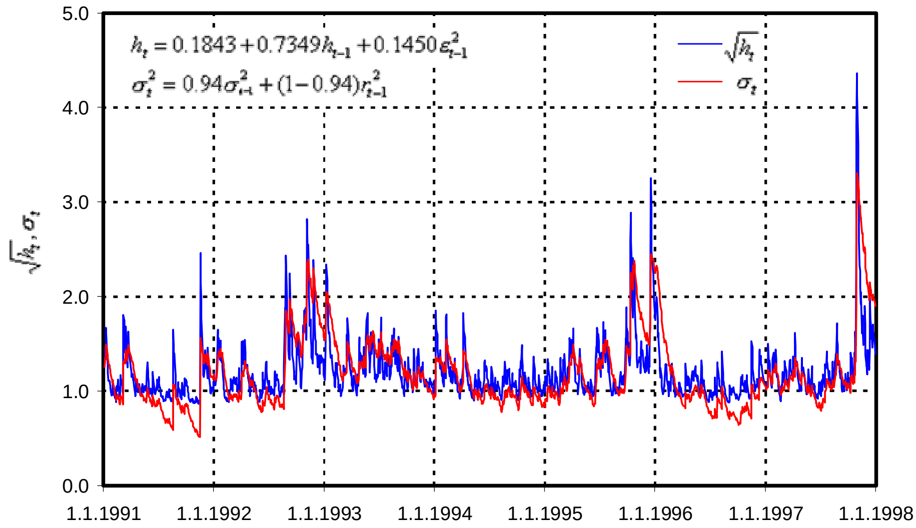


Figure 6: Time-dependent standard deviation

3.3.4 Stochastic volatility

One of the problems in (G)ARCH-type models is that the variance equation does not contain an innovation. The models of stochastic volatility assume the variance to be a random variable. The usual way to model the time-dependent variance is to assume that it follows an AR type process. One of the inherent difficulties of stochastic volatility models is that the likelihood function is substantially more difficult to handle than in ARCH type models. However, this inconvenience might be balanced by the intuitive appeal of the assumption that the volatility changes stochastically rather than deterministically. This is especially good to explain large crashes like that on black Monday in the year 1987. (See Taylor 1985, Taylor 1994 for summaries about SV)

In general, the SV type models are similar to the ARCH type model in the senses that they are also able to capture the volatility clustering observed in financial-time series, can explain the excess kurtosis, and do not necessarily violate the market efficiency. Taylor has also suggested that a judicious combination of both ARCH and SV might provide more satisfactory results than the use of a single model (Taylor 1994). An excellent discussion about the various stochastic volatility models and their relationship to ARCH type model is provided by Jacquier et. al. 1994 (Jacquier et. al. 1994).

3.4 Comparison of main stream models

3.4.1 Visual comparison

Figure 7 shows simulated return series generated according to the mainstream models: (from left to right, from top to down) S&P 500 index representing real empirical data, data simulated from kernel density estimate, normal distribution, Lévy, Student t-distribution, compound normal distribution, mixed diffusion jump, power exponential distribution, ARCH(1) process, and GARCH(1,1) process. All series have been simulated with parameter estimated from S&P 500 index and, after simulation, they have been standardised to have unit variance and zero mean. In the estimation and simulation, our MATLAB toolbox (Gillemot et. al. 2000; Töyli et. al. 2003) has been used. Note that the y-axis is truncated.

The empirical data is clearly different to all simulate processes. Especially, the normally distributed data lacks all the special characteristics that are in the empirical data. In time-independent models, the volatility clustering is missing and only GARCH process is able to capture this phenomenon but it is still too smooth – i.e., lacks large jumps and crashes. In contrast, the Lévy process seems to include too many large jumps and crashes. The compound normal distribution, Student t-distribution, and kernel density estimate seems to be rather close to real data but the assumed time-independency leads to the lack of clear volatility clustering. Mixed diffusion jump and power exponential appear too smooth since large changes are generally missing. Along with these points, the modest autocorrelation present in empirical data is naturally missing from all artificial series because it was assumed non-existent when deriving these models.

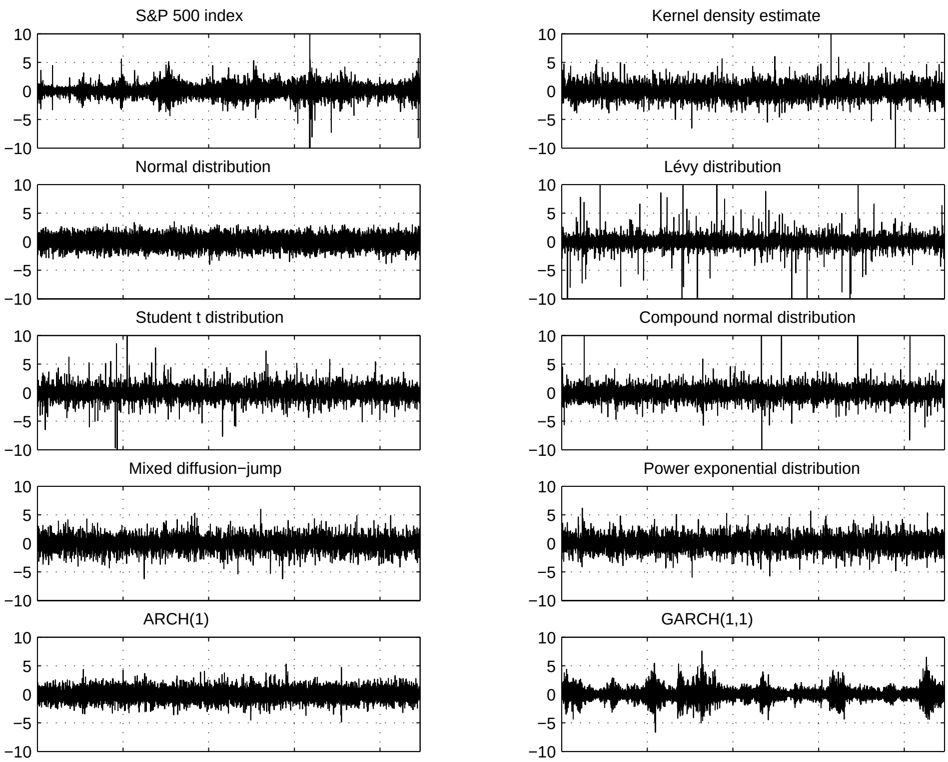


Figure 7: Comparison of mainstream models

3.4.2 Literature review

There has been an ongoing debate concerning the models of asset returns. Louis Bachelier's (1900) model, where the changes in the logarithms of bond prices were normally distributed, formed the basis for classical theory of finance that is largely based on the assumed normality. Osborne (1959) formally specified the asset return generating process as Brownian motion. However, Mandelbrot (1963) and Fama (1965a) first reported the fundamental differences from the normality: empirical return distributions are fat-tailed and peaked when compared to normal distribution (i.e., they are *leptokurtic)*. Mandelbrot (1963) and Fama (1965a) also suggested the stable distribution as a model of asset returns. Initially, the stable distribution seemed to fit the data well and be able to capture the *leptokurtic* feature of empirical distributions.

Although the stable distribution has the theoretically appealing scaling and stability under addition properties and it is able to capture the excess kurtosis, the later research has generally rejected the stable distribution as a model of asset returns (Blattberg – Gonedes 1974; Hagerman 1978). Blattberg and Gonedes argued that the Student t-distribution is a better model for daily returns than the stable distributions. They further concluded that monthly returns are roughly normally distributed. Hagerman reached a similar conclusion by observing that the return distribution approaches normal when the time interval is increased. More recently, Akgiray again reported the approximate normality of monthly returns (Akgiray 1989). Others have also reported evidence that is *not* consistent with the hypothesis means that the returns are stable under addition and show scaling behaviour. (Officer 1972; Hsu et. al. 1974; Lau – Wingender 1990). This conclusion remains even if the non-symmetric versions (GSD) of stable distributions are used (Tucker 1992).

In contrast, Mantegna and Stanley recently reached again the conclusion that stock returns follow stable distribution by using high frequency data and differences instead of logarithmic returns (Mantegna – Stanley 1995). In addition to that, Mantegna and Stanley (1995) did point out an approximately exponentially decay in the tails of the distribution. This decay would mean that the second moment exists; thus, the scaling must break down, and the distribution becomes finally Gaussian.

Along with the above points, one of the initial arguments supporting the stable distributions was their suitability for the portfolio theory. Because of the stability under addition, it should be convenient to model the portfolios. Since it is difficult to combine stocks with different characteristic exponents in portfolios, no realistic portfolio models assuming stable distribution have been suggested. The use of the same characteristic exponent for all shares in a single portfolio would in turn violate the empirical findings suggesting that assets have different characteristic exponents. The only well known attempts to solve this portfolio construction owes to Fama (Fama 1965b) and Samuelson (Samuelson 1967) but their models assume that all stocks in portfolio are distributed with the same characteristic exponent.

The estimation of the parameters of the stable distribution is also non-trivial since the density function exists in closed form only in rare special cases. At least two methods have been suggested to do this estimation. (1) The approximation of Fama and Roll and (2) Fast Fourier transformation technique of DuMouchel (see Hagerman 1978, Tucker 1992 for reviews). Researchers have applied both of these methods. For example, the approximation was used by Blattberg and Gonedes (1974), and Hagerman (1978) while the Fourier transformation was applied by Mantegna and Stanley (1995), Kullman et. al (1999), and Gillemot et. al. (2000). In addition, Tucker (1992) applied both techniques and selected the estimates that resulted in larger maximum likelihood value. Although there exist strong support for Blattberg and Gonedes's result that the Student t-distribution is better model of asset returns than the stable distribution (see Tucker 1992; Tucker - Pond 1998), there is always one minor problem when evaluating stable distribution’s goodness of fit. When researchers apply the approximation technique of Fama and Roll in parameter estimation they take the risk since this technique does not necessarily produce the maximum likelihood estimates. Thus, all comparisons based on maximum likelihood estimates could unfairly discriminate the stable distribution. However, according to research, this is not likely to bias the inference (Blattberg – Gonedes 1974).

Akgiray and Bouth (1997) provide a survey of the compound models (the mixed diffusion-jump and the compound normal distribution) of asset returns with the conclusion that for weekly returns, both models have significantly higher descriptive validity than a stationary normal distribution, and, in most cases, mixed diffusion-jump model is empirically superior to finite normal mixtures. However, according to Akgiray and Bouth (1997), the distribution of monthly returns is approximately normal and both the mixed diffusion-jump process and the compound normal distribution model converge rapidly to their limiting normal process. Akgiray and Bouth (1997) further conclude that, if one unit of time is measured as one month or longer, the normality assumption is acceptable but, otherwise, the general form of the mixed diffusion-jump process should be assumed. Kon in turn compared the compound normal distribution with the Student t-distribution with the conclusion that the compound normal distribution is better fitting (Kon 1984).

Tucker compared compound normal distribution, mixed diffusion jump model, Student t-distribution, and stable distribution with the conclusion that both the compound normal distribution and the mixed diffusion jump model outperform the Student t-distribution and the stable distributions distribution (Tucker 1992). It was also concluded that the mixed diffusion-jump model and the compound normal distribution are equally well fitting and they cannot be set in preferential order based on the goodness of fit (Tucker 1992; Tucker – Pond 1998). This conclusion can be justified the fact that the mixed diffusion-jump model can be arbitrary closely approximated by a finite mixture of normal densities (Tucker 1992, 80).

Recently, the techniques assuming time-dependent variance have enjoyed substantial success among the academic community (see Bollerslev et. al. (1992); Bera - Higgins (1993); Bollerslev et. al. (1994); Taylor 1985, Taylor 1994) because the variance of asset returns seems to be finite (recall that the general stable distribution leads to model that assumes infinite variance) but time-dependent in a complex non-linear manner (Perry 1983; see also Fama - French 1988). This time-dependency can explain the volatility clustering observed in financial time series and the leptokurtic unconditional return distributions. There is also some evidence that SV models may fit data better than ARCH-type models (Danielsson 1994). It should be noted that the ARCH type models used to filter the data might not be able to remove all non-linear dependencies while SV might perform better in this sense (Hsieh 1991).

3.4.3 Goodness of fit

Since on one hand the purpose of this research is to investigate the shape of asset return distribution, it would be convenient to assure that the models use serve also as a reasonable description of the data. It needs to be pointed out that the goodness of fit should ***not*** be interpreted so that returns are *iid* with a given probability density function (pdf). Rather that the tested model is reasonably good description of the true return distribution and, therefore, its use is not likely to bias the inference about the shape of the asset return distribution.

χ^2-test and Kolmogorov-Smirnov test are widely used to draw conclusions about the goodness of fit. These two tests are discussed next. The χ^2 goodness of the fit test statistics reads as follows:

$$S^2 = \sum_{i=1}^{r} \frac{(a_i - np_i)^2}{np_i} \tag{3.33}$$

where n is the number of data points, r is the number of groups with boundaries given a sequence $(-\infty, X_1], (X_1, X_2], (X_2, X_3], \ldots ,(X_{r-1}, \infty)$, and the matching probabilities are: $p_1 = \int_{-\infty}^{X_1} f(x)dx$, $p_i = \int_{X_{i-1}}^{X_i} f(x)dx$ for $2 \le i < r$, and $p_r = \int_{X_r}^{\infty} f(x)dx$. S^2 is asymptotically χ^2 distributed with $(r\text{-}d\text{-}1)$ degrees of freedom where d is the number of estimated parameters.

An alternative for χ^2-test is the Kolmogorov-Smirnov test that is a procedure to measure the goodness of fit in which the theoretical distribution function must be known completely for all parameters. This method is applicable only for continuous distributions but requires smaller sample size than χ^2-test. The algorithm is as follows (Harris – Stocker 1998):

1. Propose the hypothetical distribution function H_0: $F(x;W)$, and give the significance level α.
2. Use the ordered sample:
   $x_1 \le x_2 \le \ldots \le x_n$
   and its empirical distribution function

$$F_n(x) = \begin{cases} 0 & for\ x < x_1 \\ \dfrac{i}{n} & for\ x_i \le x < x_{i+1} \\ 1 & for\ x \ge x_n \end{cases} \tag{3.34}$$

3. Calculate the test quantity, maximum absolute difference: $D_n = \sup_x |F(x) - F_n(x)|$
4. Apply the decision rule: Reject H_0 if $D_n > D_{\alpha;n}$.

The original form of the Kolmogorov-Smirnov test is valid only if *all* the parameters of the proposed distribution are known and the distribution is continuous. The simple approximation for the critical value $D_{\alpha;n}$, like the algorithm and C-code suggested by Press et. al. (1992), rely on this assumption. For distributions with estimated parameters, approximations are available at least for normal, Weibull, and exponential distributions (See Law – Kelton 1991, 389-393, for references). In order to overcome this inconvenience, the critical values can be determined with simulation.

Table 4 show the critical values derived via simulation for the power exponential distribution (Töyli et. al. 2002b). When these critical values are compared to those derived under the hypothesis that all the parameters are *known*, the differences are substantial. For example, in the known parameters case, the critical value for the sample size 251 and β=0 (normal distribution) is 0.085 (see Press et. al. 1992). This is quite a substantial difference when compared to 0.056 obtained in the simulation, which is also the same value obtained by Lilliefors (1967) for normal distribution with estimated parameters. However, it needs to be pointed out that the critical values are conservative when the shape parameter β is also estimated from the sample, which is the case in our study.

Table 4: Critical values for $D_{0.05;251}$ statistic

β	Critical	β	Critical	β	Critical	β	Critical
-0.3	0.0634	0.4	0.0481	1.1	0.0552	1.8	0.0616
-0.2	0.0632	0.5	0.0479	1.2	0.0562	1.9	0.0637
-0.1	0.0581	0.6	0.0490	1.3	0.0586	2.0	0.0653
0.0	0.0564	0.7	0.0520	1.4	0.0602	2.1	0.0662
0.1	0.0516	0.8	0.0516	1.5	0.0617	2.2	0.0690
0.2	0.0496	0.9	0.0534	1.6	0.0613	2.3	0.0657
0.3	0.0485	1.0	0.0544	1.7	0.0643	2.4	0.0680

The algorithm to produce the distribution for *D-statistics* (Töyli at. al. 2002b) was as follows (repeated for each value of β of interest):

1. Simulate thousand 251 data point samples from power exponential distribution with shape parameter β (location and scale are irrelevant here).
2. Fit power exponential to each 251 points sample using maximum likelihood.

3. Calculate *D* statistics from the samples with fitted parameter values.
4. Use the achieved *D's* distribution to determine the critical value.

3.4.4 Model selection

The likelihood ratio test is one of the most widely applied methods to test the goodness of two different models. The idea of this test is that the distributions f_1 and f_2 with dimensions of parameter spaces d_1 and d_2 ($d_1 < d_2$), determine the likelihood ratio:

$$H^2 = 2\left(\ln\left(L_{f_2}\left(x \mid \Theta_{d_2}\right)\right) - \ln\left(L_{f_1}\left(x \mid \Theta_{d_1}\right)\right)\right) \tag{3.35}$$

where L_{f1} and L_{f2} are the likelihood functions; d_1 and d_2 are the respective parameter spaces, and x is the data. H^2 follows asymptotically χ^2 distribution with (d_2 - d_1) degrees of freedom. It needs to be noted here that this test necessitates that the parameter spaces are nested. This implies that the parameter space of the pdf with smaller number of independent parameters is derivable from the parameter space of the pdf with higher dimensionality. Despite of this fact there are examples of studies in which the likelihood ratio test along with Monte Carlo simulations for validity checking have been used, although the assumption of the parameter space being nested has not been checked (see Blattberg – Gonedes 1974; Merton 1976). In contrast to that, the usage of the likelihood ratio test has also been unnecessarily refused (see Akgiray - Bouth 1987).

An alternative to likelihood ratio test is the Schwarz criterion (Schwarz 1978). This test has been widely applied in financial studies (Tucker 1992; Akgiray - Bouth 1987; Jacquier et. al. 1994). Schwarz criterion is based on Bayesian approach, which says that it is most appropriate to select the model with the highest posterior probability. Since it is usually impossible to calculate posterior probabilities directly, the following approximate procedure is usually applied:

1. Calculate $SC = \log L(x \mid \Theta) - d \log T^{0.5}$ (where $L(x \mid \Theta)$ is the maximum likelihood function value, *d* the number of independent parameters, and *T* is the sample size) for each model.
2. Select the model with largest *SC*. (3.36)

The Schwarz criterion is especially applicable when the parameter spaces of the different model being compared are not nested. There is also some anecdotal evidence that the Schwarz criterion and the likelihood ration lead to similar inference when both tests are legitimate to use (see Gillemot et. al. 2000).

4. Analysis Methods

This Chapter will describe the datasets and methods used in the Papers of this bundled thesis. A data from a major market (Standard & Poor's 500), a small and quite volatile market (Helsinki Stock Exchange), and artificial data was used. We chose to use index data instead of a single asset because the index data reflects the behaviour of the marker as whole. This choice can be further argued for by the fact that, for practical purposes, the analysis of index data might be more informative because the investments should be diversified and, thus, they constitute portfolios. The indices are also available from long time periods. Nevertheless, we are not in anyway calling the importance of the analysis of single assets into question. We merely chose to start from indices.

The first sample consists of Standard & Poor's 500 (SP500a) daily index data of the New York Stock Exchange from more than 32 years (from 3 July 1962 till 29 December 1995) and contains altogether 8 431 trading days. This index is one of the most widely studied, it is available for a long time period, and reflects the behaviour of a major market. Although financial time series of this kind usually grows exponentially, the data seems to have a rather clear change in the growth rate round the end of 1970s. In other words, it seems that during the first 5 000 trading days the mean value of index seems constant but after this the mean value seems to rise. This dataset was studied in the Papers 1 and 2.

Standard & Poor's 500 (SP500b) daily index was also studied in the Papers 3, 5, and 6. However, in these Papers, a slightly longer time period (from 2 July 1962 to 31 December 1997) was analysed. The sample contained 8 939 data points.

The third dataset (analysed in the Papers 1 and 3) was an artificial data (AF) generated by the following stochastic process: $x(t) = [x(t-1) + \xi(t)] * b$ where $\xi(t)$ was distributed according to an "inflated" Gaussian process:

$$\mathrm{P}(\xi) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} e^{\left\{-\frac{\xi^2}{2\sigma^2(t)}\right\}} \tag{4.1}$$

with the standard deviation being time-dependent as $\sigma(t) = \sigma(t-1) * b$ where $b>0$. The sample contained 30 000 data points corresponding to 10 000 trading days, each day including three index points, and thus 1 000 data points would correspond to an artificial trading year. The parameters b, σ_0 and x_0

were set to their empirical values estimated from the S&P500 data: $b-1=1.78\times10^{-5}, \sigma_0=0.4, x_0=70$.

The fourth dataset was the Helsinki Stock Exchange (HEX) all shares daily index from 2 January 1991 to 30 December 1997 containing 1 751 data points. The HEX all shares index includes all stocks listed on the Helsinki Stock Exchange and the shares are weighted with their market capitalisation. The base date of the index is 28 December 1990 and the base number is 1 000. In international scale, Helsinki Stock Exchange is small and quite volatile. This data was analysed in the Papers 3, 4, 5, and 6.

The fifth dataset, studied in the Paper 6, was a high frequency data from Standard & Poor's 500 (SP500HF) index from 2 January 1997 to 31 December 1999 containing 1 081 528 data points. This data was recorded on about 20 seconds time-interval.

In the analyses, the full datasets (Papers 1-6), data split by weekdays (Papers 4 and 5), data split by years (Papers 4 and 5), 251 data points' rolling windows (Papers 4-6), and data (SP500HF) divided into 42 separate equal length bins (Paper 6) were studied. This division was applied to tick, one minute, and five minute data sets, which means that each bin included 25 700, 7 000 and 1 400 data points, respectively. The division was done because we found that the fittings for the full datasets (i.e. to very large datasets) tend to average out details. The analysis of the bins instead seems to reveal structure and thus more details about the return generating process. The binned approach was not applied for longer holding periods because of limited amount of data. In case of the rolling window, the window was rolled over the data by forwarding it every time by one data point and its length (251 data points) was set to reflect the typical length of a HEX trading year. The statistics calculated from successive rolling windows – different just by one data point – are highly dependent. However, this should not be problematic since the windows are used in the descriptive sense for investigating the changes in patterns over time. They are not used as inputs for other statistical analysis where the dependency and induced autocorrelation would be problematic. Since previous research has generally concluded that monthly returns are approximately normally distributed, the evolution of the shape of return distribution was studied in different holding periods from one day to one month (Papers 2, 4-6) and from 15 minutes to one day (Paper 6).

As the variable, we used the logarithmic returns on all papers and, in addition to that, simple price differences in the Papers 1 and 3 in order to see the effect of measurement quantity on the inference. Since financial data typically shows modest autocorrelation, its effects was studied in the Papers 4 and 5 by filtering the autocorrelation out with the help of Cochrane-Orcutt correction defined as follows: $y_t = x_t - rx_{t-1}$ where *r* is the estimated first autocorrelation,

x_t the original return at time t, and y_t the corrected series. This correction should remove the first order autocorrelation. The estimates $r_{HEX} = 0.1835$ and $r_{SP500b} = 0.114$ were obtained from the full data set and the correction was applied to the full data. All analyses were made to the corrected as well as to the raw data. Since it turned out that the estimated first autocorrelation varied substantially in different rolling windows, the Cochrane-Orcutt correction was also applied in rolling manner. However, the inference remains the same whether the rolling correction or correction to full data set was used.

In addition to the linear dependency, financial data has non-linear dependencies. These were studied in the papers 1, 5, and 6 by "pre-whitening" the data with the GARCH process as follows:

$$w_t = \frac{x_t}{\sqrt{h_t}}$$

where w_t is the whitened series, x_t is the original return series and h_t is the time-dependent variance defined as follows:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where the coefficients $\alpha_0 > 0$, $\alpha_i > 0$, $\beta_i > 0$ for all i. We used mainly the GARCH(1,1) process, i.e. chose $q=p=1$, but we also tested the IGARCH (only in case of SP500a data) and ARCH processes. The conclusions remain the same regardless of the process used to pre-whiten the data.

All models of asset returns were fitted to the data with maximum likelihood method. However, for power exponential distribution a numerical algorithm to Bayes fit it was developed and implemented as C++ code. This algorithm was used in the Papers 4 and 5. In the Paper 5, the Bayesian estimates were compared to maximum likelihood estimates whose validity was tested with the help of a simulation study.

The different models were compared using the Schwarz criteria (Papers 3, 6) and the likelihood ratio test (Paper 3). However, the likelihood ration test led to confusing results when it was applied for comparisons that are theoretically not justified. The goodness of fits were tested with χ^2-test (Papers 3 and 6) and with Kolmogorov-Smirnov test in case of power exponential distribution (Paper 5). For Kolmogorov-Smirnov test, the critical values were determined with the help of a simulation study.

5. Summary of Results

This Chapter will shortly summarise the main results presented in the Papers. The discussion is grouped under the titles of papers. It needs to be pointed out that, although the papers were published in sequential order, they were prepared partly parallel and, thus, are somewhat overlapping.

5.1 Paper 1: Characteristic times in stock market indices

In this paper, the aim was to show where the limits of scaling in the stock market data are. The SP500a and the AF datasets were analysed. The analysis was started by demonstrating with the help of a simple random walk model (AF data) that the proper quantities to look at are the logarithmic returns. Figure 8 shows Lévy distribution fitted to the AF dataset when logarithmic returns and simple price differences are used as measurement quantities. Figure 8 suggest that the use of differences instead of logarithmic returns is likely to produce Lévy like distribution although the underlying process is simple Gaussian.

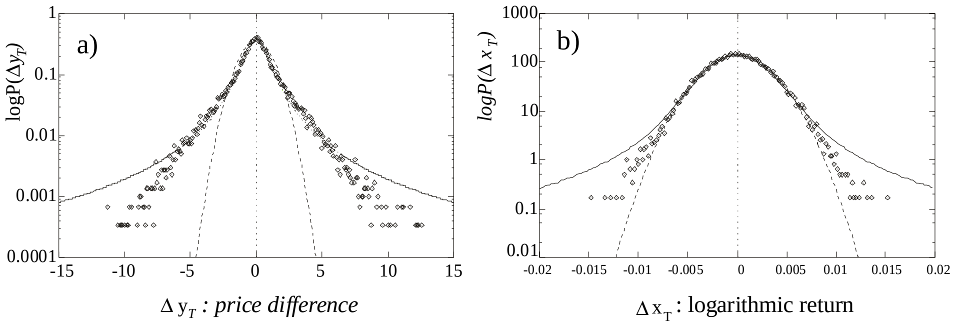
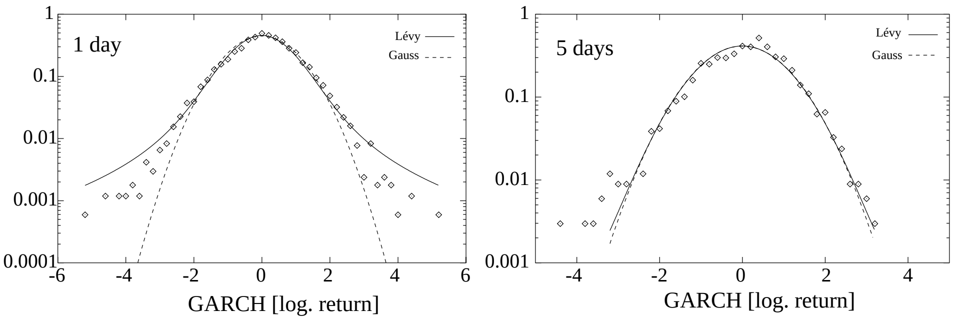


Figure 8: Effects of terms of measurement in the AF data

In the further analysis, fittings were made to the Lévy distribution either using the index data as such or pre-processing it with ARCH, GARCH or IGARCH methods, which should remove, or at least mitigate, the effects of time-dependent variance. It was found out that for short times the (truncated)

Lévy distribution fits the data relatively well. Since this truncation leads to a non-stable distribution, the scaling behaviour observed for short times should break down for longer times. Figure 9 illustrates the convergence of the return distribution from Lévy-type behaviour towards Gaussian behaviour of the GARCH treated return as a function of time step.



α=1.64 α=1.967

Figure 9: Crossover from Lévy to Gaussian behaviour in SP500a data

It was also concluded that the crossover time for scaling is definitely shorter than that reported by Mantegna and Stanley (1995) who found the scaling to hold up to 1 000 min. This contradiction was speculated to be due to the following reasons. First, Mantegna and Stanley fitted the exponent to the central part of the distribution. Second, they used a much shorter time span of the S&P500 index, which leads to a rather limited number of data points for time step Δt = 1 day and larger. Third, they studied the simple price difference that becomes more and more different from the logarithmic returns as Δt increases. We further found out that the ARCH, GARCH or IGARCH treatment of the data does not transform the data normally distributed, although the value of the exponent α increases after performing the treatment, for time step Δt = 1 day. Nevertheless, the Gaussian value is reached asymptotically when the time step increases.

In general, this inference was found to be quite robust. Thus, it did not depend on whether the raw data or the ARCH, GARCH or IGARCH treated data was being analysed or whether the data was split into two parts (the early part with fluctuations around roughly constant mean [till about the late 70's] and the later part with non-constant mean and larger fluctuations).

5.2 Paper 2: Break-Down of Scaling and Convergence to Gaussian

In this paper, we continued with the topic introduced in the Paper 1 and continued analysing SP500a data. Motivated by the fact that the truncated character of the return distribution implies that scaling must break down and that the distribution ultimately converges to a Gaussian, we concluded that at least two different characteristic times could be defined. One of them (τ_s) shows the break-down of scaling, the other one (τ_G) shows the time scale of the convergence of the kurtosis. It was also found that τ_s is of the order of one day while τ_G is in the range of few months; thus, $\tau_s \ll \tau_G$. Figure 10 illustrates the convergence of kurtosis to Gaussian value 3. Here the solid line implies the theoretical value calculated based on the fitted values.

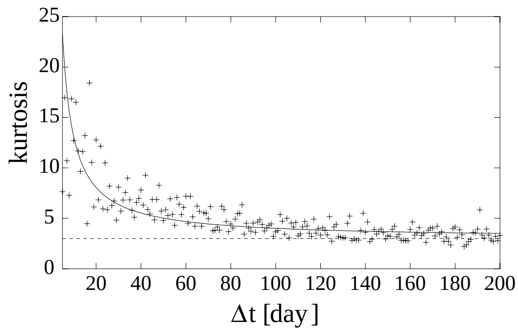


Figure 10: The convergence of the kurtosis in SP500a data

5.3 Paper 3: Time-independent models of asset returns revisited

In this paper we revisited various well-known time-independent models of asset returns. These were simple normal distribution, Student t-distribution, Lévy, truncated Lévy flight, generalised stable distribution, mixed diffusion jump, and compound normal distribution. The datasets used were SP500b, HEX, and AF.

First, we showed that the likelihood ratio test is “legitimate” to use when comparing mixed diffusion jump, compound normal distribution, and normal distribution. Although it is not known whether this test is justifiable to use to discriminate uniquely between any other models, we did so and found out that the use of the likelihood ratio test to discriminate between all the models leads to contradictory results of their ranking order. Therefore, the ranking order was decided by using the Schwarz criterion, which also gave the same results as obtained by using the likelihood ratio test for those models for which likelihood ratio test is known to be legitimate to use.

In addition, we tested whether the inference is affected if the number of normal distributions included in the compound normal distribution is decided using the Schwarz criterion instead of the maximum likelihood ratio test. Although our results showed that these two criteria occasionally result in different number of normal distributions to be included in the model, the ranking order of compound normal distribution compared to the other models did not change.

Our results also indicated that all models, excluding the simple normal distribution, are from visual perspective reasonable descriptions of the data. Furthermore, the use of differences instead of logarithmic returns tends to make the data visually more Lévy type distributed as we also concluded earlier (Paper 1). Figure 11 shows the time-independent models fitted to SP500b dataset. All other models than normal distribution seem visually very close to each other.

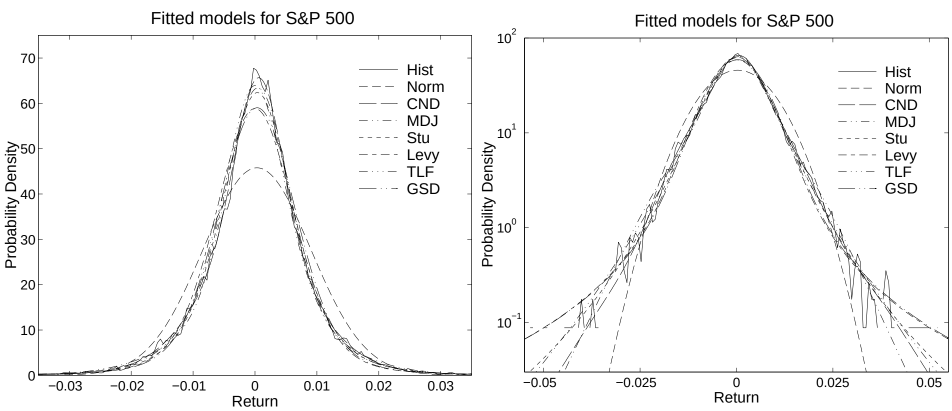


Figure 11: Time-independent models of asset returns fitted to SP500a data

When comparing the distributions statistically, the Schwarz test was used as the decision criterion. The Student t-distribution seemed to give the best description of the empirical returns (SP500b, HEX, and AF). However, this model is merely descriptive since it lacks an intuitive explanation of an underlying process, unlike most of the other models that describe the empirical returns well or reasonably well. Only the simple normal distribution seems not to be an adequate description of daily returns.

To comment about the stable distributions (the Lévy, and generalised stable distribution), they seem to be quite indistinguishable in any of the plots and underestimate the peak of the empirical return distribution while, at the same time, giving weight to the tails due to the infinite variance. In general, we concluded that the extension of the Lévy model to generalised stable distribution model by adding the location and skewness parameters did not seem in any noticeable way to improve the descriptive power of the Lévy model. Both these models demonstrate the contradiction between the fat tails and infinite variance of the data, which the truncated Lévy fight distribution resolves while keeping the quality of the fit reasonable. Finally, the result that none of the physically motivated models can be considered as very good representation of the stock index data (not as good as the merely descriptive Student t-distribution), underlines the possibility that time-dependent models may provide better representations of the return generating process.

5.4 Paper 4: Variability of Hex Return Distribution

This paper explored the possible variability in the shape of the of asset return distributions. This problem was elaborated with the help of the following questions: does the shape of distribution vary from on weekday to another, is the shape of distribution constant over time, and does the shape of the distribution vary along with the time horizon used to calculate the returns? The nature of this paper was descriptive and it seek not to give any substance or modelling based explanation of the possible time variation of the shape of return distribution.

The analyses were started with the non-parametric kernel density estimator that suggested that the return distribution is unimodal, leptokurtic, and quite symmetric. However, the limited amount of data available made the tails of the kernel density estimate messy and, thus, no inference concerning these was possible. In addition, it seems, given the sample size, that the kernel density estimator underestimates the probability mass of the central regions. These difficulties lead to the conclusion that kernel density estimation might not yield a good description of the empirical distribution – at least the statistical sample size should be substantially larger.

The left pane of Figure 12 shows the kernel density estimates that were calculated from the full HEX dataset by using the raw data and Cochrane-Orcutt corrected data. Normal distribution is plotted as a benchmark. There are no observable differences between raw data, and Cochrane-Orcutt corrected data, which indicate only modest skewness. The right pane of Figure 12 shows Kernel density estimates based on raw data grouped by weekdays. In general, these estimates are more messy than those on the left pane and unclear even in the central region of the distribution. Thus, it is not possible to do reliable inference based on these estimates. Nevertheless, one minor feature can be observed: the asset returns on Wednesdays seem to be slightly different from other weekdays. However, this has never before been reported and it is likely because of the bias in the density estimates.

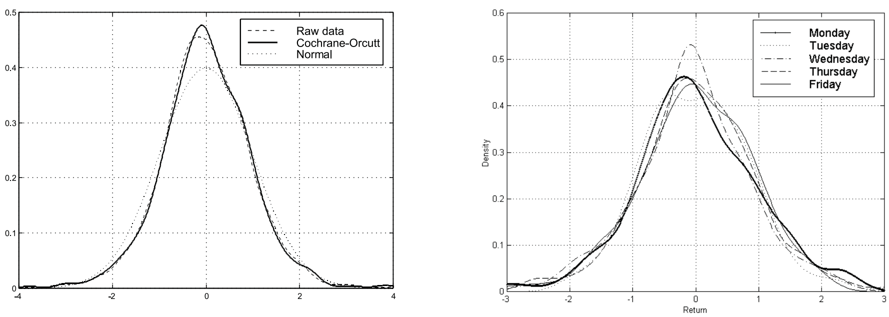


Figure 12: Kernel density estimates extracted from HEX data

The second model, we used, was the power exponential distribution whose parameters were estimated with the Bayesian approach. For the estimation, a numerical algorithm was developed and it was implemented with C++. The results show that the location parameter θ is virtually zero although, in statistical sense, small but significant deviations were observed. These deviations were concluded to be usually safe to ignore for all practical purposes since, given the transaction costs, it is hardly possible to place any profitable trading rules on them. It also seemed that in case of short-term inference, it would be justifiable to assume θ to be zero although the long-run expectation was found to be slightly positive as was expected. No significant deviations between weekdays were observed but the expectation tends to somewhat increase towards the end of the week being negative only on Mondays.

The main issue was the shape of the distribution. Since the return series generally indicate anomalies related to weekdays, the question whether the shape of the return distribution differs between weekdays was explored first. The results indicated no convincing evidence about the differences in the shape of the distribution between weekdays. This result might not hold in other markets because the weekday effects are not generally strong on Finnish market.

The variability of the shape of the distribution over time was especially interesting. The analyses provided strong evidence that the shape is not constant and, periods, when the return generating process could be assumed normal, were observed. These findings are illustrated in Figure 13 where the 90%, 50%, and 10% percentiles of marginal β calculated from 251 data point's rolling window are shown ($\beta = 0$ denotes normal distribution and $\beta = 1$ denotes double exponential distribution)

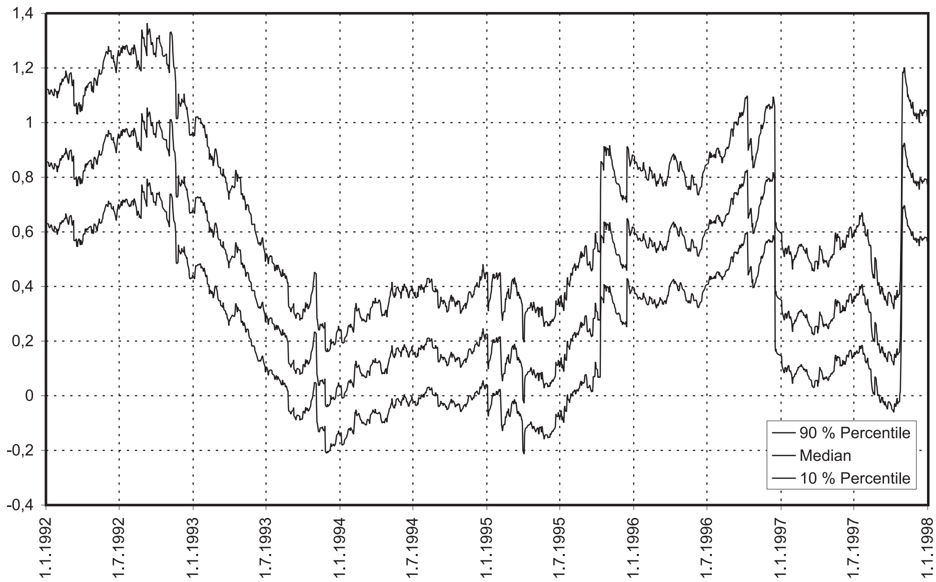


Figure 13: Percentiles of the marginal β (shape parameter) in HEX data

The observed variability of the shape might explain part of the contradiction in the previous results concerning the shape of the distribution. Nevertheless, the shape of return distribution also seemed to approach steadily normal distribution when the time interval used to calculate returns was increased from one day towards thirty days' returns. This observation was concluded to be consistent with the previous research.

5.5 Paper 5: On the shape of asset return distribution

In this paper we continued on the same path as on the Paper 4. However, we also considered data from a major market (SP500b), took into account the known time-dependencies, compared the inference with that derived from the Student t-distribution, and tested the goodness of fit. Thus, the purpose of this paper was still to explore the possible variability of the shape of the asset return distribution with a fourth research question: how the known linear and non-linear time dependencies affect the inference concerning the shape of the distribution?

In this paper, we used again the power exponential distribution as a model but compared the inference to that derived from the Student t-distribution. We also compared maximum likelihood estimates to the Bayesian estimates. The validity of maximum likelihood estimates was first tested with a simulation study. We also tested the goodness of power exponential fit with Kolmogorov-Smirnov test for which the critical values were determined with simulation.

The results indicate no convincing evidence to suggest that there are true differences in the shape of the return distribution between weekdays. In HEX data, Wednesdays appeared a bit different when raw data or Cochrane-Orcutt corrected data were analysed but, given the variance of shape parameter β, the observed difference could well be by chance. There is also no explanation why Wednesday returns should be more peaked than those of other weekdays. In SP500b data, Monday data seemed slightly more peaked than other weekdays. This finding can be partly understood or supported by the fact that most of the large stock market crashes took place on Mondays. However, Monday seems to be similar to other weekdays when the data is split into two parts approximately in the middle (the conclusion is not sensitive to the exact location of truncation point) and each part is analysed separately.

As earlier we concluded that the location of the distribution is non-stationary over time but the deviations are so small that, given the transaction costs, it is hardly possible to develop profitable trading rules based on them. We further suggested that in case of risk measurement, e.g., in value-at-risk framework, it might be justifiable to assume the location as zero. The expectation was negative for Mondays in both indices but significant only in the case of SP500b dataset. This is in line with the usual finding that the weekday effects are not strong on Finnish market. The negativity of Monday's returns in SP500b is also supported by the fact that most of the largest crashes took place on Mondays.

The main results obtained in this study were that the shape of the asset return distribution is not constant over time and that the known linear and non-linear time-dependencies cannot explain this observed variability. Thus, there are significant variations in the shape over time while periods of normality are also observed. This finding suggests that the contradiction of previous results concerning the shape of the distribution might partly be related to the different time-periods used in the analyses. The results also suggest that the known linear and non-linear time-dependencies are not enough to completely explain the observed distributions since their removal does not transform the data to normally distributed. Therefore, it might be that the shape of the distribution is time-dependent or, at least, it is non-stationary. The effect of known dependencies is illustrated on Figure 14 where the modes of the posterior distribution extracted from each 251 data point's rolling window (SP500b) using raw data, Cochrane-Orcutt corrected data, and GARCH treated data are shown.

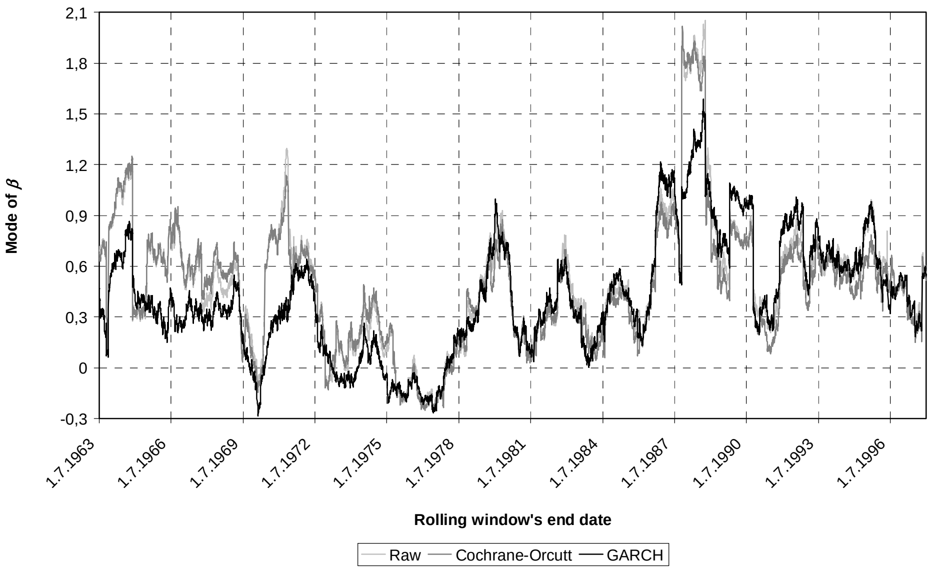


Figure 14: Effect of known time-dependencies on the shape in SP500b data

The previous research has generally concluded that monthly returns are approximately normally distributed and that by increasing the time interval or the holding period the return distribution approaches normal distribution. This tendency was observed in the HEX index and the monthly returns seemed to be quite well described by normal distribution. In contrast, although in the SP500b data the returns start to converge towards normality, this convergence seems to stops after about six days. Generally speaking the convergence of S&P 500 index towards normality is very slow being of the order of several months if it even exists. The convergence seems to be somewhat more evident if the data set is split after 5 000 trading days and the parts are analysed separately. Especially, the first part of the data indicates a tendency to converge towards normality. This might also suggest that there has been some fundamental change in the return generating process. This possible change was found also in our earlier work (Paper 1) where we observed an evident change in the trend and increased fluctuations during the latter part of the SP500a dataset. This change may be due to the development of information technology and due to the resulting increased productivity. It also seems that the

GARCH corrected data and the raw data are becoming similar when the time interval is increased although GARCH coefficients do not entirely vanish. After approximately ten days for S&P 500 and four days for HEX, the shapes of raw data and GARCH treated are virtually identical. These findings are demonstrated in Figure 15.

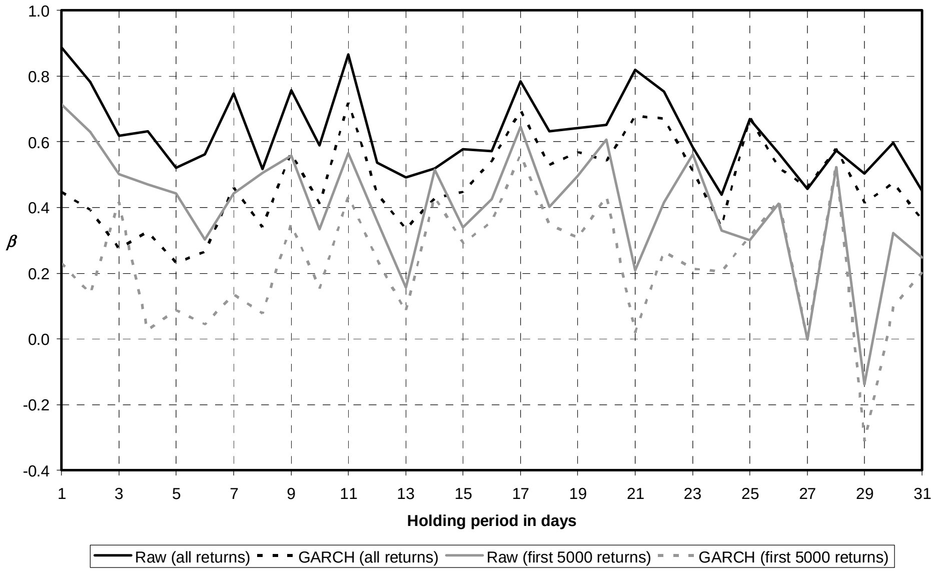


Figure 15: Effect of time intervaling on the shape in SP500b data

Finally, we crosschecked our results with the Student t-distribution and concluded that the two models lead to identical inference. Furthermore, the goodness of power exponential fit was checked with Kolmogorov-Smirnov test in order to exclude the possible bias resulting from poorly fitting model. It was concluded that the power exponential distribution is fitting the data, except at the very end of the HEX index data, and, as a result, the inference and conclusions are not likely to be biased because of a poorly fitting model. Nevertheless, these goodness-of-fit results should not be interpreted such that returns are iid with power exponential distribution function. As a matter of fact, this could not be the case since to our knowledge there is no systematic process that would generate power exponentially distributed returns and, thus, strictly speaking, the return series cannot be power exponentially distributed. However, as a descriptive model, the power exponential distribution is fitting with the data well.

5.6 Paper 6: Models of Asset Returns: Changes of Pattern from Tick by Tick to 30 Days Holding Period

This final paper of this bundled thesis summarized, clarified and extended the analyses and results of our multi-year project in exploring asset returns. This paper focused to study the effects of (i) different time periods, (ii) different holding periods and (iii) non-linear dependencies on the conclusions concerning the best fitting time-independent model of asset return distribution (iv) and to compare the time-independent models to a simple time-dependent model. A special interest was paid to the evolution of the properties of asset returns when very short holding periods (high frequency data) are being analysed and to the question whether the models of asset return distribution found good on daily data also provide good description of data for shorter time-intervals.

Here we defined SC-measure based on Schwarz criterion for visualisation purposes as follows:

$$SC_{measure} = \frac{(SC_x - SC_{Normal})}{0.5 * \log N} \tag{5.1}$$

where SC_x is the Schwarz criterion value for model x (CND, MDJ, Lévy, GSD, TLF, PED, STU, GARCH), SC_{Nomal} is the value for the normal distribution, N the number of data points, and the denominator implies the effect of one parameter on the SC-measure value. Thus, for example, a value 10 means that such a model would be equally good as normal distribution although 10 parameters were added to it (all other things being equal). In general, a larger value implies a better model.

We started our analysis with SP500HF data where we did the fittings to a sequence of subsets obtained from the tick (20 seconds), one minute, and five minutes datasets by dividing them sequentially into 42 equal length bins. The analysis of the binned datasets revealed interesting and unexpected behaviour since, although the overall trend appears to be that any of the other models seem to perform better than the normal distribution, the GARCH(1,1) model is on average the poorest performing model after normal distribution on 1- and 5-minute time intervals. In the case of tick data, the same holds for the time-independent models but now the GARCH(1,1)-model seems to perform overall similarly with, being in some bins better and in some others worse, than time-independent models. We also concluded that volatility and time-dependencies tend to grow along with the deviation from normality.

Figure 16 shows fitting for 15 minutes (N = 20 289), 1 hour (N = 4 527), 2 hour (N = 3 770), 4 hour (N = 1 513) and 1 day (N = 756) for the full SP500HF dataset. Figure 16 suggests that all time-independent models except normal distribution are close to each other, these models outperform GARCH(1,1) model for short time intervals (less than about four hours), and the normal distribution is clearly the worst model. On average, the Student t-distribution seems to be the best time-independent model and Lévy type models (Lévy, GSD, TLF) seem to perform relatively speaking better when the holding period decreases. We further concluded that the poor performance of GARCH(1,1) calls in the question the explanatory power of the time-dependent variance. Thus, it is likely that at high frequencies the time dependencies are more complex and necessitate at least that the shape of the distribution is modelled to be time dependent in addition to the second moment.

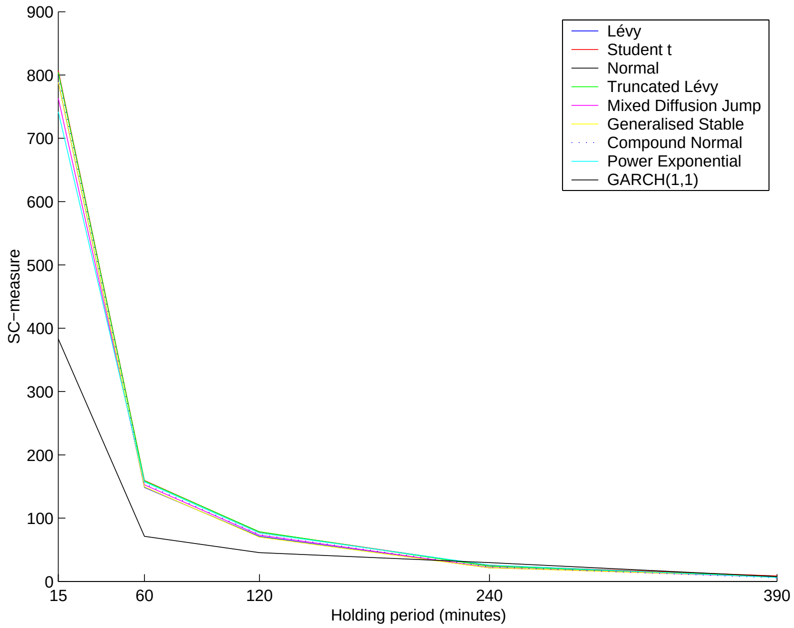


Figure 16: Ranking order for different holding periods in SP500HF data

Next we moved on to look at daily data using the 251 data points' rolling windows. The results for HEX are depicted in Figure 17 (the analysis of SP500b data yields similar inference). Figure 17 indicates that there are time-periods when the return generating process is best described by the Normal distribution (39.0 % of rolling windows, for SP500b data 19.9 % of rolling windows[11]). All other time-independent models than normal distribution seem to be relative close to each other in both datasets. The time-dependent GARCH(1,1) model is, on average, the best model although we expected it to perform better (it is best in 40.0 % of rolling windows for HEX and 30.9 % of rolling windows for SP500b). It also seems that GARCH model is usually the best model when the process cannot be assumed normal. These findings might imply that there are periods of "business as usual" when the process is described well by the Normal distribution. However, for some reason – e.g., external shock, bubble formation – every now and then periods of ferment emerge. These periods are characterised by higher volatility and increased time-dependencies. This finding needs to be tested further before more detailed conclusions can be drawn. Nevertheless, such tendency, if really the case, could help us to understand better the formation of bubbles and the crashes following them.

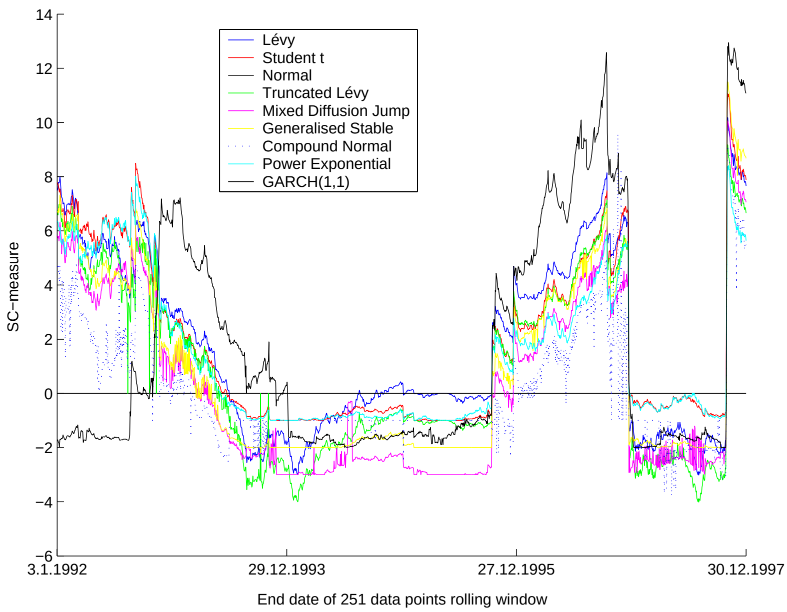


Figure 17: Ranking order of models in rolling windows in HEX data

[11] In SP500b data the Lévy distribution turns out to be the second best model, but it generally outperforms the Normal distribution when the characteristics exponent is close to 2 (=normal distribution). For such time periods the Lévy and normal models are very close to each other (the gap is typically less than "one parameter" wide). Thus, if the Normal distribution is fitted to data assuming zero mean, the GARCH model is still the best fitting model in 28.2 % of windows but the Normal distribution achieves second most first positions (in 27.2 % of windows).

Since we used earlier (Papers 1, 4, and 5) GARCH techniques to filter out non-linear dependencies, the same was done here for comparison reasons. When compared to the raw data, the processed data is clearly transformed towards normality but this normality cannot always be assumed. Typically, those parts where GARCH(1,1) model is the best fitting model are still after correction non-normal. This finding implies that the known time-dependencies do not completely explain the variability of the shape. In addition, it underlines the possibility of two kinds of periods; "business as usual" and "ferment" periods. During the ferment period the time-dependencies might become stronger as discussed above.

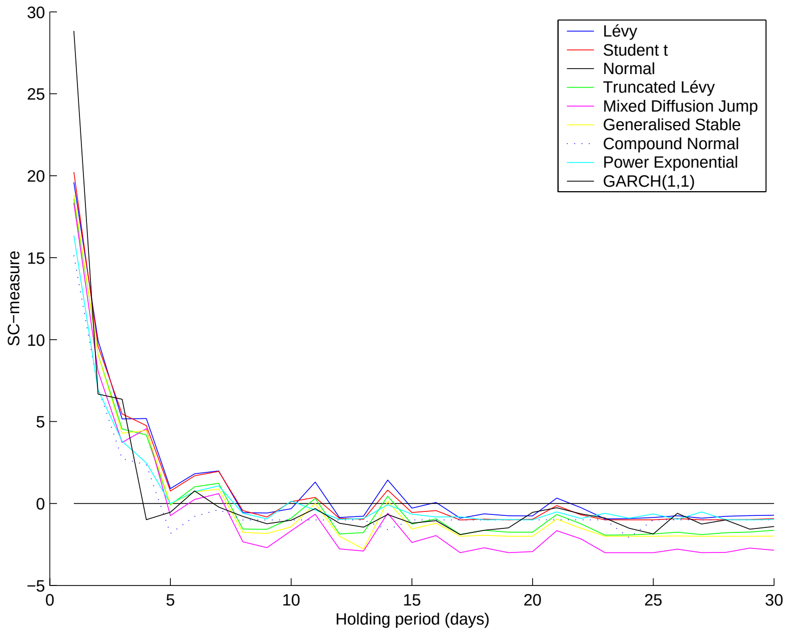


Figure 18: Ranking order for different holding periods in HEX data

Figure 18 shows the order of models extracted from different holding periods from one day to 30 days in HEX data. SP500b data behaves similarly (plot is not reported here). In both datasets, the GARCH(1,1) model is the best model for short holding periods (for 1 day in HEX and up to 8 days in S&P 500). However, when the holding period increases, superiority of GARCH model mitigates. This finding implies that the role of time dependencies is vanishing. In general, the time-dependencies vanish slower and the distribution converges slower towards normality in S&P 500 than in HEX, in which case the return distribution is well described by simple normal distribution for eight days and longer holding periods. In case of S&P 500 data, the Student t-distribution turns out to be the best model for holding periods longer than 8 days.

In summary, the results show that high frequency data is best modelled with time-independent models that are able to capture the excess kurtosis and it was a surprising finding that the GARCH(1,1) model becomes relatively poor when the holding period becomes shorter than about 4 hours. Nevertheless, the daily returns are best modelled by GARCH(1,1) while the time dependencies that GARCH model captures vanish after a few days. Thus, monthly returns are relatively well described by simple normal distribution. However, the analyses of large datasets tend to average out details in the data sets. When the data is split into the bins or rolling windows, the analyses seem to reveal structure and thus more details about the return generating process. In general, it seems that there are periods of “business as usual” when the process is described well by the Normal distribution and for some reason – e.g., external shock, bubble formation – every now and then periods of ferment emerge. These periods are characterised by higher volatility and increased time-dependencies.

6. Discussion

This research was initially motivated by Mantegna's and Stanley's results (1995) that suggested Lévy distribution as a model of asset returns and demonstrated the scaling property that seemed to be present in the data. Our journey to unknown has been challenging and difficult but rewarding. It is worth of explicitly considering the contribution of our work before closing this thesis. The Papers 1 and 2 contributed the field by identifying the characteristics times present in financial data. It was shown that the scaling tends to break down after one day and the characteristic time of the convergence of kurtosis is in the range of few months. We also demonstrated that the use of simple price changes, a practise applied by the early econophysicists, could lead to considerably different inference than the use of logarithmic return. In general, the use of logarithmic return was argued to be preferable because of their convenient theoretical and practical properties. The Papers 1 and 2 also suggested that the removal of time-dependencies with GARCH techniques might not completely transform the unconditional distribution to normal rather the unconditional distribution still shows some Lévy type behaviour.

The Paper 3 contributed the field by revisiting the well-known time-independent models of asset returns. It was shown that unjustified use of the likelihood ratio test is likely to result in confusing conclusions. It was further shown that it is legitimate to compare compound normal model and mixed diffusion jump model with likelihood ratio test. Along with that, anecdotal evidence that the Schwarz criteria leads to similar inference, when comparing two models, as likelihood ratio test, if both tests are legitimate to use, was found. In general, the Student t-distribution was found to be the best fitting time-independent model of asset returns. For the fitting of the well-known time-independent models, a Matlab toolbox was also developed.

The Papers 4 and 5 concentrated on the shape of the return distribution. The main finding was that after taking into account the know time-dependencies the shape of return distribution still remains non-stationary. This finding suggests that also the shape might be time-dependent. In contrast to temporal variation, the shape was found to be similar for different weekdays. This finding contrasts to the reported anomalies related to the mean and standard deviation. Along with that, a convergence towards normal distribution was observer when time interval was increased and the effect of time-dependencies seemed to completely vanish after a few days. Furthermore, a numerical algorithm to produce Bayesian parameter estimates for power exponential distribution was developed and implemented as C++ code. Paper 5 continued by showing that maximum likelihood estimates of power exponential distribution's parameters are unbiased with the help of a simulation study. Critical values for Kolmogorov-Smirnoff goodness of fit test for power exponential distribution were also produced with the help of a simulation study.

The final paper of this bundled thesis summarized, clarified and extended the analyses and results of the earlier papers. Thus, the findings that, on daily data the shape of asset return distribution is non-stationary, there are time periods when the Normal distribution is the best model, and the removal of known time-dependencies transfers the data towards normality although the dependencies cannot explain the non-stationary behaviour were corroborated. Here we also found out that a time-dependent GARCH(1,1) model does not necessarily outperform the time-independent models although it is on average the best fitting model for daily returns and that GARCH(1,1) model is often the best model when the process cannot be assumed normal. Especially, when full daily datasets are analysed GARCH(1,1) is clearly the best model. However, the superiority of GARCH model mitigates rapidly when the holding period increases towards a month, which implies time-dependencies are vanishing. This finding also further corroborates earlier results. We also found out that the fittings for the full datasets tend to average out details in the datasets and hide the subtle structure. Thus, the binned analysis of high frequency data revealed important phenomena: as expected the normal distribution was clearly the poorest performing model, but surprisingly, the other time-independent models seemed usually to outperform the GARCH(1,1)-model for holding periods less than four hours although the fine grained data evidently includes time dependencies. We also concluded that volatility and time-dependencies tend to grow along with the deviation from normality.

These finding led us to raise three questions. First, we speculated that there are likely periods of "business as usual" when the process is well described by normal distribution. However, for some reason – e.g., external shock, bubble formation – every now and then periods of ferment emerge. These periods are characterised by higher volatility and increased time-dependencies. Second, the poor performance of GARCH(1,1) model on high frequencies lead us to question whether the assumption of GARCH that returns are normally distributed with time-dependent parameters is reasonable and whether it should be substituted with some other model where also the shape is allowed to vary over time. Such a model could, at least in theory, capture the business as usual periods as well as periods of ferment. Third, although we were surprised by the poor performance of GARCH(1,1) on high frequencies, we were reluctant to generalise this finding before a more detailed analysis. However, if the behaviour we observed is typical for financial data, this finding would also be a source for further insight to the return generating process.

It also needs to be noted that our results in general call into question the validity of analysis methods relaying on normality assumption. This is especially crucial in risk management applications since it is likely that the normal distribution based risk management frameworks underestimate the risk during periods of ferment although they might be more adequate during business as usual periods. This could partly explain the unexpected, catastrophic losses when a ferment period emerges.

As a closing remark, it is worth to sketch the paths for future research. There are four areas both theoretically interesting and practically challenging for more detailed studies. The first area is to continue the statistical modelling of asset returns. This path includes deeper exploration of the questions raised. Thus, here the key issue might be to build a simple model that includes time-dependent interpretable shape parameter in addition to time-dependent variance. Such a model could perhaps be used to study the possible variation between periods of ferment and business as usual. The high frequency data is also relatively little studied and it deserves more attention. Here at least the surprising GARCH-finding needs to be tested further. The second, and probably the most challenging area, is to explore the operation of financial market with the help of agent based modelling. This area has a substantial potential to increase our understanding about how the market actually works since it is very challenging if not impossible to construct a model based on a simple stochastic process which could truly replicate all the empirical peculiarities. In contrast, the agent based models are very promising and the computer capacity currently available allows their effective use. The third area is to apply the results of this thesis on real world application, for example, to risk management models. Even the use of some of the time-independent models might considerably improve the contemporary solutions that are currently often based on normal distribution. And finally our Matlab toolbox for fitting the discussed distributions and sampling from them would be beneficial for the researches in the field in general if it were implemented in slightly more user-friendly way. This is worth of doing.

References

Agro, G. (1995) Maximum Likelihood Estimation for the Exponential Power Function Parameters. *Communications in statistics. Part B Simulation and Computation*. Vol. 24:2, 523-536.

Ahlstedt, M. (1998) *Analysis of Financial Risks in a GARCH Framework*. Bank of Finland studies E:11. Oy Trio-Offset Ab: Helsinki.

Akgiray, V. – Bouth, G. (1987) Compound Distribution Models of Stock Returns: An Empirical Comparison. *Journal of Financial Research*. Vol. 10:3, 269-280.

Akgiray, V. (1989) Conditional Heteroscedasticy in Time Series of Stock Returns: Evidence and Forecast. *Journal of Business*. Vol. 62:1, 55-80.

Alles, L. A. – Kling, J. L. (1994) Regularities in the variation of skewness in asset returns. *The Journal of Financial Research*. Vol. 17:3, 427-438.

Ariel, R. A. (1987) A monthly Effect in Stock Returns. *Journal of Financial Economics*. Vol. 18:1, 161-174.

Arthur, W. B. – Holland, J. H. – LeBaron, B. – Palmer, R. – Tayler, P. (1997) Asset Pricing Under Endogenous Expectations in an Artificial Stock Market. In Arthur, W. B – Durlauf S. N. – Lane, D. A. (1997) *The Economy as an Evolving Complex System II*. Addison-Wesley: Reading, Massachusetts.

Bachelier, L. (1900) *Theorie de la Speculation*. Paris: Gauthier-Villars.

Badrinath, S. G. – Chatterjee, S. (1988) On measuring skewness and elongation in common stock return distributions: the case of the market index. *Journal of Business*. Vol. 61:4, 451-472.

Badrinath, S. G. – Chatterjee, S. (1991) A data-analytic look at skewness and elongation in common-stock-return distributions. *Journal of Business and Economic Statistics*. Vol. 9:2, 223-233.

Bera, A. K. – Higgins, M. L. (1993) ARCH models: Properties, Estimation and testing. *Journal of Economic Surveys*. Vol. 7:4, 305-366.

Blattberg, R. – Gonedes, N. (1974) A comparison of Stable and Student Distributions as Statistical Models for Stock Prices. *Journal of Business***.** Vol. 47, 244–280.

Bollerslev, T. - Chou, R.Y. – Kroner, K. F. (1992) ARCH Modelling in Finance. *Journal of Econometrics*. Vol. 52:1-2, 5-59.

Bollerslev, T. – Engle, R. F (1993) Common persistence in conditional variances. *Econometrica*. Vol. 61:1, 167-187.

Bollerslev, T. – Engle, R.F. – Nelson, D.B. (1994) *ARCH models*. In Engle R.F. – McFadden D. L. (1994) *Handbook of Econometrics Volume 4*. Elsevier Science B.V: Amsterdam. 2961-3038.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*. Vol. 31:1, 307–327.

Bollerslev, T. (1987) A Conditional Heteroscedastic Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics*. Vol. 69:3, 542–547.

Bonabeau, E. (2002) Predicting the Unpredictable. *Harvard Business Review*. Vol. 80:3, 109-116.

Bookstaber, R. – McDonald, J. B. (1987) A General Distribution for Describing Security Price Returns. *Journal of Business,* Vol. 60:3, 401-424.

Booth, G. G. – Hatem, J. – Virtanen, I.– Yli-Olli, P. (1992) Stochastic Modeling of Security Returns: Evidence from the Helsinki Stock Exchange. *European Journal of Operational Research*. Vol. 56:1, 98-107.

Bouchaud, J-P. – Potters, M. (1997) Théorie des risques financiers. Commissariat aĺ'Eńergie Atomique.

Box, G. E. P. – Muller, M. E. (1958) A note on the generation of random normal deviates. *Annals of Mathematical Statistics*. Vol. 29, 610-611.

Box, G. E. P. – Tiao, G. C. (1973) *Bayesian inference in statistical analysis*. Addison-Wesley publishing company. Reading: Massachusetts.

Clark, P. K. (1973) A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*. Vol. 41:1, 135-155.

Cont, R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance. Vol. 1:2, 223-236.

Cont, R. – Potters, M. – Bouchaud, J-P. (1997) Scaling in stock market data: stable laws and beyond. In Dubrulle – Graner – Sornette (1997) Scale Invariance and Beyond (Proceedings of CNRS Workshop on Scale Invariance, Les houses, 1997). Berlin: Springer.

Danielsson, J. (1994) Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics*. Vol. 64:1-2, 375-401.

Devroye, L. (1986) Non-Uniform Random Variate Generation. Springer-Verlag.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK. *Econometrica.* Vol. 50:2, 987-1007.

Epps, T. W. – Epps, M. L. (1976) The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis. *Econometrica.* Vol. 44, 305-321.

Fama, E. (1965a) The behaviour of stock market prices. *Journal of Business.* Vol. 38, 34-105.

Fama, E. (1965b) Portfolio Analysis in a Stable Paretian Market. *Management Science*. Vol. 11.

Fama, E. (1970) Multiperiod Consumption-Investment Decisions. *American Economic Review.* Vol. 60, 163-174.

Fama, E. – French K. (1988) Permanent and Temporary Components of Stock Prices. *Journal of Political Economy.* Vol. 96:2, 246-273.

Fielitz, B. – Rozell, J. (1983) Stable Distributions and Mixtures of Distributions Hypotheses for Common Stock Returns. *Journal of the American Statistical Association.* Vol. 69, 108-113.

French, K. (1980) Stock Returns and the Weekend Effect. *Journal of Financial Economics.* Vol. 8:1, 55-70.

Gibbons, M. – Hess, P. (1981) Day of the Week Effects and Asset Returns. *Journal Business*. Vol. 54:4, 579-596.

Gillemot, L. – Töyli, J. – Kaski, K. (2000) Time Independent Models of Asset Returns Revisited. *Physica A: Statistical Mechanics and its Applications*. Vol. 282:1-2, 304-324.

Hagerman, R. L. (1978) More evidence on the distribution of security returns. *Journal of Finance.* Vol. 33:4, 1213-1221.

Harlow, W. V. (1991) Asset Allocation in a Downside-Risk Framework. *Financial Analyst Journal.* Vol. 45:5, 28-40.

Harris, J. W. – Stocker, H. (1998) *Handbook of Mathematics and Computational Science.* Springer: New York.

Hsieh, D. A. (1991) Chaos and Nonlinear Dynamics: Application to Financial Markets. *The Journal of Finance*. Vol. 46:5, 1839-1878.

Hsu, D. A. – Miller, R. – Wichern, D. (1974) On the Stable Paretian Behaviour of Stock Market Prices. *Journal of the American Statistical Association*. Vol. 69, 108-113.

Hsu, D. A. (1982) A Bayesian Robust Detection of Shift in the Risk Structure of Stock Market Returns. *Journal of the American Statistical Association.* Vol. 77, 29-39.

Hull, J. C. (1997) *Options, Futures and Other Derivatives.* 3rd edition, Prentice Hall.

Jacquier, E. – Polson, N. G. – Rossi, P. E. (1994) Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics.* Vol. 12:4, 371-389.

Janicki, A. – Weron A. (1991) Can One See α-Stable Variables and Processes? *Statistical Science.* Vol. 9:1, 109-126.

Johnson, M. E. (1979) Computational generation of the exponential power distributions. *Journal of Statistical Computation and Simulation*. Vol. 9, 239-240.

Kanto, A. J. – Kahra, H. A. – Blevins, D. R. – Schadewitz, H. J. (1998) An explanation of the Unusual Behavior of Some Market Model Residuals. *The Finnish Journal of Business Economics*. Vol. 47:3.

Keim, D. B. (1983) Size-Related Anomalies and Stock Return Seasonality: Further empirical evidence. *Journal of Financial Economics.* Vol. 12:1, 13-32.

Kon, S. J. (1984) Models of stock returns – a comparison. *The Journal of Finance*. Vol. 39:1, 147-165.

Kullman, L. – Töyli, J. – Kertesz, J. – Kanto, A. – Kaski, K (1999) Characteristic times in stock market indices. *Physica A: Statistical Mechanics and its Applications*. Vol. 269:1, 98-110.

Lau, A. H-L. – Lau H-S. – Wingender, J. R. (1990) The Distribution of Stock Returns: New Evidence Against the Stable Model. *Journal of Business & Economic Statistics*. Vol. 8:2, 217-224.

Lau, H-S. – Wingender, J. R. (1989) The Analytics of the Intervaling Effect on Skewness and Kurtosis of Stock Returns. *The Financial Review*. Vol. 24:2, 215-233.

Law, A. M. – Kelton, D. W. (1991) Simulation Modeling & Analysis. Second edition. McGraw-Hill: New York.

LeBaron, B. – Arthur, W. B. – Palmer, R. (1999) Time series properties of an artificial stock market. *Journal of Economic Dynamics & Control.* Vol. 23:9-10, 1487-1516.

LeBaron, B. (2000) Agent-based computational finance: Suggested readings and early research. *Journal of Economic Dynamics & Control*. Vol. 24:5-7, 679-702.

Lilliefors, H. W. (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of American Statistical Association.* Vol. 62, 399-402.

Litner, J. (1965) The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*. Vol. 47, 13-37

Lobato, I. N. – Savin, N. E. (1998) Real and Spurious Long-Memory Properties of Stock-Market Data. *Journal of Business & Economic Statistics*. Vol. 16:3, 261-268.

Lucas, A. – Klaassen, P. (1998) Extreme Returns, Downside Risk, and Optimal Asset Allocation. *The Journal of Portfolio Management.* Vol. 25:1, 71-79.

Lux, T. – Marchesi, M. (1999). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature.* Vol. 397:2, 498-500.

Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business.* Vol. 36, 394-419.

Mandelbrot, B. (1999) A Multifractal Walk down Wall Street. *Scientific American*. February 1999, 50-53.

Mantegna, R. N. – Stanley, E. H. (1994) Stochastic process with ultraslow convergence to a Gaussian: truncated Lévy flight. *Physical Review Letters*. Vol. 73:22, 2946-2949.

Mantegna, R. N. – Stanley, E. H. (1995) Scaling behaviour in the dynamics of an economic index. *Nature*. Vol. 376:6, 46-49.

Merton (1976) Option Pricing when underlying Stock Returns are Discontinuous. *Journal of Financial Economics.* Vol. 3, 125-144.

Officer, R. (1972) The distribution of Stock returns. *Journal of the American Statistical Association*. Vol. 67, 807-812.

Oldfield, F. S. Jr. – Rogalski R. J. – Jarrow, R. A. (1977) An autoregressive jump process for common stock returns. *Journal of Financial Economics*. Vol. 5, 389-418.

Osborne, M. F. M. (1959) Brownian Motion in the Stock Market. *Operations Research.* Vol. 7:2, 145-173.

Perry, P. R. (1983) More evidence on the nature of the Distribution of Security Returns. *Journal of Financial and Quantitative Analysis.* Vol. 18:2, 211-221.

Press, W. H. – Teukolsky, S. A. – Vetterling, W. T. – Flannery, B. P. (1992) *Numerical recipes in C the Art of Scientific Computing*. Second Edition. Cambridge university press: USA.

Rachev, S. T. – SenGupta, A. (1993) Laplace-Weibull Mixtures for Modeling Price Changes. *Management Science*. Vol. 39:8, 1029-1038.

RiskMetrics^{tm} – Technical Document (1996) *RiskMetrics^{tm} – Technical Document.* J.P. Morgan/Reuters: New York.

Ross, S. A. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*. Vol. 13, 341-360.

Samuelson, P. A. (1967) Efficient portfolio Selection for Pareto-Levy Investments. *Journal of Financial and Quantitative Analysis*. Vol. 6, 41-49.

Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics.* Vol. 6, 461-464.

Schwert, G. W. (1987) Indices of U.S. stock prices from 1802 to 1987. *Journal of Business*. Vol. 63, 399-426.

Schwert, G. W. (1989) Why does Stock Market Volatility Change over Time? *Journal of Finance*. Vol. 44:5, 1115–1153.

Sharpe, W. F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*. Vol. 19, 425-442.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.

Simons, K. (1996) Value at Risk – New approaches to risk management. *New England Economic Review*. September/October, 3-13.

Tadikamalla, P. R. (1980) Random Sampling From the Exponential Power Distribution. *Journal of the American Statistical Association*. Vol. 75:371, 683-686.

Tauchen, G. E. – Pitts, M. (1983) The price variability – volume relationship on speculative market. *Econometrica*. Vol. 51:2. 485-505.

Taylor, S. J. (1986) *Modelling Financial Time Series*. John Wiley & Sons: Chichester.

Taylor, S. J. (1994) Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*. Vol. 4:2, 183-204.

Tucker, A. L. – Pond, L. (1988) The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes. *The Review of Economics and Statistics*. Vol. 70:4, 638-648.

Tucker, A. L. (1992) A Reexamination of Finite- and Infinite-variance Distributions of Daily Stock Returns. *Journal of Business and Economic Statistics.* Vol. 10:1, 73-81.

Welsh, A. H. (1996) *Aspects of Statistical Inference*. John Wiley & Sons Inc: New York.

Appendix A: Included Papers

I. Kullman, L. – Töyli, J. – Kertesz, J. – Kanto, A. – Kaski, K (1999) **Characteristic times in stock market indices**. Physica A: Statistical Mechanics and its Applications. Vol. 269:1, 98-110.

II. Kullman, L. – Töyli, J. – Kertesz, J. – Kanto, A. – Kaski, K (2000) **Break-Down of Scaling and Convergence to Gaussian**. International Journal of Theoretical and Applied Finance. Vol. 3:3, 371-373.

III. Gillemot, L. – Töyli, J. – Kertesz, J. – Kaski, K. (2000) **Time independent models of asset returns revisited**. Physica A: Statistical Mechanics and its Applications. Vol. 282:1-2, 304-324.

IV. Töyli, J. – Kaski, K. – Kanto, A. (2002**) Variability of Hex Return Distribution**. The Finnish Journal of Business Economics. Vol. 51:1, 64-89.

V. Töyli, J. – Kaski, K. – Kanto, A. (2002) **On the shape of asset return distribution**. Communications in Statistics – Simulation and Computation. Vol. 31:4, 489-521.

VI. Töyli, J. – Sysi-Aho, M. – Kaski, K. (2002) **Models of Asset Returns: Changes of Pattern from Tick by Tick to 30 Days Holding Period**. Submitted for publication in Quantitative Finance.

ISBN 951-22-6253-3 (printed)
ISBN 951-22-6254-1 (PDF)
ISSN 1455-0474