The Skewed Perception of the Distribution of Stock Returns

Mattias Hellstrand*
Stockholm School of Economics

Robin Saliba**
Stockholm School of Economics

Abstract

This thesis investigates the sensibility of the often used simplifications of how stock returns behave in financial models by studying Swedish stock returns using data from 1979 to 2012. The data is tested for normality by using Jarque-Bera test in several steps and exogenous factors are examined for significant impact on the skewness and kurtosis of the stock returns using a non-parametric test developed for this particular purpose. The results clearly show that stock returns are not normally distributed overall. Furthermore, both skewness and kurtosis are significantly affected by exogenous factors. There is a "June effect", where the kurtosis is significantly higher during June than other months. The implication of these findings is that financial models lack in dimension by not incorporating skewness and kurtosis. It is evident that the underlying assumptions of financial models are too simplified in some cases. The thesis brings attention to the need to better understand how skewness and kurtosis varies across stocks and affect investor utility, which clearly is a subject not fully understood as of today.

Acknowledgements

We would like to express our gratitude to our tutor, Professor Francesco Sangiorgi for valuable input and comments throughout the process of writing this thesis. Furthermore, we would like to thank Erik Eklund for helping us acquire the necessary data in order to perform our study. Finally, we would like to thank Jessica Marmlöf Hellstrand for her valuable input in the final stages.

*22150@student.hhs.se

**22149@student.hhs.se

Keywords: Skewness, kurtosis, normal distribution, stock returns, stock pricing

Tutor: Francesco Sangiorgi
Stockholm School of Economics
Department of Finance
Bachelor's Thesis
Spring 2013

Contents

1. INTRODUCTION

In this section we will present the purpose and discuss the relevance of this thesis.

1.1 PURPOSE

The purpose of this thesis is to compare the empirical distribution of Swedish stock returns with the theoretical normal distribution and further investigate whether there are factors that explain deviations in skewness and kurtosis. Particularly we are to study the effects of industry and size as well as differences across time with focus on yearly and monthly deviations. The thesis aims to shed further light on how the distribution of stock returns deviate from the often simplified assumptions in financial models.

1.2 RELEVANCE

In economics and finance, professionals and academics use models to describe the world regularly. There is a trade-off between how well these models actually depict reality and how easy they are to use. To illustrate: picture a simple algorithm where one puts in a parameter and gets back some kind of output number. By making the algorithm more advanced and incorporating more parameters the quality of the output number could increase in the sense that it more accurately describes what the user actually would like to know. The problem though is that the number of parameters increases which makes the algorithm more complex to use. Economic models generally share a common structure in how to tackle this issue. By making assumptions with respect to underlying parameters one can simplify and make models easy to use even for people not well educated in the field. The problem, of course, is that the models lose in accuracy and reliability as the degree of simplicity increases.

When depicting financial markets and forecasting asset returns there are several models used; for example the Capital Asset Pricing Model, Modern Portfolio Theory and the Black-Scholes Model. All of these models make use of assumptions regarding asset returns and their distribution and/or how investors' utility is determined. What if these assumptions are not correct?

We can expect that the simplifications make the models' predictions deviate from the actual outcomes, and there is the possibility that the assumptions made have a vast impact in the decision-making. If this is in fact the case, the financial models used today are neither accurate nor reliable and could possibly produce faulty results.

An assumption that is commonly made in financial models is that stock returns behave more or less symmetrical and that the standard deviation is a complete measure of risk. If this assumption is false, what would the implications be?

Picture an important investment opportunity where a manager is to make a rational economic decision. The models say that the manager should approve since the expected return estimated is higher than the hurdle rate given from the CFO. The hurdle rate is calculated using financial models that assume that the standard deviation is a complete measure of risk, but the investment opportunity have a pay-off distribution which are both skewed and have thick tails. In situations like this the use of financial models can give results that do not fully incorporate all dimensions of risk. Especially, the financial models often neglect the skewness and kurtosis of the return distribution and instead assume that the standard deviation is a complete measure of risk.

We find it relevant to study how the distribution of stock returns actually fit into this and similar assumptions and what factors that possibly could affect the distribution. If there are clear deviations from the assumption of normal distribution of stock returns it could signal a need for further development of new models that better incorporate how stock returns behave beyond the expected return and standard deviation. Furthermore, by understanding what factors that impact the symmetry and shape of stock return distributions we can help to shed new light on necessary alterations that could be beneficial to existing financial models.

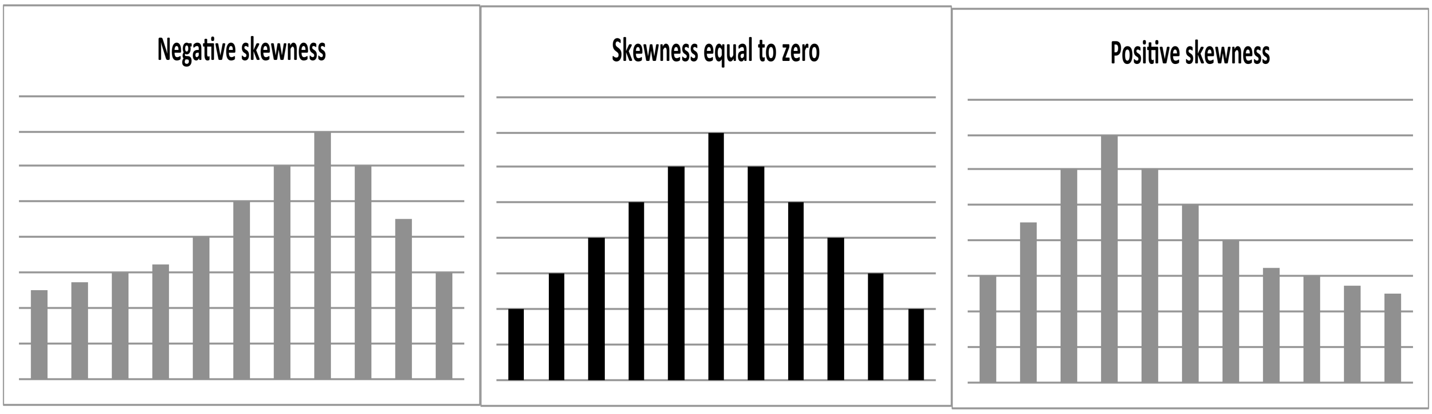
2. THEORY AND PREVIOUS LITERATURE

In this section we present relevant theory that is necessary to understand this field of research. Furthermore, this topic covers the relevant previous findings related to the field of normal distribution. This topic is an essential part of our thesis as our hypothesis and choice of methodology is based on the theory and previous literature presented.

2.1 SKEWNESS AND KURTOSIS

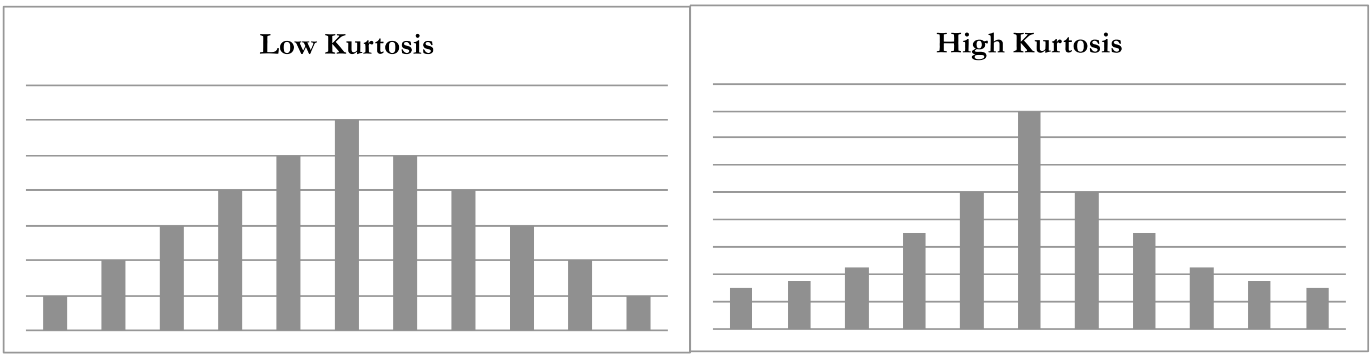
The skewness and kurtosis are characteristics of a probability distribution. More generally the skewness can be described as the parameter defining how a distribution "leans" (in the case of a distribution with only one mode). A positive skewness indicates that the distribution has thicker tails on the right side of the mean and a negative skewness indicates that the left side tails are thicker. This implies that extreme values on the right side of the mean are more likely than the corresponding extreme values on the left side of the mean. A simplified example related to stock returns would be that a negative skew implies that the stocks often yield small positive amounts relative to the expected value but quite often take big negative hits. The types of skewness are illustrated in Figure 2.1.

Figure 2.1
Different Forms of Skewness



The kurtosis is more easily understood as a measure of how thick the tails are (on both sides). A common expression is "excess kurtosis", which refers to the fact that a normal distribution always has a kurtosis of three, and thus the excess kurtosis for a distribution is the kurtosis over and above three. A distribution with positive excess kurtosis is referred to as leptokurtic and a distribution with negative excess kurtosis is referred to as platykurtic. If the kurtosis of a stock increases the probability for extreme events rises in the aspect of stock returns. This characteristic is highly relevant for an investor as a huge negative swing could mean bankruptcy and further social costs, whereas the corresponding positive swing is far away from compensating for that risk. Higher kurtosis implies "thicker tails" and "higher peaks". Figure 2.2 display low and high kurtosis relative to each other.

Figure 2.2
Different Forms of Kurtosis



There are several definitions and variants of skewness (2.1) and kurtosis (2.2), but we will solely focus on the definitions below:

(2.1) $$\frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}[(\mathrm{X} - \mu)^3]}{(\mathrm{E}[(\mathrm{X} - \mu)^2]^{3/2}}$$

(2.2) $$\frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}[(\mathrm{X} - \mu)^4]}{(\mathrm{E}[(\mathrm{X} - \mu)^2])^2}$$

By this definition the skewness is the third standardized moment and kurtosis is defined as the fourth standardized moment of a given distribution. The sample skewness (2.3) and kurtosis (2.4) is usually calculated as below:

(2.3) $$S = \frac{\widehat{\mu_3}}{\widehat{\sigma^3}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^{3/2}}$$

(2.4) $$K = \frac{\widehat{\mu_4}}{\widehat{\sigma^4}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

The above estimators are usually biased estimators for the population parameter. However, for symmetric distributions such as the normal distribution the estimators have the correct expected value and thus are not biased.[1]

[1] Bodie, Z. et al. (2008), p.133-147

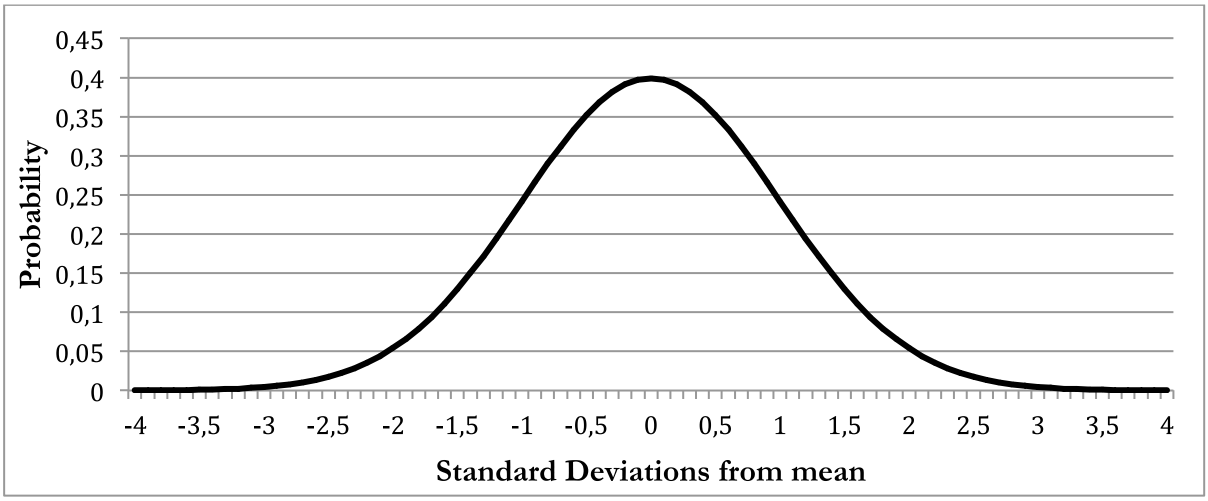
2.2 NORMAL DISTRIBUTION

The normal (or Gaussian) distribution is a two-parameter continuous probability function that is defined by the formula below:

(2.5) $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The two parameters are the mean (also called expectation) and the standard deviation. For a normal distribution the mean is also equal to the median and mode. By changing these two parameters one can obtain different normal distributions, but with all inheriting the basic characteristics of the normal distribution, figure 2.3.

Figure 2.3
Normal distribution with a mean of zero and variance of one



Two important characteristics of the normal distributions are that the skewness is equal to zero and the kurtosis equal to 3. No matter the mean and standard deviation, all normal distributions have these characteristics. This is very important for our study, as we will specifically look at how the skewness and kurtosis vary over time, across companies, industries and other characteristics.[2]

The standard normal distribution is a normal distribution with the mean of zero and variance equal to one. This special case is very useful when comparing different normal distributions as one can transform them to the standard normal distribution and thus make them more comparable as there properties become equal. The transformation formula is given below:

(2.6) $$Z = \frac{X - \mu_x}{\sigma_x} \sim \mathrm{N}(0,1)$$

[2] Bodie, Z. et al. (2008), p.138

2.3 FINANCIAL MODELS

The normality assumption is a simple way of reducing the complexity of how investors value investments and how stock returns behave. When assuming normality the standard deviation becomes the only relevant measure of risk as skewness and kurtosis thereby is constant, and thus the Sharpe Ratio becomes the only relevant measure of investment performance as it relates reward to risk. There are models that do not fully (or at all) rely on the normality assumption. One example of such a model is the Arbitrage Pricing Theory which assumes that there is at least one investor that prefers more return per risk and thus will make all arbitrage opportunities vanish over time.[3]

Apart from the below stated assumption of the models, there are several assumptions that are made, which do not affect our study significantly. Further, the degree of reliance on the normality assumption varies across models; however, there is always some kind of assumption regarding the investors' perception of risk or the distribution of stock returns.

2.3.1 THE MODERN PORTFOLIO THEORY

Henry Markowitz's paper first presented the Modern Portfolio Theory (MPT) in 1952. Given the normality assumption (and the consequence that the standard deviation is a complete measure of risk) and the assumption of risk-averse investors (an investor prefers higher expected return and less risk). The theory gives the optimal risky portfolio and thereafter the optimal allocation between this optimal risky portfolio and the risk-free asset. MPT assumes that the standard deviation captures all riskiness relevant to investors, which is achieved through the normality assumption.[4]

2.3.2. THE CAPITAL ASSET PRICING MODEL

The Capital Asset Pricing Model (CAPM) is a pricing model for assets that was independently developed by several people in a short time period, among them William Sharpe in 1964. The model divides risk into two distinct categories: idiosyncratic and systematic. The idiosyncratic risk is specific to an asset whereas the systematic is tied to the overall market. The asset specific risk can be minimized by diversification and thus does not affect the risk profile of a properly diversified portfolio. As a consequence, only bearing systematic risk should be rewarded with higher expected returns.

---

[3] Ross, Stephen A. (1976)

[4] Markowitz, Harry M. (1952)

The CAPM builds upon the assumption that investors are rational and maximize utility, and that standard deviation is the relevant measure of risk, which in turn assumes a two-parameter distribution (which implies constant skewness and kurtosis across assets as well as time) such as the normal distribution.[5] That is, the model does not assume normality but needs a distribution that satisfies the above-mentioned characteristics, which is in practice very much like the normality assumption.[6]

2.3.3 THE BLACK-SCHOLES MODEL

The Black-Scholes model for derivative instruments builds upon the assumption that stock prices follow a geometric Brownian motion with both drift and volatility constant, which in turn builds upon the normality assumption. The usage of the normal distribution is very apparent in the Black-Scholes model as the formula itself contains cumulative distributions function from the normal distribution.[7]

2.4 PREVIOUS LITERATURE

Historically there have been several studies on the normal distribution of stock returns. In the 1900 Bachelier wrote the thesis Théorie de la Spéculation (Theory of Speculation); it was the first complete development of the theory of stochastic processes in security prices. At first it received little attention, it wasn't until approximately fifty years later that his findings became relevant among economists. Bachelier was the first one to use advanced mathematics in finance. In fact, his findings in 1900 came remarkably close to the Black-Sholes-Merton option pricing formula from 1973.[8,9]

The findings of Bachelier were essential for financial models, and the assumption of normal distribution was later supported by empirical evidence. According both Kendall's and Moore's papers from 1948 and 1962 respectively there was sufficient evidence in support of the fact that the distribution of stock price changes is normally distributed.[10]

The assumption of normal distribution in stock returns wasn't seriously criticized until Mandelbrot's paper from 1963, which provided sufficient evidence disproving past findings of previous research. Instead Mandelbrot presented the stable Paretian distribution, which in the

[5] Bodie, Z. et al. (2008), p.309
[6] Sharpe, William F. (1964)
[7] Black, Fischer and Scholes, Myron (1973)
[8] http://press.princeton.edu/titles/8275.html
[9] Sullivan, Edward J. and Weithers, Timothy M. (1991)
[10] Fama, Eugene F. (1965)

author's opinion represents the distribution of stock returns better than the normal distribution. Paretian and normal distribution are both parts of the theory of stable distributions. Stable distributions that are non-normal are called stable Paretian distributions. Essentially Mandelbrot concluded that the distributions of stock price changes should have "fatter tails" than what the normal distribution represents, i.e. the distribution should contain more relative frequency in the extreme tails than in a normal distribution.[11]

Fama's paper from 1965 discusses and empirically tests the random-walk model of stock behavior with daily prices for each of the thirty stocks of the Dow Jones Industrial Average. The time periods were from the late 1950s to 1962. Fama bases his paper on Mandelbrot's findings and provides further empirical evidence that stable Paretian distribution is a better indicator than the normal distribution for changes in stock prices. Furthermore, Fama also concluded that the assumption of independence on successive price changes is valid after testing the assumption with several techniques (such as a serial correlation model) over different time intervals. Although Fama reaches these conclusions, the author also notes that there are both economic and statistical implications with the stable Paretian distribution in comparison to the Gaussian process. The paper further confirms that there is yet no exact solution for the distribution of stock returns and that there is more work to be done.[12]

Blattberg and Gonedes study the distribution of stock returns, similar to Mandelbrot and Fama the paper studies another distribution as an alternative to the normal distribution. The alternative is the t-distribution, furthermore Blattberg and Gonedes compare the fit of the t-distribution and the stable distribution. The paper from 1974 concludes that the t-distribution better describes the distribution of stock returns than the stable distribution of daily stock returns.[13]

Stanley Kon's paper from 1984 criticizes the t-distribution and suggests that a discrete mixture of normal distributions has a better descriptive validity. Like previous authors Kon wanted to test a new distribution that better explains the significant "fat tails" and the positive skewness of daily stock returns.[14]

Similar to previously mentioned papers, Mills's paper from 1995 use a new model to illustrate the distribution of stock returns on the three London Stock Exchange indices over the period 1986-1992. Mills reaches the same conclusions as previous papers, although for a

[11] Mandelbrot, B. (1963)
[12] Fama, Eugene F. (1965)
[13] Blattberg, Robert and Gonedes, Nicholas (1974)
[14] Kon, Stanley J. (1984)

different geographical region, the stock returns are not normally distributed, the distribution is both skewed and kurtotic.

Furthermore, Mills makes a noteworthy point in his paper; in the event of a significant event in one year the results can be considerably affected. In this case Mills studies the effect on the results by removing the years 1986 due to a deregulation and 1987 due to a market crash from the dataset. Before the removal the distributions were negatively skewed, after the removal the distributions were slightly positively skewed and the kurtosis had declined. Furthermore, Mills's paper note another important yet previously known fact; returns are generally different based on company size that should affect the distributions.[15]

It is evident that researchers have refined the research in this field successively since 1900. The primary method of refining is by testing a new distribution and further study whether it is a better fit than the previous. Mills's paper from 1995 studied the distribution on a new geographical region. However, there are few that have actually focused on what effect certain factors have on the 3rd and 4th standardized moments. Thus, we aim to investigate this further.

[15] Mills, Terence C. (1995)

3. HYPOTHESIS

In this section we present our hypotheses for this thesis. At first we present our main hypothesis, thereafter, we present our sub-hypotheses that are contingent on our main hypothesis being true.

3.1 MAIN HYPOTHESIS AND SUB-HYPOTHESES

There is no reason to deviate from the more recent findings in regards to our hypotheses. The hypotheses can be divided into a main hypothesis and four sub-hypotheses that are relevant if the main hypothesis is true. The main hypothesis is the following:

H1 - "Swedish stock returns are not normally distributed"

The hypothesis is in line with previous findings, for example Mandelbrot's and Fama's papers from 1963 and 1965 respectively and several other papers, yet in conflict with many financial models, which make it both interesting and relevant. To our knowledge there are no papers in recent time that oppose this hypothesis. If the hypothesis is true, then the skewness and/or kurtosis could vary across both stocks and time in contrast to the normal distribution, which has a constant skewness and kurtosis. This implies the possibility that stocks have very different distributions and vary not only in the parameters mean and standard deviation.

If our main hypothesis is true, then there are an unlimited number of factors to examine further for their impact on stock return distributions. We have chosen a set of particular factors that we hypothesize could impact the distribution of stock returns, and particularly the skewness and kurtosis. We define the sub-hypotheses one by one below as well as describe our reasoning behind our choices:

H2.1 - "The industry in which a company operates affects the skewness and/or kurtosis"

Observed returns and variation in returns varies greatly across industries, which implies that there are differences in the expected return and standard deviation of returns. Assuming that our main hypothesis is true, we find it highly likely that skewness and kurtosis vary across industries due to the fact that stock returns vary across industries.[16]

H2.2 - "Skewness and/or kurtosis varies over time"

Many macroeconomic factors such as economic growth, unemployment, interest rates and GDP gap fluctuate over time and have impact on the actual returns and variation in returns. We hypothesize that skewness and kurtosis varies significantly from year to year due to exogenous factors. This is in line with Mills's paper from 1995, which indicated that certain exogenous factors such as regulations and market crises had a significant effect on the distribution of stock returns.

[16] Hou, Kewei and Robinson David T. (2006)

H2.3 - "The relative size of a company impacts the skewness and/or kurtosis"

In academia and financial theory the impact of size on expected returns have been looked upon a vast number of times. Some financial models such as the Fama-French model predict that relative size affects the expected returns of a stock,[17,18] and we hypothesize that the relative size also affects other distribution characteristics such as skewness and/or kurtosis.

H2.4 - "The skewness and/or kurtosis varies over the calendar year"

The world is not a perfect capital market there are several frictions that affect how investors trade, such as taxation and regulation. A commonly known phenomenon is the "January effect", which indicates that there is a stock return seasonality.[19] Furthermore, based on the "January effect" we hypothesize that skewness and/ or kurtosis are affected as well, and thus vary from month to month.

We summarize our hypotheses in table 3.1.

Table 3.1 Summary of Hypotheses

Hypotheses	
H1	**"Swedish stock returns are not normally distributed"**
H2.1	**"The industry in which a company operates affects the skewness and/or kurtosis"**
H2.2	**"Skewness and/or kurtosis varies over time"**
H2.3	**"The relative size of a company impacts the skewness and/or kurtosis"**
H2.4	**"The skewness and/or kurtosis varies over the calendar year"**

[17] Fama, Eugene F. and French, Kenneth R. (1992)
[18] Fama, Eugene F. and French, Kenneth R. (1992)
[19] Keim, Donald B. (1983)

4. DATA DESCRIPTION

This section covers the data used for the analysis throughout this paper. Initially we give a general overview of the data. Secondly we cover our data collection process briefly. Lastly we evaluate the data and identify potential problems.

4.1 DATA OVERVIEW

The data used for this thesis can be divided into two parts, a primary data and a secondary data. The primary dataset was acquired from a database called Finbas that was originally composed by Nasdaq OMX. It consists of monthly stock data from all public companies on the Swedish stock market between the years 1979 and 2012.[20] The reason for this specific time period is purely based on availability. Unfortunately daily stock data was unavailable which would be in line Mandelbrot's and Fama's papers, however, we hope to remedy this by analyzing a significantly longer time period.

Furthermore, based on Mills's paper we felt that it was important to cover a few economic crises as it may give further insight on how the stock return distribution reacts. In the dataset we cover both the financial crisis in the early 1990s in Sweden, the dot-com bubble and the great recession.

In order to ensure more accurate estimations we needed to exclude companies that had too few observations from the analysis. The observation limit chosen was set to 60, since it is monthly data this implies a 5 year time period. Furthermore, we felt that it was necessary to remove time series that did not have a significant beta, i.e. an absolute t-value over two.

Table 4.1 Observations

	Observation	Percentage
Stock Prices	159 466	100%
Normalized Returns	60 404	37,9%

The final sample consists of 60404 observations. As can be observed below, the observations are relatively evenly distributed over the sample period. Further adjustments that were required to the data will be explained and motivated in the methodology section.

[20] http://houseoffinance.se/research-data-center/finbas/

Table 4.2 Yearly distribution of normalized returns

Year	Observations	Percentage
1980	750	1.24%
1981	1030	1.71%
1982	1175	1.95%
1983	1396	2.31%
1984	1661	2.75%
1985	1824	3.02%
1986	1962	3.25%
1987	2052	3.40%
1988	2062	3.41%
1989	2142	3.55%
1990	2021	3.35%
1991	2084	3.45%
1992	1809	2.99%
1993	1830	3.03%
1994	2103	3.48%
1995	2176	3.60%
1996	2278	3.77%
1997	2592	4.29%
1998	2847	4.71%
1999	3040	5.03%
2000	3295	5.45%
2001	3227	5.34%
2002	3101	5.13%
2003	2955	4.89%
2004	3023	5.00%
2005	3037	5.03%
2006	2932	4.85%
Total	60404	100%

The secondary data has been collected from two separate sources; it is basic information that complements the primary data. The collected data are monthly risk-free interest rates and market data for the time period chosen.

4.2 DATA COLLECTION

The primary data that was needed for this paper were information on public companies on the Swedish stock market. This can be obtained from several sources; we chose a rather unique database called Finbas. It was donated to the Stockholm School of Economics in 2011 from

Nasdaq OMX.[21] Erik Eklund from the Swedish House of Finance (SHOF) was an essential part for the acquisition of this dataset; Mr. Eklund helped us with access and the collection of data.

The secondary datasets were collected from the Swedish Riksbank[22] and SIX Financial Information[23] separately. The risk-free rate was acquired from the Swedish Riksbank and the market data from SIX Financial Information.

4.3 DATA EVALUATION

As the data used for this thesis is collected from highly reliable sources we deem the sample data to be accurate in order to draw informing conclusions.

However, with over 150 000 observations originally there were unfortunately companies that had missing values and errors; this has lead to the exclusion for certain companies. One could then argue that this data set is not a complete representation of the Swedish stock market, however, we judge this complication to be negligible.

Furthermore, a complication regarding the data that should be mentioned is that the data is divided into time series, which has the implication that each company could have several time series. This is due to the fact that some companies have several classes of stock (A, B, preference etcetera) and that stocks that have been listed on several Nordic stock exchanges have separated into several series (due to multiple listings). We chose to not "fix" these matters as we argue that the implication for our study is very small and therefore not worth and that the risk of manually making mistakes too high. Furthermore, the fact that some companies have several classes of stock is something that is not in any way "wrong", and therefore cannot be fixed.

Note that we use the terms "time series" and "stock" rather interchangeable throughout this thesis as the two overlap. The reason for this is that our dataset is built on time series, but these time series represent stocks and to make it easier for the reader we use the term stock mostly even though the term time series would be more correct.

As we are dealing with stock data, we can never fully escape the possibility of a slight survivorship bias, thus our data can be biased in its selection. Thus, failed companies are not included in the dataset, which can lead to more optimistic results as the dataset only includes survivors.[24]

---

21 http://houseoffinance.se/research-data-center/finbas/
22 http://www.riksbank.se/sv/Riksbanken/Forskning/Historisk-monetar-statistik/Rantor-och-aktieavkastningar/
23 http://www.six-telekurs.se/sv/se/Produkter-och-tjanster/Nordiska-innehallsprodukter-/Index/SIX-Index/
24 Banz, Rolf W. and Breen William J. (1986)

5. METHODOLOGY

Throughout this section we will describe the methodology for this thesis, which we use to test our main hypothesis and the eventual sub-hypotheses. As our methodology uses several methods we have divided into separate parts to make it more pedagogical. We will present our methodology in three steps and thereby create an understanding of each step of the methodology. The steps are as follows:

Step 1: *Normalization of stock returns*

Step 2: *Test of normality using the Jarque-Bera test*

Step 3: *Test of impact from other factors using non-parametric tests*

5.1 NORMALIZATION OF STOCK RETURNS

The first step is to obtain normalized returns, which are essential for both testing for normality and to study the distribution of several stocks in aggregate. That is, the normalized returns will be used in both parts of the study but in different ways.

To test our main hypothesis we are using the fact that if all stock returns are normally distributed then all observations should follow the below distribution:

(5.1)
$$R_{i,t} \sim N[E(R_{i,t}); Var(R_{i,t})]$$

We can thereby use the following normalization to obtain normalized returns that follow a standard normal distribution (given that the stocks are normally distributed) with:

(5.2)
$$\frac{[R_{i,t} - E(R_{i,t})]}{SD(R_{i,t})} \sim \mathrm{N}(0,1)$$

This is an essential process as it allows us to transform each and every stock return to a homogenous distribution, namely the standard normal distribution with mean of zero and variance of one. Thereafter we can use the normalized returns to test for normality.

Even if we are to conclude that the stocks (or some sub-sample of them) are not normally distributed we can use these normalized returns when aggregating stocks into groups when testing our sub-hypotheses. No matter the actual distribution the stocks will then have the same mean and standard deviation and thus we are able to study the distribution patterns.

But first we have to obtain all the necessary data: actual returns, expected returns as well as the standard deviation of the expected returns. It should be mentioned that what we refer to as "returns" hereafter are the excess returns, which is equal to the nominal return minus the risk-free rate.

5.1.1 ACTUAL RETURNS: DATASET

The first piece is the actual (excess) returns, which we calculate from the stock price data. There are two widely used ways of calculating returns: arithmetic returns (5.3) and logarithmic/ continuously compounded returns (5.4):

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} - r_f \tag{5.3}$$

$$R_t = \ln(S_t) - \ln(S_{t-1}) - r_f \tag{5.4}$$

We chose to use the arithmetic returns (5.3), as the method is widely used in financial reporting and media.

5.1.2 EXPECTED RETURNS: THE CAPM

In contrast to the actual returns the expected returns are unobservable and therefore have to be estimated in some way. There are several ways, all with advantages and disadvantages. We have chosen to use the Capital Asset Pricing Model (CAPM) to obtain estimates of the expected returns, which is one of the mostly used and accepted ways of estimating expected returns, primarily due to its simplicity.

The CAPM is a model that divides risk into idiosyncratic risk (specific to a firm) and market risk (affecting the whole market). An investor can diversify its holdings and thus eliminate (or at least minimize) the idiosyncratic risk of the portfolio. Hence an investor should not be rewarded for bearing firm specific risk. Instead, an investor should only be paid for holding risk associated with the market. The resulting equation below states that the expected return of a stock is determined by the risk-free rate, the stock's sensitivity to the market portfolio (and thus the market risk) and the expected excess return of the market portfolio (which is the market portfolio's expected return minus the risk-free rate): [25]

$$E(R_{i,t}) = r_f + \beta_{i,t}[E(R_{m,t}) - r_f] \tag{5.5}$$

Since all right hand side variables are unknown and unobservable, we have to either estimate them or use a proxy for them. We will hereafter walk through the choices and why we have made them.

[25] Bodie, Z. et al. (2008), p.295-300

In theory the risk-free rate is the return of an investment one can yield by investing in an asset that pays the same payoff in every possible state of the world. In practice there are no true risk-free rate and to complicate the matter even more one has to take into consideration that even if it would be possible to construct a risk-free asset, that asset would only be risk-free in the specific currency in which the payoffs are defined.[26]

We have chosen to use data from the Swedish central bank, Riksbanken, which approximates the short-term risk-free rate as Riksbanken's diskonto for the period 1856 to 1982 and the Swedish 30-day Treasury Bill thereafter. This dataset is put together by Daniel Waldenström and contains an approximation of Swedish risk-free rate for each month (on a yearly basis).[27]

As we are studying monthly stock returns we had to adjust the data from yearly rates to monthly, which is done by taking the 12[th] square root of one plus the rate (in decimal form) and then subtracting one.

$$R_{monthly} = (1 + R_{yearly})^{\frac{1}{12}} - 1 \tag{5.5}$$

The rationale for using this dataset is that we are studying Swedish stocks and thus the returns are in Swedish currency for each stock and each point in time and that the rate is monthly and backed by the Swedish government (which arguably is close to risk-free), thus being the closest to risk-free rate in Swedish currency.

When using models such as the Capital Asset Pricing Model one has to decide what proxy to use for the market portfolio as the portfolio as well as its return is not observable. Further, in theory the portfolio should contain value-weighted fractions of all existing assets, from human capital to stocks to complex derivatives. This is of course not obtainable in practice since all assets are not publically traded (or even possible to trade). Thus one has to use a proxy instead of the actual market portfolio.

As we are studying Swedish stocks (and consequently Swedish currency returns) we have used the SIXRX index which pictures the return of the Swedish Stock Exchange with dividends reinvested to give a more correct picture of the returns compared to SIXGX, which does not adjust for dividends. As we have used monthly stock prices adjusted for dividends as well, the choice of SIXRX is rather trivial. Another argument for using the SIXRX is that all the stocks

[26] Bodie, Z. et al. (2008), p.318
[27] http://www.riksbank.se/sv/Riksbanken/Forskning/Historisk-monetar-statistik/Rantor-och-aktieavkastningar/

that we study are included in the index, thus better reflecting the property of the theoretical market portfolio containing all possible assets.[28]

The core of the CAPM is the beta, the sensitivity to the return of the market portfolio. Like the other parameters the beta is unobservable and must be either estimated or replaced by some proxy. The most common way is to estimate the beta by regressing actual stock returns on some proxy for the market portfolio return and risk-free rate.

$$R_{i,t} - r_f = \beta[E(R_{m,t}) - r_f] \tag{5.6}$$

The model then uses assumptions regarding ordinary-least square regressions.[29]

In theory the beta for a given company will vary, as capital structure, strategic and operational decisions influence the sensitivity to the overall market and state of the economy. We have simplified the matter by assuming that the beta is constant over time. There is a trade-off between observations per regression and flexibility and we chose to maximize the amount of observations arguing that most companies' betas are fairly constant over time as long-term strategic and financial goals are rather constant. We have chosen to exclude all time series for which the absolute value of the t-value is less or equal to two.

As proxy for the expected market portfolio return we have used the actual market portfolio return and for the risk-free rate we have used the same method as above. These are both common methods when conducting studies like this, and arguably rather adequate for this particular methodology.

Then, we regressed the excess returns from each time series on the actual market portfolio excess return and obtained the betas. From this regression we also estimated the alphas, which is the extra return over what CAPM predicts. Since we want to estimate the expected return we do not use the alphas, as the expected alpha in every time should be zero if investors are close to rational. If that was not true, then some stocks would be expected to yield higher than their fair return, and there would be an arbitrage opportunity – which a rational investor would likely use to its advantage.[30]

From these components we consequently calculated the expected returns for each time period and each time series using the CAPM formula.

---

[28] http://www.six-telekurs.se/sv/se/Produkter-och-tjanster/Nordiska-innehallsprodukter-/Index/SIX-Index/
[29] Woolridge, Jeffrey M. (2008), p.801-807
[30] Bodie, Z. et al. (2008), p.296-297

5.1.3 STANDARD DEVIATION

The standard deviation of the expected value is also somewhat complicated, as it cannot be observed. We chose between two options: use the sample standard deviation from the time series or the sample standard deviation from the predicted expected returns obtained from CAPM.

The advantages with using the first option is that the values are not estimated, the values are actual observations, in contrast to being estimated and thus of higher degree of uncertainty. The disadvantage is that the observations and their difference from the expected values are what we study, and thus the use of their sample standard deviation might be inappropriate. On the other hand, the estimated values do not contain the idiosyncratic variance as the values have been estimated from the CAPM and thus only incorporate the variation derived from the market portfolio. We argue that the first option is better since the second one do not incorporate the idiosyncratic variation and therefore we chose the sample standard deviation from the actual returns.

After obtaining the actual returns, the expected returns and lastly the standard deviation for each time series we calculate the normalized returns. The normalized return will be used in the subsequent parts of the methodology, firstly as the input for the Jarque-Bera test of normality.

5.2 THE JARQUE-BERA TEST OF NORMAILTY[31]

The Jarque-Bera test is a goodness-of-fit test that examines whether the skewness and kurtosis in a sample of data matches the normal distribution. We have chosen this test since it is recommended by several sources.[32] The test uses a joint null hypothesis that the skewness and excess kurtosis is zero, which is what the normal distribution has. That is, given our hypothesis that the stock returns are not normally distributed we hypothesize that we will reject the null hypothesis (which is that the skewness and kurtosis jointly match the normal distribution). The test statistic (often referred to as *JB*) is:

$$JB = \frac{n}{6}[S^2 + \frac{1}{4}(K-3)^2] \tag{5.7}$$

Where n is sample size, S is skewness and K is kurtosis (and since the normal distribution has a kurtosis of three the K−3 equals the excess kurtosis). Some argue that the statistic should be adjusted when used in combination with regressions, but given our sample size this would not affect the results substantially. The test statistic is asymptotically chi-squared distributed with two

[31] Jarque, Carlos M. And Bera Anil K. (1987)
[32] Gujarati, Damodar N. (2002), p. 147-148

degrees of freedom if the sample comes from a normal distribution. The sample skewness, S and kurtosis, K are obtained with equations 5.8 and 5.9 below.

(5.8)
$$S = \frac{\widehat{\mu_3}}{\widehat{\sigma^3}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^{3/2}}$$

(5.9)
$$K = \frac{\widehat{\mu_4}}{\widehat{\sigma^4}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

The test statistic is compared with the critical value and the null hypothesis is either accepted (and the sample is therefore assumed to come from a specific normal distribution) or rejected (and thus assumed not from a normal distribution). We have chosen a significance level of 0.1% for this test and the critical value therefore is 13.82. If we obtain a value at least as high as 13.82 we will therefore reject the null hypothesis and conclude that the tested sample is not from a normal distribution. We argue that based on previous literature and the nature of the test that a rather low significance level (and thus high confidence when rejecting) is appropriate.

If we reject the null hypothesis of the test we can conclude that the data sample does not come from a normal distribution. This implies that at least some sub-sample is not normally distributed (and thus makes the total sample not normally distributed). We cannot however conclude that no stock has a normally distributed return. We will therefore perform two more steps of Jarque-Bera tests: one on industry level and one on individual stock level. By concluding that no industry and thereafter that no stock individually has a normally distributed return we can conclude that stock returns are not normally distributed overall. However, if we fail to reject the null hypothesis in any of these two steps (and particularly the last one) we are to accept that we cannot fully conclude our main hypothesis.

There are several risks and problems with this method, below we present the potential risks with using this method and how to mitigate them.

- Not representative sample: as we do not have a complete sample of each and every stock we cannot be sure that any conclusion is more than partially correct. We have used a sample as big as possible to minimize this risk.
- Risk of accepting a distribution that is similar to the normal distribution: as the test only testes whether the skewness and kurtosis jointly match the normal distribution one could accept a different distribution that in these characteristics matches the normal distribution.

- Risk of accepting the null hypothesis when it is false: there is always a risk of having to accept a faulty null hypothesis and we have used a low significance level to minimize this risk.
- Risk of rejecting the null hypothesis when it is true: the Jarque-Bera test is rather sensitive to sample size and we have therefore minimized the risk by using samples that are as big as possible.

5.3 NON-PARAMETRIC TEST

The second step of our methodology is to test whether our four sub-hypotheses are true or not:

H2.1 - "The industry in which a company operates affects the skewness and/or kurtosis"

H2.2 - "Skewness and/or kurtosis varies over time"

H2.3 - "The relative size of a company impacts the skewness and/or kurtosis"

H2.4 - "The skewness and/or kurtosis varies over the calendar year"

There is no simple and clear-cut way to test these hypotheses as regular parametric methods either need assumptions regarding the distributions or knowledge of the actual distribution; neither which fits our purposes. We therefore have developed a non-parametric test to use to test our hypotheses.

The idea behind the test is that if there is no difference between, for example, industries' skewness and/or kurtosis then the probability that a given industry has higher skewness and/or kurtosis in a given time period is equal for all industries. The same is clearly the case for years and months as well, if the skewness and/or kurtosis are the same the probability should be equal.

We will use the normalized returns from prior steps as it gives us the opportunity to group stocks into different groups without altering the expected return or standard deviation. To show that this works even without the normality assumption think of a stock that is distributed by the unknown distribution X with an expected value of E(R) and variance VAR(R):

$$R_{i,t} \sim X[E(R_{i,t}); Var(R_{i,t})] \tag{5.10}$$

If we subtract the expected return on each side and divide by the standard deviation we get following:

$$\frac{[R_{i,t} - E(R_{i,t})]}{SD(R_{i,t})} \sim \mathrm{X}(0,1) \tag{5.11}$$

Clearly, we can use the same normalization to obtain aggregates of variables with the same expected return and the same standard deviation (and thus variance). The difference compared to when assuming normality is that we now do not make any assumption regarding the different distributions of stock returns, only that the distributions have the same expected return and variation.

We will hereafter call the variable that we are studying for "the impacting variable" and the variable that we use to obtain several observations "the segment variable". The logic behind the terminology is that the impacting variable is what we are testing for impact on the skewness and kurtosis, whereas the segment variable is the variable that is used to segment the impacting factor to obtain more than one observation. It is important to note that the impacting variable contain several values as well, and the total amount of observations used is equal to the number of values of the impacting variable times the number of values of the segment variable.

By obtaining an observation for each value of the impacting variable for each and every value of the segment variable we can rank the impacting variables' values for each segment from highest to lowest in both skewness and kurtosis. These ranks can thereafter be summed to a total sum of ranks for each value of the impacting variable. If there are no differences in the underlying skewness and/or kurtosis all the impacting variable's values' sum of ranks should be distributed in the same way. If we can obtain the distribution of the impacting variable's values under the assumption that there are no differences we can test whether the assumption is true or not by calculating how extreme an obtained value is given that it comes from the distribution.

The underlying distribution under the assumption that there are no differences (hereafter called *the null distribution*) is obtainable, but as the number of values of the impacting variable and the segment variable increases the complexity of the distribution increases rapidly. We will therefore program functions in Excel and Visual Basic that makes it much easier to simulate these distributions and thus obtain approximations of the distributions. By simulating enough series we can minimize the risk of using a faulty estimate of the distribution, and our simulated distribution will in all significant ways be close enough to the null distribution.

After obtaining the sum of ranks for each value of the impacting variable, we can calculate the absolute difference between each sum of ranks and the expected value under the null distribution. We will thereafter compare the value with the simulated distribution to obtain a p-value. We will use the absolute deviations from the simulated distribution, as it simplifies the actual testing (not having to test both tails of the distribution). We can use the absolute deviations only if the distributions' skewness is zero, and we will therefore only accept simulated distributions with an absolute skewness of less than 0.1.

It is only necessary to prove that one of the impacting variable's values is significantly different from what it should be under the null distribution to disprove the null hypothesis that the skewness and/or kurtosis are the same.

The test is done for skewness and kurtosis separately and it is fully possible that one of them is in fact impacted by the impacting variable where the other is not. A short step-by-step summary is given in table 5.1:

Table 5.1 Summary of process

Non Parametric Test	
Step 1	Simulate null distribution under the assumption that there are not differences across the impacting variable's values
Step 2	Obtain values for skewness/kurtosis for each intersection of the impacting variable and segment variable
Step 3	Sum the ranks over segments for each value of the impacting variable
Step 4	Compute the absolute deviations from the expected value
Step 5	Compare to the null distribution of absolute deviations

As our test is designed to test the null hypothesis that a factor does not have significant impact our thesis' hypotheses (H2.1, H2.2, H2.3 and H2.4) will be that we will reject the null hypothesis of the actual test.

To fully explain the method in practice we have included a short example, which is fairly easy both in computation and understanding:

Assume that Mattias and Robin are to compete in running and that we want to test whether Mattias is significantly faster than Robin. If the null hypothesis is that they are equally probable to win a race then the chance to win a given race is equal to 50% for both candidates. After performing 10 experiments (or races in this case) Mattias has won 9 out of 10.
The results, displayed as first and second place can be seen in the below table:

Table 5.2
Placements for each race and the sum of placements

Name/Race	1	2	3	4	5	6	7	8	9	10	Sum
Mattias	2	1	1	1	1	1	1	1	1	1	11
Robin	1	2	2	2	2	2	2	2	2	2	19

The next question is: given that they are both equally likely to win any given race, what is the probability to obtain the above results.

The mean value of sum is 15 under the null hypothesis as the mean value per race is 1.5 (the weighted average of 1 and 2) and the number of races is 10. Each race has two possible outcomes and there are 10 races, thus the combinations are 2^10 = 1024. There are two possible outcomes that give the sum of 10, when either Mattias or Robin wins all the races. There are 20 different outcomes that give the sum of 11: when Mattias win 9 out of 10 races (which give 10 possible combinations) and when Robin wins 9 out of 10 races (which add 10 more possible combinations). In total there are 22 different combinations that give at least as extreme outcome as the actual result. The probability of each combination is the same and thus the probability is 22/1024 = 2.15%. On a significance level of 5% the rule to reject the null hypothesis is to reject whenever the probability to obtain at least as extreme result is lower than the significance level. Therefore we would conclude that Mattias is more probable to win a given race given the hypothesis.

The three sub-hypotheses that we are to test will be further explained in detail below.

Table 5.3

Summary of hypotheses impacting and segmenting variable

Hypothesis	Impacting variable	Segment variable
Industry impact	Industry (15 industries)	Year (1989-2006)
Time impact	Year (1981-2006)	Industry (14 industries)
Size impact	Size (three cohorts)	Year (1994-2006)
Calendar year impact	Month (Jan-Dec)	Year (1981-2006)

The different values that each variable contain for each hypothesis is listed in the Appendix, but we will give a short description of the three setups.

When studying the effect of industry on skewness and kurtosis we will use the sample skewness/kurtosis for each industry in our sample in each year and rank the industries per year. We will only use data for the years 1989-2006 since our data did not include any observations for some industries before 1989 ("Services" only has data from 1989 and "Transport" only has data from 1981).

When looking at the yearly effect we will transpose the data and instead rank the years for each industry. So, for each industry we will rank the years from highest to lowest skewness/kurtosis. We have excluded the industry "Services" as it was not included in the data for all the years. We argue that the incremental value from getting more years (about a decade) is higher than the lost value from excluding the industry. The time span we have used is 1981-2006 and we have excluded the year 1980 as the industry "Transport" did not have any data for that year.

When studying the effect of size we will divide the stocks into three cohorts (small, medium and large) based on the average market capitalization during each year. A stock that is included in one cohort can therefore change cohort in the next year. A third of the total stocks will be allocated to each cohort so stocks moving between cohorts between years do not affect the result in any way. We will study the period of 1994-2006 as we could not get data for more years.

Lastly, when studying the monthly impact we divide the data into cohorts on a monthly basis as well as a yearly basis, thus getting a value for each month and year. We will thereafter rank the months for each year. We have excluded the year 1980 as we did not have data for the full year, missing observations; January and February.

We will throughout this method use a significance level of 5%. If the obtained p-value is 5% or lower we will reject the null hypothesis and if the obtained p-value is higher we will accept the null hypothesis.

There are some flaws and risks with every method, below we present the major risks and what we will do to mitigate them:

- Biased estimator for skewness and kurtosis: as pointed out in section 2 the estimator of the third and fourth standardized moment is usually biased. Our way of reducing the impact of this is to use very large samples, which minimizes the problem effectively.
- Biased simulated null distribution: naturally a simulation is not perfect, but by using a large amount of simulated series we minimize the risk of using a simulated null distribution that deviates much from the theoretical distribution. We will also test the simulated distribution and reject and do another simulation if the mean value deviates more than 0.1% from the theoretical, as well if the skewness is more than 0.1 in absolute terms.
- Accepting the null hypothesis when it is false and rejecting the null hypothesis when it is true: there is a trade-off between these two problems and we have chosen to use a significance level of 5% as a try to mitigate both problems at the same time.

6. RESULTS & EMPIRICAL ANALYSIS

We have divided the results and analysis into three sections to separate the different methods used and to more easily relate to our hypothesis structure. First, we present the results from the Jarque-Bera tests and our conclusions thereof regarding our main hypothesis, then we move onto the results and analysis from the non-parametric testing and the related sub-hypothesizes. Lastly we look further into some interesting topics that have emerged during our study.

6.1 TEST OF NORMALITY USING THE JARQUE-BERA TEST

The first step in testing our main hypothesis is to test that whether all Swedish stocks' in aggregate have normally distributed returns. If all stocks' returns are normally distributed then the aggregate of normalized returns should pass the Jarque-Bera test. If the sample fails the test we can conclude that not all Swedish stocks' returns are normally distributed. We cannot, however, conclude that none are normally distributed; only that at least some are not. To conduct the test we have calculated the sample skewness and sample kurtosis as well as the total number of observations from the aggregate normalized returns. The below table 6.1 gives the sample numbers as well as the calculated Jarque-Bera statistic:

Table 6.1
Statistics used for Jarque-Bera test

Statistics	Values
Observations	60 404
Skewness	1.0362
Kurtosis	10.1134
Jarque-Bera	138 164.7

If the sample is from a normal distribution the test statistic should be Chi-square distributed with two degrees of freedom.[33] For a significance level of 0.1% the critical level of the statistic is 13.82 (see Appendix for chi-square distribution), which is clearly much lower than our obtained statistic. We can therefore with confidence conclude that not all Swedish stock returns are normally distributed.

It is worth noting that the skewness is positive and given the huge sample size of more than 60 000 observations we can therefore suspect that stock returns are positively skewed on average.

To further investigate whether Swedish stock returns are normally distributed we are to divide the sample into categories based on the industries in which the companies operate in and

[33] Jarque, Carlos M. And Bera Anil K. (1987)

thereafter each and every industry is tested for normality. The results can be seen in table 6.2 below:

Table 6.2
Statistics used for Jarque-Bera test grouped by industry

Industry	Sample Size	Skewness	Kurtosis	Jarque-Bera
Capital Goods	6435	0.897816	11.00492	18046
Chemicals & Energy	1502	0.507354	5.751775	538
Financial Services	2713	2.510025	27.10498	68532
Forest Industry	2111	0.423129	6.244738	989
Health Care & Pharmaceutical	3431	1.404231	9.937503	8008
Household Goods	3032	0.716512	9.320319	5306
Investment & Holding	4569	1.046467	15.48911	30528
Consulting Services	1711	0.658283	6.212103	859
Media	801	1.208876	7.770833	955
Real Estate	4377	1.003497	10.13117	10009
Retail	1047	0.394863	7.035873	738
Services	1042	0.824604	6.223679	569
Technology	3028	1.531334	11.7211	10779
Transport	1896	0.668497	6.213282	957
Vehicles & Machines	2016	1.305993	13.47382	9788

From the test statistics one can see that there are clear deviations across industries, ranging from more than 68 000 for financial companies such as banks down to 538 for chemical & energy. It is worth nothing that the Jarque-Bera test by its definition gets very sensitive as the sample size increases, and therefore gets very powerful which results in generally high values of the test statistic for the industries with bigger samples. None the less, 538 is far away from passing the test on a 0.1% significance level and we can conclude that even though there are clear deviations across industries when it comes to skewness, kurtosis and distribution none of the industries are close to pass the normality test. We therefore conclude that no industry as a group has normally distributed returns.

We can further see that there are noticeable differences across industries both regarding skewness and kurtosis, and that skewness for all industries is positive (as for the market overall). We will look into the differences across industries in more detail in the next part, as we will only cover the normality test in this part.

To fully test for normality we also conducted the Jarque-Bera test on each and every stock in our sample. Of the 266 times series we could on a 0.1% significance level disprove normality for 202 (75.95%). It should be noted that this result not necessarily mean that the other 64 stocks are normally distributed, only that we did not have enough evidence to reject the null hypothesis of normality on our rather powerful significance level. Had we lowered the bar by increasing the significance level we would have successfully rejected a lot more of the stocks. In our appendix we have included a list with the stocks that passed the test and their respective industry as well as their Jarque-Bera test statistic. We conclude that at least ¾ of Swedish stocks do not have normally distributed returns. We summarize the findings from the first part of the study below:

- **"Not all Swedish stock returns are normally distributed"**
- **"No industry as a group has normally distributed returns"**
- **"At least ¾ of Swedish stocks' returns are not normally distributed"**

6.2. IMPACT OF EXOGENOUS FACTORS ON SKEWNESS AND KURTOSIS

6.2.1 INDUSTRY IMPACT ON SKEWNESS AND KURTOSIS

We tested whether the industry in which a company operates in have significant impact on either the skewness or kurtosis (or both). We therefore chose to use industry as impacting variable and year as segment variable. With 15 industries and 18 years (1989-2006) there are a total of 270 sample skewness and sample kurtosis respectively. We have included these numbers as well as the rankings in the appendix and are only to present the final results here. Our simulated distribution has a mean value of 143.9996 (as the mean value should be 144, equal to 8 times 18) and skewness of 0.0020239 and thus passes our test of appropriateness. Summary statistics can be seen in the appendix. The below table 6.3 shows the results for skewness:

Table 6.3
Sum of ranks, deviation form average, absolute deviation
from average and p-value for each industry with respect to skewness

Industry	SoR	Deviation	AbsDev	P-value
Capital Goods	144	0	0	100,00%
Chemicals & Energy	194	50	50	8,32%
Financial Services	127	-17	17	99,82%
Forest Industry	159	15	15	99,97%
Health Care & Pharmaceutical	102	-42	42	27,84%
Household Goods	168	24	24	95,92%
Investment & Holding	122	-22	22	98,05%
Consulting Services	141	-3	3	100,00%
Media	151	7	7	100,00%
Real Estate	143	-1	1	100,00%
Retail	150	6	6	100,00%
Services	142	-2	2	100,00%
Technology	102	-42	42	27,84%
Transport	160	16	16	99,89%
Vehicles & Machines	155	11	11	100,00%

The lowest p-value (which is the relevant one for the test) is that for Chemicals & Energy, which is 8.32%. The value is clearly not low enough to reject the null hypothesis at our predetermined level of 5%. Therefore we cannot conclude any significant impact of industry on skewness.

In table 6.4 one can see the corresponding data for kurtosis, with sum of ranks, deviation from average, absolute deviation as well as p-value:

Table 6.4
Sum of ranks, deviation form average, absolute deviation
from average and p-value for each industry with respect to kurtosis

Industry	**SoR**	**Deviation**	**AbsDev**	**P-value**
Capital Goods	126	-18	18	99,67%
Chemicals & Energy	206	62	62	0,70%
Financial Services	132	-12	12	100,00%
Forest Industry	173	29	29	83,66%
Health Care & Pharmaceutical	120	-24	24	95,92%
Household Goods	120	-24	24	95,92%
Investment & Holding	77	-67	67	0,16%
Consulting Services	158	14	14	99,98%
Media	193	49	49	9,83%
Real Estate	104	-40	40	35,59%
Retail	195	51	51	7,10%
Services	165	21	21	98,72%
Technology	96	-48	48	11,67%
Transport	149	5	5	100,00%
Vehicles & Machines	146	2	2	100,00%

From these results we can clearly conclude that kurtosis varies across industries, as the p-value of the Investment & Holding is as low as 0.16%, which passes the test at a significance level of 5%.

6.2.2 SKEWNESS AND KURTOSIS OVER TIME

When we test whether skewness and/or kurtosis varies over time the impacting variable is year and the segment variable industry. We used the time period 1981-2006 which consists of 26 years and used 14 industries to segment the sample. We excluded the industry "Services" as it did not contain any data for the first several years.

Our simulated distribution had a mean of 189 (clearly passing the test as the mean should be 189, the number of industries times the average rank per industry) and a skewness of 0.0042 and thus being appropriate. Descriptive statistics can be found in the appendix. Table 6.5 shows the results of our findings.

Table 6.5
Sum of ranks, deviation form average, absolute deviation from average and p-value for each year with respect to skewness

Year	SoR	Deviation	AbsDev	P-value
1981	161	-28	28	99.99%
1982	171	-18	18	100%
1983	194	5	5	100%
1984	208	19	19	100%
1985	215	26	26	100%
1986	159	-30	30	99.95%
1987	236	47	47	93.05%
1988	175	-14	14	100%
1989	177	-12	12	100%
1990	312	123	123	0%
1991	263	74	74	18.41%
1992	213	24	24	100%
1993	91	-98	98	0.77%
1994	117	-72	72	22.48%
1995	210	21	21	100%
1996	148	-41	41	98.67%
1997	139	-50	50	87.99%
1998	193	4	4	100%
1999	166	-23	23	100%
2000	159	-30	30	99.95%
2001	213	24	24	100%
2002	222	33	33	99.88%
2003	183	-6	6	100%
2004	173	-16	16	100%
2005	202	13	13	100%
2006	214	25	25	100%

We can see that the value for 1990 is further away from the average than any of the values from our simulated data sample, hence being very significant (having a p-value of 0%). The year 1993 also have a significant deviation, which further enforces our conclusion that skewness varies over the years. Table 6.6 shows the same test for kurtosis:

Table 6.6
Sum of ranks, deviation form average, absolute deviation from average and p-value for each year with respect to kurtosis

Year	SoR	Deviation	AbsDev	P-value
1981	247	58	58	65.02%
1982	218	29	29	99.99%
1983	259	70	70	27.14%
1984	214	25	25	100.0%
1985	256	67	67	35.13%
1986	206	17	17	100.0%
1987	234	45	45	95.49%
1988	182	-7	7	100.0%
1989	180	-9	9	100.0%
1990	184	-5	5	100.0%
1991	156	-33	33	99.88%
1992	142	-47	47	93.05%
1993	154	-35	35	99.77%
1994	154	-35	35	99.77%
1995	223	34	34	99.86%
1996	157	-32	32	99.92%
1997	162	-27	27	100.0%
1998	161	-28	28	99.99%
1999	182	-7	7	100.0%
2000	111	-78	78	12.02%
2001	220	31	31	99.94%
2002	190	1	1	100.0%
2003	186	-3	3	100.0%
2004	168	-21	21	100.0%
2005	189	0	0	100.0%
2006	179	-10	10	100.0%

None of the years above show low enough p-value to disprove the hypothesis that kurtosis is constant over time. We cannot show any significant variation in kurtosis over the years.

6.2.3 SKEWNESS AND KURTOSIS OVER MONTHS

To test whether the skewness and/or kurtosis varies over the calendar year we have conducted a test using the same method as before. If we can prove that some months have significantly higher/lower skewness/kurtosis we can conclude that there is variation across months. The impacting variable thus is month and the segment variable is year. Our simulated distribution has a mean of 169 and skewness of 0.0060, which both is in the acceptable range. In table 6.7 one can find our results with respect to skewness:

Table 6.7
Sum of ranks, deviation form average, absolute deviation
from average and p-value for each month with respect to skewness

Month	SoR	Deviation	AbsDev	P-value
January	143	-26	26	83.1%
February	172	3	3	100%
Mars	147	-22	22	93.2%
April	160	-9	9	100%
May	146	-23	23	91.3%
June	161	-8	8	100%
July	177	8	8	100%
August	197	28	28	75.8%
September	177	8	8	100%
October	199	30	30	67.1%
November	192	23	23	91.3%
December	157	-12	12	100%

Clearly there are no significant differences across months. The lowest p-value is 67.1% which is very far from being a significant deviation given our significance level of 5%. We can therefore conclude that skewness does not vary over the calendar year.

The results from studying the kurtosis are shown in table 6.8:

Table 6.8
Sum of ranks, deviation form average, absolute deviation
from average and p-value for each month with respect to kurtosis

Month	SoR	Deviation	AbsDev	P-value
January	147	-22	22	93.2%
February	201	32	32	57.1%
Mars	145	-24	24	89.1%
April	194	25	25	86.9%
May	145	-24	24	89.1%
June	116	-53	53	2.1%
July	188	19	19	98.1%
August	194	25	25	86.9%
September	177	8	8	100%
October	158	-11	11	100%
November	199	30	30	67.1%
December	164	-5	5	100%

Here we can see that June has a p-value of 2.1%, which on our significance level of 5% passes the test. Since the deviation is negative the month has a significantly higher kurtosis. This is very interesting as it implies that the distribution during June has thicker tails than other months. We conclude that kurtosis varies over the calendar year.

6.2.4 THE IMPACT OF SIZE ON SKEWNESS AND KURTOSIS

The impacting variable is size and the segment variable is year. We divided the stocks into three cohorts based on average total market value: large, medium and small, for each year. We used data for the years 1994 to 2006 (13 years total) as those are the ones that we have market capitalization data for. The simulated distribution has a mean of 26 (which is equal to the correct average of 13 times 2) and skewness of -0.0143 and thus is in the acceptable range.

The results for skewness can be seen below:

Table 6.9
Sum of ranks, deviation form average, absolute deviation from average and p-value for each size cohort with respect to skewness

	SoR	Deviation	AbsDev	P-value
Large	22	-4	4	46.06%
Medium	23	-3	3	67.34%
Small	33	7	7	6.49%

None of the categories have a p-value low enough to pass the test on a significance level of 5%. The small cohort though are quite close and given a more fine-tuned method we could possibly have found evidence that skewness varies with size of the company. We cannot, however, given our methodology do so and conclude that size not impact skewness. The following table shows the results for kurtosis:

Table 6.10
Sum of ranks, deviation form average, absolute deviation from average and p-value for each size cohort with respect to kurtosis

	SoR	Deviation	AbsDev	P-value
Large	28	2	2	86.71%
Medium	20	-6	6	14.65%
Small	30	4	4	46.06%

Given the above results we cannot conclude that there is significant impact of size on kurtosis.

7. CONCLUSIONS AND DISCUSSION

In this section we will present our conclusions based on our findings. Furthermore, we will discuss the relevance for academia, financial models and professionals as well as what the impact is on future research. We will start off with the findings related to our main hypothesis and then go through the findings related to each and every sub-hypothesis one by one.

7.1 ALL STOCK RETURNS ARE NOT NORMALLY DISTRIBUTED

The main conclusion from testing our main hypothesis is that there is a significant discrepancy between the common assumption of normality in stock returns and how the stock returns actually behave. The main findings are:

- *All Swedish stocks' returns are not normally distributed*
- *None of our selected industries are normally distributed as group*
- *At least ¾ of Swedish stocks' returns are not normally distributed*

Aggregated we can conclude that using the normality assumption in a model is at best a negligible deviation that simplifies the derivation and usage of the model. At worst, these results calls for major redevelopment of the models as the models produce faulty results due to false assumptions. Our study has only shown that there is a discrepancy; we have not studied the effect of using this faulty simplification. Therefore we can only point out that further research has to be done to determine if this deviation is worth the risk, and thus study the important trade-off between mitigating complexity and producing reliable results. But nonetheless, our findings provide evidence that the common assumption of normality in reality is incorrect.

7.2 INDUSTRY IMPACTS KURTOSIS

The finding that kurtosis varies across industries is very interesting as it implies that not only the industry in which a company operates impacts the kurtosis but also that each stock could have a unique kurtosis which is determined by the specific company's operations. However, the finding brings up the matter of compensating investors for holding more kurtosis as it implies more extreme deviations. A common assumption in economics is "the law of diminishing marginal utility". An investor would possibly lose utility by increasing the kurtosis as it increases the risk of incurring big deviations in returns, which impacts the utility negatively as losing an amount is worse than gain from winning the same amount (given that the law of diminish marginal utility holds). The question then is, what amount of kurtosis maximizes utility? The redevelopment of financial models and how the models perceive investors would be necessary to incorporate the impact of kurtosis. Possibly investors should be compensated for higher kurtosis in the same sense as some models compensate for low liquidity.[34]

7.3 SKEWNESS VARIES OVER THE YEARS

The finding that skewness varies significantly over years have several important implications. First of all it means that a substantial portion of the stocks are affected in similar way by similar factors that in turn varies over time. We have not investigated any time-varying macroeconomic factors, such as economic growth and interest rates, and its impact as it would vastly extend the scope of our thesis. However, we can conclude that there are factors that impact and we would expect that these factors are mainly of macroeconomic nature. The rationale is that macroeconomic factors both have the broadness and the economic characteristic needed to satisfy our findings. The fact that the years of 1990 and 1993 both had substantial deviations and are associated with the financial crisis in Sweden during the early 1990s further supports this view.

The effect itself is very much comparable to the often used systematic risk, and possibly should be treated in the same way in financial models. Holding a portfolio carrying much exposure (assuming that an investor actually can alter the exposure to changing skewness) to changes in skewness could warrant a premium. We therefore encourage future research to center around these macroeconomic factors and how investors are impacted by different exposures to skewness.

7.4 THE JUNE-EFFECT: SUMMER BRINGS HIGHER KURTOSIS

In our study of the variation across months we found significant deviations and the most apparent one is that kurtosis is high (and thus generate low rankings in our test) in June. The implication of this finding can be divided into two parts: there is monthly variation in kurtosis and June has higher kurtosis than other months. The first implication is relevant to investors as it means that extreme deviations are more likely in some months than others, which further enforces the need to incorporate a more complete measure of risk as the standard deviation clearly cannot capture these differences. We cannot answer why the kurtosis is higher in June than other months, but we speculate that it could be due to lower turnover on capital markets which makes extreme deviations more likely. However, the finding in itself is relevant as it makes the need for forecasting kurtosis apparent to investors, as otherwise investors risk to not

[34] Fama, Eugene F. and French, Kenneth R. (1992)

internalize the differences in kurtosis (and thus the differences in actual risk compared to model risk such as standard deviation).

7.5 THE SIZE-EFFECT: STOPS AT EXPECTED RETURNS?

Our study has not provided evidence enough to show significant deviations across size cohorts. However, the p-values were quite close to be significant on a significance level of 5% (and would have passed the test on a 10% level, thus it is hard to draw any relevant conclusions from the test. Given more data and a more fine-tuned method one could possibly show significant impact, which we failed to do in the case of this factor.

7.6 REFLECTION AND CRITIQUE

To fully understand the findings in a study one has to first understand the limitations of the study. Our study has focused on Swedish stocks only and therefore it could lack in relevance in other markets and geographic areas. We argue that this risk is small as capital markets have become more homogenous and globalized over the years and should not differ that much.

8. REFERENCES

8.1 LITERATURE

Banz, Rolf W. and Breen William J. (1986), “Sample-Dependent Results Using Accounting and Market Data: Some Evidence”, *The Journal of Finance*, Vol. 41, No. 4, pp. 779-793

Black, Fischer and Myron Scholes (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, Vol. 81, No. 3, pp. 637–654

Blattberg, Robert, Gonedes, Nicholas (1974), “A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices”, *The Journal of Business*, Vol. 47, No. 2, pp. 244-280

Bodie, Z., Kane, Alex and Markus, Alan J. (2008), “Investments”, *McGraw-Hill Higher Education*, 8th edition, ISBN: 9780077261450

Fama, Eugene F. (Jan., 1965), “The Behavior of Stock-Market Prices”, *The Journal of Business*, Vol. 38, No. 1, pp. 34-105

Fama, Eugene F. and French, Kenneth R. (1992), ”The Cross- Section of Expected Stock Returns”, *The Journal of Finance*, Vol. 47, No. 2, pp. 427-465

Fama, Eugene F., and French Kenneth R. (1993), ”Common Risk Factors in the Returns on Stocks and Bonds”, *Journal of Financial Economics*, Vol. 33, pp 3-56

Gujarati, Damodar N. (2002), ”Basic Econometrics”, *McGraw-Hill Higher Education*, 4th edition, ISBN: 9780071230173

Hou, Kewei and Robinson David T. (2006), ”Industry Concentration and Average Stock Returns”, *The Journal of Finance*, Vol. 61, No. 4, pp. 1927-1956

Jarque, Carlos M. and Bera, Anil K.(1987), ”A Test for Normality of Observations and Regression Residuals”, *International Statistical Review*, Vol. 55, No. 2, pp. 163-172

Keim, Donald B. (1983), ”Size-related anomalies and stock return seasonality: Further empirical evidence”, *Journal of Financial Economics*, Vol. 12, No. 1, pp.13-32

Kon, Stanley J. (1984), “Models of Stock Returns - A Comparison”, *The Journal of Finance*, Vol. 39, No. 1, pp. 147-165

Mandelbrot, B. (Oct., 1963), ” The Variation of Certain Speculative Prices”, *Journal of Business*, Vol. 36, No. 4, pp. 394-419

Markowitz, Harry M. (1952), "Portfolio Selection", *The Journal of Finance*, Vol. 7, No. 1, pp. 77–91

Mills, Terence C. (1995), “Modelling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions”, *Journal of the Royal Statistical Society*, Vol. 44, No. 3, pp. 323-332

Ross, Stephen A. (1976), ”The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory,* Vol. 13, No. 3, pp. 341-360

Sharpe, William F. (1964), “Capital asset prices: A theory of market equilibrium under conditions of risk”, *Journal of Finance*, Vol. 19, No. 3, pp. 425-442

Sullivan, Edward J. and Weithers, Timothy M. (1991), “Louis Bachelier: The Father of Modern Option Pricing Theory”, *The Journal of Economic Education*, Vol. 22, No. 2, pp. 165-171

Woolridge, Jeffrey M. (2008), “Introductory Economics - A Modern Approach”, *South Western College,* 4th edition, *ISBN*: 9780324788907, p.801-807

8.2 ELECTRONIC SOURCES

Nasdaq OMX.
http://www.nasdaqomx.com/listing/europe/primarylisting/Industriessegmentandindexes/
[Collected 2013-02-06]

Princeton University. http://press.princeton.edu/titles/8275.html
[Collected 2013-02-06]

SIX Financial Information. http://www.six-telekurs.se/sv/se/Produkter-och-tjanster/Nordiska-innehallsprodukter-/Index/SIX-Index/[Collected 2013-02-14]

Swedish House of Finance. http://houseoffinance.se/research-data-center/finbas/ [Collected 2013-02-06]

The Swedish Riksbank. http://www.riksbank.se/sv/Riksbanken/Forskning/Historisk-monetar-statistik/Rantor-och-aktieavkastningar/ [Collected 2013-02-14]

9. APPENDIX

Table 9.1
Descriptive statistics for normalized returns used in Jarque-Bera test for industries

Data	
Observations	60 404
Mean	0,0436
Standard Deviation	0,8632
Variance	0,7451
Skewness	1,0362
Kurtosis	10,1134
Jarque-Bera	138 164,7

Table 9.2
Sample size, skewness, kurtosis and Jarque-Bera test statistic per industry

INDUSTRY	**n**	**S**	**K**	**JB**
Capital Goods	6 435	0,8978	11,0049	18 045,6274
Chemicals & Energy	1 502	0,5074	5,7518	538,3355
Financial Services	2 713	2,5100	27,1050	68 531,6193
Forest Industry	2 111	0,4231	6,2447	989,0454
Health Care & Pharmaceutical	3 431	1,4042	9,9375	8 008,0121
Household Goods	3 032	0,7165	9,3203	5 305,9985
Investment & Holding	4 569	1,0465	15,4891	30 528,2002
Consulting Services	1 711	0,6583	6,2121	859,1326
Media	801	1,2089	7,7708	954,7377
Real Estate	4 377	1,0035	10,1312	10 009,0337
Retail	1 047	0,3949	7,0359	737,7833
Services	1 042	0,8246	6,2237	569,2790
Technology	3 028	1,5313	11,7211	10 779,3672
Transport	1 896	0,6685	6,2133	956,9061
Vehicles & Machines	2 016	1,3060	13,4738	9 787,9636

Table 9.3
Number of stocks that passed the individual Jarque-Bera test or did not pass the test

pass	Freq.	Percent	Cum.
NO (0)	202	75,94	75,94
YES (1)	64	24,06	100
Total	266	100	

Table 9.4
Distribution of stocks that is tested for normality using the Jarque-Bera test, grouped by industry

desc	Freq.	Percent	Cum.
Capital Goods	43	16,17	16,17
Chemicals & Energy	9	3,38	19,55
Financial	15	5,64	25,19
Forest Industry	12	4,51	29,7
Health/Pharma	23	8,65	38,35
Household Goods	22	8,27	46,62
Invest/Hold	27	10,15	56,77
Konsulter	13	4,89	61,65
Media	7	2,63	64,29
Real Estate	30	11,28	75,56
Retail	7	2,63	78,2
Services	7	2,63	80,83
Technology	23	8,65	89,47
Transport	16	6,02	95,49
Vehicles and Machines	12	4,51	100
Total	266	100	

Table 9.5
Distribution of stocks that passed the Jarque-Bera test for normality grouped by industry

desc	Freq.	Percent	Cum.
Capital Goods	13	20,31	20,31
Chemicals & Energy	2	3,13	23,44
Financial	3	4,69	28,13
Forest Industry	3	4,69	32,81
Health/Pharma	3	4,69	37,5
Household Goods	7	10,94	48,44
Invest/Hold	5	7,81	56,25
Konsulter	4	6,25	62,5
Media	2	3,13	65,63
Real Estate	4	6,25	71,88
Retail	3	4,69	76,56
Services	2	3,13	79,69
Technology	3	4,69	84,38
Transport	6	9,38	93,75
Vehicles and Machines	4	6,25	100
Total	64	100	

Table 9.6
List of time series that passed the individual Jarque-Bera test and their industry and Jarque-Bera test statistic

timeserieid	desc	jb
3	Chemicals & Energy	0,7745
4	Chemicals & Energy	9,1178
20	Capital Goods	1,6447
83	Health Care & Pharmaceutical	3,3117
84	Health Care & Pharmaceutical	9,4442
303	Forest Industry	0,0510
442	Media	6,5652
450	Household Goods	4,0583
580	Invest/Hold	4,3452
585	Invest/Hold	12,8684
694	Vehicles and Machines	3,6245
800	Real Estate	7,6239
829	Capital Goods	2,1529
1049	Forest Industry	12,6634
1160	Capital Goods	2,8001
1247	Capital Goods	0,5900
1248	Capital Goods	11,5162
1430	Capital Goods	3,9311
1529	Vehicles and Machines	9,0927
1855	Invest/Hold	4,3709
2034	Real Estate	7,1005
2114	Financial	4,6122
2404	Health/Pharma	2,5596
2425	Financial	3,1699
2827	Transport	7,1880
2831	Transport	12,3821
2897	Invest/Hold	8,2531
2923	Transport	2,3537
3068	Transport	0,3321
3102	Capital Goods	0,9850
3103	Capital Goods	9,0381
3158	Forest Industry	1,9876
3251	Invest/Hold	2,9457
3266	Household Goods	13,4998
3267	Household Goods	11,1998
3287	Konsulter	13,3648
3449	Technology	12,9263
3457	Transport	2,8106
3731	Vehicles and Machines	1,3405
4075	Transport	2,4212
4445	Capital Goods	8,4863
4466	Capital Goods	7,8560
4589	Technology	12,3494
4606	Services	1,5055
4624	Capital Goods	2,5726
4658	Retail	11,9871
4714	Konsulter	4,4531
4825	Services	8,8136
4947	Real Estate	0,1390
5015	Technology	8,4042
5250	Konsulter	3,4423
5258	Financial	4,1567
5283	Household Goods	4,8932
5295	Retail	4,0644
5404	Real Estate	2,0939
5405	Konsulter	2,5438
5411	Household Goods	12,7077
5415	Capital Goods	2,7457
5437	Household Goods	0,0091
5464	Media	0,7438
5469	Household Goods	9,3908
5474	Capital Goods	5,2552
5549	Retail	7,3087
5596	Vehicles and Machines	13,1562

Table 9.7

Table of critical values of the chi-square distribution
for two degress of freedom, generated by STATA

Sign. Level	Critical value
10%	4,61
5%	5,99
1%	9,21
0,10%	13,82

Table 9.8
Distributions of absolute deviations for simulated distribution for industry over years

AbsDev	Freq	Cum freq	Cum %
13	2	10000	1
14	1	9998	0,9998
15	8	9997	0,9997
16	7	9989	0,9989
17	15	9982	0,9982
18	20	9967	0,9967
19	29	9947	0,9947
20	46	9918	0,9918
21	67	9872	0,9872
22	83	9805	0,9805
23	130	9722	0,9722
24	172	9592	0,9592
25	183	9420	0,942
26	242	9237	0,9237
27	292	8995	0,8995
28	337	8703	0,8703
29	381	8366	0,8366
30	422	7985	0,7985
31	408	7563	0,7563
32	401	7155	0,7155
33	459	6754	0,6754
34	508	6295	0,6295
35	479	5787	0,5787
36	456	5308	0,5308
37	436	4852	0,4852
38	436	4416	0,4416
39	421	3980	0,398
40	385	3559	0,3559
41	390	3174	0,3174
42	373	2784	0,2784
43	305	2411	0,2411
44	278	2106	0,2106
45	239	1828	0,1828
46	202	1589	0,1589
47	220	1387	0,1387
48	184	1167	0,1167
49	151	983	0,0983
50	122	832	0,0832
51	116	710	0,071
52	94	594	0,0594
53	79	500	0,05
54	70	421	0,0421
55	70	351	0,0351
56	48	281	0,0281
57	38	233	0,0233
58	40	195	0,0195
59	40	155	0,0155
60	22	115	0,0115
61	23	93	0,0093
62	13	70	0,007
63	11	57	0,0057
64	11	46	0,0046
65	11	35	0,0035
66	8	24	0,0024
67	2	16	0,0016
68	3	14	0,0014
69	3	11	0,0011
70	1	8	0,0008
71	1	7	0,0007
72	0	6	0,0006
73	2	6	0,0006
74	0	4	0,0004
75	0	4	0,0004
76	2	4	0,0004
77	1	2	0,0002
78	1	1	0,0001

Table 9.9
Sample skewness for each industry per year

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	1,0232	-0,6790	-0,2056	-0,5288	1,2461	0,0857	0,9349	1,4096	1,3034	0,5395	0,6189	0,1462	0,2890	0,1246	0,3019	0,4061	0,8218	0,5265
Chemicals & Energy	0,1564	-0,1971	-0,0540	-0,4916	0,4301	0,3125	0,6962	0,7090	0,4429	0,0432	0,1182	-0,7547	0,9752	-0,0297	0,4035	-0,2551	0,2486	-0,4929
Financial Services	0,3712	-0,1487	0,8167	2,9047	2,0456	0,8532	0,6627	0,5909	1,3071	0,4472	0,0400	0,4012	-0,2783	0,5732	0,4080	0,5835	2,1843	0,2023
Forest Industry	0,1767	-0,4966	0,6552	-0,8938	1,6933	1,3301	0,1346	0,2926	0,7721	0,8125	0,8580	0,1956	0,2811	-0,1862	0,7920	0,3686	0,2043	0,1629
Health Care & Pharmaceutical	0,6678	1,5555	-0,1682	-0,3371	0,6722	1,1125	0,8742	1,1499	1,3511	0,9188	0,4163	1,4984	0,6776	0,4713	2,1531	0,8876	0,2347	1,6519
Household Goods	-0,7331	-0,7352	-0,4060	1,7693	2,0987	1,0348	-1,2235	1,5948	0,7886	0,2496	0,0485	-0,2007	0,1646	0,0125	-0,0859	0,5531	0,6677	0,3932
Investment & Holding	1,4794	-1,0944	0,2145	0,9106	0,6014	2,1489	0,8437	1,2809	1,4875	0,2056	0,2608	4,0096	0,4157	0,7989	0,0250	0,2302	0,4689	0,9947
Consulting Services	0,9772	0,2155	-1,3394	-0,7379	0,8821	0,2356	1,6664	0,7955	0,5328	0,3979	1,1846	0,8373	-0,0897	0,3829	0,2309	1,5879	0,8471	0,7237
Media	0,0161	-0,8267	0,3571	0,5221	0,8112	1,1963	-0,0219	0,3660	0,3716	0,5756	0,5354	1,7440	0,1059	0,7386	1,9024	0,7568	0,5804	-0,2752
Real Estate	2,0394	-1,1747	0,3062	0,6954	1,4295	1,6297	0,7092	0,4722	-0,2065	0,0618	0,3409	2,0404	-0,1157	0,6802	1,2687	0,2873	-0,1950	1,0264
Retail	0,3226	-0,0960	0,6576	-1,1999	1,3847	0,9293	0,0525	0,5484	0,1975	0,7524	0,7906	0,7649	0,8786	0,3985	0,8009	0,2014	0,3109	-0,1831
Services	-0,1391	1,2860	-1,1708	0,8981	1,3834	1,0271	0,5362	0,2773	0,6327	1,2304	1,0845	0,8302	0,2867	0,2666	0,1988	2,3626	0,8263	-1,0029
Technology	2,7588	-0,6143	0,2127	0,6201	1,8955	0,3078	-0,0174	1,3984	1,0097	1,3706	1,5593	0,9792	1,2590	1,9174	0,4331	3,7334	0,3733	-1,7316
Transport	0,3815	0,2307	-1,1826	0,5253	1,3602	1,8792	0,3203	0,3601	0,5875	-0,1509	0,4827	0,9251	0,1492	0,1320	0,3303	0,9569	0,1701	0,3811
Vehicles & Machines	0,6697	-0,1593	-0,4839	2,3706	0,6411	0,8919	-0,1686	0,1018	1,1489	0,9818	2,4150	0,7508	0,8264	-0,3055	0,1270	0,3300	0,2124	1,3399

Table 9.10
Ranking of sample skewness

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of Ranks
Capital Goods	4	11	10	12	9	15	2	2	4	8	7	13	7	11	10	9	4	6	**144**
Chemicals & Energy	12	8	8	11	15	12	6	7	12	14	13	15	2	13	8	15	10	13	**194**
Financial Services	9	6	1	1	2	11	7	8	3	9	15	11	15	5	7	7	1	9	**127**
Forest Industry	11	9	3	14	4	4	10	13	8	5	5	12	9	14	5	10	13	10	**159**
Health Care & Pharmaceutical	7	1	9	10	12	6	3	5	2	4	10	4	5	6	1	5	11	1	**102**
Household Goods	15	12	11	3	1	7	15	1	7	11	14	14	10	12	15	8	5	7	**168**
Investment & Holding	3	14	6	4	14	1	4	4	1	12	12	1	6	2	14	13	7	4	**122**
Consulting Services	5	4	15	13	10	14	1	6	11	10	3	7	13	8	11	3	2	5	**141**
Media	13	13	4	9	11	5	13	11	13	7	8	3	12	3	2	6	6	12	**151**
Real Estate	2	15	5	6	5	3	5	10	15	13	11	2	14	4	3	12	15	3	**143**
Retail	10	5	2	15	6	9	11	9	14	6	6	9	3	7	4	14	9	11	**150**
Services	14	2	13	5	7	8	8	14	9	2	4	8	8	9	12	2	3	14	**142**
Technology	1	10	7	7	3	13	12	3	6	1	2	5	1	1	6	1	8	15	**102**
Transport	8	3	14	8	8	2	9	12	10	15	9	6	11	10	9	4	14	8	**160**
Vehicles & Machines	6	7	12	2	13	10	14	15	5	3	1	10	4	15	13	11	12	2	**155**

Table 9.11
Sample kurtosis for each industry per year

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	9,0393	5,0634	7,4011	4,3131	5,5364	3,2318	4,2326	8,3929	7,0676	4,1935	3,9091	4,7628	3,7244	3,7163	4,0726	5,0565	11,1441	3,4404
Chemicals & Energy	3,3576	3,7896	4,6977	3,4200	2,1626	2,6116	4,3346	4,0473	3,4652	3,3124	4,6778	5,4435	4,4060	2,0504	2,8202	3,4368	2,3220	2,3631
Financial Services	3,3892	3,9637	15,1052	16,7196	9,8341	5,6918	3,2057	4,6658	7,8329	4,4946	3,2743	4,1240	5,6358	5,6200	3,1963	3,5733	12,2873	3,2529
Forest Industry	3,2044	6,0582	5,2988	5,4390	7,4228	8,4981	3,1868	2,5908	4,1071	4,1099	4,0119	2,8952	3,5293	3,0612	3,5219	4,8981	2,5336	2,7987
Health Care & Pharmaceutical	2,7734	6,1071	2,7706	3,0541	3,9476	5,8681	3,2672	7,2220	5,7930	10,2521	4,8929	7,0073	6,8472	5,1999	10,6804	4,9994	4,2751	8,8048
Household Goods	3,9778	3,6401	5,7945	12,1759	11,2160	5,7707	11,9649	6,6897	4,7460	4,2425	5,0394	4,4924	2,9967	6,8322	3,0403	3,9419	4,6726	5,4018
Investment & Holding	9,4795	6,5957	5,3909	9,7277	3,4130	16,1685	4,3589	7,8599	12,3532	4,7717	4,5908	39,9654	4,5814	6,1303	5,2391	6,1639	6,8291	4,9291
Consulting Services	5,8155	2,9918	4,5981	3,6639	4,0787	3,6070	6,7677	4,6944	3,0158	3,7626	4,6843	6,4785	3,1247	5,4546	2,6179	9,6725	4,6951	4,2674
Media	1,7804	4,0956	2,4865	3,3312	3,4870	5,2931	2,2917	3,6357	2,9435	3,9310	2,8299	9,4820	2,1620	3,2041	7,3788	4,1839	5,1400	3,1880
Real Estate	10,8153	6,6768	4,0561	3,8298	5,4835	7,6867	4,0159	4,7496	8,4351	4,3973	4,3002	13,4397	3,6469	5,8998	11,3958	3,9418	4,8893	8,7612
Retail	3,1245	2,7926	3,8021	4,7536	4,8795	3,3259	2,0122	3,0931	2,2937	3,7579	3,4266	3,0198	4,6580	3,8959	9,2434	4,8253	3,5896	3,9076
Services	1,5000	3,4416	3,8496	4,0813	4,2530	4,5698	3,8280	3,4475	3,3045	8,2393	4,8980	4,1910	2,6383	2,7433	3,8981	13,7937	4,6735	9,2051
Technology	11,8616	3,0214	3,7043	5,1144	9,7434	3,9079	3,7179	5,1852	4,3806	12,5388	8,1757	6,5357	5,9790	10,0663	4,3652	30,0904	4,5126	14,5645
Transport	3,6038	4,0923	5,8799	3,9161	6,4294	11,5443	5,4473	3,5803	3,7790	4,8261	4,0923	6,5237	3,0007	2,6847	4,1250	3,9581	2,1230	3,7886
Vehicles & Machines	4,3433	3,9754	4,9952	17,5491	3,0111	3,4826	2,7000	4,4083	7,2039	5,9642	12,0648	6,4591	4,1664	3,1418	4,0327	2,4731	2,6170	6,7820

Table 9.12
Ranking of sample kurtosis

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of ranks
Capital Goods	4	5	2	8	6	14	6	1	5	10	12	10	8	9	8	5	2	11	**126**
Chemicals & Energy	10	10	8	13	15	15	5	10	11	15	7	9	6	15	14	14	14	15	**206**
Financial Services	9	9	1	2	2	7	11	8	3	7	14	13	3	5	12	13	1	12	**132**
Forest Industry	11	4	6	5	4	3	12	15	9	11	11	15	10	12	11	7	13	14	**173**
Health Care & Pharmaceutical	13	3	14	15	11	5	10	3	6	2	5	4	1	7	2	6	10	3	**120**
Household Goods	7	11	4	3	1	6	1	4	7	9	3	11	13	2	13	11	8	6	**120**
Investment & Holding	3	2	5	4	13	1	4	2	1	6	8	1	5	3	5	4	3	7	**77**
Consulting Services	5	14	9	12	10	11	2	7	13	13	6	7	11	6	15	3	6	8	**158**
Media	14	6	15	14	12	8	14	11	14	12	15	3	15	10	4	9	4	13	**193**
Real Estate	2	1	10	11	7	4	7	6	2	8	9	2	9	4	1	12	5	4	**104**
Retail	12	15	12	7	8	13	15	14	15	14	13	14	4	8	3	8	11	9	**195**
Services	15	12	11	9	9	9	8	13	12	3	4	12	14	13	10	2	7	2	**165**
Technology	1	13	13	6	3	10	9	5	8	1	2	5	2	1	6	1	9	1	**96**
Transport	8	7	3	10	5	2	3	12	10	5	10	6	12	14	7	10	15	10	**149**
Vehicles & Machines	6	8	7	1	14	12	13	9	4	4	1	8	7	11	9	15	12	5	**146**

Table 9.13
Descriptive statistics for simulated distribution (industry)

Distribution test	Percentiles	Smallest		
1%	102	66		
5%	114	68		
10%	120	68	Obs	150000
25%	131	74	Sum Wgt.	150000
50%	144		Mean	143,9996
		Largest	Std. Dev.	18,35466
75%	156	215		
90%	168	217	Variance	336,8936
95%	174	217	Skewness	0,0020238
99%	186	221	Kurtosis	2,912538

Table 9.14
Descriptive Statistics Random Values

Descriptive statistics of Random Values	
Count	10 000
Max	78
Average	36,94
Min	13
STDEV	8,69

Table 9.15

Distributions of absolute deviations for simulated distribution for years over industry

AbsDev	Freq	Cum freq	Cum %
27	1	10000	1
28	0	9999	0,9999
29	4	9999	0,9999
30	1	9995	0,9995
31	2	9994	0,9994
32	4	9992	0,9992
33	2	9988	0,9988
34	9	9986	0,9986
35	9	9977	0,9977
36	6	9968	0,9968
37	15	9962	0,9962
38	23	9947	0,9947
39	29	9924	0,9924
40	28	9895	0,9895
41	62	9867	0,9867
42	78	9805	0,9805
43	83	9727	0,9727
44	95	9644	0,9644
45	107	9549	0,9549
46	137	9442	0,9442
47	153	9305	0,9305
48	170	9152	0,9152
49	183	8982	0,8982
50	210	8799	0,8799
51	251	8589	0,8589
52	245	8338	0,8338
53	299	8093	0,8093
54	293	7794	0,7794
55	311	7501	0,7501
56	321	7190	0,7190
57	367	6869	0,6869
58	347	6502	0,6502
59	343	6155	0,6155
60	371	5812	0,5812
61	350	5441	0,5441
62	327	5091	0,5091
63	351	4764	0,4764
64	315	4413	0,4413
65	300	4098	0,4098
66	285	3798	0,3798
67	285	3513	0,3513
68	264	3228	0,3228
69	250	2964	0,2964
70	215	2714	0,2714
71	251	2499	0,2499
72	221	2248	0,2248
73	186	2027	0,2027
74	180	1841	0,1841
75	163	1661	0,1661
76	149	1498	0,1498
77	147	1349	0,1349
78	113	1202	0,1202
79	113	1089	0,1089
80	96	976	0,0976
81	120	880	0,0880
82	87	760	0,0760
83	77	673	0,0673
84	92	596	0,0596
85	73	504	0,0504
86	58	431	0,0431
87	53	373	0,0373
88	36	320	0,0320
89	40	284	0,0284
90	28	244	0,0244
91	21	216	0,0216
92	24	195	0,0195
93	20	171	0,0171
94	15	151	0,0151
95	18	136	0,0136
96	24	118	0,0118
97	17	94	0,0094
98	7	77	0,0077
99	10	70	0,0070
100	7	60	0,0060
101	6	53	0,0053
102	15	47	0,0047
103	8	32	0,0032
104	5	24	0,0024
105	5	19	0,0019
106	4	14	0,0014
107	0	10	0,0010
108	1	10	0,0010
109	1	9	0,0009
110	0	8	0,0008
111	2	8	0,0008
112	0	6	0,0006
113	1	6	0,0006
114	1	5	0,0005
115	1	4	0,0004
116	1	3	0,0003
117	1	2	0,0002
118	0	1	0,0001
119	0	1	0,0001
120	0	1	0,0001
121	1	1	0,0001

Table 9.16
Samples skewness for each year by industry

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	0,7898	0,6598	1,1992	-0,4590	-0,4966	1,8688	5,0885	1,8316	1,0232	-0,6790	-0,2056	-0,5288	1,2461	0,0857	0,9349	1,4096	1,3034	0,5395	0,6189	0,1462	0,2890	0,1246	0,3019	0,4061	0,8218	0,5265
Chemicals & Energy	1,3893	0,9132	0,5922	0,7652	0,8019	0,6594	0,2970	0,9969	0,1564	-0,1971	-0,0540	-0,4916	0,4301	0,3125	0,6962	0,7090	0,4429	0,0432	0,1182	-0,7547	0,9752	-0,0297	0,4035	-0,2551	0,2486	-0,4929
Financial Services	0,5531	0,6214	1,0256	-0,5636	-0,3666	0,7387	0,4466	-0,3032	0,3712	-0,1487	0,8167	2,9047	2,0456	0,8532	0,6627	0,5909	1,3071	0,4472	0,0400	0,4012	-0,2783	0,5732	0,4080	0,5835	2,1843	0,2023
Forest Industry	-0,1263	0,4395	0,2512	0,0362	0,9332	1,2923	-1,2717	0,9987	0,1767	-0,4966	0,6552	-0,8938	1,6933	1,3301	0,1346	0,2926	0,7721	0,8125	0,8580	0,1956	0,2811	-0,1862	0,7920	0,3686	0,2043	0,1629
Health Care & Pharmaceutical	0,4672	0,6850	0,1280	1,1855	0,1278	2,1453	-0,3339	0,6103	0,6678	1,5555	-0,1682	-0,3371	0,6722	1,1125	0,8742	1,1499	1,3511	0,9188	0,4163	1,4984	0,6776	0,4713	2,1531	0,8876	0,2347	1,6519
Household Goods	0,7630	-0,1617	1,9078	1,5244	0,5631	0,7010	0,2982	1,6367	-0,7331	-0,7352	-0,4060	1,7693	2,0987	1,0348	-1,2235	1,5948	0,7886	0,2496	0,0485	-0,2007	0,1646	0,0125	-0,0859	0,5531	0,6677	0,3932
Investment & Holding	0,6845	1,3556	1,3561	0,0383	-0,2247	0,2396	0,0874	1,4523	1,4794	-1,0944	0,2145	0,9106	0,6014	2,1489	0,8437	1,2809	1,4875	0,2056	0,2608	4,0096	0,4157	0,7989	0,0250	0,2302	0,4689	0,9947
Consulting Services	0,3981	-0,2156	-0,0058	1,2346	-0,0435	1,2187	0,2570	-0,3974	0,9772	0,2155	-1,3394	-0,7379	0,8821	0,2356	1,6664	0,7955	0,5328	0,3979	1,1846	0,8373	-0,0897	0,3829	0,2309	1,5879	0,8471	0,7237
Media	0,5265	1,5640	-0,2315	-0,0096	0,1248	-0,4914	0,3123	0,0597	0,0161	-0,8267	0,3571	0,5221	0,8112	1,1963	-0,0219	0,3660	0,3716	0,5756	0,5354	1,7440	0,1059	0,7386	1,9024	0,7568	0,5804	-0,2752
Real Estate	0,2832	0,1576	-0,5143	-0,2062	1,6066	0,0769	-0,6048	-0,0356	2,0394	-1,1747	0,3062	0,6954	1,4295	1,6297	0,7092	0,4722	-0,2065	0,0618	0,3409	2,0404	-0,1157	0,6802	1,2687	0,2873	-0,1950	1,0264
Retail	0,9809	-0,7119	0,3940	0,9418	0,7499	0,2040	0,2550	0,1005	0,3226	-0,0960	0,6576	-1,1999	1,3847	0,9293	0,0525	0,5484	0,1975	0,7524	0,7906	0,7649	0,8786	0,3985	0,8009	0,2014	0,3109	-0,1831
Technology	1,8595	1,4400	-0,2081	-0,4906	0,2380	-0,0624	0,6374	0,1200	2,7588	-0,6143	0,2127	0,6201	1,8955	0,3078	-0,0174	1,3984	1,0097	1,3706	1,5593	0,9792	1,2590	1,9174	0,4331	3,7334	0,3733	-1,7316
Transport	0,2318	1,2611	1,2096	0,2325	0,2931	1,2597	0,5557	2,1502	0,3815	0,2307	-1,1826	0,5253	1,3602	1,8792	0,3203	0,3601	0,5875	-0,1509	0,4827	0,9251	0,1492	0,1320	0,3303	0,9569	0,1701	0,3811
Vehicles & Machines	1,0313	0,4874	0,4287	1,1539	0,9266	0,4667	0,2779	0,7345	0,6697	-0,1593	-0,4839	2,3706	0,6411	0,8919	-0,1686	0,1018	1,1489	0,9818	2,4150	0,7508	0,8264	-0,3055	0,1270	0,3300	0,2124	1,3399

Table 9.17
Ranking of sample skewness

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	11	12	7	23	24	2	1	3	8	26	22	25	6	21	9	4	5	14	13	19	18	20	17	16	10	15
Chemicals & Energy	1	4	10	6	5	9	15	2	17	22	21	24	12	14	8	7	11	19	18	26	3	20	13	23	16	25
Financial Services	14	10	5	26	25	8	16	24	19	22	7	1	3	6	9	11	4	15	21	18	23	13	17	12	2	20
Forest Industry	22	11	15	21	5	3	26	4	18	24	10	25	1	2	20	13	9	7	6	17	14	23	8	12	16	19
Health Care & Pharmaceutical	19	13	22	7	23	2	25	17	16	4	24	26	15	9	12	8	6	10	20	5	14	18	1	11	21	3
Household Goods	9	21	2	6	12	10	15	4	24	25	23	3	1	7	26	5	8	16	18	22	17	19	20	13	11	14
Investment & Holding	13	7	6	23	25	18	22	5	4	26	20	10	14	2	11	8	3	21	17	1	16	12	24	19	15	9
Consulting Services	13	23	20	3	21	4	16	24	6	19	26	25	7	17	1	10	12	14	5	9	22	15	18	2	8	11
Media	11	3	23	21	17	25	16	19	20	26	15	12	5	4	22	14	13	9	10	2	18	7	1	6	8	24
Real Estate	15	16	24	22	4	17	25	19	2	26	13	9	5	3	8	11	23	18	12	1	20	10	6	14	21	7
Retail	2	25	14	3	10	18	17	21	15	23	11	26	1	4	22	12	20	9	7	8	5	13	6	19	16	24
Technology	5	7	23	24	18	22	13	20	2	25	19	14	4	17	21	8	11	9	6	12	10	3	15	1	16	26
Transport	20	4	6	19	18	5	10	1	13	21	26	11	3	2	17	15	9	25	12	8	23	24	16	7	22	14
Vehicles & Machines	6	15	17	4	8	16	19	12	13	23	26	2	14	9	24	22	5	7	1	11	10	25	21	18	20	3
Sum of ranks	**161**	**171**	**194**	**208**	**215**	**159**	**236**	**175**	**177**	**312**	**263**	**213**	**91**	**117**	**210**	**148**	**139**	**193**	**166**	**159**	**213**	**222**	**183**	**173**	**202**	**214**

Table 9.18
Sample kurtosis for each year by industry

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	3,5006	4,6962	5,0007	6,2111	4,6837	11,7576	52,3130	12,4272	9,0393	5,0634	7,4011	4,3131	5,5364	3,2318	4,2326	8,3929	7,0676	4,1935	3,9091	4,7628	3,7244	3,7163	4,0726	5,0565	11,1441	3,4404
Chemicals & Energy	4,4096	6,2819	2,7722	4,1048	3,8204	3,9224	4,7855	4,7314	3,3576	3,7896	4,6977	3,4200	2,1626	2,6116	4,3346	4,0473	3,4652	3,3124	4,6778	5,4435	4,4060	2,0504	2,8202	3,4368	2,3220	2,3631
Financial Services	4,9053	3,8549	5,8549	3,1624	3,9054	4,0358	4,0616	6,4898	3,3892	3,9637	15,1052	16,7196	9,8341	5,6918	3,2057	4,6658	7,8329	4,4946	3,2743	4,1240	5,6358	5,6200	3,1963	3,5733	12,2873	3,2529
Forest Industry	2,0391	2,9541	2,3500	2,7233	5,9199	5,6607	6,2287	4,6629	3,2044	6,0582	5,2988	5,4390	7,4228	8,4981	3,1868	2,5908	4,1071	4,1099	4,0119	2,8952	3,5293	3,0612	3,5219	4,8981	2,5336	2,7987
Health Care & Pharmaceutical	1,9087	2,9145	1,9825	5,5189	2,2818	9,3198	2,6753	2,6557	2,7734	6,1071	2,7706	3,0541	3,9476	5,8681	3,2672	7,2220	5,7930	10,2521	4,8929	7,0073	6,8472	5,1999	10,6804	4,9994	4,2751	8,8048
Household Goods	3,4491	3,0837	7,4036	7,3804	3,2089	3,3193	3,0005	9,0714	3,9778	3,6401	5,7945	12,1759	11,2160	5,7707	11,9649	6,6897	4,7460	4,2425	5,0394	4,4924	2,9967	6,8322	3,0403	3,9419	4,6726	5,4018
Investment & Holding	4,7635	9,4131	5,5608	4,1774	3,5835	3,9156	4,0150	7,8242	9,4795	6,5957	5,3909	9,7277	3,4130	16,1685	4,3589	7,8599	12,3532	4,7717	4,5908	39,9654	4,5814	6,1303	5,2391	6,1639	6,8291	4,9291
Consulting Services	1,7606	2,4082	2,3601	3,6183	2,3613	3,4555	2,7624	3,5396	5,8155	2,9918	4,5981	3,6639	4,0787	3,6070	6,7677	4,6944	3,0158	3,7626	4,6843	6,4785	3,1247	5,4546	2,6179	9,6725	4,6951	4,2674
Media	3,7268	4,5066	2,7125	2,1524	2,2586	3,0942	1,7766	2,4899	1,7804	4,0956	2,4865	3,3312	3,4870	5,2931	2,2917	3,6357	2,9435	3,9310	2,8299	9,4820	2,1620	3,2041	7,3788	4,1839	5,1400	3,1880
Real Estate	2,8860	2,8074	3,2462	4,8115	6,4702	4,6751	5,7154	3,3809	10,8153	6,6768	4,0561	3,8298	5,4835	7,6867	4,0159	4,7496	8,4351	4,3973	4,3002	13,4397	3,6469	5,8998	11,3958	3,9418	4,8893	8,7612
Retail	3,8101	2,8683	2,2586	5,4715	2,9469	2,1821	2,4745	1,7418	3,1245	2,7926	3,8021	4,7536	4,8795	3,3259	2,0122	3,0931	2,2937	3,7579	3,4266	3,0198	4,6580	3,8959	9,2434	4,8253	3,5896	3,9076
Technology	6,5035	4,9166	1,6946	3,0434	2,5243	2,7316	4,0575	2,6972	11,8616	3,0214	3,7043	5,1144	9,7434	3,9079	3,7179	5,1852	4,3806	12,5388	8,1757	6,5357	5,9790	10,0663	4,3652	30,0904	4,5126	14,5645
Transport	3,2944	3,5797	3,8417	2,0656	2,4336	8,5887	3,1306	12,9066	3,6038	4,0923	5,8799	3,9161	6,4294	11,5443	5,4473	3,5803	3,7790	4,8261	4,0923	6,5237	3,0007	2,6847	4,1250	3,9581	2,1230	3,7886
Vehicles & Machines	3,8413	4,3342	3,0821	4,9777	4,2668	2,5072	2,3799	3,8605	4,3433	3,9754	4,9952	17,5491	3,0111	3,4826	2,7000	4,4083	7,2039	5,9642	12,0648	6,4591	4,1664	3,1418	4,0327	2,4731	2,6170	6,7820

Table 9.19
Ranking of sample kurtosis

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Capital Goods	24,0000	15	13	9	16	3	1	2	5	11	7	17	10	26	18	6	8	19	21	14	22	23	20	12	4	25
Chemicals & Energy	7	1	21	10	13	12	3	4	18	14	5	17	25	22	9	11	15	19	6	2	8	26	20	16	24	23
Financial Services	11	19	7	26	18	16	15	6	21	17	2	1	4	8	24	12	5	13	22	14	9	10	25	20	3	23
Forest Industry	26	19	25	22	5	6	3	10	16	4	8	7	2	1	17	23	12	11	13	20	14	18	15	9	24	21
Health Care & Pharmaceutical	26	19	25	11	24	3	22	23	20	8	21	18	16	9	17	5	10	2	14	6	7	12	1	13	15	4
Household Goods	20	23	5	6	22	21	25	4	17	19	9	1	3	10	2	8	13	16	12	15	26	7	24	18	14	11
Investment & Holding	18	6	13	22	25	24	23	8	5	10	14	4	26	2	21	7	3	17	19	1	20	12	15	11	9	16
Consulting Services	26	23	25	14	24	17	21	16	4	20	9	13	11	15	2	7	19	12	8	3	18	5	22	1	6	10
Media	9	5	18	24	22	15	26	19	25	7	20	12	11	3	21	10	16	8	17	1	23	13	2	6	4	14
Real Estate	25	26	24	13	8	15	10	23	3	7	18	21	11	6	19	14	5	16	17	1	22	9	2	20	12	4
Retail	9	19	23	2	18	24	21	26	15	20	10	5	3	14	25	16	22	11	13	17	6	8	1	4	12	7
Technology	9	13	26	21	25	23	17	24	4	22	20	12	6	18	19	11	15	3	7	8	10	5	16	1	14	2
Transport	20	19	14	26	24	3	21	1	17	10	6	13	5	2	7	18	16	8	11	4	22	23	9	12	25	15
Vehicles & Machines	17	11	20	8	12	24	26	16	10	15	7	1	21	18	22	9	3	6	2	5	13	19	14	25	23	4
Sum of ranks	**247**	**218**	**259**	**214**	**256**	**206**	**234**	**182**	**180**	**184**	**156**	**142**	**154**	**154**	**223**	**157**	**162**	**161**	**182**	**111**	**220**	**190**	**186**	**168**	**189**	**179**

Table 9.20

Descriptive statistics for simulated distribution (years)

Distribution test	Percentiles	Smallest		
1%	125	68		
5%	143	73		
10%	153	74	Obs	260 000
25%	170	75	Sum of Wgt.	260 000
50%	189		Mean	189
		Largest	Std. Dev.	28,0829
75%	208	298		
90%	225	300	Variance	788,6501
95%	235	300	Skewness	0,0042
99%	254	306	Kurtosis	2,9020

Table 9.21

Distributions of absolute deviations for simulated distribution for months over years

AbsDev	Freq	Cum freq	Cum %
13	1	1000	1
14	3	999	0,9990
15	3	996	0,9960
16	3	993	0,9930
17	2	990	0,9900
18	7	988	0,9880
19	9	981	0,9810
20	16	972	0,9720
21	24	956	0,9560
22	19	932	0,9320
23	22	913	0,9130
24	22	891	0,8910
25	38	869	0,8690
26	39	831	0,8310
27	34	792	0,7920
28	50	758	0,7580
29	37	708	0,7080
30	43	671	0,6710
31	57	628	0,6280
32	44	571	0,5710
33	52	527	0,5270
34	49	475	0,4750
35	43	426	0,4260
36	42	383	0,3830
37	42	341	0,3410
38	41	299	0,2990
39	30	258	0,2580
40	24	228	0,2280
41	32	204	0,2040
42	20	172	0,1720
43	22	152	0,1520
44	20	130	0,1300
45	21	110	0,1100
46	19	89	0,0890
47	10	70	0,0700
48	5	60	0,0600
49	10	55	0,0550
50	12	45	0,0450
51	5	33	0,0330
52	7	28	0,0280
53	4	21	0,0210
54	2	17	0,0170
55	1	15	0,0150
56	2	14	0,0140
57	2	12	0,0120
58	1	10	0,0100
59	3	9	0,0090
60	0	6	0,0060
61	2	6	0,0060
62	0	4	0,0040
63	0	4	0,0040
64	1	4	0,0040
65	0	3	0,0030
66	1	3	0,0030
67	0	2	0,0020
68	0	2	0,0020
69	1	2	0,0020
70	0	1	0,0010
71	0	1	0,0010
72	1	1	0,0010

Table 9.22

Sample skewness for months over years

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
January	1,1220	1,5012	2,1477	0,1257	0,6089	1,5100	-0,8026	0,2338	0,4204	0,3570	-0,2963	-0,6340	1,2516	0,9899	1,1633	0,7739	0,6378	0,1633	2,3568	1,7621	1,0122	0,7831	0,6248	2,9294	1,0524	0,2236
Feb	0,5350	0,6899	0,4078	-0,0808	0,3672	0,7949	-0,6010	1,0910	0,9189	0,6758	-0,0355	-0,2473	0,5147	1,0006	0,4431	0,3392	1,4392	0,8439	0,6630	1,3548	-0,0017	0,2211	-0,7074	0,6966	0,9962	0,3323
Mars	0,4211	-1,1136	1,0680	-0,0407	0,6096	-0,0479	1,0003	0,4515	0,0941	-0,0831	-0,4167	0,3358	1,0543	0,1989	-0,2144	1,0349	1,6611	2,1187	1,1855	2,5198	1,3818	1,1347	2,2806	0,7369	0,8636	1,3551
April	1,0730	0,2854	0,7996	1,3286	0,7156	-0,2130	0,1361	1,8046	0,7585	0,2395	1,7952	-0,6224	1,3913	0,0559	0,6242	1,2238	0,0573	1,5029	0,6114	0,6767	0,4581	0,1719	0,4380	0,8068	0,4776	0,5084
May	1,6185	-0,5952	1,8197	0,4956	0,4062	1,7876	1,8118	0,5052	1,3404	1,1252	-0,1401	2,9120	2,8209	0,6080	0,2416	0,8435	0,2016	0,7615	0,3779	0,8696	0,8049	0,9927	1,3906	0,6469	0,0099	-0,5957
June	1,5728	-0,5303	0,1372	1,5497	0,3485	1,8725	-0,0964	0,9708	2,1255	0,9997	0,2273	-0,2741	1,0766	1,0555	1,1476	1,4865	0,5245	0,8620	-0,6381	1,2058	-0,0472	0,0352	1,6139	0,2259	-0,8584	0,1283
July	0,1916	0,5608	1,2668	0,2449	-0,2412	0,8723	0,7184	0,4081	0,8120	0,0334	1,9900	-1,4661	1,3163	-0,3890	1,1910	1,9160	0,6406	1,0871	0,8064	-0,3101	0,0487	0,9517	-0,2125	-0,2157	0,2408	0,1950
August	0,2308	0,1742	0,9891	0,0179	0,0244	0,6923	0,5254	-0,1409	1,3615	-0,4391	0,0331	-0,0973	0,2934	0,0129	1,1595	0,9932	1,1018	-0,1889	0,3404	1,1183	-0,7474	-0,0597	0,4404	0,5858	1,2636	2,5621
September	1,7609	0,7157	0,2880	-0,1296	-0,2736	0,8077	4,8194	0,4874	2,2151	-0,3118	-0,1015	-0,7533	1,1393	0,4422	-0,2056	0,9788	0,6719	-0,5341	2,0840	0,8566	0,0109	0,0320	3,1524	0,4423	0,6111	0,9333
October	0,7142	0,8006	-0,8869	-0,6915	-0,4170	0,5367	0,1874	2,7233	-0,7767	-0,2700	-0,9298	1,8533	1,0667	0,6658	0,4826	0,9496	0,4410	0,3277	0,4316	-0,0413	0,4793	1,0316	0,8558	0,9938	0,9090	-1,1125
November	-0,1523	0,2028	0,5936	-0,2131	0,3633	1,4508	-0,9779	2,8455	1,2927	-0,8262	-0,2264	0,2960	0,3164	-0,3001	0,1818	1,3336	0,9597	0,7545	0,5996	0,5234	0,6746	1,1174	0,9555	0,6304	0,7014	-0,1829
December	1,5353	1,3194	1,1997	0,3587	1,1870	2,4455	-0,2085	0,2511	0,0744	0,0324	-0,7021	0,6316	0,8202	1,8262	0,4895	1,2273	1,2132	0,5588	0,5654	0,1154	0,3791	-0,1421	0,8675	0,9108	0,2642	0,5357

Table 9.23

Ranking of sample skewness

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of ranks
January	5	1	1	6	4	4	11	11	9	4	9	10	4	4	2	11	8	10	1	2	2	6	8	1	2	7	143
Feb	8	5	9	9	6	8	10	4	6	3	5	7	10	3	8	12	2	5	5	3	10	7	12	6	3	6	172
Mars	9	12	5	8	3	11	3	8	10	8	10	4	8	8	12	6	1	1	3	1	1	1	2	5	5	2	147
April	6	7	7	2	2	12	7	3	8	5	2	9	2	9	5	5	12	2	6	8	6	8	10	4	8	5	160
May	2	11	2	3	5	3	2	6	4	1	7	1	1	6	9	10	11	6	10	6	3	4	4	7	11	11	146
June	3	10	11	1	8	2	8	5	2	2	3	8	6	2	4	2	9	4	12	4	11	9	3	11	12	9	161
July	11	6	3	5	10	6	4	9	7	6	1	12	3	12	1	1	7	3	4	12	8	5	11	12	10	8	177
August	10	9	6	7	9	9	5	12	3	11	4	6	12	10	3	7	4	11	11	5	12	11	9	9	1	1	197
September	1	4	10	10	11	7	1	7	1	10	6	11	5	7	11	8	6	12	2	7	9	10	1	10	7	3	177
October	7	3	12	12	12	10	6	2	12	9	12	2	7	5	7	9	10	9	9	11	5	3	7	2	4	12	199
November	12	8	8	11	7	5	12	1	5	12	8	5	11	11	10	3	5	7	7	9	4	2	5	8	6	10	192
December	4	2	4	4	1	1	9	10	11	7	11	3	9	1	6	4	3	8	8	10	7	12	6	3	9	4	157

Table 9.24

Sample kurtosis for months over years

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
January	4,9505	7,6489	9,4700	3,5152	3,5140	6,3423	6,8239	2,9180	3,1918	2,7154	4,0950	3,8184	5,6445	4,5203	5,1571	4,9574	3,6857	5,8382	14,6814	7,7451	6,1757	4,8563	6,8994	19,5906	6,5247	7,1942
Feb	3,2707	4,3610	3,5691	3,8178	3,1693	4,1829	3,5643	5,5717	6,8043	4,5589	4,5560	3,8210	3,3114	4,7018	4,4853	3,5331	6,7241	6,0188	3,5398	6,0862	2,6766	2,5228	4,5480	4,3777	4,9307	3,4480
Mars	2,9326	5,2961	4,6332	3,1771	2,9047	5,1658	6,7936	4,2511	4,2951	3,7045	6,3606	3,2257	4,3531	4,0234	5,8618	4,9449	9,6595	13,1783	5,8130	20,2108	10,1654	6,1536	19,1430	4,5452	5,5130	6,8732
April	3,4239	2,5100	4,0491	8,7498	3,3229	2,8901	3,0895	10,1422	4,1392	4,4009	12,4943	7,1037	8,2500	3,4385	3,5062	7,3001	4,6792	10,7957	3,3092	5,2431	4,3813	3,5365	2,9955	4,3124	4,5403	4,2940
May	5,9860	3,8194	11,1700	3,3711	3,2672	10,2297	8,3562	2,4038	6,6598	5,7816	4,5759	22,5540	12,5133	3,8512	3,4051	6,2014	11,3880	4,2708	3,0759	5,7815	5,0474	6,6980	6,9788	6,6613	4,2232	4,6719
June	6,7692	5,3066	2,4640	8,8932	4,4377	8,0102	4,9183	4,3268	11,8052	6,9726	11,7728	10,6185	5,0854	6,0994	4,4631	6,8583	3,6221	6,6317	6,1084	8,1241	4,6071	3,5165	12,4772	3,7720	8,2269	7,0659
July	1,8166	3,3498	5,6568	3,1347	3,5485	4,2408	3,2869	3,6916	3,9845	5,8839	12,2079	10,5643	4,4704	3,2517	5,9162	11,3770	4,1551	5,3928	4,0290	4,6634	4,0059	5,0575	5,5399	5,0238	3,7362	3,3646
August	2,0341	2,5544	4,5075	4,3925	3,3373	3,5707	3,2949	3,2528	6,1932	3,6327	2,7682	2,8260	3,9029	3,5687	5,7734	7,1217	6,7865	3,2867	4,1789	6,3622	4,2682	3,6050	4,6625	4,0181	12,6284	14,3309
September	8,0909	2,7081	2,5766	2,7090	3,0160	5,3016	35,6361	3,3768	10,4361	4,9382	4,6444	3,5971	6,1552	3,7730	5,4377	4,1408	5,6584	4,4674	12,2526	4,8502	3,3453	3,1602	15,7842	4,5526	4,4666	8,8665
October	4,4758	3,7790	5,2740	9,2416	5,6674	3,4377	4,1796	15,8178	4,3628	3,4932	5,2462	8,7255	4,4969	7,6367	4,1061	5,6964	3,5078	3,1679	5,3672	5,2371	3,5647	6,4816	4,9518	11,5659	5,6056	8,8608
November	2,0766	3,2963	4,2798	2,7620	2,5426	7,2076	3,5174	16,0837	8,9342	3,6410	3,4368	4,4198	4,0125	4,6653	3,9031	6,4833	4,6404	5,9485	4,1124	5,0262	4,2202	5,6449	4,4622	3,7825	4,5886	6,5560
December	9,0404	4,5662	5,9645	3,7409	5,0909	16,5167	5,1273	3,7568	4,5754	4,6556	4,8215	9,1889	4,1721	8,6906	3,8550	5,0973	5,8417	4,3565	3,5861	3,2099	3,6684	2,9066	6,8820	5,4741	3,8240	4,3201

Table 9.25

Ranking of sample kurtosis

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of ranks
January	5	1	2	7	5	5	3	11	12	12	10	9	4	6	5	9	10	6	1	3	2	6	5	1	3	4	**147**
Feb	8	5	10	5	9	9	8	4	4	6	9	8	12	4	6	12	4	4	10	5	12	12	10	8	6	11	**201**
Mars	9	3	6	9	11	7	4	6	9	8	4	11	8	7	2	10	2	1	4	1	1	3	1	7	5	6	**145**
April	7	12	9	3	7	12	12	3	10	7	1	6	2	11	11	2	7	2	11	7	5	8	12	9	8	10	**194**
May	4	6	1	8	8	2	2	12	5	3	8	1	1	8	12	6	1	10	12	6	3	1	4	3	10	8	**145**
June	3	2	12	2	3	3	6	5	1	1	3	2	5	3	7	4	11	3	3	2	4	9	3	12	2	5	**116**
July	12	8	4	10	4	8	11	8	11	2	2	3	7	12	1	1	9	7	8	11	8	5	7	5	12	12	**188**
August	11	11	7	4	6	10	10	10	6	10	12	12	11	10	3	3	3	11	6	4	6	7	9	10	1	1	**194**
September	2	10	11	12	10	6	1	9	2	4	7	10	3	9	4	11	6	8	2	10	11	10	2	6	9	2	**177**
October	6	7	5	1	1	11	7	2	8	11	5	5	6	2	8	7	12	12	5	8	10	2	8	2	4	3	**158**
November	10	9	8	11	12	4	9	1	3	9	11	7	10	5	9	5	8	5	7	9	7	4	11	11	7	7	**199**
December	1	4	3	6	2	1	5	7	7	5	6	4	9	1	10	8	5	9	9	12	9	11	6	4	11	9	**164**

Table 9.26
Distributions of absolute deviations for simulated distribution for months over years

Distribution test	Percentiles	Smallest		
1%	128	94		
5%	140	95		
10%	146	97	Obs	120 000
25%	157	97	Sum of Wgt.	120 000
50%	169		Mean	169
		Largest	Std. Dev.	17,5941
75%	181	238		
90%	192	238	Variance	309,5529
95%	198	240	Skewness	0,0060
99%	210	243	Kurtosis	2,9655

Table 9.27
Distributions of absolute deviations for simulated distribution for size over years

AbsDev	Freq	Cum freq	Cum %
0	217	10 000	1
1	1 112	9 783	0,9783
2	1 937	8 671	0,8671
3	2 128	6 734	0,6734
4	1 840	4 606	0,4606
5	1 301	2 766	0,2766
6	816	1 465	0,1465
7	396	649	0,0649
8	179	253	0,0253
9	52	74	0,0074
10	15	22	0,0022
11	7	7	0,0007

Table 9.28
Sample skewness for size over years

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Large	0,8689	0,4688	0,7989	0,9738	1,7978	1,4574	1,6882	1,1960	0,9447	0,9326	1,4113	1,5107	1,5900
Medium	1,2290	0,3298	1,2353	1,1070	0,5901	1,8888	1,3094	1,3091	0,4409	0,5146	2,1705	1,4669	0,8530
Small	1,6082	0,5464	0,7205	0,4348	0,3451	0,9784	0,9508	0,2482	0,2582	1,1934	0,6432	0,3407	0,3218

Table 9.29
Ranking of sample skewness

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of ranks
Large	3	2	2	2	1	2	1	2	1	2	2	1	1	**22**
Medium	2	3	1	1	2	1	2	1	2	3	1	2	2	**23**
Small	1	1	3	3	3	3	3	3	3	1	3	3	3	**33**

Table 9.30
Sample kurtosis for size over years

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Large	5,4494	6,2418	4,6418	4,6554	12,8229	7,2637	9,4135	7,3597	5,0868	4,9352	6,6202	8,0651	8,6552
Medium	7,0628	4,5546	6,7813	5,5034	5,1724	12,2318	7,7350	11,3570	6,0492	7,9059	18,9689	10,7720	8,0122
Small	11,2704	4,4801	4,9166	5,4958	4,3131	6,2561	11,1266	4,3865	5,9907	11,5290	5,3976	5,7534	7,9623

Table 9.31

Ranking of sample kurtosis

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Sum of ranks
Large	3	1	3	3	1	2	2	2	3	3	2	2	1	**28**
Medium	2	2	1	1	2	1	3	1	1	2	1	1	2	**20**
Small	1	3	2	2	3	3	1	3	2	1	3	3	3	**30**

Table 9.32

Descriptive statistics for simulated distribution (size over years)

Distribution test	Percentiles	Smallest		
1%	19	15		
5%	21	15		
10%	22	15	Obs	30 000
25%	24	15	Sum of Wgt.	30 000
50%	26		Mean	26
		Largest	Std. Dev.	2,9423
75%	28	36		
90%	30	36	Variance	8,6569
95%	31	36	Skewness	-0,0143
99%	33	37	Kurtosis	2,8890

THE PROCESS OF DIVIDING STOCKS INTO INDUSTRIES

In order to properly analyze the effects of industries on stock return distribution we had to adjust the dataset given. Initially there were over 500 defined industries in the dataset; this is due to the fact that industry definitions have changed during the time period 1979-2010. The problem with this was that many of the industries could be similar with only the formulation that separated them from each other, so we had to decrease the number of industries.

Nasdaq OMX is currently using the Industry Classification Benchmark (ICB) that is maintained by the FTSE Group, it consists of 10 industries[35] The same list could be used for this paper, however, it would become problematic since the industries have changed over time and the risk of misplacing companies would increase. The ICB might not be relevant for past data, so in order to solve this problem we chose the most frequently used industry definitions in our dataset and divided the companies accordingly. The final result was 15 industries, which are listed below with the ICB:

[35] http://www.nasdaqomx.com/listing/europe/primarylisting/Industriessegmentandindexes/

Table 9.33
Industry Definitions

OUR INDUSTRIES	NASDAQ OMX INDUSTRIES
Capital Goods	Oil & Gas
Chemicals & Energy	Basic Materials
Financial Services	Industrials
Forest Industry	Consumer Services
Health Care & Pharmaceutical	Consumer Goods
Household Goods	Health Care
Investment & Holding	Financials
Consulting Services	Technology
Media	Telecommunications
Real Estate	Utilities
Retail	
Services	
Technology	
Transport	
Vehicles & Machines	

As can be seen from the list the industry definitions are quite similar, thus we deem that our list is equally applicable to make inference. However, it is important to note that the ICB only have 10 industry definitions in comparison to the 15 most frequent industries, which does affect sample sizes and distributions across industries.