MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

"The distribution of stock returns and its implications"

verfasst von / submitted by

Mag. rer. nat. Richard Artner, Bsc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. Oec.)

Wien, 2017/ Vienna, 2017

Table of contents

List of figures

List of tables

1. Introduction

Ever since the first stock exchanges opened in the 18th century[1], traders, in order to make money, as well as mathematicians, statisticians and economists, out of academic interest, wanted to understand the behavior of stock indices. A very important hypothesis that has its roots in the 19th century[2] and was famously formalized by Eugene Francis Fama is the *Efficient‑market hypothesis* (Fama, 1970). It states that the current prize of a stock is fair in the sense that one cannot expect to win or lose by buying or selling shares of it. Hence, if this hypothesis holds true, it is not possible to outperform the market. Models in agreement with the Efficient-market hypothesis are the *random walk* models developed by Bachelier (1900) and expanded and refined by Osborne (1959) and the *Martingale* theory developed by Pearson (1905) and taken up by Samuelson (1973). Many random walk models assume that the random variation, governing the price changes of stock indices, is (a continuous function of) *Gaussian white noise*, the most famous being the *Black-Scholes-Merton model* (Black and Scholes, 1973). This model is still used these days to calculate the price of a European call option and it assumes that stock indices follow a geometric *Brownian motion*. The existence of the underlying stochastic process of Bachelier´s as well as Black, Scholes and Merton´s model, called *Brownian motion* or *Wiener process,* was proven by Wiener (1923).

Most random walk models are good approximations of the behavior of stock indices, however, they cannot adequately describe the behavior of stocks because future price changes are not independent of past changes, the underlying stochastic process is not stationary (in particular the volatility is not constant in time) and stock price changes are seldom log-normally let alone normally distributed. There is a huge body of literature on the random walk hypothesis (and the efficient market hypothesis) and the evidence against it grows each year (e.g. Lo and MacKinlay 1988, 2002). The first major alternative hypothesis, the *stable (Paretian) distribution hypothesis*, was developed by Mandelbrot (1963) and Fama (1965). Mandelbrot noticed, while analyzing cotton prices, that it is virtually impossible to distinguish the stock charts of daily, weekly, monthly and annual prices if the axes were unlabeled. This fractional behavior of stock indices led him to use the *Levy-alpha-stable* distribution family, first studied by Levy (1928), with their intrinsic property that sums of independent and identically distributed random variables out of that family have the same distribution (up to scale and location) as the respective random

---

1   The London Stock Exchange opened in 1977, the New York Stock Exchange in 1792.

2   For a detailed summary of the Efficient-Market hypothesis the reader is referred to the well-researched article by Martin Sewell (Sewell, 2011).

variables, to fit the logarithmic price changes (i.e. log returns). The Normal distribution is the only member of that family that has a defined second moment. All other members are leptokurtic, which means that they have more probability mass in the center and in the tails than a Normal distribution. One advantage of the Levy-alpha-stable distribution family is that it allows for skewness in the data, an important improvement over the random walk models since the distribution of negative price returns seems to differ from the distribution of positive ones (e.g. Sheik and Qiao, 2010). Although Fama (1965), Mandelbrot (1967) and Roll (1970) provided some more empirical evidence in the 60´s, more and more evidence piled up against that hypothesis, for instance in the works of Akgiray and Booth (1988), Lau, Lau and Wingender (1990) and Kunst, Reschenhofer and Rodler (1991). Just as the simpler random walk models, the stable distribution model cannot explain dependencies (and correlations) between price changes. Along with the fact, that an undefined second moment is implausible, this model is just another good approximation of the behavior of stocks. Longin (1996) provides evidence that the true (unknown and permanently changing) stock return generating process lies somewhere between the Bachelier and the Mandelbrot model.

At the core of all random walk models lies the distribution of stock returns. This work analyzes the daily returns of two major stock indices (Chapter 2). Read (1906, page 310) famously wrote: "It is better to be vaguely right than exactly wrong." The main approach of this work is to be approximately right and to avoid being precisely wrong, an approach certainly not pursued by a large portion of the scientific community, particularly in the social sciences. Following the Great Benoit B. Mandelbrot, graphical analysis forms the core of this work. After a systematic dissection of the properties of the two stocks (Chapter 2.1 - 2.3), continuous distributions were fitted to their daily returns (Chapter 2.4). Furthermore, in order to assess the implications of the properties of the daily returns, a simulation study was conducted (Chapter 3).

2. Analysis of stock returns

2.1. Datasets

This work provides an in-depth analysis of two major stock indices, the Standard & Poor 500 (S&P 500) and the Nikkei 225. Together with the Dow Jones Industrial Average the former is considered as the most important compound stock index because they both represent the biggest companies of the biggest stock exchange of the world, the American Stock Exchange ("wall street" in colloquial speech) with a market capitalization of about 18.5 trillion US dollars (Visual Capitalist, 2017). The S&P 500 also contains companies listed on NASDAQ, the second biggest stock exchange in the world with a market capitalization of about $ 7.5 trillion (Visual Capitalist, 2017).

The S&P 500 contains (as the name suggests) 500 large US companies. Those 500 companies satisfy among others the following criteria (Investopedia, 2017):

- A market capitalization of at least $ 5.3 billion
- The headquarter has to be in the United States of American
- Most of its shares have to be in the public´s hands

The S&P 500 contains among others companies from the following sectors (CNN Money, 2017):

- Healthcare
- Information technology
- Financials
- Real Estate
- Energy

The Nikkei 255 of Tokyo´s Stock Exchange, the third largest stock exchange with a market capitalization of $ 7.5 trillion, is considered as the most important Asian compound stock index. It includes 225 Japanese companies in a multitude of different sectors, for instance (CNN Money, 2017):

- Foods
- Automotive
- Chemicals
- Financials

The daily closing prices (P_t) of the S&P 500 (from 1950-11-20 to 2016-11-18) and the Nikkei 225 (from 1984-01-05 to 07-12-2016) have been downloaded from Yahoo (Yahoo Finance, 2017). Those two datasets were then imported into R (R Core Team, 2016) and the **daily returns (R_t)**

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

as well as the **daily log returns (L_t)** were calculated.

$$L_t := \log(R_t + 1) \overset{(1)}{=} \log\left(\frac{P_t}{P_{t-1}}\right) = \log(P_t) - \log(P_{t-1}) \tag{2}$$

L_t is equal to R_t if the latter is zero and smaller otherwise. However, for R_t values close to zero, the difference between L_t is very small, which can readily be seen via Taylor series decomposition of the natural logarithm centered at one. The daily returns of the S&P 500 and the Nikkei 225 are all smaller than 0.25 and mostly smaller than 0.05. Hence, in all subsequent analysis, the **daily percentage changes** (i.e. $100*R_t$) are used instead of the daily returns to simplify the representation and the interpretation of the relative index changes.

2.1.1. S&P 500

Figure 1 shows the S&P 500 from the 20th of November, 1984 to the 18th of November, 2016 as well as the daily percentage changes. For the plot of the S&P 500 a semi-logarithmic scale is used, hence a linear increase of the trajectory corresponds to an exponential increase of the S&P 500. There is huge variation in the daily percentage changes (Figure 1.b) with both large positive and large negative changes. In particular, the 20.47% decrease on the 19th of October, 1987 sticks out. On this infamous day, which is often called "Black Monday", stock markets around the world experienced dramatic losses which almost all financial experts at that time did not consider possible.

2.1.2. Nikkei 225

Figure 2 shows the Nikkei 225 from the 5th of January, 1984 to the 7th of December, 2016 as well as the daily percentage changes. In contrast to the S&P 500, no clear upward trend between 1984 and 2016 is visible. The peak in 1989 was never reached again and the stock index at the end of this time series is less than half as big as it was at its peak. Just like in the case of the S&P 500, Figure 2.b shows very strong daily percentage changes every now and then. The strongest percentage change (a loss of 14.9%) occurred on the 20th of October, in the events of the aforementioned "Black Monday"[3].

―――――――――――――――

3   Due to time lag the worst one day percentage losses of the S&P 500 and the Nikkei 225 differ by less than 24 hours.

Standard & Poor 500 Index
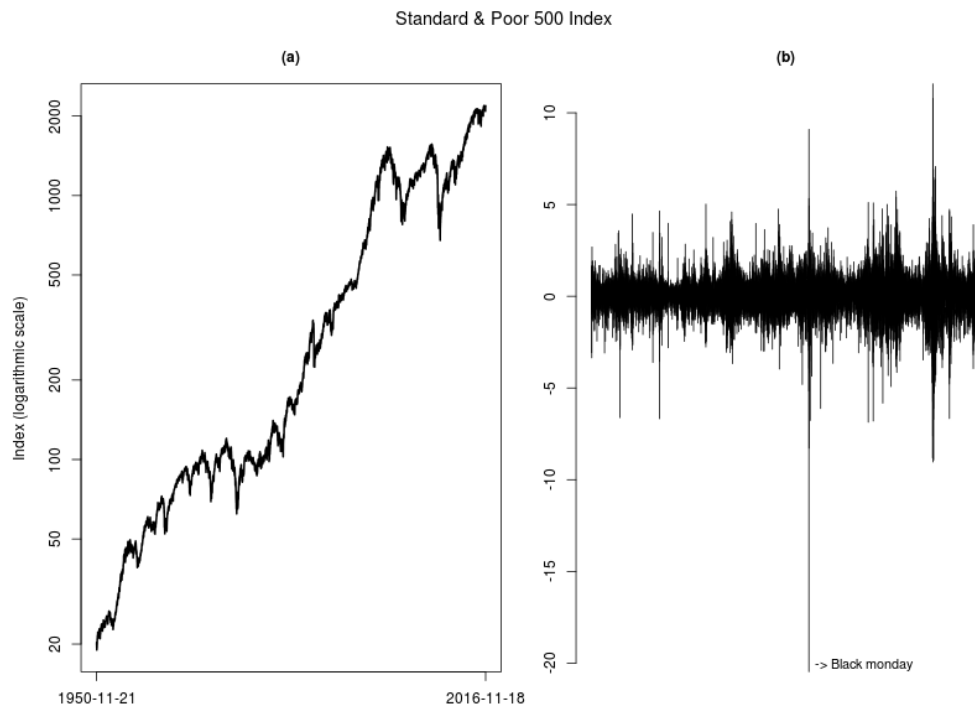
Figure 1: The S&P 500 index from 1950-11-21 to 2016-11-18 (a) and its daily percentage changes (b).
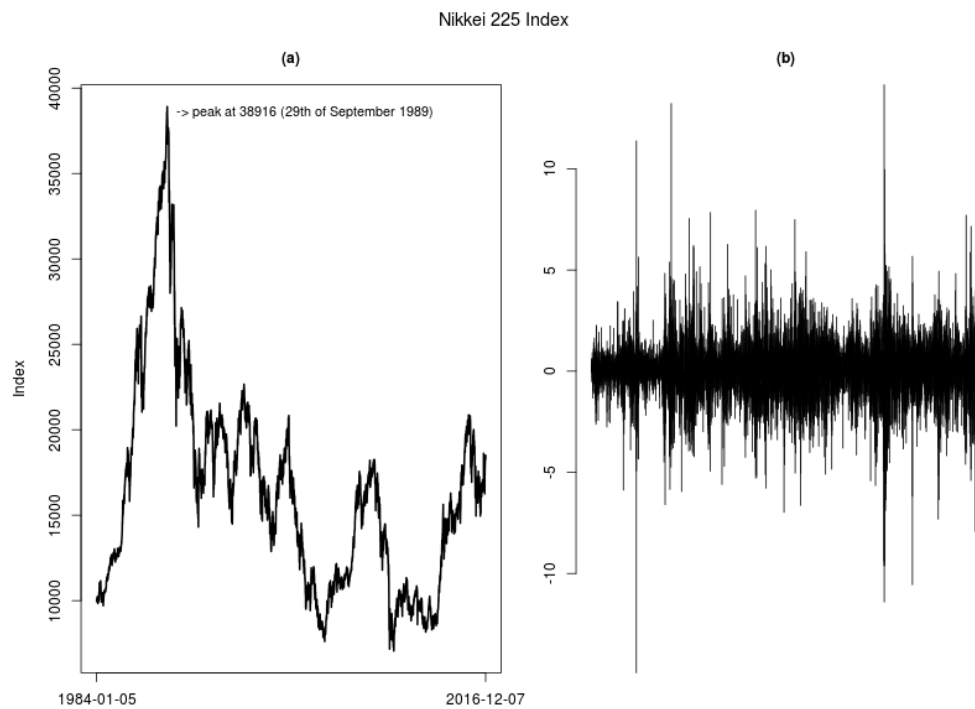
Nikkei 225 Index

Figure 2: *The Nikkei 225 index from 1984-01-05 to 2016-12-07 (a) and its daily percentage changes (b).*

2.2. Are the daily percentage changes i.i.d.?

How plausible is the assumption that the behavior of some stock index on any given day is independent of its changes in the antecedent days, weeks or months? How about the independence of a stock index and some other stock index which may or may not come from a different stock exchange, maybe even from a different country? The answer to the former question is beyond reasonable doubt: NO! One way to rebut the assumption of independence is the detection of cycles. A multitude of cycles have been analyzed in the scientific literature and some even created catchphrases such as "Sell in May and go away!" The interested reader is referred to Andrade, Chhaochharia and Fuerst (2012) for this particular seasonal effect. Other well-known seasonal dependencies are the "January effect" analyzed among others by Haugen and Jorion (1996) and the "presidential cycle effect" (Gärtner & Wellershoff, 1995). As shown by Peiro (1994) cycle effects can also persist for short periods.

There also exist various dependencies among different stock indices around the globe and in turbulent times like in the events around "Black Monday" or the financial crisis of 2008, almost all stock indices lose in value. Furthermore, there exist various negative dependencies were an increase of some asset comes along with a decrease in another asset. The fact that the prize of commodities, gold in particular, and the exchange rate of the US dollar are negatively correlated is often used in hedging strategies and is analyzed in great detail by Ciner, Gurdgiev and Lucey (2013).

Equation (3) shows that due to definition (1) the closing price P_t is fully determined by some starting value P_1 and all daily returns between 1 and t:

$$P_t \overset{(1)}{=} (R_t+1)\cdot P_{t-1} \overset{(1)}{=} (R_t+1)\cdot(R_{t-1}+1)\cdot P_{t-2} \overset{(1)}{=}...= P_1\cdot\prod_{i=2}^{t}(R_i+1) \qquad (3)$$

Given the current stock value P_t, the random variable P_{t+k} is the product of random increments (R_i+1) for i=t+1,t+2,…,t+k. Hence, the stochastic process {P_{t+k}: k=1,2,3,…} is a *multiplicative random walk* in discrete time if the daily returns R_i are independent and identically distributed (i.i.d.) random variables. Certain properties of the daily returns transfer to the log returns:

$$(4.1): R_i \; i.i.d. \; \Rightarrow L_i \overset{(2)}{=} \log(R_i+1) \; i.i.d. \; since \; f(x)=\log(x+1) \; is \; bijective \; \forall \, x>0$$
$$(4.2): \forall i\in\mathbb{N}: E(R_i)<\infty \Rightarrow E(L_i)=: \mu<\infty \; since \; 0<\log(x+1)<x \; \forall \, x>0$$
$$(4.3): \forall i\in\mathbb{N}: Var(R_i)<\infty \Rightarrow Var(L_i)\overset{def}{=}E(L_i^2)-\mu^2=: \sigma^2<\infty \; since \; 0<\log(x+1)<x \; \forall \, x>0$$
$$Furthermore: \forall \, i\neq j: Cov(L_i,L_j):=E((L_i-E(L_i))\cdot(L_j-E(L_j)))\overset{(4.2)}{=}E(L_i\cdot L_j)-\mu^2\overset{(4.1)}{=}0$$

(4)

In order to analyze the plausibility of (4), two important functions are hereby defined:

$$S_p(t) := \sum_{i=2}^{t-p} L_{i+p} \cdot L_i \qquad cS_p(t) := \sum_{i=2}^{t-p} \left(L_{i+p} - \frac{\sum_{k=2}^{t} L_k}{(t-1)}\right) \cdot \left(L_i - \frac{\sum_{k=2}^{t} L_k}{(t-1)}\right) \qquad \forall \, 0 \le p \le t-2 \qquad (5)$$

The process $cS_p(t)$ is the sample autocovariance estimator of order p multiplied by the number of used log returns (i.e. t-p-1). If p is set to zero, $cS_p(t)$ is equal to the sample variance of the first (t-1) log returns multiplied by (t-1). $S_p(t)$ is a non-centralized modification of $cS_p(t)$.

To compare the fluctuations of the S&P 500 with those of the Nikkei 225 dataset a scale invariant measure of variation is of need. For a random variable (r.v.) X the **squared coefficient of variation** is defined as:

$$CV^2(X) := Var\left(\frac{X}{E(X)}\right) = \frac{Var(X)}{(E(X))^2} \quad if \ Var(X) < \infty \qquad (6)$$

If properties (4.1), (4.2) and (4.3) are satisfied, results (7.1) – (7.14) follow!

Result (7.1) shows that the summed log returns are equal to the logarithm of P_t minus the constant $\log(P_1)$ at each point in time. Hence, the process $\{\log(P_{t+k}): k=1, 2, 3,...\}$ is an *additive random walk* in discrete time if properties (4.1), (4.2) and (4.3) are satisfied. Result (7.10) shows that $cS_p(t)$ divided by (t-p-1) would be a biased estimator for the true autocovariance of order p, which is zero no matter the p because of (4). Whereas the expected value of $cS_p(t)$ converges to $-\sigma^2$ (7.10), the expected value of the off-centered process $S_p(t)$ diverges to $+\infty$ if μ is unequal to zero (7.9). The variance of $cS_p(t)$ only depends on the first and second moment of L_i if p is unequal to zero (7.12). The variance of $cS_0(t)$, on the other hand, also depends on the fourth moment of L_i (7.8). In order to assess the plausibility that the daily returns are i.i.d. random variables with finite second moments the cumulative sums of the log returns, as well as $S_p(t)$ and $cS_p(t)$ for p=0,1,2,3, are plotted for the two indices (Figures 3-6).

(7.1): $\sum_{i=2}^{t} L_i \overset{(2)}{=} \sum_{i=2}^{t}(\log(P_i)-\log(P_{i-1})) \overset{(2)}{=} \log(P_t)-\log(P_1) \overset{(2)}{=} \log(\frac{P_t}{P_1})$

(7.2): $E(\sum_{i=2}^{t} L_i) \overset{(4.2)}{=} (t-1)\cdot\mu$

(7.3): $E(S_0(t))=\sum_{i=2}^{t} E(L_i^2) \overset{(4.3)}{=} (t-1)\cdot(\sigma^2+\mu^2)$

(7.4): $E((\sum_{i=2}^{t} L_i)^2)=\sum_{i=2}^{t} E(L_i^2)+\sum_{i=2}^{t}\sum_{k=2,k\neq i}^{t} E(L_i\cdot L_k) \overset{(4,7.3)}{=} (t-1)\cdot(\sigma^2+\mu^2)+(t-1)\cdot(t-2)\cdot\mu^2$

$=(t-1)\cdot\sigma^2+(t-1)^2\cdot\mu^2$

(7.5): $Var(\sum_{i=2}^{t} L_i) \overset{def}{=} E((\sum_{i=2}^{t} L_i)^2)-(E(\sum_{i=2}^{t} L_i))^2 \overset{(7.2,7.4)}{=} (t-1)\cdot\sigma^2+(t-1)^2\cdot\mu^2-(t-1)^2\cdot\mu^2=(t-1)\cdot\sigma^2$

(7.6): $E(cS_0(t))=E(\sum_{i=2}^{t} L_i^2-\frac{(\sum_{i=2}^{t} L_i)^2}{t-1}) \overset{(7.3,7.4)}{=} (t-1)\cdot(\sigma^2+\mu^2)-\frac{(t-1)\cdot\sigma^2+(t-1)^2\cdot\mu^2}{t-1}=(t-2)\cdot\sigma^2$

(7.7): $E((S_0(t))^2)=E(\sum_{i=2}^{t} L_i^2\cdot\sum_{k=2}^{t} L_k^2)=\sum_{i=2}^{t} E(L_i^4)+\sum_{i=2}^{t}\sum_{k=2,k\neq i}^{t} E(L_i^2\cdot L_k^2)$

$\overset{(4)}{=} (t-1)\cdot E(L_1^4)+(t-1)\cdot(t-2)\cdot(\sigma^2+\mu^2)^2$

(7.8): $Var(S_0(t)) \overset{def}{=} E((S_0(t))^2)-(E(S_0(t)))^2 \overset{(7.3,7.7)}{=} (t-1)\cdot[E(L_1^4)-(\sigma^2+\mu^2)^2]$

(7.9): $E(S_p(t))=\sum_{i=2}^{t-p} E(L_{i+p}\cdot L_i) \overset{(4)}{=} (t-p-1)\cdot\mu^2 \quad \forall\, 1\leq p\leq t-2$

(7.10): $E(cS_p(t))=\sum_{i=2}^{t-p}[E(L_{i+p}\cdot L_i)-\frac{1}{(t-1)}\sum_{k=2}^{t} E(L_{i+p}\cdot L_k)-\frac{1}{(t-1)}\sum_{k=2}^{t} E(L_i\cdot L_k)+\frac{1}{(t-1)^2} E((\sum_{k=2}^{t} L_k)^2)]$

$\overset{(4,7.4)}{=} \sum_{i=2}^{t-p}[\mu^2-\frac{2}{(t-1)}\cdot((t-2)*\mu^2+(\sigma^2+\mu^2))+\frac{\sigma^2}{(t-1)}+\mu^2]=(t-p-1)\cdot\frac{-\sigma^2}{(t-1)} \quad \forall\, 1\leq p\leq t-2$

(7.11): $E((S_p(t))^2)=E((\sum_{i=2}^{t-p} L_{i+p}\cdot L_i)\cdot(\sum_{k=2}^{t-p} L_{k+p}\cdot L_k))$

$=\sum_{i=2}^{t-p} E(L_{i+p}^2\cdot L_i^2)+2\sum_{i=2+p}^{t-p} E(L_{i+p}\cdot L_i^2\cdot L_{i-p})+2\sum_{i=3}^{t-p}\sum_{j\neq k,i\neq k+p,k=2}^{t-p-1} E(L_{i+p}\cdot L_i\cdot L_{k+p}\cdot L_k)$

$\overset{(4)}{=} (t-p-1)\cdot(\sigma^2+\mu^2)^2+2\cdot(t-p-p-1)\cdot\mu\cdot(\sigma^2+\mu^2)\cdot\mu+((t-p-1)\cdot(t-p-4)+2p)\cdot\mu^4$

$=\sigma^4\cdot(t-p-1)+\sigma^2\cdot\mu^2\cdot(4t-6p-4)+\mu^4\cdot[(t-p-1)\cdot((t-p-4)+1+2)+2p-2p] \quad \forall\, 1\leq p\leq\frac{(t-2)}{2}$

(7.12): $Var(S_p(t)) \overset{def}{=} E((S_p(t))^2)-(E(S_p(t)))^2 \overset{(7.9,7.11)}{=} (t-p-1)\cdot(\sigma^4+4\sigma^2\cdot\mu^2)-2p\cdot\sigma^2\cdot\mu^2$

$+\mu^4\cdot(t-p-1)^2-(t-p-1)^2\cdot\mu^4=(t-p-1)\cdot(\sigma^4+4\sigma^2\cdot\mu^2)-2p\cdot\sigma^2\cdot\mu^2 \quad \forall\, 1\leq p\leq\frac{(t-2)}{2}$

(7.13): $CV^2(\log(\frac{P_t}{P_1})) \overset{(6,7.1,7.2,7.5)}{=} \frac{(t-1)\cdot\sigma^2}{(t-1)^2\cdot\mu^2}=\frac{\sigma^2}{\mu^2}\cdot\frac{1}{(t-1)}$

(7.14): $CV^2(S_p(t)) \overset{(6,7.9,7.12)}{=} \frac{(t-p-1)\cdot(\sigma^4+4\sigma^2\cdot\mu^2)-2p\cdot\sigma^2\cdot\mu^2}{(t-p-1)^2\cdot\mu^4}=(\frac{\sigma^4}{\mu^4}+\frac{4\sigma^2}{\mu^2})\cdot\frac{1}{(t-p-1)}-\frac{2p\cdot\sigma^2}{\mu^2(t-p-1)^2}$

The sum of the log returns increases over time in the case of the S&P 500 (Figure 3.a). An underlying random walk with a small positive drift μ seems certainly possible. Figure 3.b shows that $cS_0(t)$ and $S_0(t)$ are indistinguishable by the naked eye. This means that the variance σ^2 is relatively strong in comparison to the drift μ since the relative distance of the expectations of $S_0(t)$ (7.3) and $cS_0(t)$ (7.6) depends on the ratio of σ^2 and μ^2 (Table 1). $cS_0(t)$ (and $S_0(t)$) are a lot less stochastic than the sum of the log returns (Figure 3.a) which means that the 4th moment, responsible for the variance of the squared log returns (7.8), is relatively weak in comparison to σ^2. The average increase of $cS_0(t)$ in the last three decades is stronger as in the 50´s, 60´s and 70`s. This indicates that σ^2 is not constant over the whole time frame, but bigger in the second half of the time series.
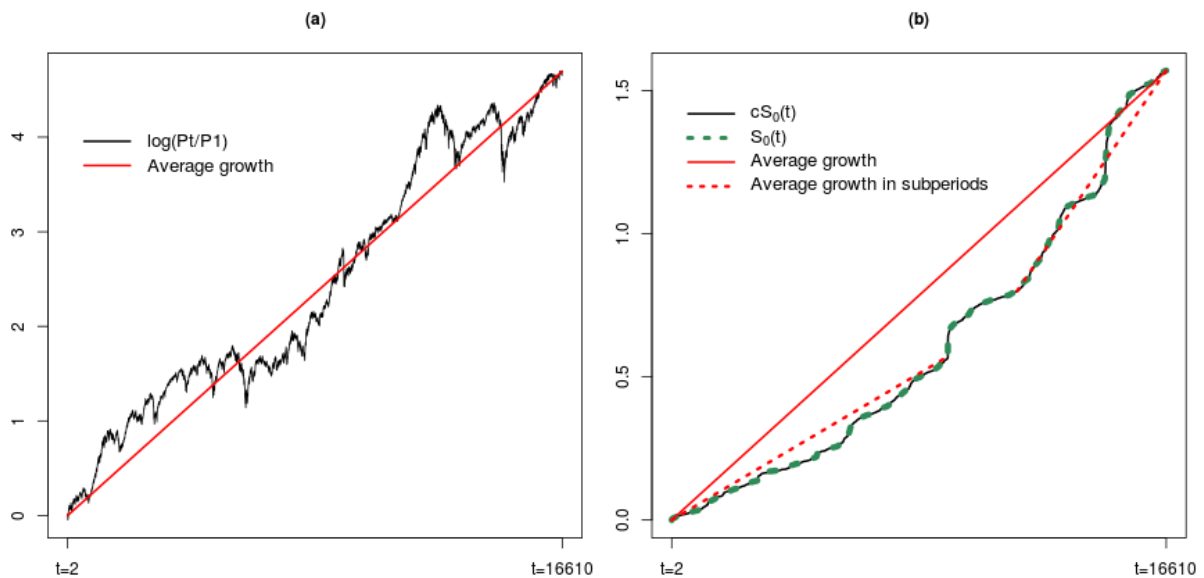
Figure 3: *The sum of the daily log returns (a) and the sum of the (centralized) squared log returns (b) of the S&P 500.*

A constant drift term μ seems more unlikely for the Nikkei 225 (Figure 4.a). The sum of the log returns rises relatively constant for the first 1000 days or so, after which it starts to wildly fluctuate until day 8109. It seems as though the drift is positive in the beginning and in the end, but negative in the middle of this time series. Just as in the case of the S&P 500, $cS_0(t)$ and $S_0(t)$ are practically equal at each time point t (Figure 4.b). Hence, σ^2 is relatively strong in comparison to μ (Table 1). $cS_0(t)$ (and $S_0(t)$) rise relatively constant as (7.3) and (7.6) would predict under assumptions (4). The sharp increase of $cS_0(t)$ during the 2008 financial crisis distorts the picture. Such a "volatility cluster" is not uncommon in stock markets. A "volatility cluster" is a period of time where large price changes occur more (in)frequently as usual or in the words of the great Benoit Mandelbrot (1963): "...large changes tend to be followed by large changes-of either sign-and small changes tend to be followed by small changes..." A good way

to analyze volatility clustering in financial data is to substitute the log returns L_i with $|L_i|$ or L_i^2 in (5). The interested reader is referred to Cont (2007) who offers a detailed analysis of volatility clustering in financial time series.
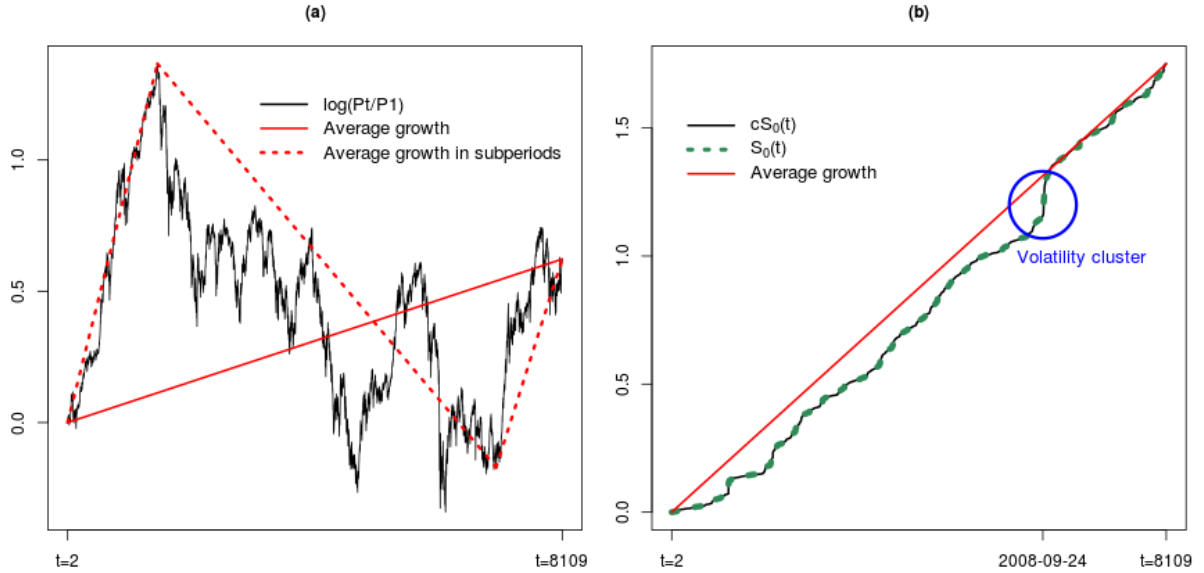
Figure 4: *The sum of the daily log returns (a) and the sum of the (centralized) squared log returns (b) of the Nikkei 225.*

For p=1,2,3, $S_p(t)$ and $cS_p(t)$ are practically equal for both indices (Figures 5 & 6). Given (4) it was shown in (7.9) and (7.10), that the expectation of $cS_p(t)$ converges to $-\sigma^2$ and the expectation of $S_p(t)$ diverges to $+\infty$ if μ is unequal to zero. However, both μ and σ^2 are probably very close to zero and, hence, the expected value of $cS_p(t)$ and $S_p(t)$ is almost identical given the finite sample sizes (Table 1). μ and σ^2 were estimated via the sample mean and the sample variance of the log returns:

$$\hat{\mu} := \frac{\sum\limits_{i=2}^{t_{max}} L_i}{(t_{max}-1)} \quad \hat{\sigma}^2 := \frac{cS_0(t_{max})}{(t_{max}-1)} = \frac{\sum\limits_{i=2}^{t_{max}} L_i^2}{(t_{max}-1)} - (\hat{\mu})^2 \quad where \ t_{max} := number\ of\ days \quad (8)$$

The CV² of the summed log returns (7.13) and $S_p(t)$ (7.14) can now be estimated by replacing μ and σ^2 with the respective estimator (8). Table 1 shows the computation of (8) for the S&P 500 and the Nikkei 225 as well as the estimated decisive components of (7.13) and (7.14). Whereas the estimated σ^2 of the Nikkei 225 is just about twice as big as those of the S&P 500, the estimated (7.13) is more than 30 times as big for the same time point t and the estimated decisive factor of (7.14) is almost 1000 times as big.

	S&P 500	Nikkei 225
t_{max}	16610	8109
$\hat{\mu}$	$2.83*10^{-4}$	$7.68*10^{-5}$
$\hat{\sigma}^2$	$9.45*10^{-5}$	$2.16*10^{-4}$
$\dfrac{\hat{\sigma}^2}{\hat{\mu}^2}$	$1.18*10^{3}$	$3.67*10^{4}$
$\dfrac{\hat{\sigma}^4}{\hat{\mu}^4}+4\cdot\dfrac{\hat{\sigma}^2}{\hat{\mu}^2}$	$1.40*10^{6}$	$1.34*10^{9}$

Table 1: *Mean and variance estimations of the log returns as well as an estimation of the decisive constant factors in the CV² of log(Pt_{max}/P_1) and $S_p(t)$.*

The behavior of the $cS_1(t)$ (and the $S_1(t)$) seems to change after "Black Monday" for the S&P 500 dataset. All (centralized) autocovariance processes drastically change in the events around "Black Monday" and during the financial crisis of 2008. Those drastic changes clearly argue against a random walk hypothesis with properties (4).
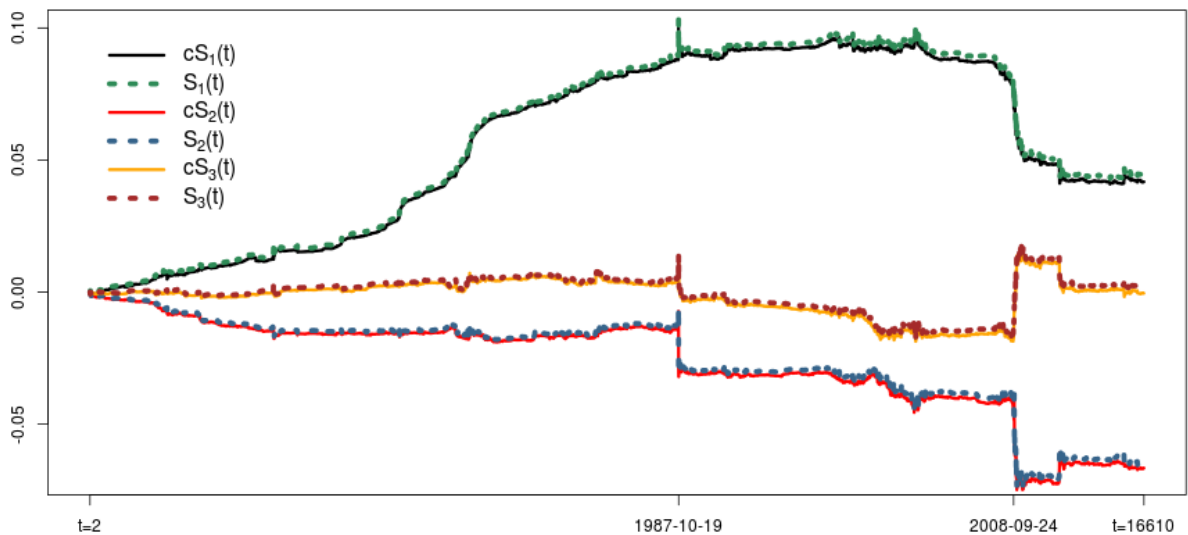
Figure 5: *The (centralized) autocovariance process (5) of the S&P 500 for p=1,2,3.*

11

In accordance with their estimated CV² (Table 1), the (centralized) autocovariance processes fluctuate much stronger in the case of the Nikkei 225 (Figure 6). Similar to the S&P 500, the fluctuation of the processes is greatly increased around "Black Monday" and the 2008 financial crisis. Unlike the S&P 500 (Figure 5), $cS_1(t)$ is negative for later time points. This means that a positive change on any given day makes a negative change on the subsequent day more likely and vice versa.
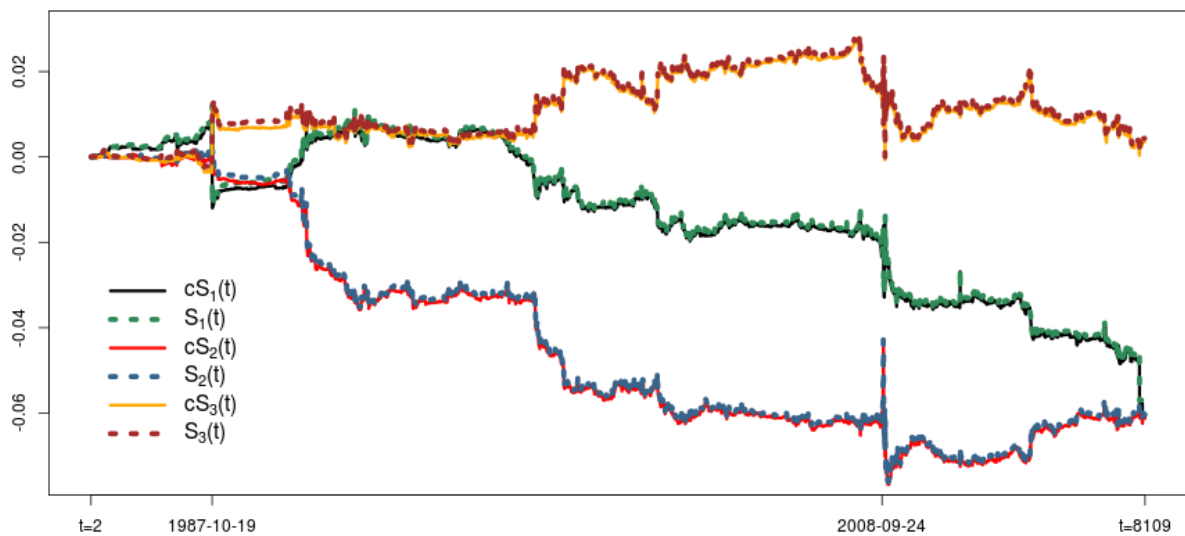
Figure 6: *The (centralized) autocovariance process (5) of the Nikkei 225 for p=1,2,3.*

This chapter can be concluded with the conviction that the (log) returns of the S&P 500 and the Nikkei 225 (and any other stock for that matter) are not independent realizations of some random variable. However, in statistics, particularly in the field of stochastic processes, i.i.d. realizations seldom arise and the question of interest is how good an i.i.d. approximation is. The goodness of fit of certain i.i.d. approximations for the daily percentage changes of the S&P 500 and the Nikkei 225 will be evaluated in Chapter 2.4.

2.3. How fat are the tails of the distribution?

2.3.1. Law of large numbers

One important, if not the most important, family of statistical theorems is the family of laws of large numbers (LLN). The generalization of this family, ergodic theory, constitutes an important mathematical branch. All theorems in that family show that, if certain properties are met, some time average converges (in some stochastic way) to some ensemble (space) average (e.g. the expected value of some random variable) as the sample size increases. The most important theorem in this family is the **Strong Law of Large Numbers (SLLN)** for i.i.d. random variables:

$$If\ X_i\ i.i.d.\ with\ E(X_1) < \infty\ then\colon\ Probability\left[\frac{\sum\limits_{i}^{n} X_i}{n} \to E(X_1)\ for\ n \to \infty\right] = 1 \qquad (9)$$

The power of (9) is that an approximation of the unknown expected value via sample mean is sound and the precision of this approximation can be increased ad infinitum by incorporating new data (i.e. increasing the sample size). The process of the **pth (raw) sample moment** of the daily percentage changes is given by:

$$M_p(t) := \frac{\sum\limits_{i=2}^{t} (100 \cdot R_i)^p}{(t-1)}\quad for\ p \in \mathbb{N} \qquad (10)$$

Figures 7 & 8 show the progression of (10) with p=1,2,3,4 for the S&P 500 (Figure 7) and the Nikkei 225 (Figure 8). For both indices a clear decrease of variation (i.e. convergence) can only be observed for the first sample moment (Figures 7.a & 8.a). The impact of "Black Monday" on (10) is gigantic, especially for the respective 3rd and 4th sample moments.

Does the size of the relative index change differ on days with negative and positive returns? Are large losses in value more or less likely than large gains? In order to compare the distribution of negative and positive daily percentage changes, the two datasets were respectively divided into all days with a positive return (**increase subsample**) and all days with a negative return. The absolute values of all negative returns then constituted the respective **decrease subsample**. The 115 days of the S&P 500 with a return equal to zero were discarded. The Nikkei 225 sample only contains nonzero returns.
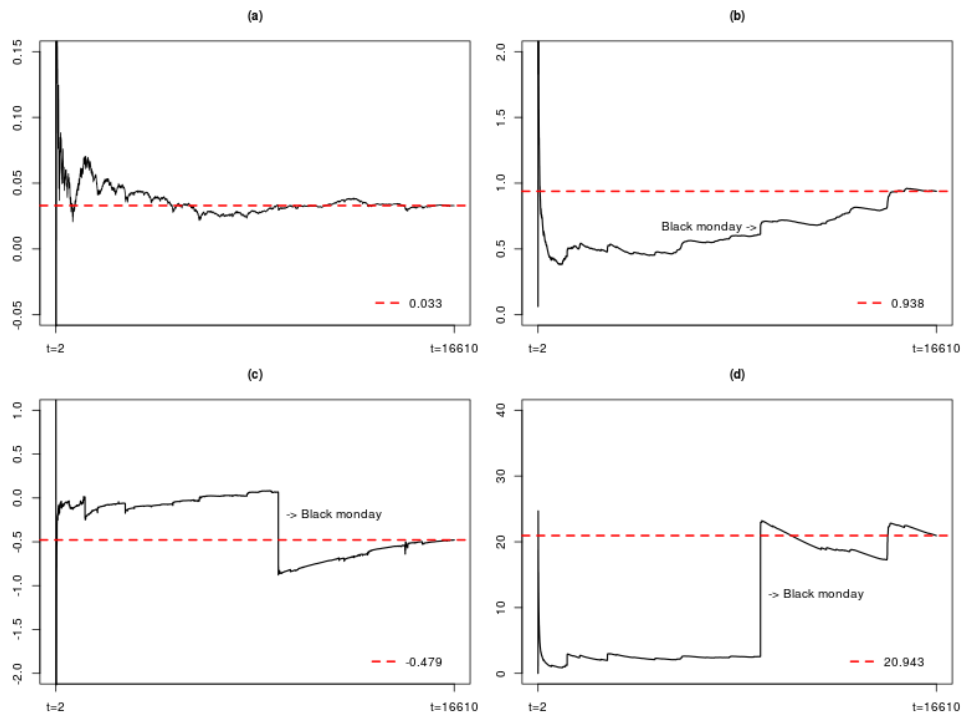
Figure 7: Process (10) with p=1 (a), p=2 (b), p=3 (c) and p=4 (d) for the S&P 500.
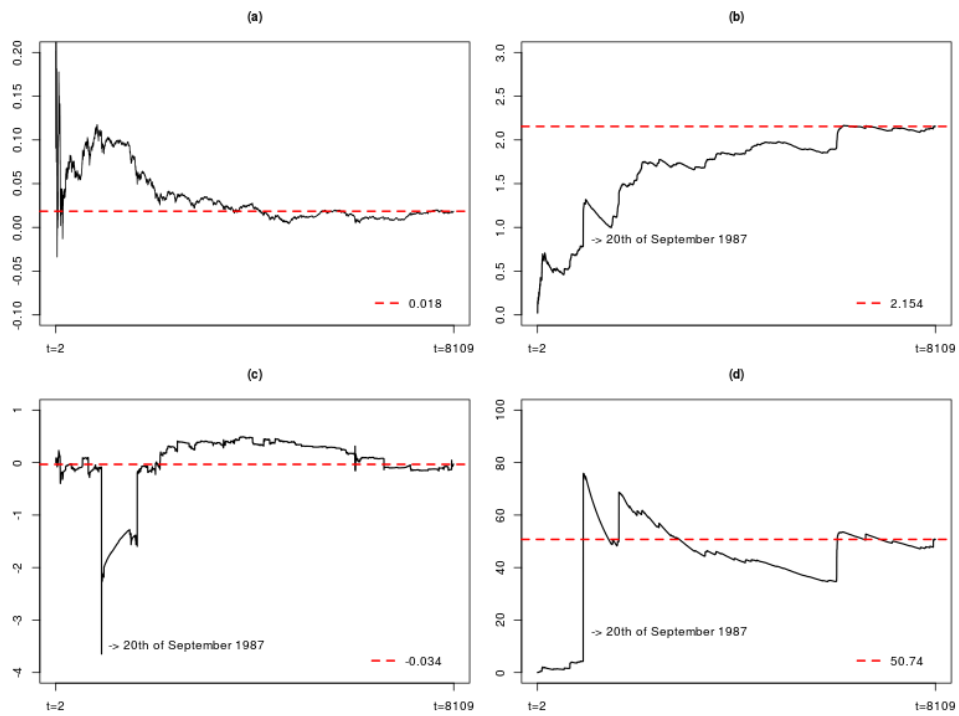
Figure 8: Process (10) with p=1 (a), p=2 (b), p=3 (c) and p=4 (d) for the Nikkei 225.

Can a convergence of higher moments, as in (9), be ascertained for these subsamples? According to Cirillo and Taleb (2016), the following holds true:

"For a sequence $X_1, X_2, \ldots, X_n$ of nonnegative i.i.d. random variables, if $E[X^p] < \infty$ for

p = 1, 2, 3..., then $Q_n^p = M_n^p/S_n^p \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$, where $S_n^p = \sum_{i=1}^{n} X_i^p$ is the partial sum,

and $M_n^p = max(X_1^p, \ldots, X_n^p)$ the partial maximum." (*Result 1*)[4]

In other words, the largest observations share of the sum of all instances gets smaller (not necessarily monotonically) with an increasing sample size. One diagnostic tool that makes use of this fact is the *Maximum-to-Sum* plot, a two-dimensional plot with the sample size n on the abscissa and Q_n^p (for some p = 1, 2, 3..) on the ordinate. If no clear convergence to zero for Q_n^p can be observed one possible explanation is that the pth moment $E[X^p]$ does not exist. Another explanation is that the pth moment exists (i.e. $E[|X^p|] = E[X^p] < \infty$) but that the i.i.d. assumption is violated quite considerably. In any case, it is valid to conclude that $(1/N)*S_N^p$ is a bad estimator for $E[X^p]$ given the sample size N, if no clear convergence to zero is observed for Q_n^p with n=1,...,N. Figures 9 & 10 show the Maximum-to-Sum plots with p=1,2,3,4 for the respective increase and decrease subsamples of the S&P 500 and the Nikkei 255. Note that the nonnegative assumption of *Result 1* is always met, due to the construction of the decrease subsample.

The Maximum-to-Sum plots of the decrease subsample (N=7721) only show a clear convergence for Q_n^1 (Figure 9). Since all data points are non-negative M_N^1, M_N^2, M_N^3 and M_N^4 always realize their maximum on the same day for all n=1,...,N. In the S&P 500 dataset (Figure 9) the sample maxima of the first four moments (M_N^1, M_N^2, M_N^3, M_N^4) in case of a percentage decrease are equivalent to $((-X_{n*})^1, (-X_{n*})^2, (-X_{n*})^3, (-X_{n*})^4)$ where X_{n*} is equal to -20.47, the percentage change on the 19th of September, 1987 ("Black Monday"). This drop was so huge that it makes up more than 90% of S_{n*}^4 and approximately 70% of S_N^4, which is quite astonishing since the latter tells us that more than two-thirds of the 4th raw sample moments value comes from just one out of the 7721 days. The four Maximum-to-Sum plots corresponding to the increase subsample are smoother than those corresponding to a decrease (Figure 9). The biggest one-day jump (+11,58% on the 13th of October, 2008) has no big impact on Q_N^1 and Q_N^2 and its impact on Q_N^3 can be compared to "Black Monday´s" impact on Q_N^2 in the decrease subsample. Figure 9 suggests that a LLN like (9) seems to apply for the negative, the positive and the squared positive daily percentage changes. Estimations based on the 3rd and 4th sample moment (10) of the daily percentage changes (e.g. skewness and kurtosis) should only be used with extreme caution due to their strong dependence on just a few days.

---

4   The letter R was replaced by the letter Q in order to avoid a clash of notation in this work.
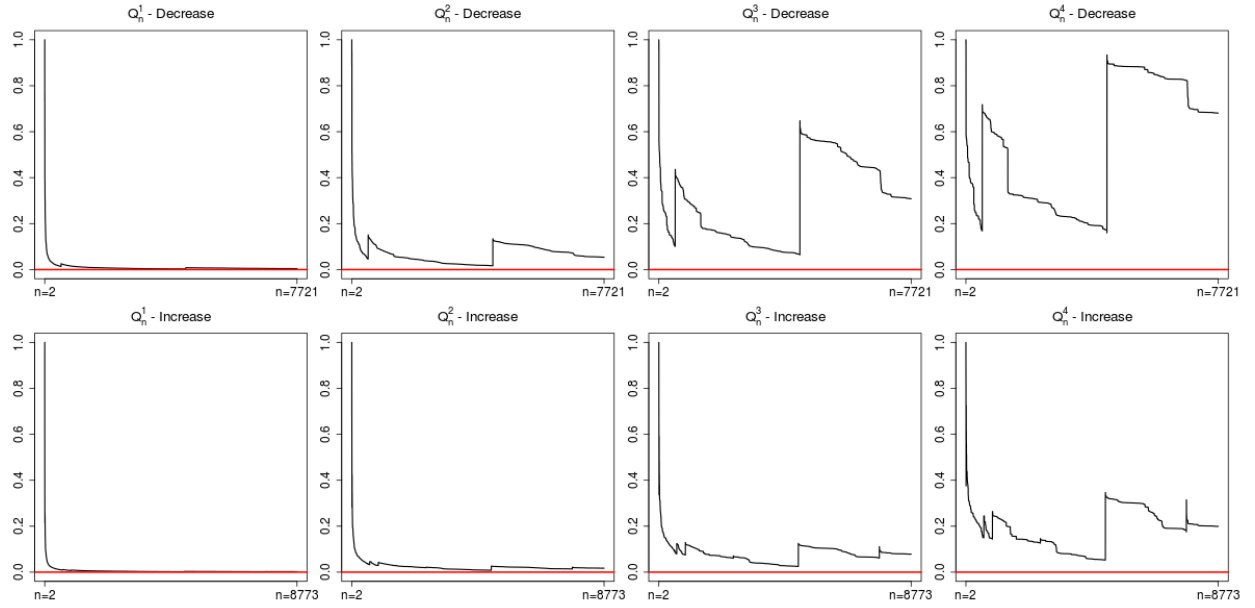
Figure 9: *Maximum-to-sum plots for the decrease and the increase subsample of the S&P 500 with p=1,2,3,4.*
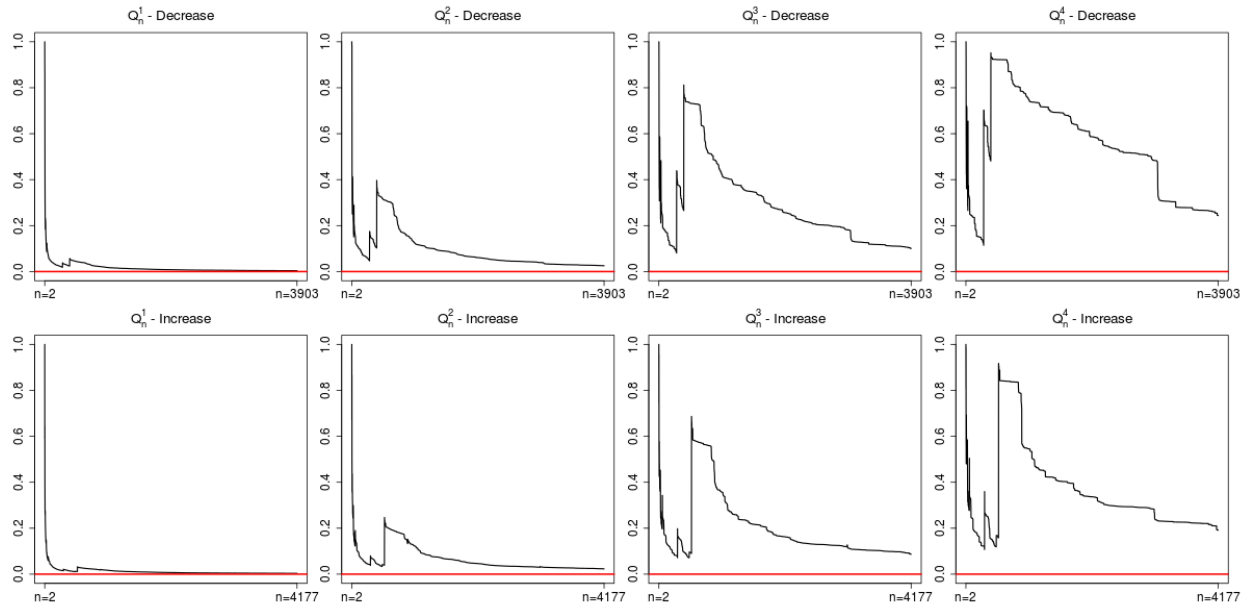
Figure 10: *Maximum-to-sum plots for the decrease and the increase subsample of the Nikkei 225 with p=1,2,3,4.*

The difference between the decrease and the increase subsample is smaller in the case of the Nikkei 225 (Figure 10). Just as in the case of the S&P 500 there are gigantic jumps[5] of Q_n^3 and Q_n^4 in both subsamples. After the first 1000 days, Q_n^2 converges relatively smoothly to values smaller than 0.05 and estimations based on S_N^2 (e.g. the sample variance) should, therefore, be considered as valuable in both subsamples. Just as in the case of the S&P 500, estimations based on the 3rd and 4th sample moment should only be used with extreme caution, if at all.

2.3.2. Zipf plots

One diagnostic tool regularly used to analyze the tail behavior of a univariate distribution has its roots in the works of the linguist George Kingsley Zipf. Zipf (1949) analyzed the frequency of different words and found that the probability of a certain word was proportional to the inverse of its rank in the frequency table. That means that the most common word is used about twice as often as the second and about three times as often as the third most common word. Similar observations in the studies of population ranks of cities, the distribution of wealth and the size of corporations were made and let to the development of discrete power law distributions. Only continuous probability distributions are used in this work (either as a fit for the datasets or for comparison reasons) and, therefore, only continuous power law distributions and their properties are of concern.

A r.v. X follows a **Pareto Type** I distribution if it has the following density function (pdf):

$$f(x) = \frac{k \cdot a^k}{x^{k+1}} \ 0 < a \leq x, x \in \mathbb{R} \ with \ parameter \ a > 0 \ (scale), k > 0 \ (shape); \ a, k \in \mathbb{R} \tag{11}$$

By integrating (11) over (a, x) we get the **cumulative distribution function (cdf)**

$$F(x) = 1 - \left(\frac{a}{x}\right)^k \ 0 < a \leq x \tag{12}$$

and the **survival function** of X:

$$\bar{F}(x) \overset{\text{def}}{=} 1 - F(x) = \left(\frac{a}{x}\right)^k = a^k \cdot x^{-k} \ 0 < a \leq x, \ a \in \mathbb{R} \tag{13}$$

---

5   A big jump down means that the percentage change on that day was almost but not quite as big as the currently biggest one.

The shape parameter k in (11) governs the thickness of the right tail. The scale parameter a is the left endpoint of the distribution. The smaller k, the thicker the right tail due to the slower decay of the survival function (13).

A distribution function F is called **scale-invariant** if there exist a continuous function g and a real number $x_0 > 0$ such that the survival function satisfies:

$$\bar{F}(c \cdot x) = g(c) \cdot \bar{F}(x) \quad \forall x, c \in \mathbb{R}: x_0 \leq x, \ x_0 \leq c \cdot x \tag{14}$$

By using (13) and setting x_0 equal to a in (14) we see that the Pareto Type I, sometimes referred to as the *standard Pareto distribution*, is scale-invariant:

$$\bar{F}(c \cdot x) \overset{(13)}{=} \left(\frac{a}{cx}\right)^k = a^k \cdot x^{-k} \cdot c^{-k} \overset{(13)}{=} \bar{F}(x) \cdot g(c) \ \text{ where } g(c) := c^{-k} \tag{15}$$

The standard Pareto (11) belongs to the class of **regularly varying distributions** where the survival function has the form

$$\bar{F}(x) = L(x) \cdot x^{-k} \quad k > 0, \ k \in \mathbb{R} \tag{16}$$

where L(x) is a so-called **slowly varying function** with the following property:

$$\lim_{x \to \infty} \frac{L(c \cdot x)}{L(x)} = 1 \quad \forall c > 0, \ c \in \mathbb{R} \tag{17}$$

Prominent members of this class are the Burr, the Log-gamma and the Student`s t distribution. In the case of (13), L(x) is the constant function a^k. Whereas the standard Pareto (11) is scale-invariant over its whole support $(a, +\infty)$ all distributions satisfying (16) with a non-constant slowly varying function are only asymptotically scale-invariant:

$$\lim_{x \to \infty} \frac{\bar{F}(c \cdot x)}{\bar{F}(x)} \overset{(16)}{=} \lim_{x \to \infty} \frac{L(c \cdot x) \cdot (cx)^{-k}}{L(x) \cdot x^{-k}} = c^{-k} \cdot \lim_{x \to \infty} \frac{L(c \cdot x)}{L(x)} \overset{(17)}{=} c^{-k} \tag{18}$$

If we take the logarithm (any logarithm will do) on both sides of (13) we get:

$$\log(\bar{F}(x)) = k \cdot \log(a) - k \cdot \log(x) \quad 0 < a \leq x \tag{19}$$

Equation (19) tells us that the logarithm of the survival function is a linear function of the logarithm of x with a slope of -k in the case of the standard Pareto (11).

If the slowly varying function in (16) is not a constant function this relationship is non-linear:

$$\log\left(\bar{F}(x)\right) \overset{(16)}{=} \log(L(x)) - k \cdot \log(x) \tag{20}$$

However, the bigger x, the lesser log(L(x)) changes with the input x compared to log(x) due to (17) and, hence, the curvature of (20) converges to zero (i.e. the farther in the tails, the more (20) mimics a linear relationship)[6]. The relations (19) and (20) give rise to the *zipf plot*, a simple two-dimensional plot with the logarithm of the empirical survival function on the ordinate and the logarithmized order statistics on the abscissa. The stronger the concavity of the Figure for the higher order statistics, the thinner the tail of the distribution. A linear behavior after a certain threshold is evidence for a Paretian right tail, i.e., underlying i.i.d. random variables from the class of regularly varying distributions (16).

Figure 11 shows the *zipf plots* for the S&P 500 dataset. We can clearly see that the body of the distribution (decrease as well as increase) behaves quite different from the tail (e.g. the top 10% of the data) and, hence, it can be ruled out that the whole distribution follows a standard Pareto distribution. However, if only the top 10% of data points are considered, a linear trend seems quite reasonable. The red lines in Figure 11 are the ordinary least squares (OLS) regression fits for the respective top 10% data points. The intercept of the regression line multiplied by minus one can be used to estimate the shape parameter *k* due to (20). This method of fitting a Paretian tail has two downsides. Firstly, the OLS assumption of homoscedasticity is not met since there are fewer data points in the tails and, hence, the empirical survival function has a higher variance in the tails. Secondly, one can generate different regression lines (with different slopes) depending on the chosen threshold (e.g. the 90% quantile in Figure 11). The OLS estimate of *k* is a bit smaller in the decrease subsample (3.012 vs 3.061) which indicates that the left tail (i.e. losses) of the S&P 500 daily percentage changes is a little thicker than the right tail (i.e. wins). This matches the observations in Chapter 2.3.1 where the convergence of the first four sample moments was worse in the decrease subsample.

---

6   All functions that converge to a real-valued constant are slowly varying functions, however, the reverse does not necessarily hold true. For instance, any logarithmic function fulfills (17):

$$\lim_{x\to\infty} \frac{\log(c \cdot x)}{\log(x)} = \log(c) \cdot \lim_{x\to\infty} \left(\frac{1}{\log(x)}\right) + \lim_{x\to\infty} \frac{\log(x)}{\log(x)} = \log(c) \cdot 0 + 1 = 1 \ \ \forall\, c > 0,\ c \in \mathbb{R}$$
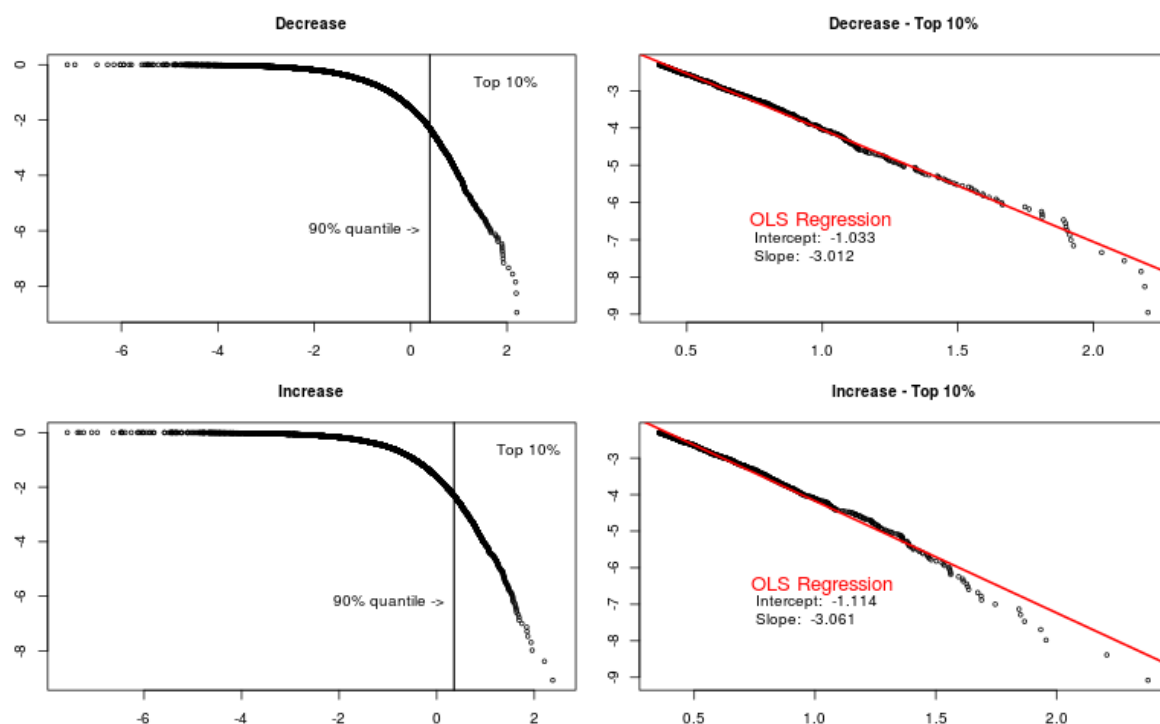
Figure 11: *Zipf plots for the decrease and the increase subsample of the S&P 500.*
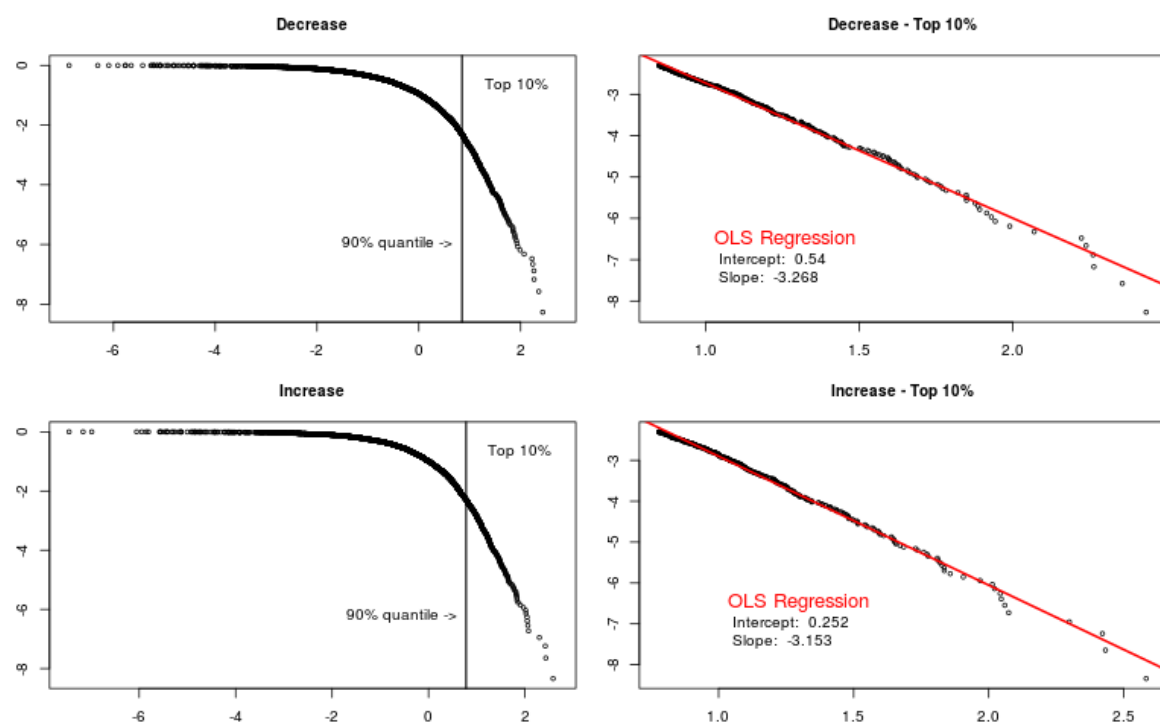
Figure 12: *Zipf plots for the decrease and the increase subsample of the Nikkei 225.*

The two OLS fits for the top 10% of the Nikkei 225 subsamples seem to fit quite well and a linear tail behavior seems plausible (Figure 12). In contrast to the S&P 500 dataset, the OLS estimates of *k* (3.268 vs 3.153) suggest that the right tail (i.e. wins) of the Nikkei 225 is thicker than its left tail (i.e. losses). Given the facts that the 90% quantile was chosen somewhat arbitrary as the threshold for the Paretian tail and that better estimators, the maximum likelihood estimator in particular, exist, it is emphasized that these OLS tail estimations should only be taken as rough guidelines. For more details, the reader is referred to De Haan and Ferreira (2007) as well as Sousa and Michailidis (2004).

One property of the class of regularly varying distributions (16) is that only those moments smaller than *k* are defined. In the case of a standard Pareto distributed r.v. X with density (11), the derivation of the **p^th moment** is relatively straightforward:

$$E\left(X^p\right) \overset{(11)}{=} \int_a^\infty x^p \frac{k \cdot a^k}{x^{k+1}} dx = k \cdot a^k \int_a^\infty x^p x^{-(k+1)} dx = \frac{k \cdot a^k \cdot x^{p-k}}{p-k}\bigg|_a^\infty = \frac{k \cdot a^p}{k-p} \ if \ p < k, \ \infty \ if \ p \geq k \quad (21)$$

It is essential that p is smaller than the shape parameter *k* in the last step. The four estimates of *k* (Figures 11 & 12) are all between three and four and, therefore, suggest that the respective 4th moment is not defined. At least for the decrease subsamples, this is simply impossible because a stock cannot lose more than 100% of its value and, hence, there is no such thing as a negative value for a stock index.

2.3.3. Mean excess function

One important characterization of (continuous) probability distributions can be made with the help of the **mean excess function** (*mef*):

$$mef(u) := E\left(X - u | X > u\right) = \frac{\int_u^\infty (x-u) \cdot f(x) dx}{\bar{F}(u)} \quad for \ 0 < u < Supremum\left[x \in \mathbb{R} : F(x) < 1\right] \quad (22)$$

In this work, only continuous probability distributions with no finite right endpoint (i.e. Supremum in (22) equal to +∞) are of concern. In the case of a standard Pareto random variable with density (11), the mean excess function has the following form:

$$mef(u) = \frac{\int_u^\infty x \cdot f(x) dx}{\bar{F}(u)} - \frac{u \cdot \int_u^\infty f(x) dx}{\bar{F}(u)} \overset{(13,\,15)}{=} \frac{\frac{a^k \cdot k \cdot u^{(1-k)}}{k-1}}{a^k \cdot u^{-k}} - \frac{u \cdot a^k \cdot u^{-k}}{a^k \cdot u^{-k}} = \frac{k \cdot u}{k-1} - u = \frac{u}{k-1} \ \forall \ u > a . \quad (23)$$

In the second step of (23) the results from (13) and (15), with p equal to one, were used. (23) shows that the expected excess increases in a linear fashion with the threshold u. A linear relationship also holds true for the Pareto Lomax and the Generalized Pareto distribution (GPD) although with different slopes and intercepts. Other distributions in the class of regularly varying distributions (16) have more complicated *mefs*, however, they can always be reduced to a linear part and an error term that loses its importance with an increasing threshold. More details are given by Johnson, Klotz and Balakrishnan (1970).

A r.v. X follows an **Exponential** distribution if it has the following density function (pdf):

$$f(x) = \lambda\, e^{-\lambda x} \quad x \in \mathbb{R},\ x > 0 \ \text{with parameter } \lambda > 0\, (scale);\ \lambda \in \mathbb{R} \tag{24}$$

A r.v. X follows a **Gamma** distribution if it has the following density function (pdf):

$$f(x) = \frac{\beta^{\alpha}\, x^{\alpha-1}\, e^{-x\beta}}{\Gamma(\alpha)} \quad x \in \mathbb{R},\ \text{with parameters } \alpha > 0\, (shape), \beta > 0\, (scale);\ \alpha, \beta \in \mathbb{R}$$
$$\text{where } \Gamma(z) := \int_0^{\infty} x^{(z-1)} e^{-x} dx \tag{25}$$

A r.v. X follows a **Log-normal** distribution if it has the following density function (pdf):

$$f(x) = \frac{1}{x\, \sigma \sqrt{2\pi}}\, e^{-\frac{(\log(x)-u)^2}{2\sigma^2}} \quad x \in \mathbb{R},\ \text{with parameters } u,\ \sigma > 0\, (shape);\ u, \sigma \in \mathbb{R} \tag{26}$$

Figure 13 shows the shape of the mean excess function for different continuous distribution functions. A well-known property of the Exponential distribution (24) is the "memoryless property", which is just a fancy word for a constant *mef*. The Exponential distribution is very useful to categorize the thickness of the tails of other continuous distributions. Those distributions with an increasing *mef* (e.g. the Log-normal (26) and the Pareto Lomax) have thicker tails than the Exponential distribution, those with a decreasing *mef* like the Gamma distribution (25) have thinner one`s[7]. The **Pareto Lomax** (Y is a Pareto Lomax r.v. with shape *k* and scale a, if and only if, X := Y+a is a standard Pareto r.v. with shape *k* and scale a) was chosen instead of (11) in order to compare only distributions with support on the complete positive real axis.

——————————————

[7] The characterization of probability distributions into the subexponential and the superexponential class is closely related to the behavior of the mean excess function. The interested reader is referred to Wierman (2014) who gives a great introduction into the properties, the emergence and the identification of heavy tails.

The Log-normal distribution is a very important continuous distribution due to the following relationship: Y is a Log-normal r.v. with parameters s and u if and only if $X := \log(Y)$ is a Normal r.v. with mean u and variance s^2 whose density is shown in formula (29). The shape parameter of Y is the scale parameter of X. A substitution in (26) reveals that e^u is the scale parameter of Y:

$$\frac{1}{x\,\sigma\sqrt{2\pi}}\,e^{-\frac{(\log(x)-u)^2}{2\sigma^2}} = \frac{1}{x\,\sigma\sqrt{2\pi}}\,e^{-\frac{\left(\log\left(\frac{x}{s}\right)\right)^2}{2\sigma^2}} \quad with \ s := e^u > 0 \tag{27}$$

The Log-normal has an increasing *mef*, but in contrast to the Pareto Lomax, its shape is concave (Figure 13). The higher the shape parameter in (26), the fatter the (right) tail of the Log-normal. Anyhow, no matter how big the shape parameter, all moments of the Log-normal r.v. are defined (i.e. finite). This fact connotes a decisive difference between a Log-normal r.v. and any r.v. from the class of regularly varying distributions, with all moments higher than the respective shape parameter undefined. This fact was shown in (21) for the standard Pareto distribution.

In the case of the Pareto Lomax, the shape parameter solely determines the slope of the *mef* whereas the intercept (i.e. the mean excess function evaluated at zero) is determined by both the shape and the scale parameter. Note that *mef*(0) is equivalent to the mean of the respective distribution for all distributions in Figure 13 because they all have a support that starts at zero. Every Exponential distribution has a *mef* with slope zero and an intercept that is solely determined by the respective scale parameter. The mean excess functions are nonlinear in case of the Log-normal (increasing) and the Gamma distributions (decreasing) and their shapes, the respective *mef*(0) in particular, are determined by both shape and scale.
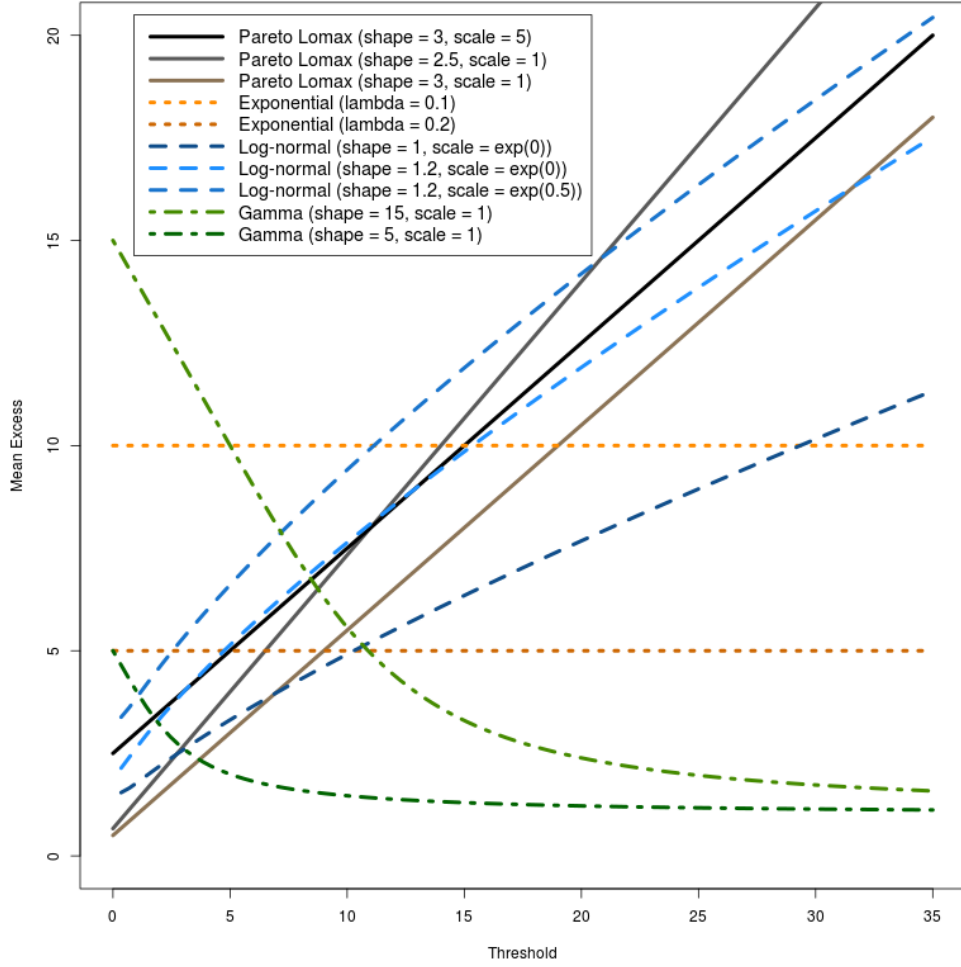
Figure 13: M*ean excess functions of four important families of continuous distributions.*

The **empirical mean excess function** (*emef*) of a sample of random variables X_1, X_2,…, X_N at threshold u is calculated as:

$$emef(u) = \frac{\sum_{i=1}^{N}(X_i - u)\cdot(I\{Xi>u\})}{\sum_{i=1}^{N}(I\{Xi>u\})} \quad where \; I\{Xi>u\}=1 \; if \; Xi>u, \; 0 \; if \; Xi\leq u \tag{28}$$

In order to assess the tail behavior of the two datasets, the empirical mean excess functions were plotted for the respective decrease and increase subsamples (Figures 14 & 15). Figure 14 shows that the shape of the *emef* in the body (low threshold values) differs quite substantially from its shape in the tail (high threshold values) in both S&P 500 subsamples. In the decrease subsample, the *emef* stays almost constant for threshold values smaller than 1.5. Somewhere

around 1.5, the *emef* starts to increases in a convex fashion up until a threshold of about 3. The *emef* continues to increase between threshold values of 3 and 5, although now approximately linearly. For threshold values bigger than 5 it is nonsensical to analyze the shape of the *emef* in too much detail because only a few data points are bigger than this threshold. Figure 14 suggests, however, that the (true) *mef* continues to rise for values higher than 5. For small values, the *emef* of the S&P 500 increase subsample behaves similar to the decrease subsample. The increase between the threshold values 2 and 4, however, is considerably smaller. Around a threshold value of 4 the slope of the *emef* increases constantly (convex behavior) up to a value of approximately 5.5.

A comparison of Figure 14 with the theoretical mean excess functions in Figure 13, clearly shows that the S&P 500 has heavier tails than the Exponential distribution. Furthermore, the shapes of both e*mefs* in Figure 14 are more complex than those of a Pareto Lomax or a Log-normal due to changes in their curvatures. These changes in curvature suggest that the daily percentage changes of the S&P 500 come from a mixture of different distributions. A linear (true) *mef* after a certain threshold (e.g. 3 for the decrease subsample), which would be equivalent to a Paretian tail, cannot be ruled out for either subsample. In the case of the increase subsample, the *emef* decreases after a threshold of 7.07. This should not be seen as a clear indication of a decreasing true *mef* since only three data points are bigger than this threshold leading to highly uncertain *mef* estimates (i.e. *emef* values) above this value. In agreement with the analysis in Chapter 2.3.2, it seems as though the daily percentage changes of the S&P 500 have a thicker left then right tail since the *emef* of the decrease subsample is higher than its increase counterpart for almost all threshold values and the last ten order statistics (i.e. the 10 largest data points) of the decrease subsample are all higher than their increase counterparts.

The *emefs* are approximately constant at 1 until a threshold value of about 2 for both subsamples of the Nikkei 225 dataset (Figure 15). Somewhere around this threshold value, the *emefs* start to increase linearly with a slope of approximately 0.25. In the case of the increase subsample, this linear trend does not change strongly until threshold values somewhere around 8, after which a drastic increase followed by a drastic decrease distorts the picture. Once again it is stressed that the *emef* is particularly unreliable for large (i.e. large in relation to the highest order statistics of the given dataset) threshold values. The *emef* of the decrease subsample decreases with increasing threshold values between approximately 4.5 and 5. This could very well be not just an anomaly but a real trend reversal since quite a view data points happen to be larger than 4.5.
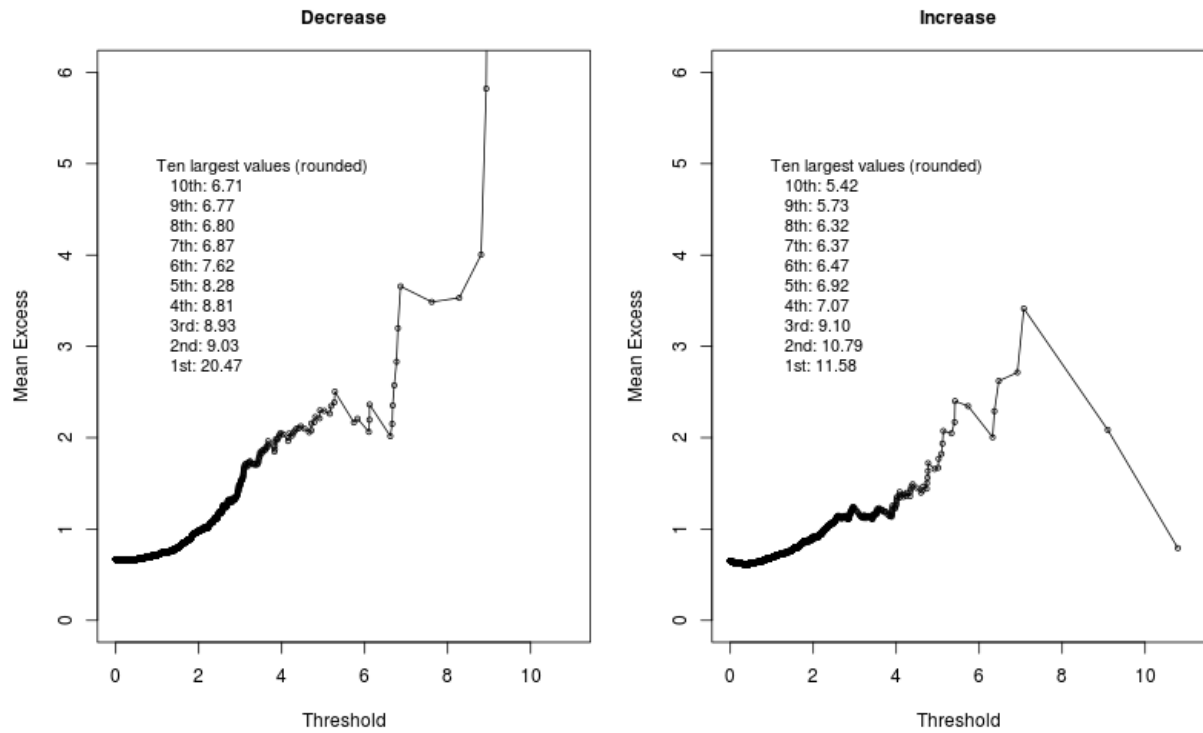
Figure 14: *Mean excess plots for the decrease and the increase subsample of the S&P 500.*
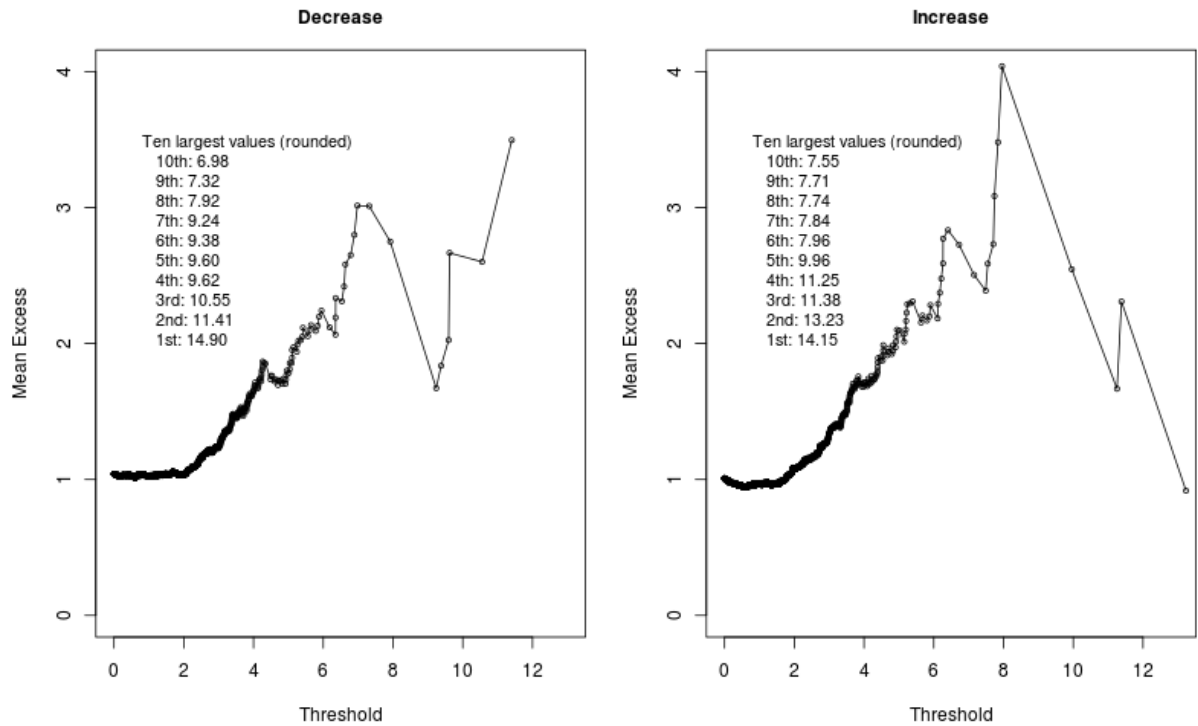
Figure 15: *Mean excess plots for the decrease and the increase subsample of the Nikkei 225.*

Showing close resemblance to the S&P 500 dataset (Figure 14), the two *emefs* in Figure 15 strongly suggest heavy tails for both the lower (decrease) and the upper (increase) tail of the Nikkei 225 daily percentage changes. Paretian tails, starting somewhere around a value of 2, are quite plausible. That being said, a concave shape of the true *mefs*, characteristic for a Log-normal distribution, cannot be ruled out either given these *emefs*. The two tails of the Nikkei 225 appear to be more similar than their S&P 500 counterparts. In sharp contrast to the S&P 500 dataset, six out of the last ten order statistics in the decrease subsample are smaller than their respective increase counterparts.

2.4. Fitting distributions to the daily percentage changes

The analysis thus far revealed that the two stock indices of interest, the S&P 500 and the Nikkei 225 do have fat tails. Furthermore, Chapter 2.2 concluded that our datasets most definitely are not i.i.d. realizations of some random variable. However, maybe an i.i.d. approximation works quite well for our two indices and daily data from several decades. The aim of this chapter is to find probability distributions with known properties that describe the behavior of the two indices fairly well. Such a parametric description is the basis of probabilistic inference and the estimation of important risk management quantities like Value at Risk (V@R) and Expected Shortfall (ES). In Chapter 2.4.1 well-known probability distributions are fitted to the complete S&P 500 dataset as well as the complete Nikkei 225 dataset. In Chapter 2.4.2, inspired by the results from Chapter 2.3, different distributions are fitted for the left (decrease) and the right tails (increase) of the two stock indices.

2.4.1. Continuous and symmetric distributions for the complete datasets

A random variable X follows a **Normal** distribution if it has the following density function (pdf):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\sigma^2}} \quad x \in \mathbb{R} \text{, with parameters } u\,(location),\ \sigma > 0\,(scale)\text{; } u, \sigma \in \mathbb{R} \qquad (29)$$

A r.v. X follows a **Logistic** distribution if it has the following density function (pdf):

$$f(x) = \frac{e^{-\frac{x-u}{s}}}{s\left(1 + e^{-\frac{x-u}{s}}\right)^2} \quad x \in \mathbb{R} \text{, with parameters } u\,(location),\ s > 0\,(scale)\text{; } u, s \in \mathbb{R} \qquad (30)$$

A r.v. X follows a **Laplace** distribution if it has the following density function (pdf):

$$f(x) = \frac{1}{2b} e^{-\frac{|x-u|}{b}} \quad x \in \mathbb{R} \text{ , with parameters } u \, (location), \, b > 0 \, (scale); \, u, b \in \mathbb{R} \tag{31}$$

A r.v. X follows a **Student`s t** distribution if it has the following density function (pdf):

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{\left(\frac{x-u}{\sigma}\right)^2}{v}\right)^{-\frac{v+1}{2}} \quad x \in \mathbb{R} \text{ , with parameters } u \, (location),$$

$$\sigma > 0 \, (scale), \, v > 0 \, (shape); \, u, \sigma, v \in \mathbb{R}; \text{ where } \Gamma(z) := \int_0^{\infty} x^{(z-1)} e^{-x} \, dx \tag{32}$$

Formulas (29) – (32) show the density functions of four important families of continuous distributions with support on the complete real axis. All four densities are symmetric around their respective location parameter u, although this is not immediately obvious for (30). The University of Alabama (2017) offers a short proof of the symmetry of the Logistic distribution on its website. Furthermore, the location parameter happens to be the mean (only if v>1 in the case of the Student`s t since it is a member of the class of regularly varying distributions), the median and the mode for these four distribution families. Out of (29), (30) and (31), the Normal distribution has the thinnest tails since the input is squared in (29) and the Laplace has the thickest tails. The Laplace distribution is often called Double-Exponential distribution since its density has the shape of an Exponential distribution (multiplied with 0.5) with its support shifted from $(0, \infty)$ to (u, ∞) for values greater than u and a vertically mirrored shape for values smaller than u. This can easily be seen if (31) is compared to (24):

$$\text{set } \lambda := \frac{1}{b} \text{ then } \frac{1}{2b} e^{-\frac{|x-u|}{b}} = \frac{1}{2} \cdot (\lambda \, e^{-\lambda |x-u|}) = \frac{1}{2} \cdot (\lambda \, e^{-\lambda y}) \text{ for } y := |x-u| \geq 0 \tag{33}$$

The densities (29) – (31) all decrease exponentially the further the distance from their respective location parameter. Out of (29) – (32), the Student`s t is the only distribution with a density that does not decrease exponentially and the only one with a shape parameter which is often called the "degrees of freedom" parameter. This shape parameter v is equivalent to *k* in (16) and the smaller v, the slower the decrease of the density and the fatter the tails. Whereas all moments are defined for (29), (30) and (31) only those moments smaller than v exist for (32).

A very important figural tool to assess the distribution of some numerical data is the histogram. Figures 16 & 17 show the respective histograms for our two indices. 150 breaks were chosen for our two datasets because that value seemed to achieve better noise to signal ratios than smaller or higher numbers. Furthermore, the histograms were normed such that the respective area under all bins equals 1. This ensures meaningful comparisons with theoretical probability distributions. Since the goal in this chapter is to treat our datasets as if they would come from some unknown continuous probability distribution, kernel density estimates were plotted in addition. It is stressed once more, that the goal in this chapter is not to find the perfect parametric fit for our two datasets but the assessment of the accuracy of different approximations via some continuous probability distributions[8].

The parameters of four continuous distributions (Normal, Logistic, Laplace, Students´ t) were estimated via maximum likelihood for both indices (Table 2) and its densities plotted in Figures 16 & 17. The daily percentage changes of the S&P 500 (Figure 16) seem to be unimodal and almost symmetrical, although not around a value of zero but a slightly bigger one – of which more later. The bins around the middle of the histogram are clearly higher than the density of the fitted Normal distribution which, coupled with the fact that the tails of the S&P 500 daily percentage changes are quite fat (Chapter 2.3), is a clear sign of a leptokurtic shape coming from a kurtosis greater than 3[9]. Whereas the Normal and the Logistic distribution fit the data very poorly, the densities of the fitted Laplace and Student`s t manage to stay close to the kernel density estimate. The Laplace density fit stays closer to the kernel density than the Student`s t density fit for negative values but further away for positive values. Note that the two histograms (Figures 16 & 17) are only plotted between -4 and 4 and, therefore, the goodness of fit in the tails cannot be assessed graphically just now.

---

8   Note that, for instance, the S&P 500 index did not change at all for 116 out of the 16609 days in our sample. Hence, a underlying continuous probability distribution can safely be ruled out.

9   3 is the kurtosis (i.e. the 4th standardized moment) of a Normal distribution and all distributions with a kurtosis lower than the Normal are categorized as platykurtic whereas those with a higher kurtosis are labeled as leptokurtic.

Values rounded to three decimal places!		Normal	Logistic	Laplace	Student`s t
S&P 500	**shape**	-	-	-	3.152
	scale	0.968	0.478	0.654	0.611
	location	0.033	0.041	0.045	0.045
Nikkei 225	**shape**	-	-	-	3.475
	scale	1.468	0.747	1.021	0.989
	location	0.018	0.031	0.040	0.039

Table 2: *Maximum likelihood parameter estimates for different distributions and the daily percentage changes of the Standard & Poor 500 as well as the Nikkei 225.*
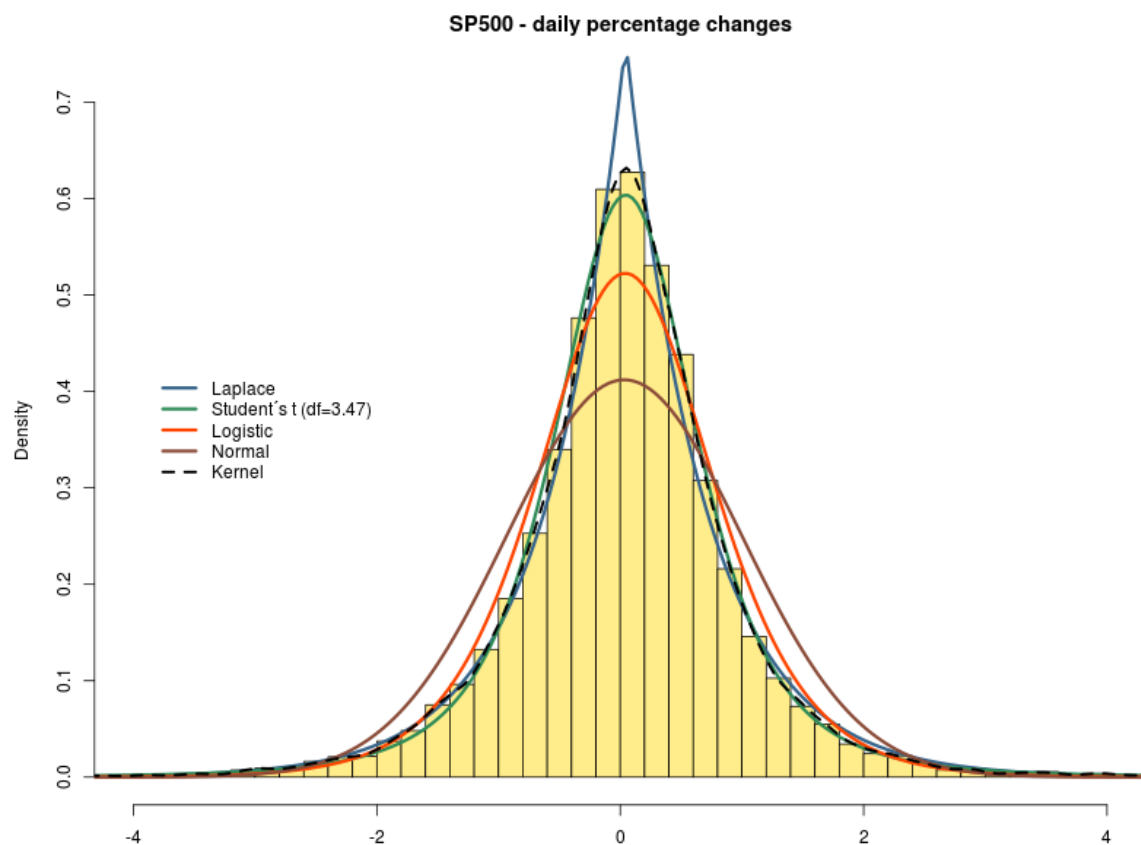
Figure 16: *The histogram of the daily percentage changes of the S&P 500 index, a kernel density estimate and fitted densities of four important continuous probability distributions.*

The Nikkei 225´s daily percentage changes seem to be unimodal and symmetric (Figure 17). The kernel density is leptokurtic, although not quite as much as those of the S&P 500 since its peak is at approximately 0.4 as opposed to approximately 0.6 in Figure 16. Just as in the case of the S&P 500 dataset, the Normal and the Logistic fit the daily percentage changes very poorly. The Laplace as well as the Student`s t fit is quite good, and once again, the former is better for values between -2 and 0 whereas the latter is better for values between 0 and 2. In comparison to Figure 16, the bins are higher between -4 and -2 as well as between 2 and 4 indicating that the Nikkei 225 experienced more medium sized index changes than the S&P 500.
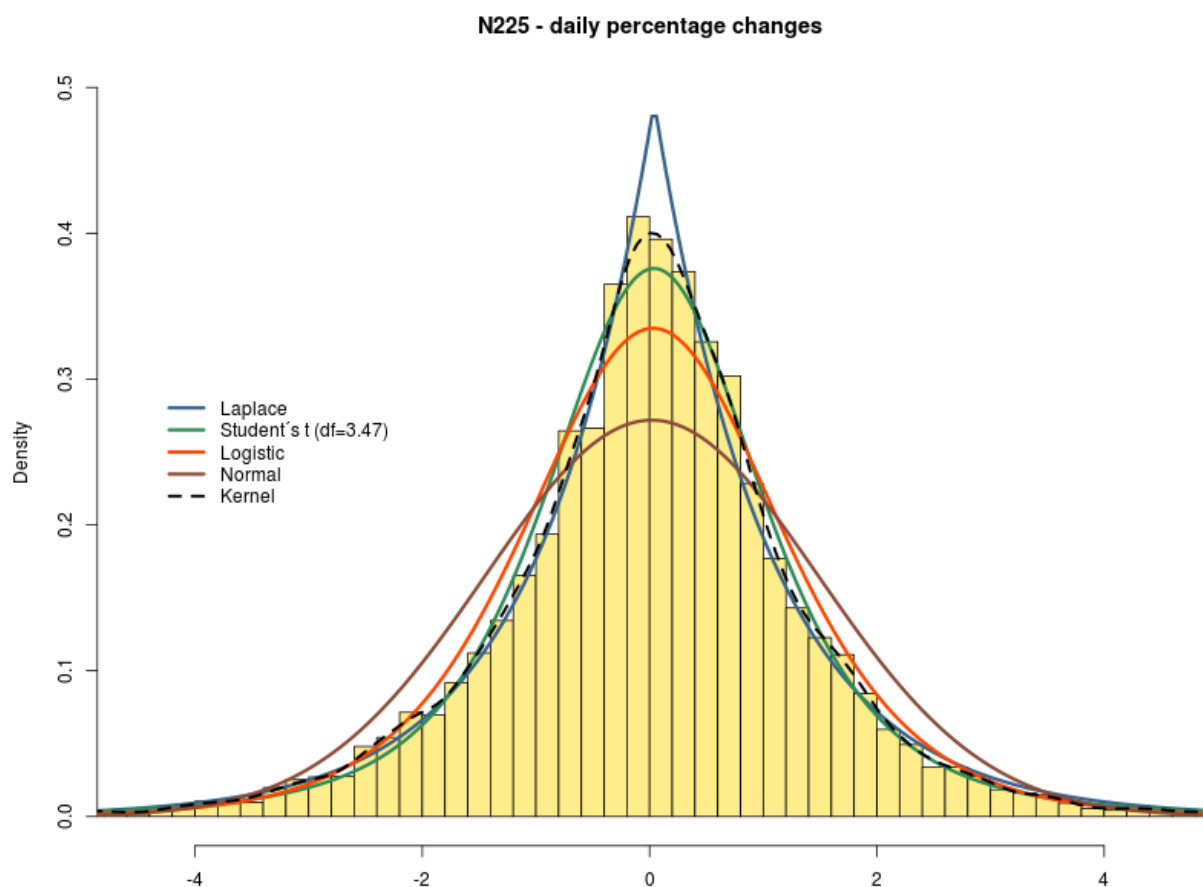
Figure 17: *The histogram of the daily percentage changes of the Nikkei 225 index, a kernel density estimate and fitted densities of four important continuous probability distributions.*

For both the S&P 500 and the Nikkei 225 the Normal and the Logistic density fits are not analyzed any further due to their strong deviations from the respective histograms (Figures 16 & 17). The Laplace as well as the Student`s t describe the body of the two distributions reasonably well and in order to compare their shapes in the tails, Figures 18 & 19 zoom in on the respective left and right tails. In addition to the kernel density estimator used in the histograms (Kernel 1) a second kernel estimator, with a four times larger bandwidth is

plotted (Kernel 2). The small hills in the tails of Kernel 1 (Figure 18&19) look implausible and the larger bandwidth of Kernel 2 leads to smoother densities and more reasonable fits in the tails. However, any nonparametric estimate (e.g. a kernel density) for the tail behavior of empirical data is problematic for heavy-tailed phenomena because of the small number of data points in the tails.

Figure 18 shows that in the case of the S&P 500 dataset the fitted Laplace density lies above the fitted Student`s t density until approximately values of -3.5 and 3.5. Both the Laplace (31) and the Student`s t (32) are symmetric around their location parameter u, hence, their respective density at (u - x) has the same value as their respective density at (u + x) for all real numbers x[10]. For values larger than 6 (or smaller than -6) the Laplace density fit is essentially zero. The considerably slower decreasing densities of Kernel 1, Kernel 2 and the Student`s t fit, make values larger than 6 neither likely nor astronomically unlikely as the Laplace fit does.
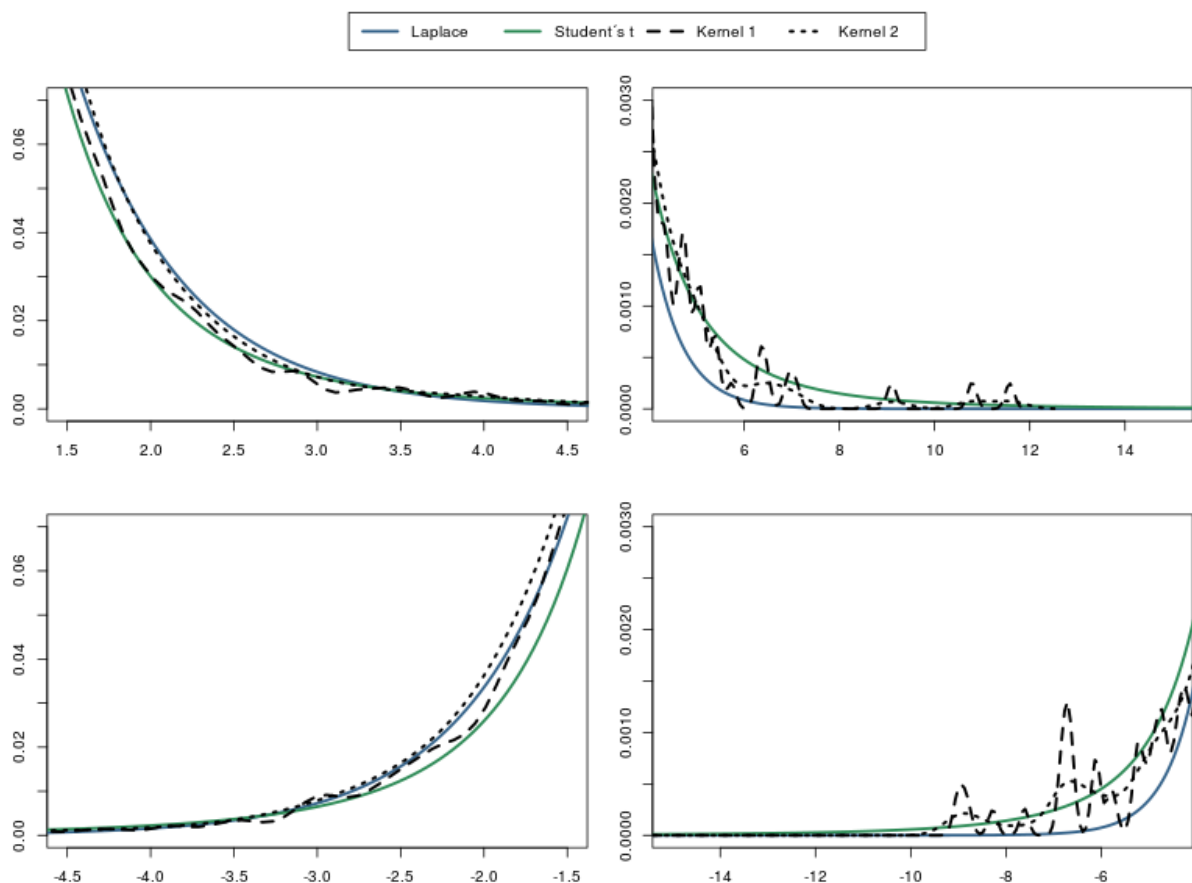
Figure 18: *S&P 500: Zoom of two fitted theoretical distributions and two kernel densities.*

---

10 The mean, if defined, is equivalent to the median for all unimodal symmetric distributions.

In the case of the Nikkei 225 dataset (Figure 19), the densities of the Laplace and the Student`s t fit cross later as in the S&P 500 dataset, at approximately values of 5 and -5. The tails of the Laplace fit are thicker as those of its S&P 500 counterpart, yet they are not thick enough to describe the distribution of the Nikkei 255 accurately for values larger than 6 (or smaller than -6).
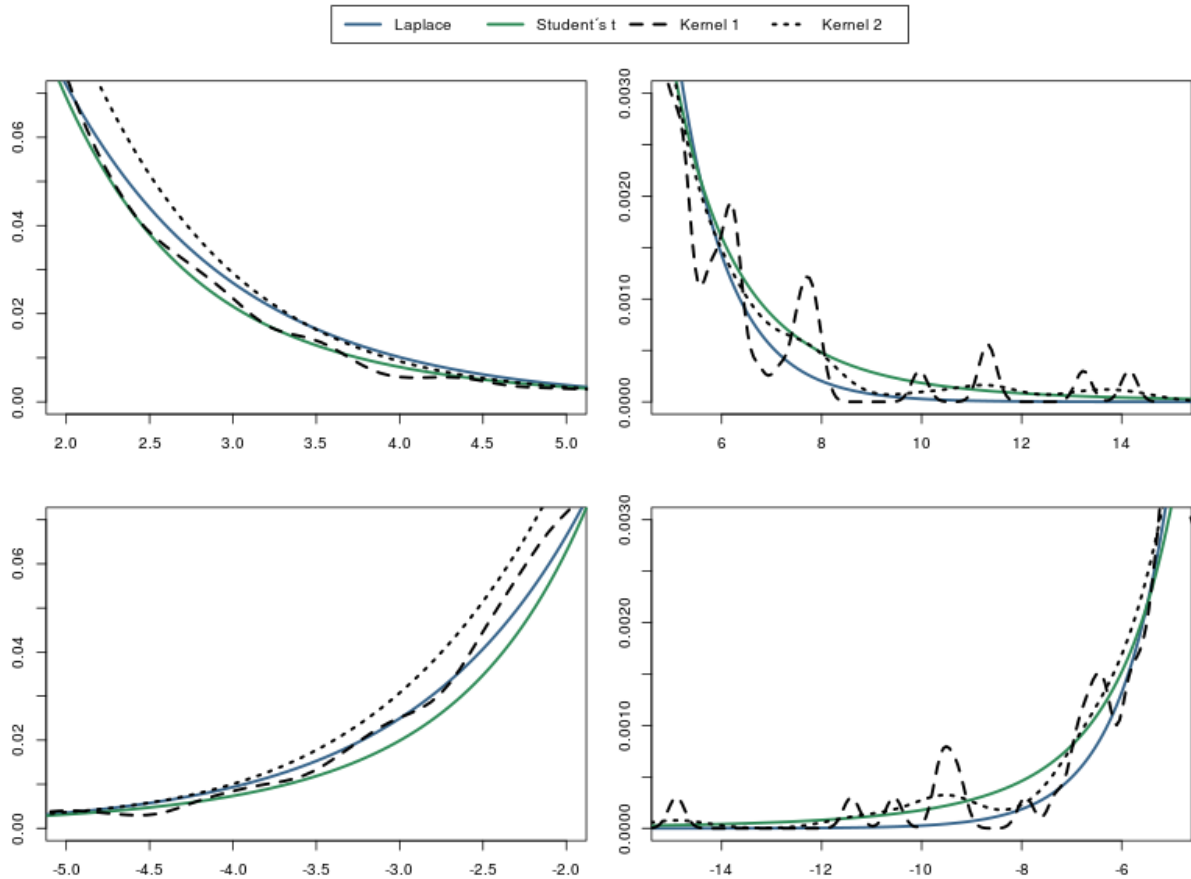
Figure 19: *Nikkei 225: Zoom of two fitted theoretical distributions and two kernel densities.*

The most rigorous tool for a graphical analysis of some fitted distribution is the quantile-quantile (Q-Q) plot. In the case of a univariate r.v., a Q-Q plot is a two dimensional plot with the empirical quantiles (i.e. the order statistics of the sample) on the ordinate and the theoretical quantiles on the abscissa. For each order statistic (the y-coordinate) a point is plotted with the inverse of the theoretical distribution function evaluated at the empirical distribution function of that order statistic as the x-coordinate[11]. The closer the N points lie to the 45° line, the more similar the theoretical and the empirical distribution.

---

11 This is actually not precise, since the y-coordinate is the empirical distribution function multiplied with N/(N+1), a correction factor. Otherwise the x-coordinate would be +∞ for the Nth order statistic (i.e. the maximum) for all theoretical distributions with an unbounded support on the positive real axis.

Figures 20 & 21 show the Q-Q plots for the Laplace and the Student`s t fits of the two indices. For each of these four Q-Q plots, three additional plots (zoom in the center, the left, and the right tail) are shown in order to better detect deviations from the respective theoretical distribution.
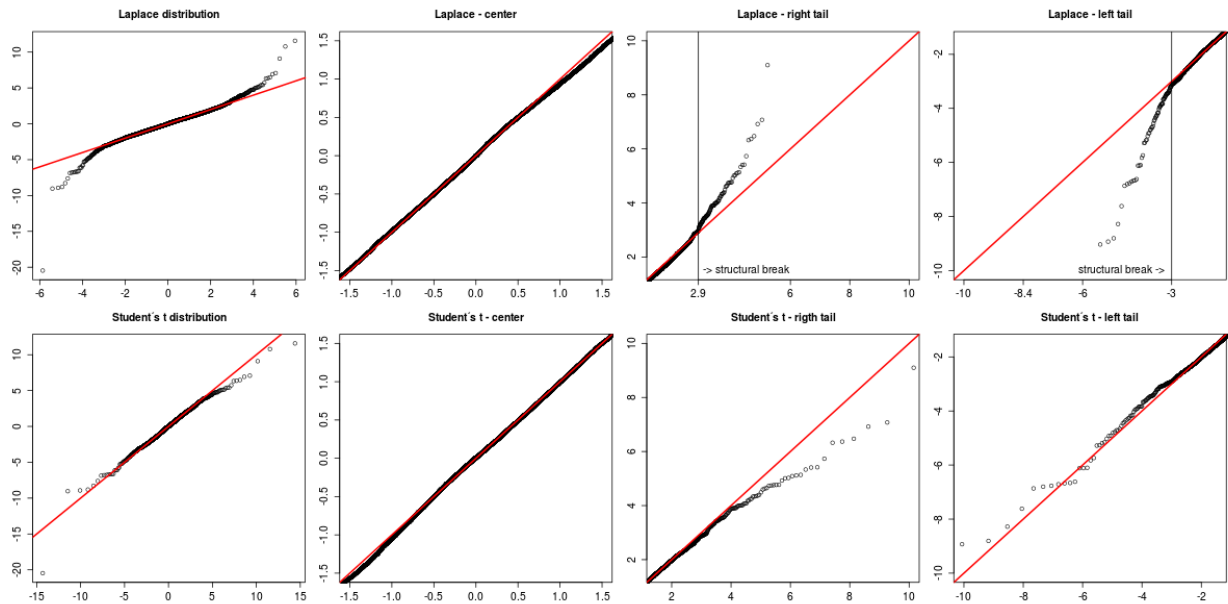
Figure 20: *S&P 500: Quantile-Quantile plots for the Student`s t and the Laplace fit.*
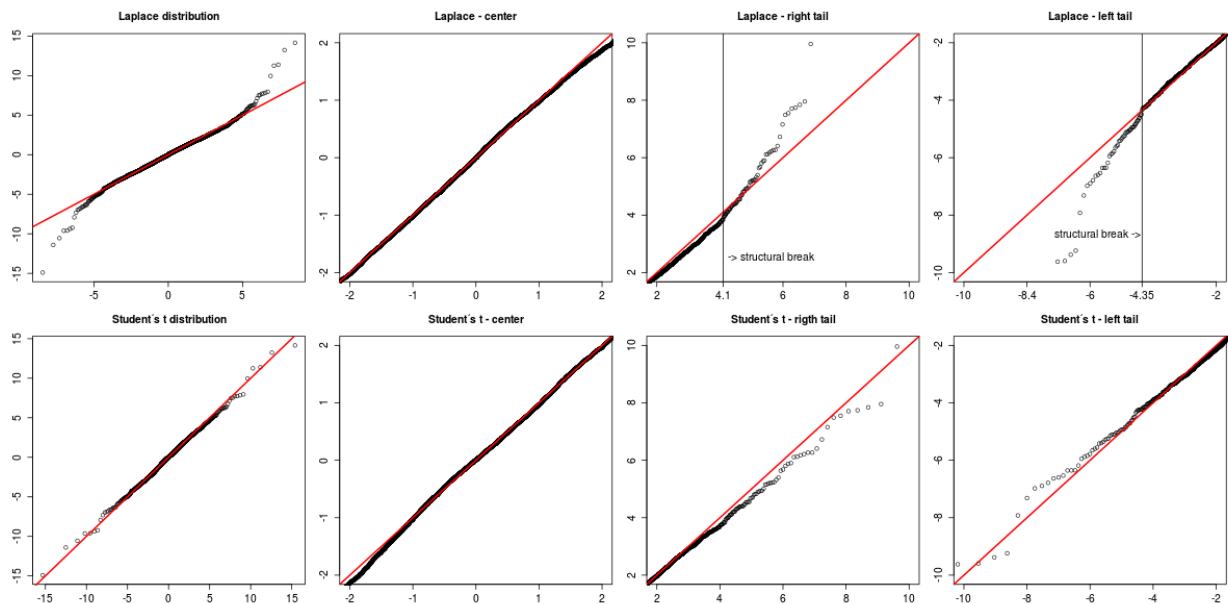
Figure 21: Nikkei 225: Quantile-Quantile plots for the Student`s t and the Laplace fit.

The Q-Q plots for the S&P 500 (Figure 20) clearly show that the Laplace fit performs poorly in the tails. There seems to be a point in both tails where the quantiles of the Laplace fit are starting to increase at a higher rate than the empirical quantiles. These two structural breaking points are different for the right (2.5) and the left tail (3). After these two breaking points, the respective theoretical (Laplace fit) quantiles start to diverge from their respective empirical quantiles, at an approximately constant rate, the further in the tails. In the case of the Student`s t fit, no structural breaking points can be detected. All points with an x-coordinate bigger than 4 lie under the 45° line, which connotes that the actual right tail is not as heavy as the theoretical fit. However, in contrast to the Laplace fit, the points do not diverge from the 45° line, the further in the tails. Instead, the increasing distance between the points and the 45° line between approximately 4 and 7 is followed by a short decrease, another increase after approximately 7.5 and yet another decrease at around 10. The Student`s t fit is better in the left than in the right tail since the points in the left tail stay close to the 45° line until values of around -9. Of course, "Black Monday" steps out of line with a theoretical quantile of around 15. Both the Laplace and the Student`s t fit are quite good in the center of the distribution.

The Q-Q plots for the Nikkei (Figure 21) show a similar picture. The Laplace fit cannot adequately describe the tails and there seem to exist structural breaking points once again. The Student`s t fit describes this dataset pretty well in both tails, as no big deviations from the 45° line can be found in the plots. This fits with the previous observation that the difference in the thickness of the respective tails is smaller in the case of the Nikkei 225. The Student`s t distribution is symmetric and, therefore, the more asymmetric a dataset, the poorer the fit. Both the Laplace and the Student`s t fit are quite good. The former is inferior at around a value of 2, the latter around -2.

2.4.2. Different fits for the left and the right tail

Chapter 2.4.1 revealed that the Students´s t distribution fits the daily percentage changes better than the Normal, the Logistic and the Laplace. The respective fits of the Student`s t via maximum likelihood estimation (MLE) describe the two datasets actually pretty good given the fact that they do not constitute i.i.d. realizations of some random variable (Chapter 2.2). In Chapter 2.3 it was discovered that the datasets are somewhat asymmetric and that the left tail, particularly in the case of the S&P 500 dataset, seems to be thicker than the respective right tail. Therefore, in this chapter, the datasets are split at their respective median (0.04480137 for the S&P 500 and 0.04035366 for the Nikkei 225) and separate distributions are fitted to both parts of the respective index[12].

---

12 Unlike the Nikkei 225 dataset, the S&P 500 dataset consists of an uneven number of daily percentage changes (16609) and therefore the median is equal to one data point, namely the $[(16609+1)/2]^{th}$ order statistic. This data point was not considered in the subsequent Double distribution fits.

The following steps describe the creation of the artificial fits for both indices (Figures 22-27):

- Firstly, the median was subtracted for each and every data point. After these subtractions, all values greater than zero (n_{right}) were used to fit distributions for values greater than the median and values smaller than zero were multiplied by -1 (n_{left}) and then used to fit distributions for values smaller than the median.

- The Gamma (25) and the Log-normal (26) distribution were fitted to n_{right} as well as to n_{left} via MLE. The density fits for n_{right} were multiplied by 0.5 and shifted to the right by the value of the median. The density fits for n_{left} were also multiplied by 0.5 and shifted to the right by the value of the median. Additionally, the variable x in (25) and (26) was replaced by -x in order to get a vertically mirrored density shape on the left side of the median. The **Double Log-normal** and the **Double Gamma** densities (Figures 22-25) are just the respective density fits for the left and the right part, plotted with the same color.

- The **Double Student`s t** fit was created in a different way since unlike the Log-normal and the Gamma, the Student`s t distribution with density (32) has support on the complete real axis. For each value in n_{right} as well as in n_{left} a negative counterpart (i.e. -1 times the value) was added and, hence, their respective sizes doubled and both n_{right} and n_{left} became perfectly symmetric around the value 0. Afterward, a Student`s t was respectively fitted to $n_{right}$ and $n_{left}$ via MLE. Both distributions were then shifted to the right by the value of the median. The Double Student`s t density (Figures 22-25) is the n_{left} density fit until the median and the n_{right} fit after the median.

It is important to note that because of this construction, all three *Double Distributions* have a discontinuity point at the median of the respective dataset. Furthermore, the Double Log-normal and the Double Gamma densities get infinitely small, the nearer to the respective discontinuity point. This is certainly not a realistic description of the two datasets. However, if one or more of the Double Distribution fits were to describe the dataset(s) well for all values not too close to the median, modifications to smooth out the transitions from the left to the right side of the median are certainly possible. The parameter estimates for the respective left and right parts of the three artificially created Double distributions can be found in Table 3. A smaller shape parameter indicates a fatter tail in case of the Gamma and the Student`s t and a thinner tail in case of the Log-normal. In agreement with the analysis in Chapter 2.3, the left tail of the two indices seems to be thicker than the right tail, although not by much. The differences in shape and scale are smaller between the left and the right tail of both indices than they are between the indices and the same tails for all three Double distributions.

(Left / Right) Values rounded to three decimal places!		Double Log-normal	Double Gamma	Double Student`s t
S&P 500	shape	1.252 / 1.210	1.012 / 1.069	3.094 / 3.219
	scale	0.377 / 0.376	0.658 / 0.601	0.619 / 0.604
Nikkei 225	shape	1.245 / 1.223	1.036 / 1.071	3.399 / 3.588
	scale	0.599 / 0.585	1.006 / 0.932	1.006 / 0.973

Table 3: *Maximum likelihood parameter estimates for three artificially constructed distributions in the S&P500 as well as the Nikkei 225 dataset.*
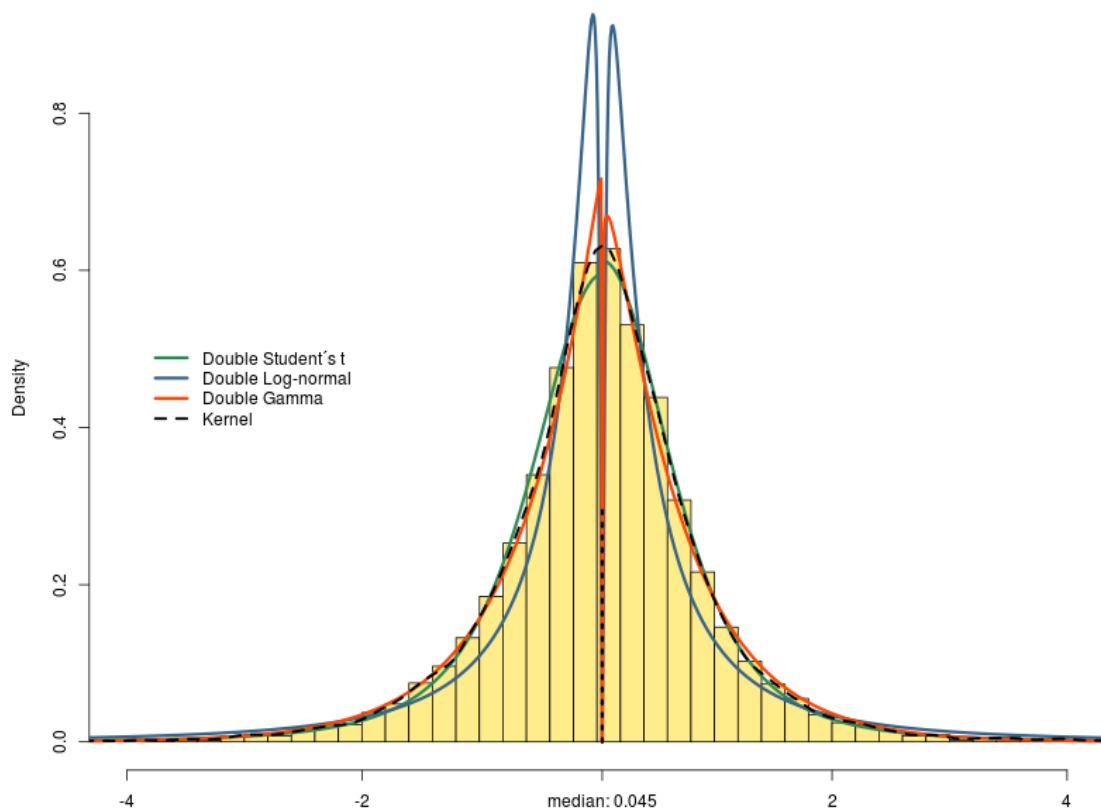
Figure 22: *The histogram of the daily percentage changes of the S&P 500 index, a kernel density estimate and the densities of three artificially constructed probability distributions.*
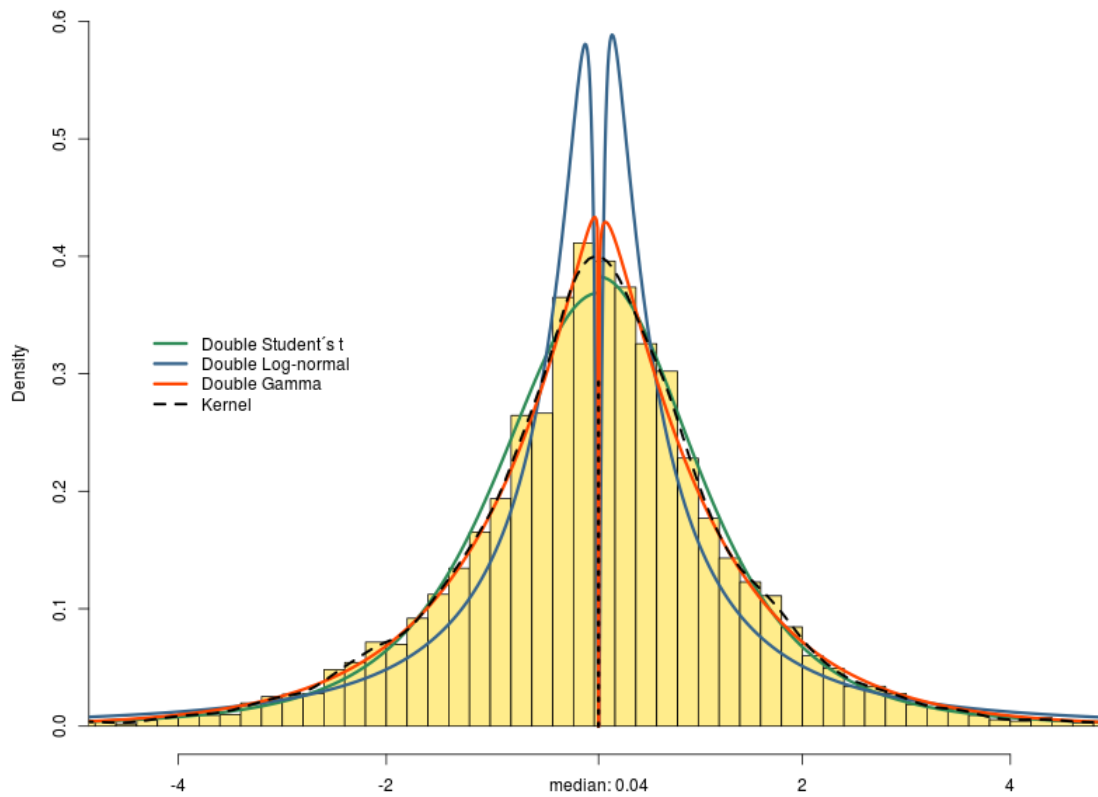
Figure 23: *The histogram of the daily percentage changes of the Nikkei 225 index, a kernel density estimate and the densities of three artificially constructed probability distributions.*

In the case of the S&P 500, the Double Log-normal is a poor fit for the daily percentage changes (Figure 22). The blue density is too leptokurtic, that is too high for very small values, too low for values between approximately -2 and -0.5 as well as approximately 0.5 and 2, and again, too high for values farther than 2 from the median. The Double Gamma and the Double Student`s t both fit the body of this dataset quite well. The Double Gamma lies closer to the kernel density than the Double Student`s t in the left part and farther away in the right part. Everything said about the performance of the three artificial Double distributions also holds true for the Nikkei 225 (Figure 23).

In order to compare the Double distribution fits in the tails, Figures 24 & 25 zoom in on the densities in the tails of the respective distributions. For both the S&P 500 as well as the Nikkei 225, the Double Log-normal density is too high for small and for large values. The Double Gamma densities, on the other hand, are too small for values far away from the respective median. For both datasets, only the respective Double Student`s t density fit lies close to Kernel 2 for values farther than 6 away from the respective median (Figures 24 & 25).
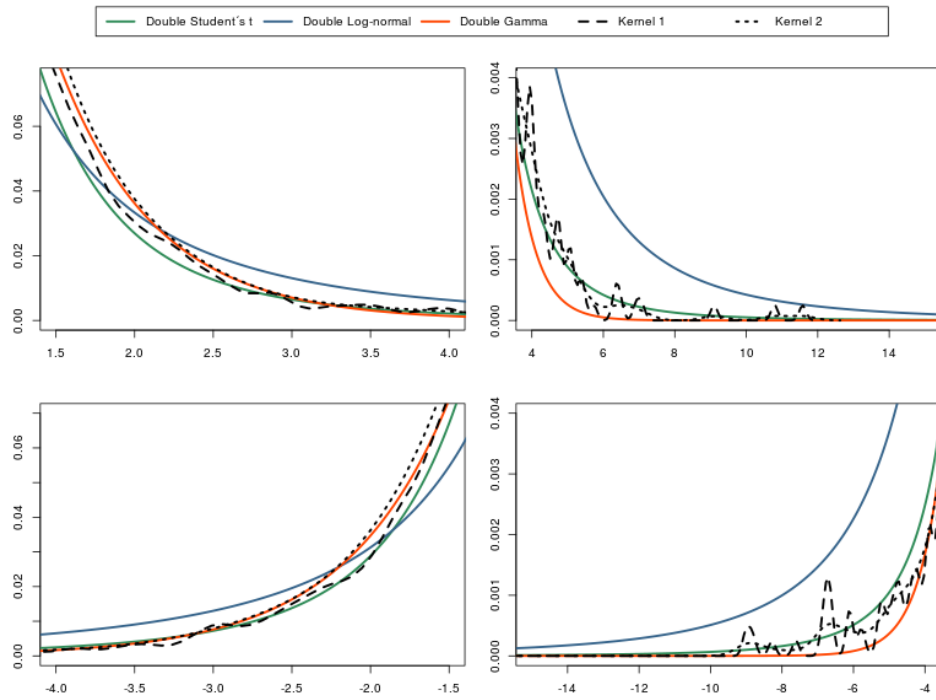
Figure 24: *S&P 500: Zoom of the densities of two artificially fitted theoretical distributions as well as two kernel densities.*
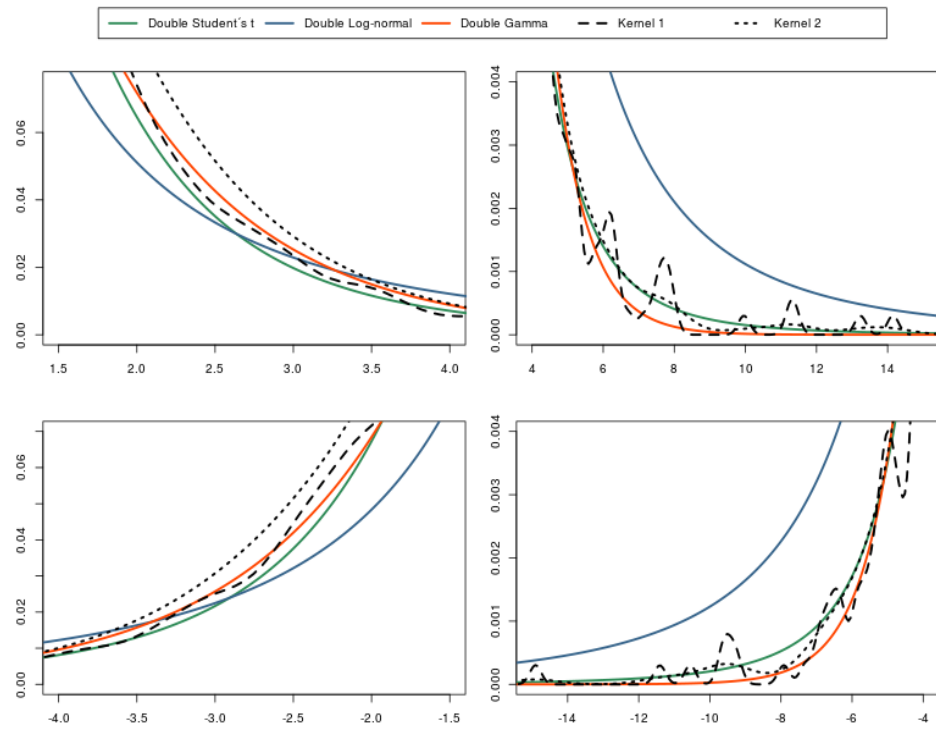
Figure 25: *Nikkei 225: Zoom of the densities of two artificially fitted theoretical distributions as well as two kernel densities.*

The Double Student`s t fits are clearly superior to the other two artificially fitted Double distributions. The next obvious question is: How big is the improvement in describing the two datasets with the artificial Double Student`s t with its five parameters (2x shape, 2x scale, and the median) instead of the three parameter Student`s t fit (Chapter 2.4.1). To answer this question, Figures 26 & 27 show the Q-Q plots for the Double Student`s t and the Student`s t.
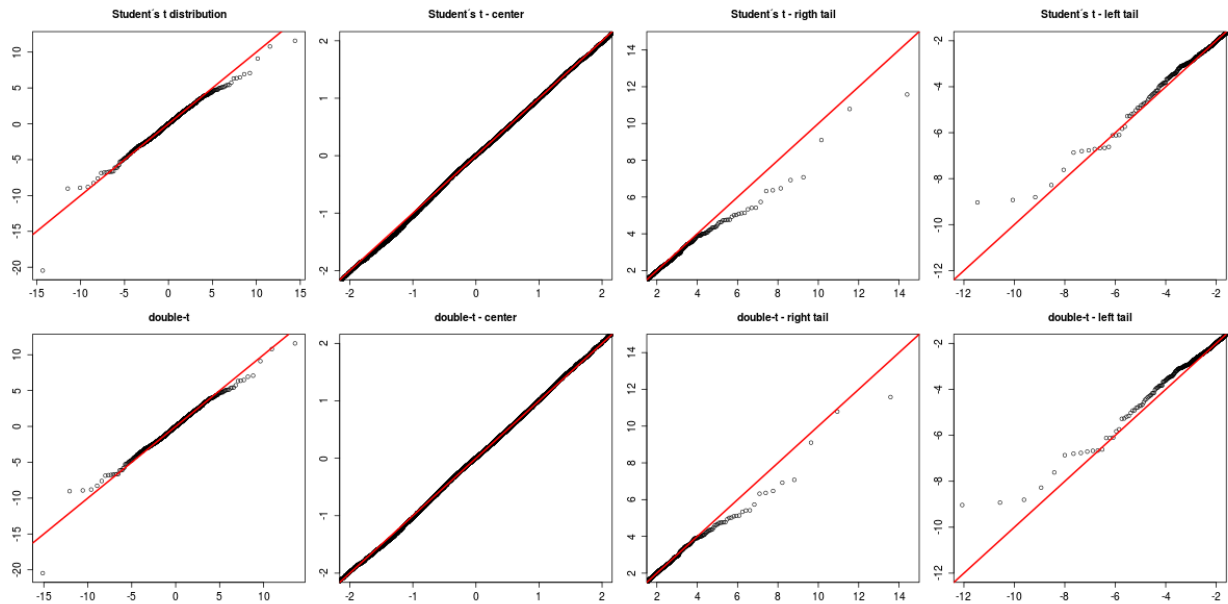
Figure 26: *S&P 500: Quantile-Quantile plots for the Student`s t and the Double Student`s t fit.*
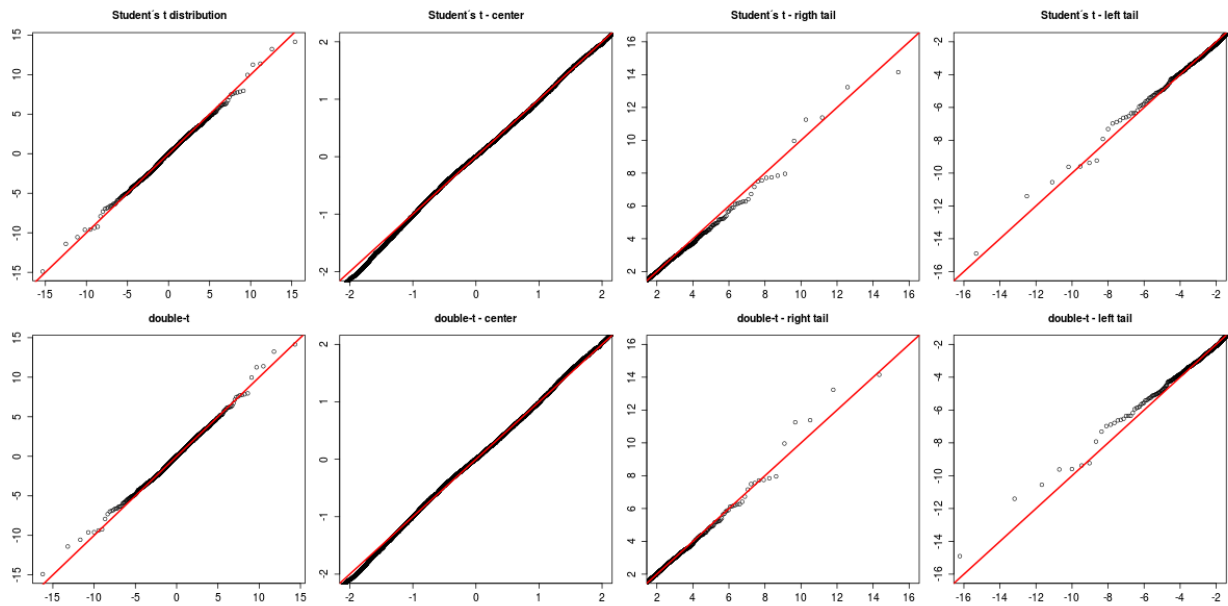
Figure 27: *Nikkei 225: Quantile-Quantile plots for the Student`s t and the Double Student`s t fit.*

40

In the case of the S&P 500 (Figure 26), there seems to be a slight improvement in the right tail. On the other hand, the Double distribution seems to be a poorer fit in the left tail, at least until values smaller than -10. This is clearly owed to "Black Monday" and its huge impact on the shape parameter in the left tail. Neither improvement nor debasement in the center of the distribution can be spotted for the Double distribution. There also seems to be a slight improvement in the right and a slight debasement in the left tail for the Nikkei 225 (Figure 27). The Double distribution of the Nikkei 225 improves the fit between values of approximately -1 and -2.

3. Monte Carlo simulations

3.1. Design and methods

Just as the analysis of the S&P 500 and the Nikkei 225 index in Chapter 2 (including Figures 1-27), this simulation was programmed in R (R Core Team, 2016). The complete R code including all non-basic functions of the simulation as well as the code of the (graphical) analysis is available from the author upon request. Exact reproducibility was established since random variables were generated and stored in a first step, and loaded and analyzed later on.

In 50 scenarios i.i.d. Student`s t random variables (32), with a location parameter of 0 and a scale parameter of 1, were generated. Two parameters were varied in order to generate these 50 scenarios, namely the degrees of freedom (shape) parameter and the number of i.i.d. realizations. The *degrees of freedom* (v) parameter was either 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7 or 3.8. These values were chosen because they cover the range of maximum likelihood estimates of the degrees of freedom parameters for the S&P 500 and the Nikkei 225 dataset (Tables 2&3) well. The *sample size* (n) was either 100, 500, 1000, 5000 or 10000. The largest value (i.e. n=10000) lies between the sample size of the S&P 500 dataset (16609) and the sample size of the Nikkei 225 dataset (8108). Since an analysis of daily financial data is particularly interesting for time frames of only a few years or even just a few months, smaller sample sizes were investigated as well. A fully crossed design with these two parameters was chosen and, therefore, 50 (5x10) different scenarios were constituted. Each scenario was simulated 1000 times in order to assess the distribution of the estimators of interest.

For each scenario the following quantities were computed:

- Maximum likelihood estimate of the location parameter
- Maximum likelihood estimate of the scale parameter
- Maximum likelihood estimate of the degrees of freedom parameter
- Sample variance
- Sample mean
- Sample median
- Sample 99% quantile
- Sample 99.9% quantile

The main goal of this simulation is to answer the following questions:

- How well can five important quantities, namely **mean**, **variance**, **median**, **99% quantile** and **99.9% quantile**, be estimated?
- Which estimation technique (**MLE** or **nonparametric**) fairs better for which quantity and which scenario?
- How strong is the influence of the **sample size** and the **degrees of freedom** parameter on the precision of the estimators?

The sample quantiles (50% (median), 99% and 99.9%) are computed with the function quantile () of the *stats* package (R Core Team, 2016) with the type parameter equal to 7. Three other non-basic R packages were used in this simulation: *MASS* (Venables & Ripley, 2002), *LaplacesDemon* (Statisticat, Llc., 2016) and *matrixStats* (Bengtsson, 2016).

The maximum likelihood (ML) estimates for the mean and the median are equivalent to the ML estimate of the location parameter due to the symmetry of the Student`s t distribution. The ML estimates of the 99% and the 99.9% quantile are just the respective theoretical quantiles of a Student`s t distribution with the ML estimates of the location, scale and degrees of freedom parameter as the respective parameter. The ML estimate of the variance is computed by replacing the scale and the degrees of freedom parameter with the respective ML estimate in the following equation:

$$If\ r.v.\ X\ has\ density\ (16),\ then\ Variance(X) = \frac{v}{v-2}\sigma^2,\ \sigma>0\,(scale),\ v>2\,(shape) \qquad (34)$$

Due to (34), no variance can be estimated via ML if the estimated degrees of freedom parameter is less than or equal to 2.

The following metric was used to compare the performance of the ML and the nonparametric estimators in this simulation:

$$\hat{MAD}(\hat{\theta}) := \hat{E}(|(\hat{\theta}-\theta)|) = \frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i-\theta|, \quad \theta\,(\textit{true quantity}),\, \hat{\theta}\,(\textit{estimator for }\theta). \tag{35}$$

This metric is an estimator for the **mean absolute deviation (MAD)** of an estimator from the true quantity. Obviously, the smaller the MAD, the better the estimator.

An often used alternative metric is the estimated **mean squared error (MSE)**:

$$\hat{MSE}(\hat{\theta}) := \hat{E}((\hat{\theta}-\theta)^2) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i-\theta)^2 \quad \theta\,(\textit{true quantity}),\, \hat{\theta}\,(\textit{estimator for }\theta). \tag{36}$$

Note that in this simulation N is equal to 1000 in (35) and (36) since each estimator is computed in each of the 1000 iterations of a certain scenario. Just as in the case of the MAD, the smaller the MSE, the better the estimator. Since the differences between the estimates and the true values are squared in (36), strong deviations have a greater impact on the MSE than on the MAD. A well-known relation between the MSE and the variance of an estimator exists:

$$\begin{aligned} Bias(\hat{\theta}) &:= E(\hat{\theta})-\theta \;\rightarrow\; MSE(\hat{\theta}) := E((\hat{\theta}-\theta)^2) = E((\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2) \\ &= E((\hat{\theta}-E(\hat{\theta}))^2) - 2\cdot(E(\hat{\theta})-\theta)\cdot E(\hat{\theta}-E(\hat{\theta})) + E((E(\hat{\theta})-\theta)^2) \\ &= Var(\hat{\theta}) - 2\cdot Bias(\hat{\theta})\cdot(E(\hat{\theta})-E(\hat{\theta})) + E((Bias(\hat{\theta}))^2) = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2 \end{aligned} \tag{37}$$

Both the MAD and the MSE of some estimator are scale-dependent. The MSE of the sample mean of n i.i.d. random variables with density (32) and v>1 is:

$$MSE\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \overset{(31)}{=} Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) \overset{(28)}{=} \frac{1}{n}\frac{v}{v-2}\sigma^2 \tag{38}$$

The first equality in (38) holds because the sample mean of i.i.d. realization is an unbiased estimator for the true mean since v > 1 and, hence, the mean is finite. The second equality holds because the random variables are independent and, therefore, all covariances are equal to zero. Result (38) shows that if the scale parameter is 2 instead of 1, as in this simulation, the MSE of the sample mean would be $2^2=4$ times as great for a given shape parameter and a given sample size. Result (39) shows that the MAD of the sample mean increases linearly with the scaling factor a:

$$\begin{aligned} a\in\mathbb{R}, a>0;\; Y_i &:= a\cdot X_i \; then\colon\; MAD\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) := E\left(\left|\frac{1}{n}\sum_{i=1}^{n}Y_i - E(Y_1)\right|\right) \\ &= E\left(\left|\frac{1}{n}\sum_{i=1}^{n}a\cdot X_i - E(a\cdot X_1)\right|\right) = |a|\cdot E\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - E(X_1)\right|\right) \overset{\text{def}}{=} a\cdot MAD\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \end{aligned} \tag{39}$$

No MSE estimator (36) was used in this simulation study due to the heavy tails of the simulated Student`s t distributions. (36) would be particularly bad if the variance is the quantity of interest since the degrees of freedom parameter is smaller than 4 in all 50 scenarios and, hence, the 4th moment is always undefined. This poses a huge problem because the expected value of the square of a variance estimator (e.g. the sample variance) depends on the 4th moment of the underlying random variable. Due to the scale dependence of the MAD, the relative MAD or **mean absolute percentage error (MAPE)** was estimated as well, at least for the variance, the 99% quantile and the 99.9% quantile:

$$M\hat{A}PE(\hat{\theta}) = 100 \cdot \hat{E}\left(\left|\frac{\hat{\theta}-\theta}{\theta}\right|\right) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\hat{\theta}_i-\theta}{\theta}\right| = \frac{100}{|\theta|}\cdot M\hat{A}D(\hat{\theta}) \ if \ \theta \neq 0. \tag{40}$$

The MAPE cannot be estimated if the true parameter happens to be 0 such as in the case of the mean and the median in this simulation study. The estimated MAPE reveals how big the relative distance between an estimator and the true quantity is on average for a given scenario.

At this point, it seems important to note that (34) should not be confused with MAPE estimations often used as an alternative metric to *least squares* in forecasting models. These estimations differ from (40) insofar that each estimator (e.g. estimated y_i given x_i in a regression model y ~= f(x)) has a different true quantity (e.g. the actual value y_i). However (40) is a questionable metric in such a case because the smaller the actual value, the bigger the size of the absolute relative distance between the estimated and the true value. Hence, this metric puts weights on the estimates that are inversely proportional to the size of the true quantity. The interested reader is referred to Foss et. al. (2003) and Tofallis (2015) for an in-depth analysis of the MAPE metric as well as alternative metrics. In each of the 1000 iterations of a certain scenario, the true quantity is the same and, therefore, no pitfalls other than the fact that it cannot be estimated if the true quantity happens to be zero exist.

3.2. Results

Whereas all nonparametric (NP) estimates are valid for each and every scenario and iteration, the ML estimates failed some of the time for n=100. In 59 out of these 10000 (10x1000) samples, the MLE of the 3-dimensional parameter vector (location, scale, and v) failed because no maximum could be found. Moreover, the ML estimate of v was less than or equal to 2 in 319 of the remaining 9941 samples. Hence the variance could not be estimated via (34) in these 319 samples. For all scenarios with n>100, the ML estimation never failed and all estimates of v were higher than 2.

Figure 28 shows the estimated (sample) MAD´s (35) of all estimators and all scenarios with n>100. With only a single exception, namely the variance estimates for n=500, the ML estimates had a lower sample MAD than their NP counterparts (The green points in Figure 28

lie below the blue points with the same shape). In accordance with the Strong Law of Large Numbers (9) and the (asymptotic) unbiasedness of the used estimators, the larger n, the smaller the sample MAD. The increase in precision between 5000 and 10000 i.i.d. realizations is quite small, the differences between 500 and 1000 as well as 1000 and 5000, on the other hand, seem to be rather substantial. The influence of v on the sample MAD´s depends strongly on the estimated quantity. No influence can be spotted for the ML estimates of the median and the expected value and for the NP median estimates. The sample MAD´s increase roughly linear with v for both the NP and the ML estimates of the 99% and the 99.9% quantile. The larger n, the lesser the slopes of these "lines", suggesting a constant percentage decrease of the (true) MAD´s with increasing v. In case of the variance, the convex shapes of the blue and green lines suggest that the percentage decrease of the MAD is not constant with increasing v. The sample MAD´s for v=2.9 are between two and three times the size of the respective sample MAD´s for v=3.8.
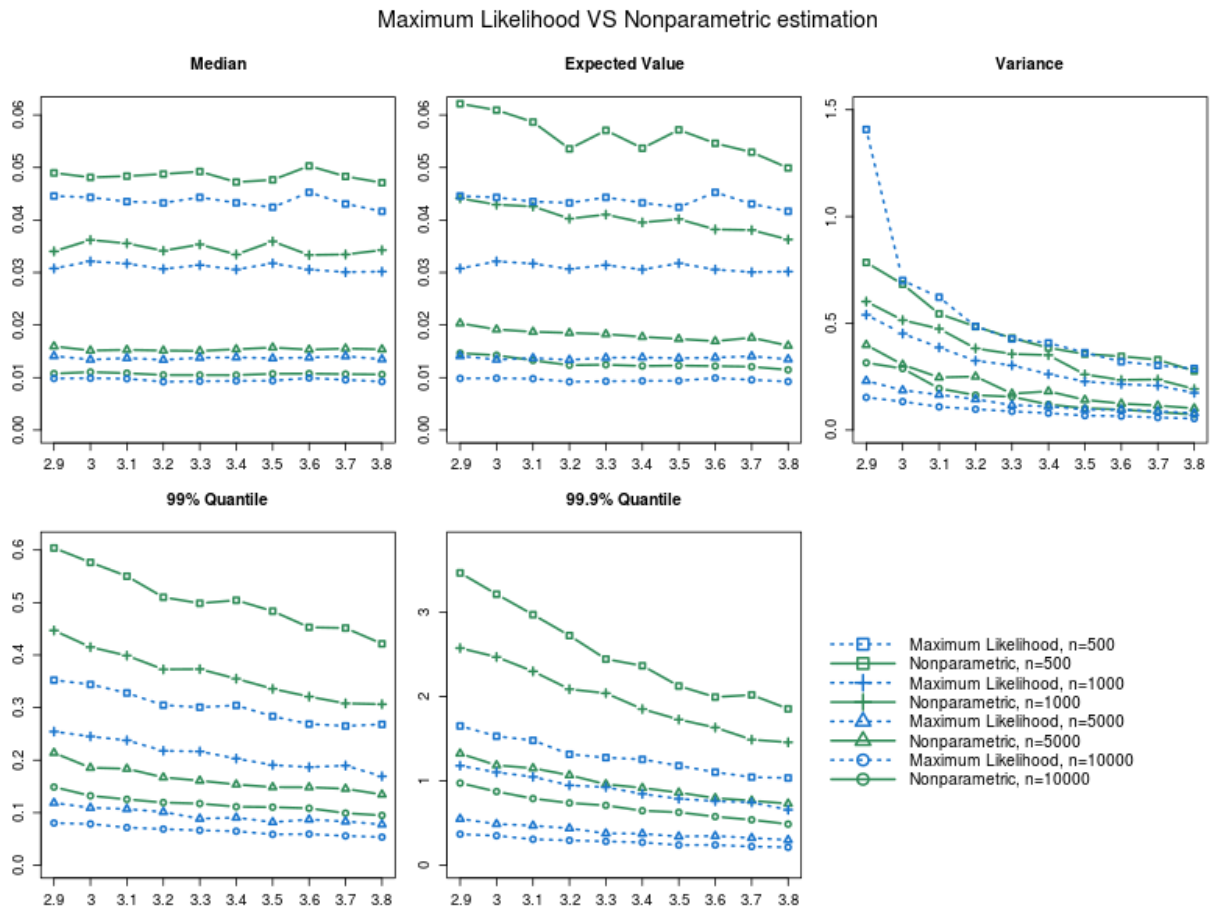

Figure 28: *Sample mean absolute deviations (29) of the ML (blue) and the NP (green) estimates of five important quantities (median, mean, variance, 99% quantile and 99.9% quantile) for different sample sizes and degrees of freedom.*

The used metric (35) shines a light on the question how far an estimator is on average away from the quantity it is estimating and the results in Figure 28 clearly favor the use of MLE over NP estimations for n >= 500. However, the MAD is certainly not the only interesting property of an estimator[13] and hence, further analysis was conducted. Figures 29-32 show the histograms of the respective 1000 estimates for different scenarios and quantities. These histograms shed light on the (true) distribution of the respective estimators. Figure 29 reveals a clear right-skewness of the NP variance estimator (The histograms have more area left of the red dotted line – the true variance). The larger n for a given v, the lesser the skewness of the sample variance´s histogram. For n=500 and n=1000 there are cases where the sample variance is more than ten times as big as the true variance. This shows that a NP variance estimate can be horribly bad at times when the sample size of i.i.d. realizations happens to be small. The higher v, the lesser the skewness of the histograms in Figure 29.

Figure 30 shows the histograms of the ML estimates of the variance for the same scenarios as in Figure 29. The histograms are vastly less skewed as their NP counterparts (Figure 29) and are almost symmetric for n=5000 and n=10000. Furthermore, the worst estimate (i.e. the estimate farthest away from the true variance) is smaller than its NP counterpart for all twelve scenarios. In case of n=1000 and v=3.5, no estimate departs more than 15% from the true variance and the histogram looks very much like coming from i.i.d. realizations of some normal distribution with a density as in (29).

Figure 31 shows the NP and Figure 32 the ML estimates of the 99.9% quantile for the same scenarios as for the variance (Figures 29 & 30). Once again, the NP estimates are smaller than the true quantity (dotted red line) most of the time and occasionally far too big, particularly for n=500 and n=1000. Just like for the variance, the larger n, the lesser the skewness and the lesser the difference between the worst estimate and the true value. Once again, the histograms of the ML estimates (Figure 32) are vastly less skewed than their NP counterparts and the worst estimate is better than its NP counterpart for all twelve scenarios. Unlike ML variance estimates (Figure 30), the histograms even look quite symmetric for n=1000. Whereas the histograms of the ML estimates (Figure 32) all resemble the shape of a normal distribution for n=5000 and n=10000, not one histogram of NP estimates resembles a normal distribution (Figure 31). For a given n, the shapes of the histograms are lesser skewed the bigger v for both the NP and the ML estimates. However, the differences between v=3.1, v=3.3, and v=3.5 are quite small.

---

13 Two estimators A & B could have the same MAD, lets say 1, but behave very differently: Whereas A estimates the true quantity (=50) perfectly 9 out of 10 times, but fails horrible every 10th time where A is equal to 40, estimator B is either 49 or 51 with a probability of ½. Depending on the situation one of these two estimators is to be preferred.
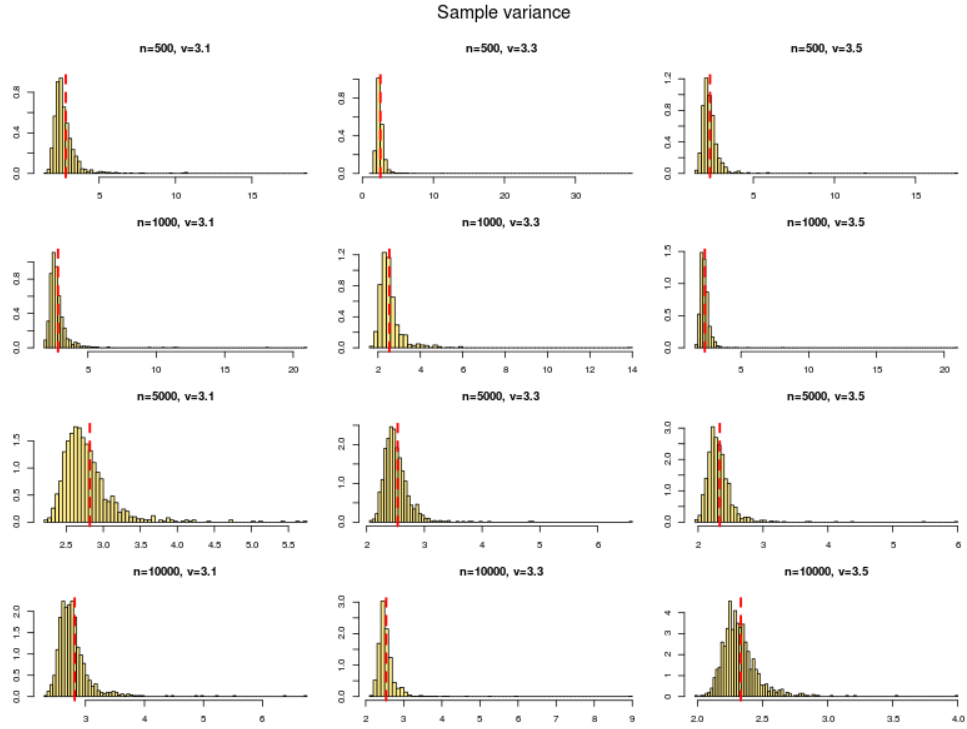
Figure 29: *Histograms of the respective 1000 sample variances and the true variance (red line) for different sample sizes and different degrees of freedom.*
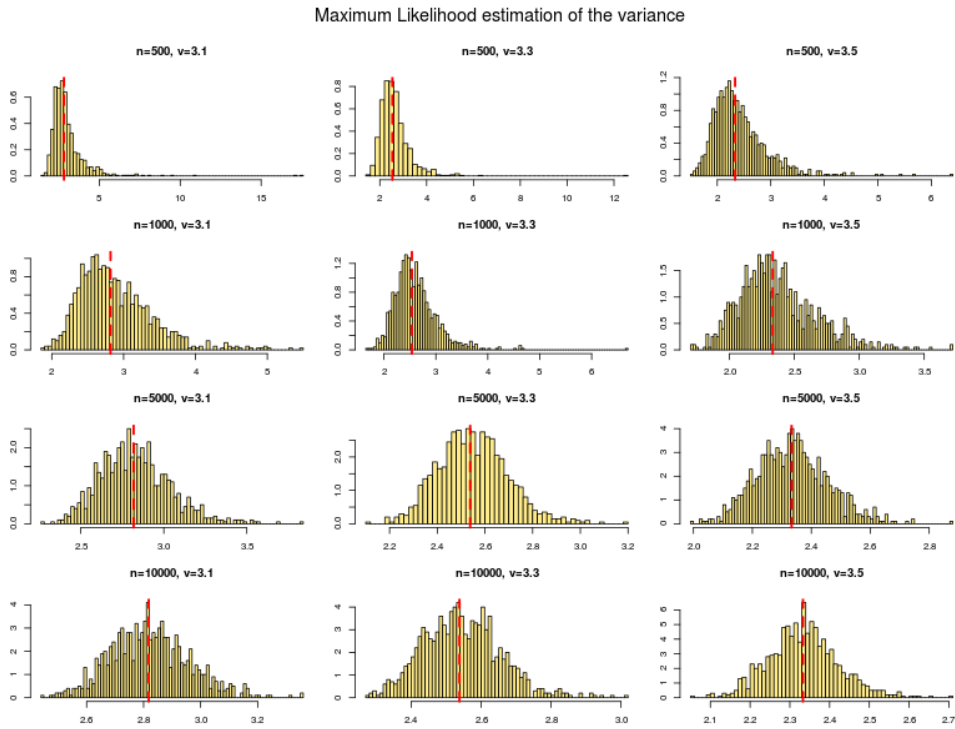
Figure 30: *Histograms of the respective 1000 ML variance estimates and the true variance (red line) for different sample sizes and different degrees of freedom.*
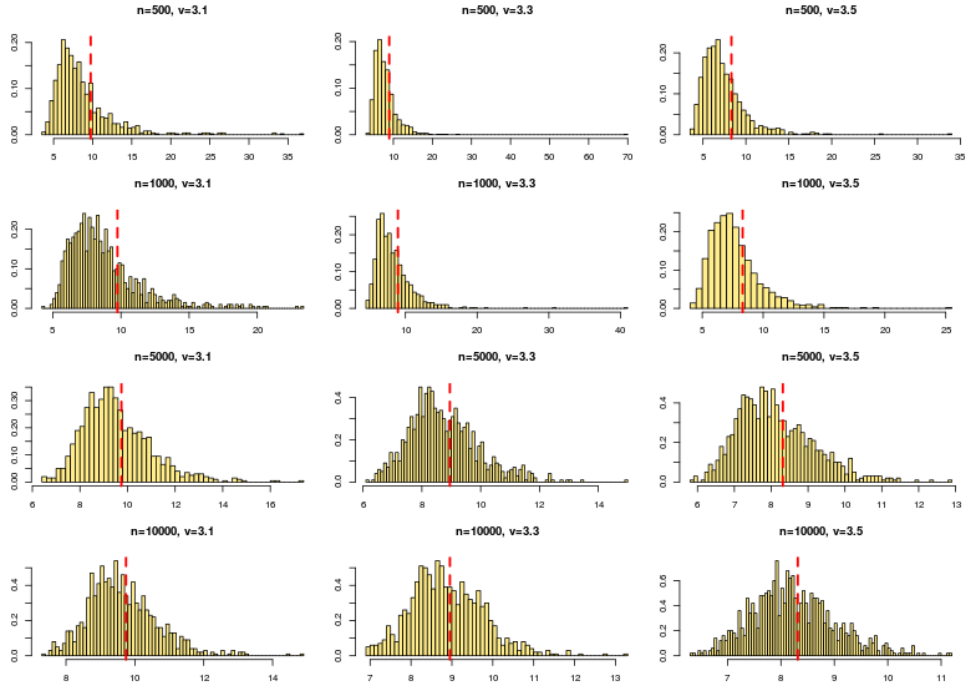
Figure 31: *Histograms of the respective 1000 nonparametric 99.9% quantile estimates and the true 99.9% quantile (red line) for different n and v.*
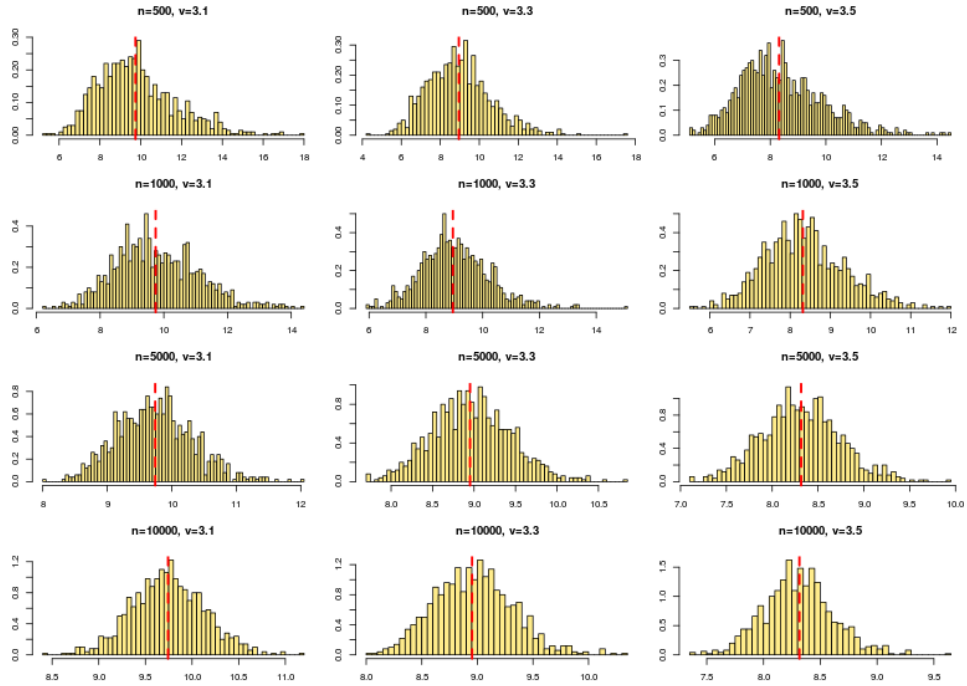
Figure 32: *Histograms of the respective 1000 ML 99.9% quantile estimates and the true 99.9% quantile (red line) for different n and v.*

Table 4 shows the estimated MAPE´s (40) of the NP as well as the ML estimates of the variance, the 99% quantile and the 99.9% quantile. No MAPE´s could be estimated for the ML estimates if the sample size was equal to 100 since the maximization of the likelihood failed at least once in each of these ten scenarios. Table 4 is useful to assess the precision of a certain estimator given some n and v. For instance, for 1000 i.i.d realizations of a Student`s t distribution with v=3.5 the sample variance is on average 11.2% of the true variance´s value away from the true variance, regardless of the location and the scale parameter. The bigger n, the smaller the average estimation error, for both the NP and the ML estimates. However, whereas the improvement for the three estimated quantities is very much alike in case of the ML estimates, the NP quantile estimates benefit more from an increased sample size than the sample variance. While the ML quantile estimates are clearly more precise than the respective NP estimates for n=500, n=1000, n=5000 and n=10000, the ML variance estimates really only excel the respective sample variances for n=5000 and n=10000. For both the ML and the NP estimates it applies that the smaller v, the bigger the estimated MAPE. The degrees of freedom parameter has a considerable bigger impact on the accuracy of the variance estimates than on the accuracy of the quantile estimates.

Sample size	Estimated quantity	Estimator	2.9	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8
n=100	Variance	NP	41.6	37.1	32.2	31.9	27.0	27.6	29.3	27.3	25.8	24.4
		ML	-	-	-	-	-	-	-	-	-	-
	99% quantile	NP	26.8	25.0	26.2	24.5	23.2	24.1	24.1	23.1	22.2	23.5
		ML	-	-	-	-	-	-	-	-	-	-
	99.9% quantile	NP	49.3	48.7	47.5	47.0	46.5	45.7	45.4	43.7	43.2	42.6
		ML	-	-	-	-	-	-	-	-	-	-
n=500	Variance	NP	24.4	22.8	19.3	18.2	17.0	15.9	15.3	15.4	15.2	13.1
		ML	43.7	23.4	22.1	18.2	16.8	16.8	15.5	14.3	13.9	13.7
	99% quantile	NP	12.9	12.7	12.4	11.8	11.8	12.2	11.9	11.4	11.5	10.9
		ML	7.6	7.6	7.6	7.1	7.1	7.3	7.0	6.7	6.8	6.9
	99.9% quantile	NP	32.3	31.5	30.5	29.2	27.3	27.5	25.6	24.8	25.9	24.5
		ML	15.4	15.0	15.2	14.1	14.3	14.6	14.2	13.7	13.4	13.7
n=1000	Variance	NP	18.7	17.2	16.8	14.3	14.1	14.4	11.2	10.4	10.9	9.1
		ML	16.8	15.1	13.7	12.2	12.0	10.8	9.8	9.6	9.6	8.3
	99% quantile	NP	9.6	9.1	9.0	8.6	8.8	8.6	8.3	8.0	7.9	7.9
		ML	5.5	5.4	5.4	5.0	5.1	4.9	4.7	4.7	4.8	4.4
	99.9% quantile	NP	24.0	24.2	23.6	22.4	22.8	21.5	20.8	20.3	19.1	19.3
		ML	11.0	10.8	10.8	10.2	10.3	9.8	9.5	9.4	9.6	8.7
n=5000	Variance	NP	12.4	10.2	8.8	9.4	6.7	7.5	6.1	5.5	5.3	4.8
		ML	7.2	6.2	5.9	5.4	4.6	4.6	4.1	4.3	4.1	3.8
	99% quantile	NP	4.6	4.1	4.2	3.9	3.8	3.7	3.7	3.7	3.7	3.5
		ML	2.5	2.4	2.4	2.3	2.1	2.2	2.0	2.2	2.1	2.0
	99.9% quantile	NP	12.3	11.6	11.8	11.5	10.7	10.6	10.4	9.9	9.8	9.6
		ML	5.1	4.8	4.8	4.7	4.2	4.3	4.1	4.3	4.2	4.0
n=10000	Variance	NP	9.8	9.6	6.9	6.1	6.1	4.9	4.4	4.3	3.9	3.5
		ML	4.8	4.5	3.9	3.7	3.5	3.3	2.9	2.9	2.7	2.6
	99% quantile	NP	3.2	2.9	2.8	2.8	2.8	2.7	2.7	2.7	2.5	2.5
		ML	1.7	1.7	1.6	1.6	1.6	1.6	1.4	1.5	1.4	1.4
	99.9% quantile	NP	9.1	8.5	8.1	7.9	7.9	7.5	7.5	7.2	6.9	6.4
		ML	3.4	3.4	3.2	3.2	3.2	3.1	2.9	3.0	2.8	2.8

Table 4: *The estimated mean absolute percentage error of the NP and the ML estimates of the variance, the 99% quantile and the 99.9% quantile, in each and every scenario.*

4. Concluding remarks

The daily percentage changes and, hence, the daily returns are leptokurtic. Both, the S&P 500 and the Nikkei 225 have fat tails and Pareto tails cannot be ruled out for either of them. The daily returns are not symmetric and in the case of the S&P 500, the left tail is thicker than the right. Stock returns are seldom symmetric, however, the difference in tail thickness varies from stock to stock and from time period to time period. Whereas Sheikh and Qiao (2010) present clear negative skewness (i.e. fatter left tails) for various monthly stock returns, Jondeau and Rockinger (2003) as well as Koronokiewicz and Jamroz (2014) find no significant differences in tail thickness of daily returns. The approach of this work was to fit the same distribution to the left and the right tail and, therefore, the parametric fits of the two tails only differ by their estimated parameters. A more flexible approach would be to use more complicated distributions for the whole dataset which allow for different shapes in the left and the right tail. The *Double Student`s t* and the traditional three parameter Student`s t fit clearly outperformed all other distributions analyzed in this work (Chapter 2.4) and, hence, future research may focus on extensions of the Student`s t distribution. Aas and Haff (2006) give an excellent overview of the various skew extensions of the Student`s t distribution and find that the five parameter *Generalized Hyperbolic skew Student`s t distribution* fits the log returns of Norwegian stocks well. Another possibility, not analyzed in this work, is an artificial "Triple Distribution" fit. One could, for instance, fit the body with a symmetric distribution such as the Laplace and the left as well as the right tail with some distribution from the class of regularly varying distributions. Such a fit would probably be quite good since the analysis in Chapter 2.3 showed that Paretian tails are plausible and Chapter 2.4.1 showed that the Laplace distribution fits the body of the two datasets quite well. One big downside of such a fit is the requirement to estimate at least ten parameters - namely, location (1x), scales (3x), shapes (2x), breaking points (2x) and proportion of data in each part (2x). Reschenhofer (2013) proposed a seven parameter extension of the *wizard´s hat distribution* and future research could address how many parameters are optimal insofar that peculiarities of stock returns are adequately modeled and overfitting is avoided at the same time. All artificially constructed distributions (Chapter 2.4.2) have a discontinuity point at the respective median. This is not necessarily a weakness of this approach. Reschenhofer (2013) provides evidence that a discontinuity point at zero exists for the S&P 500 log returns. Future research may analyze discontinuity points of stock returns and whether it is advantageous to split the dataset at zero as opposed to at the median as in this work.

The results of the simulation (Chapter 3.2) showed that the maximum likelihood estimations of five important quantities outperformed the respective nonparametric estimations, for a big enough sample size. This is certainly no surprise as every statistician knows that parametric estimates outperform nonparametric if the true model is known. However, given real (financial) data, it is impossible to know the (true) model that generated the dataset. Although the

Student`s t distribution seems to be a good approximation (Chapter 2.4), the daily percentage changes of the S&P 500 as well as the Nikkei 225 index are certainly not i.i.d. realizations of some continuous distribution (Chapter 2.2). If the assumed parametric model is wrong, the parametric estimates are biased. In case of severe differences between the assumed and the true model, parametric estimates can be horribly wrong. Nonparametric estimates, on the other hand, are much more robust and future research may focus on the question how big the difference between the assumed and the true model can be for the parametric approach to be superior to the nonparametric. An often used metric in finance and insurance is the so-called *p-Value at Risk* (p-V@R). The p-V@R is the value which is exceeded with a probability of p. If the daily percentage change of some stock happens to be Student`s t distributed with location=0, scale=1 and v=3.5 then the 0.99-V@R is the 99% quantile of this distribution and Table 4 tells us how precise we can estimate this value for a given sample size. Alongside the p-V@R, volatility estimates, the variance in particular, are the most important quantities in finance, insurance, and risk analysis. In case of financial data or any other data coming from areas with equally heavy tails (e.g. natural disasters like earthquakes, floods, and operational losses) Table 4 suggests that additional data benefits nonparametric p-V@R estimations more than nonparametric volatility estimations. Table 4 also suggests that the thicker the tail of a certain dataset supposedly is, the better the accuracy of p-V@R estimates compared to variance estimates. Nonparametric p-V@R and variance estimates tend to underestimate the respective true values, particularly in small samples. However, every once in a while they hugely overestimate the true value.

References

Aas, K., & Haff, I. H. (2006). The generalized hyperbolic skew student's t-distribution. *Journal of financial econometrics*, *4*(2), 275-309.

Akgiray, V., & Booth, G. G. (1988). The stable-law model of stock returns. *Journal of Business & Economic Statistics*, *6*(1), 51-57.

Andrade, S. C., Chhaochharia, V., & Fuerst, M. E. (2012). 'Sell in may and go away" just won't go away.

Bachelier, L. (1900). Théorie de la spéculation. *Gauthier-Villars.*

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637-654.

Ciner, C., Gurdgiev, C., & Lucey, B. M. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates. *International Review of Financial Analysis*, *29*, 202-211.

Cirillo, P., & Taleb, N. N. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, *452*, 29-45.

CNN Money http://money.cnn.com/data/markets/sandp/, 2017.

Cont, R. (2007). Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics* (pp. 289-309). Springer Berlin Heidelberg.

De Haan, L., & Ferreira, A. (2007). Extreme value theory: an introduction. *Springer Science & Business Media.*

De Sousa, B., & Michailidis, G. (2004). A diagnostic plot for estimating the tail index of a distribution. *Journal of Computational and Figureical Statistics*, *13*(4), 974-995.

Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Probability Theory and Related Fields*, *55*(1), 119-122.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, *38*(1), 34-105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383-417.

Foss, T., Stensrud, E., Kitchenham, B., & Myrtveit, I. (2003). A simulation study of the model evaluation criterion MMRE. *IEEE Transactions on Software Engineering*, *29*(11), 985-995.

Gärtner, M., & Wellershoff, K. W. (1995). Is there an election cycle in American stock returns?. *International Review of Economics & Finance*, *4*(4), 387-410.

Haugen, R. A., & Jorion, P. (1996). The January effect: Still there after all these years. *Financial Analysts Journal*, 27-31.

Henrik Bengtsson (2016). matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.51.0.

Investopedia, http://www.investopedia.com/ask/answers/130.asp, 2017

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1970). Distributions in statistics: continuous univariate distributions, Vol. 2. *NY: Wiley*.

Jondeau, E., & Rockinger, M. (2003). Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, *10*(5), 559-581.

Kingsley, Z. G. (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Addison-Wesley Press.

Koronkiewicz, G., & Jamroz, P. (2014). Comparison of the tails of market return distributions.

Kunst, R. M., Reschenhofer, E., & Rodler, K. (1991). Analysis of Austrian stocks: Testing for stability and randomness. *Empirical Economics*, *16*(4), 465-477.

Lau, A. H. L., Lau, H. S., & Wingender, J. R. (1990). The distribution of stock returns: New evidence against the stable model. *Journal of Business & Economic Statistics*, *8*(2), 217-223.

Lévy, P. (1928). Calcul des probabilités.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies*, *1*(1), 41-66.

Lo, A. W., & MacKinlay, A. C. (2002). *A non-random walk down Wall Street*. Princeton University Press.

Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns. *Journal of business*, 383-408.

Mandelbrot B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394-419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, *40*(4), 393-413.

Mandelbrot, B., & Hudson, R. L. (2007). The Misbehavior of Markets: A fractal view of financial turbulence. *Basic books.*

Osborne, M. M. (1959). Brownian motion in the stock market. *Operations research*, 7(2), 145-173.

Peiro, A. (1994). Daily seasonality in stock returns: Further international evidence. *Economics Letters*, *45*(2), 227-232.

Random, Probability, Mathematical Statistics, Stochastic Processes, *http://www.math.uah.edu/stat/special/Logistic.html*, 2017

Read, C. (1906). *Logic deductive and inductive*. A. Moring, Limited.

Roll, R. (1970). The behavior of interest rates: An application of the efficient market model to US treasury bills (Vol. 1). *Basic Books (AZ).*

Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, 369-374.

Sewell, M. (2011). History of the efficient market hypothesis. *RN*, *11*(04), 04.

Sheikh, A. Z., & Qiao, H. (2010). „Non-Normality of Market Returns: A Framework for Asset Allocation Decision Making (Digest Summary)". *Journal of Alternative Investments*, *12*(3), 8-35.

Statisticat, LLC. (2016). LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com. R package version 16.0.1

Team, R. C. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, *66*(8), 1352-1362.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. *Springer, New York.* ISBN 0-387-95457-0

Visual Capitalist, http://money.visualcapitalist.com/all-of-the-worlds-stock-exchanges-by- size/, 2017.

Wiener, N. (1923). Differential-Space. *Studies in Applied Mathematics*, *2*(1-4), 131-174.

Wierman, A., CATASTROPHIES, CONSPIRACIES, AND SUBEXPONENTIAL DISTRIBUTIONS (PART III),*Rigor+Relevance,https://rigorandrelevance.wordpress.com/2014/01/09/*catastrophes-conspiracies-and-subexponential-distributions-part-iii/, 2017

Yahoo Finance, https://finance.yahoo.com/, 2017

Abstract / Zusammenfassung

Abstract:

The daily returns of the S&P 500 index from 20.11.1950 to 18.11.2016 and the Nikkei 225 index from 5.01.1984 to 7.12.2016 are analyzed in great depth with a focus on graphical analyses. The analyses revealed that both datasets have fat-tails and that estimations of skewness and kurtosis are problematic because the estimates vary widely in different time periods and are significantly influenced by just a few days with strong index fluctuations. Parametric analysis of the tails clearly suggests that the tails of stock returns are so thick that no 4th moment exists. Various symmetric continuous probability distributions were fitted to the datasets and, building on the assessment that the properties of negative and positive returns differ, artificial distributions were constructed out of different fits for the left and the right tail for both indices. The three parameter Student`s t distribution fits both datasets reasonable well and, hence, samples of different sizes of independent and identically Student`s t distributed random variables with shape parameters in the range of [2.9, 3.8] were simulated. Five important quantities, namely mean, median, variance, 99% quantile and 99.9% quantile were then estimated parametrically via maximum likelihood as well as nonparametrically and the mean absolute deviation was computed for each estimator in each scenario. The parametric estimations did not turn out to be satisfactory for a sample size of 100 because the estimations of the shape parameter occasionally deviated strongly from the true parameter and the maximization of the likelihood even failed in some of the iterations. For sample sizes larger than 500, the parametric estimates are clearly superior to their nonparametric counterparts. The nonparametric quantile and variance estimates are much more right-skewed than their parametric counterparts, particularly for smaller sample sizes. The sample quantiles (99% and 99.9%) benefit more from an increase in sample size and are harmed less from a decrease of the shape parameter (i.e. a thicker tail) than the sample variance.

Keywords: Fat-tails, Mean Absolute Deviation, Simulation Study, Stock Returns, Student`s t Distribution

Zusammenfassung:

Die täglichen Renditen des Standard & Poor`s 500 Aktienindexes vom 20.11.1950 bis zum 18.11.2016 und die des Nikkei 225 Aktienindexes vom 5.01.1984 bis zum 7.12.2016 wurden auf Herz und Nieren untersucht, und zwar mit einem starken Fokus auf graphischen Analysemethoden. Beide Datensätze haben sogenannte „fat-tails" und Schätzungen des standardisierten dritten (Schiefe) und vierten (Kurtosis) Moments sind problematisch da sich die Schätzwerte je nach Zeitperiode stark unterscheiden und von einigen wenigen Tagen mit starken Kursschwankungen maßgeblich beeinflusst werden. Die durchgeführten parametrischen Analysen legen zudem den Schluss nahe, dass kein viertes Moment existiert. Verschiedene symmetrische und stetige Wahrscheinlichkeitsverteilungen wurden an die beiden Datensätze angepasst um weitere Aufschlüsse über das Verhalten der Renditen zu gewinnen. Aufgrund der Erkenntnis, dass sich positive und negative Renditen unterscheiden, wurden des Weiteren artifizielle Verteilungen durch ein separates Anpassen von Verteilungen an die oberen und unteren 50% der Daten kreiert. Die (drei Parameter) studentsche t-Verteilung beschreibt die Renditen der beiden Aktienindexe recht gut und aus diesem Grund wurden mithilfe der Programmiersprache R unterschiedlich große Stichproben aus unabhängigen und identisch t-verteilten Zufallsvariablen, mit Freiheitsgraden aus dem Intervall [2.9, 3.8], erzeugt. Fünf essentielle statistische Größen, nämlich Erwartungswert, Median, Varianz, 99% Quantil und 99.9% Quantil wurden anschließend mittels parametrischer Maximum-Likelihood-Methode sowie nicht-parametrisch geschätzt. In weiterer Folge wurde die Genauigkeit der Schätzungen mithilfe der mittleren absoluten Abweichung ermittelt. Die parametrischen Schätzungen waren für eine Stichprobengröße von 100 nicht zufriedenstellend, da die Schätzungen der Freiheitsgrade teilweise sehr schlecht waren und in manchen Iterationen sogar die Maximierung der Likelihood scheiterte. Ab einer Stichprobengröße von 1000 zeigte sich, dass die parametrischen Schätzungen genauer als die nicht-parametrischen sind. Die nicht-parametrischen Quantils- und Varianzschätzer, also die Stichprobenquantile sowie die Stichprobenvarianz, sind rechts-schief und weniger symmetrisch als die entsprechenden parametrischen Schätzer, insbesondere bei einer Stichprobengröße von 100 oder 500. Die Stichprobenquantile (99% und 99.9%) profitieren stärker von einer Erhöhung der Stichprobengröße und weniger stark von einer Verringerung der Freiheitsgrade als die Stichprobenvarianz.