ADVANCED TEXT MINING METHODS FOR THE FINANCIAL MARKETS AND FORECASTING OF INTRADAY VOLATILITY

Zur Erlangung des akademischem Grades eines

DOKTORS DER WIRTSCHAFTSWISSENSCHAFTEN

(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
des Karlsruher Institut für Technologie

genehmigte

DISSERTATION

von

DIPL.-PHYS. MICHAEL J. PIEPER
aus Essen

Tag der mündlichen Prüfung: 8.12.2011
Referent: PROF. DR. SVETLOZAR T. RACHEV
Korreferent: PROF. DR. MICHAEL FEINDT

Karlsruhe, September 2011

MICHAEL J. PIEPER

ADVANCED TEXT MINING METHODS FOR THE FINANCIAL MARKETS AND FORECASTING OF INTRADAY VOLATILITY

To Vanessa Kim

ERKLÄRUNG

Ich versichere wahrheitsgemäß, die Dissertation bis auf die in der Abhandlung angegebene Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig angegeben und genau kenntlich gemacht zu haben, was aus Arbeiten anderer und aus eigenen Veröffentlichungen unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, September 2011

Michael J. Pieper

PREFACE

The flow of information is a fundamental aspect of the dynamics of financial markets. In particular, Ross [82] showed that the flow of information can be measured by an asset's realized volatility. Thus, the relation between the release of information and realized return volatility with respect to various asset classes as well as different sources of information were investigated.

For example, Ederington and Lee [29] notice that macroeconomic news announcements account for much of the intra-day as well as inter-day volatility of interest and foreign exchange futures markets. Their findings are in line with the efficient market hypothesis. Fleming and Remolona [35] argue that the release of macroeconomic news triggers a two stage process in the US treasury market with nearly instantaneous price changes, a reduction in trading volume and widening spreads in the first step and persisting volatility as well as increased trading volumes in the second step. Concerning the stock market Nikkinen and Sahlström [71] explore the impact of scheduled domestic and US macroeconomic news announcements on the implied volatility of the German and Finnish markets. They find that German and Finnish news have a low signal to noise ratio whereas the US employment report and the meeting of the Federal Open Market Committee have significant impact.

Consequently, the first part of this thesis focuses on generic approximations of intra-day volatility using high-frequency data. This includes explicit estimates for the immediate adjustments of volatility as well lifetimes for various event classes. The corresponding high-order profile of intraday dispersion is merged with classical GARCH models in order improve the forecasting of intraday volatility, in particular, with respect to backtesting Value-at-Risk which lies at the core of current regulatory frameworks.

More precisely, the intraday dispersion as measured by realized absolute returns on one minute time intervals of the US and Europe is examined. For Europe, market fragmentation is taken into account by analysing patterns on Euronext, London Stock Exchange and XETRA. Moreover, both single stock as well as market patterns are investigated. For the latter, two measures are applied. The first considers the realization of various indices such as the Standard & Poors 500 (S&P500) as well as the German blue chip index, the DAX. The second representation averages over the single stock universe.

A rank-k approximation of the intraday data is performed. This allows to model pure intraday effects independently of inter-day effects. For example, in the US the well-known profile which shows increased volatility near the beginning and end of the continuous trading session is modified by the release of macroeconomic news. Considering Europe further mechanisms related to the opening of the US market as well as market mechanics are described.

Moreover, different types of *information shock* exist. Announced events such as macroeconomic indicators or earning reports are anticipated by investors and the corresponding volatility response may be estimated in a robust fashion. However, without certain event-specific, exogenous variables it is largely impossible to forecast unexpected events such as the so-called *Flash-Crash* as well as their effects on financial markets. Therefore, a method is proposed to identify effects triggered by external events in the data and to remove them in an automatic fashion from the generic intraday volatility profile.

Response patterns of the S&P500 and the DAX to macroeconomic news releases are explored. To this end, immediate adjustments to the typical intraday profile as well as characteristic lifetimes for specific events are stated explicitly. In general, it is discovered that immediate adjustments are stronger and lifetimes are longer with respect to the S&P500 as compared to the DAX.

Subsequently, the high-order profile of intraday return dispersion is constructed using no future information in order to boost the performance of classical GARCH models which were originally introduced by Engle [31], and extended by Bollerslev [12]. To this end, realizations of the S&P500 and the DAX on time horizons ranging from 15 seconds to 10 minutes are fitted to various GARCH models. Furthermore, boosted variants called SVD-GARCH, SVD-EGARCH and SVD-GJR are introduced.

Thereby, the goodness-of-fit of various statistical distributions such as the normal, the Student-t, the classical tempered stable and the variance-gamma distribution of the innovation process are examined. In particular, hetereoskedasticity of financial time series in part yields leptokurtic, unconditional return distributions as noticed by Mandelbrot [63]. However, especially in the context of high-frequency data, the innovation process still exhibits considerably high excess kurtosis. Stable Paretian distributions were successfully applied in finance, see, for example, Rachev and Mittnik [77]. Both the need for finite second moments, e. g., in the context of mean-variance portfolio theory [65], as well as debates over the actual tail-behaviour of the empirical distributions have led to the introduction of new models such as the classical tempered stable [81] and variance gamma distributions [61]. In general, it is found that both the classical tempered

stable and the variance gamma distribution describe high-frequency data rather well.

In addition, the scaling behaviour of parameters of the variance gamma distribution are investigated. It is found that the extraordinary increase of kurtosis of the innovation process corresponding to the high-frequency end of the spectrum, i. e., all data on time horizons below 1 minute, can be explained by a non-vanishing autocorrelation function of the standardized innovations.

Finally, one-step-ahead forecasts of volatility for the various models on 1 minute time horizons are calculated. Concerning the SVD-GARCH and SVD-GJR models a significant improvement with respect their classical variants as measured by the mean absolute percentage error can be reported. However, for the EGARCH model no such improvement is observed. Moreover, a Value-at-Risk backtest using Kupiec's proportion of failures and is conducted on the 1 minute time horizon. It is found that SVD variants in general improve the acceptance rate whereas classical GARCH formulations overestimate risk and, thus, are rejected.

The second part of this thesis examines the flow of information from a point of view that focuses more on the actual information contained in financial news. In 2007, mankind was able to communicate almost $2 \cdot 10^{21}$ entropy maximized bytes [42]. Moreover, the globally stored information grew by roughly 23%. As both the capabilities to store vast amounts of data as well as the capacities to communicate increases so does the need to process the information.

Over the last decade, text mining of news and its application to finance were a vibrant topic of research starting with the pioneering work of Wüthrich [90]. Generally, one transforms the unstructured data into a set of numbers such that methods from statistical learning theory can be applied. A concise review of the existing prototypes was written by Mittermayer and Knolmayer [68].

However, the existing prototypes typically lack on one of the following dimensions. In some cases features are used on the basis of expert-compiled dictionaries. However, only rarely a proof of significance is given. Moreover, no systematic investigation of the actual time scales of returns that are related to ad-hoc news reporting on individual stocks exists. In some cases the forecast of the respective machine learning is not even compared to a random classification that uses only the inclusive target probabilities. Even worse, future information cannot be excluded. Finally, the profitability of various approaches is not stated.

The proposed framework tackles all of the aspects above. To this end, times series of stocks that can non-ambiguously be linked to certain news are investigated. The analysis is restricted to the European market. In particular, appropriately normalized returns as well as return dispersions

on one minute time horizons are considered. A low-rank approximation of the data is performed that exhibits typical response patterns. It is found that volatility adjustments are largest within the minute of release of new information. However, for the dataset of roughly 1.4 million text-messages there appears to be evidence that volatility adjustments both following as well as proceeding the actual news release exists. No price adjustments are observed on time horizons such as hours or days.

The response behaviour is consistent both for news that are released during the continuous trading as well as for news published when markets are closed. Naturally, for information related to nightly news, the corresponding response of the return can no longer be linked solely to information contained within the single news but the effect is rather a measure of the collective information aggregated over the non-trading period.

With respect to a suitable representation of the news, an approach is pursuit that allows to assess the relevance of individual features in an automatic fashion. A transformation is applied that aims to preserve as much of the original information while enhancing the statistical weight of individual features. Ultimately, this leads to the term-frequency inverse-document-frequency measure. Thereby, the importance of individual features is measured leading and the most important words are stated explicitly.

Finally, various formulations of Support Vector Machines with varying dimensions of the feature space as well as different time horizons following the minutes of news publications are trained. Moreover, both linear as well as non-linear kernels are considered. Compared to a random classification that takes into account the global target distribution, it is found that the out-of-sample accuracy of the classification is enhanced. Subsequently a simple strategy is proposed that generates on average between 17.5bps and 46.3bps per round-trip.

The remainder of this thesis is structured as follows. Chapter 1 reviews some basic concepts of statistics which are applied in the thesis. A brief introduction to statistical learning theory is given. Chapter 2 introduces financial econometrics. In light of the vast amount of work conducted over the last decades, we restrict the discussion to return distributions and classical approaches of volatility modeling. Chapter 3 analyses patterns of intraday volatility and applies the findings in order to enhance classical GARCH models. Chapter 4 introduces a framework to forecast short-term price movements on the basis of unstructured information contained in financial news. The appendix reviews the most important probability distributions as utilized throughout this thesis as well as various statistical tests.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AD Anderson-Darling

BLUE Best Linear Unbiased Estimator

CAPM Capital Asset Pricing Model

CDF Cumulative Distribution Function

CGMY Carr-Geman-Madan-Yor

CTS Classical Tempered-Stable

CRB Cramér Rao Bound

CTS Classical Tempered Stable

DJIA Dow Jones Industrial Average

EMA Exponential Moving Average

FOMC Federal Open Market Committee

IID Independent and Identically Distributed

IR Information Retrieval

ISIN International Securities Identification Number

KS Kolmogorov-Smirnov

LSE London Stock Exchange

MAPE Mean Absolute Percentage Error

MLE Maximum Likelihood Estimator

PDF Probability Density Function

POF Proportion of Failures

RDTS Rapidly Decreasing Tempered-Stable

SVD Singular Value Decomposition

TF-IDF Term Frequency-Inverse Document Frequency

VaR Value-at-Risk

VC Vapnik-Chervonenkis

VG Variance Gamma

Part I

INTRODUCTION

1 A BRIEF REVIEW OF STATISTICS

This chapter on statistics and probability theory focuses on the most important properties and ideas as employed in later chapters. We discuss basic concepts of probability theory as well as quantities to characterize probability distributions. This is followed by sections on stochastic processes, the central limit theorem and stability, basics of information theory and parameter estimation. The chapter concludes with a compact introduction to supervised learning theory including a discussion of model selection, the Vapnik-Chervonenkis dimension and Support Vector Machines.

1.1 PROBABILITY

Probability theory is based on investigations regarding the game of chance which date back to early publications by Cardano [16], Huygens [43] and Galilei [37]. At its core lies the notion of a random experiment, i. e., given a number of a priori known outcomes the result of a single experiment is not deterministic. However, one can repeat it under identical conditions. Modern probability theory is predicated on Kolmogorow's [53] axiomatic formulation[1] which introduces a non-empty set, the sample space Ω, of possible outcomes ω. This set can be either countable, e. g., rolling a dice, uncountable, e. g., the position of an elementary particle, or decomposable into both. The sample space Ω is supplemented by a σ-field $\mathcal{F}$, i. e., a set of all possible unions of $A_i \subset \Omega$ called events. For example, all even numbers of a dice might belong to an event.

Given a σ-field $\mathcal{F}$ defined by the following properties

$$\Omega \in \mathcal{F} \tag{1.1}$$

$$A \in \mathcal{F} \rightarrow (\Omega \backslash A) \in \mathcal{F} \tag{1.2}$$

$$A_i \in \mathcal{F} \rightarrow (\cup_i A_i) \in \mathcal{F} \tag{1.3}$$

1 Other theories on probability exist, among them Fuzzy sets [91] and negative probability [25].

the space $(\Omega, \mathcal{F})$ is measurable [8]. Equipped with a probability measure P one obtains a probability space $(\Omega, \mathcal{F}, P)$. Hereby, the probability measure P assigns a real number between 0 and 1 to every event A_i, such that

$$P(A_i) \geqslant 0, \ A_i \in \mathcal{F} \tag{1.4}$$

$$P(\Omega) = 1 \tag{1.5}$$

$$P(\cup_i A_i) = \sum_i P(A_i) \text{ if } A_i \cap A_j = \emptyset, \tag{1.6}$$

where $\{A_i\}$ is countable. However, in practice one is faced with the need to fix individual probabilities for certain events A_i. Unfortunately, different interpretations of probability exist. *Frequentists* view probability to be the result of infinite, independent repetitions of a random experiment under identical conditions. That is, given k occurrences of an event A_i after n trials the true probability emerges under $\lim_{n\to\infty} k/n$. Typically, the assumption of identical and infinite repeatability are questioned by the *Bayesian* approach that considers probability to be the result of real data modified by a prior. Thus, probability reflects the available information of the given system. This induces the possible notion of personal belief [46].

1.1.1 *Discrete Probabilities*

For a countable sample space Ω probability is assigned to every atomic outcome ω of the random experiment. Thus, the probability of an event $A_i \in \mathcal{F}$ is given by

$$P(A_i) = \sum_\omega P(\omega), \ \omega \in A_i. \tag{1.7}$$

The conditional probability of two events $A_i, A_j \in \mathcal{F}$, that is the probability of an event A_i given another event A_j has occurred, is defined by

$$P(A_i|A_j) = \frac{P(A_i \cap A_j)}{P(A_j)}, \tag{1.8}$$

with $P(A_j) > 0$. Two events are independent if $P(A_i \cap A_j) = P(A_i)P(A_j)$, with $P(A_j) > 0$, i. e., $P(A_i|A_j) = P(A_i)$.

From equation 1.8 one then can deduce Bayes' theorem using the symmetry property of sets

$$P(A_i|A_j) = \frac{P(A_j|A_i)P(A_i)}{P(A_j)}, \ P(A_i), P(A_j) > 0, \tag{1.9}$$

where $P(A_i|A_j)$ is the posterior probability, $P(A_j|A_i)$ the likelihood and $P(A_{i,j})$ the prior probability of the respective event. Thus, Bayes' theorem essentially allows the reversal of conditioning, i. e., given the probability for

A_i under the condition that A_j has occurred one can deduce the probability for A_j under the condition that A_i has occurred. However, there is long lasting dispute about what are the statistical problems where one can apply Bayes theorem (1.9). For a more exhaustive discussion we refer to Carlin and Louis [17].

1.1.2 *Continuous Probabilities*

A continuous random variable maps the outcome of a random experiment to the real numbers. Thus, the random variable $X(\omega)$ is a function $X : \omega \mapsto \mathbb{R}$ such that

$$\{\omega : X(\omega) \leqslant x'\} \in \mathcal{F}, \quad x' \in \mathbb{R}. \tag{1.10}$$

The inverse $X^{-1}((-\infty, x'])$ is an event, i. e., the corresponding measure is induced by a Borel σ-algebra. For the space $(\Sigma, \mathcal{F}, P)$ the cumulative distribution function (CDF) or probability mass function of the random variable $X(\omega)$ is given by

$$F_X(x') = P(X \leqslant x') = \int_{-\infty}^{x'} dF_X(x) \tag{1.11}$$

where the integral is defined in the Lebesgue sense. Equation (1.4) to (1.6) imply that the CDF is a càdlàg[2] function. Furthermore, it is monotonically non-decreasing and fulfills

$$\lim_{x' \to -\infty} F_X(x') = 0 \quad \text{and} \quad \lim_{x' \to +\infty} F_X(x') = 1. \tag{1.12}$$

The inverse CDF[3] reads

$$Q_X(p) \equiv \inf_{x \in \mathbb{R}} \{p \leqslant F_X(x)\}, \quad p \in (0, 1). \tag{1.13}$$

In particular, the median $Q_X(1/2)$ is a robust parameter of the location that divides the lower and upper half of the probability mass.

Due to the continuity of the CDF derivatives exist which results in the definition of the probability density function (PDF) $f_X(x) = dF_X(x)/dx$. Given the PDF $f_X(x)$ the corresponding PDF of $Z = h(X)$ is given by

$$f_Z(z) = \left| \frac{dx}{dz} \right| f_X(x) = \sum_{i=1}^{n(z)} \left| h'(h_i^{-1}(z)) \right|^{-1} f_X(h_i^{-1}(z)) \tag{1.14}$$

with the Jacobian $|dx/dz|$ and where $h(x) = z$ has $n(z)$ solutions.

2 Continu à droite, limite à gauche

3 Also called quantile function

The generalization to n random variables is straightforward. For details concerning measure theory we refer to [8]. For notational simplicity we state the two dimensional case given random variables X and Y. The joint probability is defined by

$$P(X \leqslant x', Y \leqslant y') \equiv F_{X,Y}(x', y') = \int_{-\infty}^{x'} \int_{-\infty}^{y'} f_{X,Y}(x, y) dx dy, \tag{1.15}$$

with the two dimensional density distribution $f_{X,Y}(x, y)$. Unconditional probabilities for events from X average over outcomes of Y

$$P(X \leqslant x') \equiv F_X(x') = \int_{-\infty}^{x'} \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy}_{f_X(x)} dx, \tag{1.16}$$

with the marginal CDF $F_X(x)$ and the marginal density function $f_X(x)$ of X. Furthermore, given (1.15) and (1.16) the conditional density $f_X(x|y)$ is defined by

$$f_X(x|y) \equiv \frac{f_{X,Y}(x, y)}{f_Y(y)}. \tag{1.17}$$

However, the Borel-Kolmogorow paradox [44] – conditioning on events of zero probability – implies that conditional density distributions are not invariant under coordinate transformations of the *dependent* variable.

1.1.3 *Expected Value and Higher Moments*

The expected value $\mathbb{E}[X]$ of a random variable X is interpreted as a localization parameter of the probability distribution

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{1.18}$$

For discrete variants one replaces the integral over the distribution by a weighted sum over the probabilities. The definition (1.18) implies that it is a linear operator, i. e., given two random variables X and Y it holds

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c, \quad a, b, c \in \mathbb{R}. \tag{1.19}$$

Furthermore, the expectation operator does not commute[4] with arbitrary (albeit convex) functions ψ

$$\psi(\mathbb{E}[X]) \leqslant \mathbb{E}[\psi(X)]. \tag{1.20}$$

4 This is Jensen's inequality.

The conditional expectation of a random variable X conditioned on Y that has taken a value y is given by

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x f(x|y) dx. \tag{1.21}$$

Using equation (1.17) and (1.21) one arrives at the law of iterated expectation

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[X|Y=y] f_Y(y) dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y) dx dy = \mathbb{E}[X].\end{aligned} \tag{1.22}$$

Furthermore, the expected value of $\exp(itX),\ t\in\mathbb{R}$, is the characteristic function

$$\phi_X(t) \equiv \mathbb{E}[e^{itX}] \equiv \int_{-\infty}^{\infty} e^{itx} f_X(x) dx \tag{1.23}$$

which is just the Fourier transform $\mathfrak{F}(f_X(x))$ of the density function $f_X(x)$ and thus completely describes the random variable X. From a theoretical point of view the characteristic function is very interesting as it is uniformly continuous on $\mathbb{C}$ which is very useful when considering sequences of random variables (section 1.3).

The covariance of two random variables X and Y describes their *linear* dependence and is defined by

$$\mathrm{Cov}(X,Y) \equiv \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{1.24}$$

Thus, the expected value of the product of two random variables it not equal to the product of the expectation values of two random variables unless the probability density factorizes, i. e., the variables are independent. Furthermore, covariance essentially can be interpreted as an inner product, i. e., it is a symmetric and positive semi-definite bilinear form with the following properties

$$\mathrm{Cov}(aX+bY+c, Z) = a\mathrm{Cov}(X,Z) + b\mathrm{Cov}(Y,Z) \tag{1.25}$$

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) \tag{1.26}$$

$$\mathrm{Cov}(X,X) \geqslant 0 \tag{1.27}$$

with random variables X, Y and Z and $a, b, c \in \mathbb{R}$. The covariance of a random variable with itself is the variance

$$\sigma_X^2 = \mathrm{Var}[X] \equiv \mathrm{Cov}(X,X), \tag{1.28}$$

with the standard deviation $\sigma_X = \sqrt{\mathrm{Var}[X]}$ of the random variable X. It measures the spread of the probability distribution.

In order to make the linear dependence of different random variables comparable one introduces Pearson's correlation coefficient defined by

$$\rho_{XY} = \mathrm{Corr}(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{1.29}$$

Equation (1.29) can be interpreted as the covariance of two standartized random variables. i. e., ρ_{XY} takes values between -1 and 1. While it is true that the correlation coefficient is a good measure of linear dependence for Gaussian random variables other measures are needed for random variables from arbitrary distributions.

Some additional notes are in order. Given a conditional distribution $f_X(X|Y)$ that has a nonlinear dependence on Y Pearson's correlation coefficient cannot capture the whole dependence. Even worse, for nonlinear relationships different distributions might be mapped to the same correlation coefficient. From (1.24) it follows that the covariance and hence the correlation of statistically independent random variables is zero. However, the reverse conclusion is not valid: vanishing correlation does not induce statistical independence except for Gaussian random variables.

Expected value, variance and covariance are the leading (mixed) moments of a probability distribution. The expected value is the first absolute moment which are generally defined by

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx. \tag{1.30}$$

Central moments adjust for the expected value, i. e., $\mu_k \equiv \mathbb{E}[(X - \mathbb{E}[X])^k]$. It can easily be seen that the probability density function is uniquely determined by all of their moments using the characteristic function (1.23)

$$\mathbb{E}[X^k] = i^{-k} \left(\frac{d^k}{dt^k} \phi_X(t) \right)_{t=0} \tag{1.31}$$

given the k-th moment exists, that is the absolute moment $\mathbb{E}[|X|^k]$ is finite.

Particular importance has the third and fourth standardized moments. The former is also called skewness

$$\gamma_1 = \mathrm{Skew}[X] \equiv \frac{\mathbb{E}[(X - \mathbb{E}[X])^3]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^{3/2}} = \frac{\mu_3}{\sigma^3} \tag{1.32}$$

which reveals asymmetries of the probability mass with respect to its expected value, whereas the latter is the kurtosis of the probability distribution

$$\beta_2 = \mathrm{Kurt}[X] \equiv \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{(\mathbb{E}[(X - \mathbb{E}[X])^2])^2} = \frac{\mu_4}{\sigma^4} \tag{1.33}$$

In order to account for the notion that normal distributions are neither leptokurtic, i. e., heavy-tailed, nor platykurtic, i. e., thinner tailed, one introduces the excess kurtosis $\gamma_2 \equiv \beta_2 - 3$ which is zero for normal distributed random variables.

Certain important probability distributions and their properties can be found in appendix A.

1.2 STOCHASTIC PROCESSES

A stochastic process over a probability space $(\Omega, \mathcal{F}, P)$ is a collection of random variables

$$\{X_t | t \in T\}. \tag{1.34}$$

Every $\omega \in \Omega$ has a trajectory $t \mapsto X_t(\omega)$. The parameter set T may be discrete, e. g., $T \subset \mathbb{Z}$ or continuous $T \subset \mathbb{R}$. If t represents the time we refer to the former one as a time series.

Furthermore, given the probability space $(\Omega, \mathcal{F}, P)$ and the finite and ordered subset $\tau \subset T$ the distribution $F_X(\tau)$ is called the finite dimensional marginal distribution. The existence of a stochastic process given a family of finite marginal distributions was proven by Kolmogorow [4].

The mean and covariance function are $\mu_X(t) \equiv \mathbb{E}[X_t]$ and

$$\Gamma_X(s,t) \equiv \mathbb{E}[(X_s - \mu_X(s))(X_t - \mu_X(t))]. \tag{1.35}$$

A stochastic process is said to be strictly stationary if a constant shift in the indexed subset does not alter the result of the finite dimensional marginal distribution. Thus, given the two subsets $\{t_1, \ldots, t_n\}$ and $\{t_1 + t_0, \ldots, t_n + t_0\}$ it holds

$$F_X(t_1, \ldots, t_n) = F_X(t_1 + t_0, \ldots, t_n + t_0). \tag{1.36}$$

A stochastic process is weakly stationary if mean and covariance function are time independent. In this case the mean is constant $\mu_X(t) = \mu_X(t + t_0)$ and the covariance function depends only on time diferences $\Gamma_X(s,t) = \Gamma_X(s + t_0, t + t_0)$.

Examples of some important stochastic processes are in order. For example, a stochastic process $\{B_t | t \in \mathbb{R}_+\}$ is a Brownian motion if $B_0 = 0$ almost surely, increments $B_t - B_s$, $0 \leqslant s \leqslant t$ are IID, stationary and normally disitributed $B_t - B_s \sim \mathcal{N}(0, t-s)$. Furthermore, the process has to exhibit continuous trajectories.

Another important example is the homogenous Poisson process $\{N_t | t \in \mathbb{R}_+\}$ with rate λ defined by $N_0 = 0$ almost surely, IID and stationary increments $k \equiv N_t - N_s$, $0 \leqslant s \leqslant t$ that are Poisson distributed, i. e.,

$$P(k) = \frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}, \ \lambda > 0, \tag{1.37}$$

The Poisson process is memoryless which is easily seen by noting

$$P(T > s + t|T > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t). \tag{1.38}$$

Given a Poisson process $\{N_t|t \in \mathbb{R}_+\}$ and a set of IID random variables $\{Y_n|n \geqslant 1\}$ the process $Z_t = \sum_{i=1}^{N_t} Y_i$ with $Z_0 = 0$ is called a compound Poisson process.

A continuous time process $\{M_t|t \in \mathbb{R}_+\}$ is a martingale with respect to a family of σ-fields $\{\mathcal{F}_t|t \in \mathbb{R}_+\}$ if for every $t \in \mathbb{R}_+$ it holds that M_t is measurable with respect to $\mathcal{F}_t$, $\mathbb{E}[|X_t|] < \infty$ and $\mathbb{E}[M_t - M_s|\mathcal{F}_s] = 0$, $0 \leqslant s \leqslant t$. The discrete case follows analogously.

Brownian motion, compound Poisson process and martingales are special instances of a Lévy process which is a continuous-time process with $T = \mathbb{R}_+$ for which $X_0 = 0$ with independent and stationary increments $X_t - X_s$, $0 \leqslant s \leqslant t$ and where the trajectory is continuous.

All Lévy processes are infinite divisible and, vice versa, for every infinite divisible process there exists a Lévy process. The Lévy-Khintchine theorem establishes a relation for the characteristic function $\phi_{X_t}(u) = e^{\Psi(u)}$ of every infinite divisible and, hence, for every Lévy process [4]

$$\Psi(u) = imu - \frac{1}{2}\sigma^2u^2 + \int_{\mathbb{R}^*}(e^{iux} - 1 - iux\mathbb{1}_{|x|\leqslant 1})\nu(dx), \tag{1.39}$$

with $m \in \mathbb{R}$, $\sigma^2 \geqslant 0$, $\mathbb{R}^* \equiv \mathbb{R}\backslash\{0\}$ and the indicator function $\mathbb{1}$. The corresponding Lévy measure ν must fulfill

$$\int_{\mathbb{R}^*}(1 \wedge x^2)\nu(dx) < \infty. \tag{1.40}$$

Moreover, a theorem due to Lévy and Itô states [4] that every Lévy process can be decomposed

$$X_t = \alpha t + \sigma B_t + J_T + M_t \tag{1.41}$$

where B_t is a Brownian motion, J_t is an independent compound Poisson process and M_t is a square-integrable martingale such that the Lévy process can uniquely be specified by a triplet (α, σ^2, ν), which can easily be seen by reorganizing $\Psi(u)$ such that each summand belongs naturally to the exponent of the characteristic function of the corresponding process.

1.3 STABILITY

A set of IID random variables $X_1, \ldots X_n$ is said to be stable if it holds for any positive integer n

$$X \stackrel{d}{=} a_n(X_1 \circ \cdots \circ X_n) + b_n, \tag{1.42}$$

with $a_n > 0$ and $\stackrel{d}{=}$ indicating equality in distribution. The componentwise composition $\circ$ can be either additive, multiplicative or may involve minimization or maximization [51]. Additive schemes lead to stable Paretian laws for which $a_n = n^{-1/\alpha}$, with $\alpha \in (0, 2]$.

Furthermore, multiplicative schemes induce multiplication-stable distributions whereas minimization and maximization yield extreme value distributions.

A special case of the additive schemes is given if the number of terms is itself a random variable $n = \nu_p$ geometrically distributed with mean $1/p$, i. e.,

$$P(\nu_p = k) = (1-p)^{k-1}p, \; k = 1, 2, \ldots. \tag{1.43}$$

For example, generalized Laplace distributions are known to follow geometrically stable laws [55].

Moreover, a sequence of IID random variables $X_1, \ldots, X_n$ is said to be in the domain of attraction of a random Variable Z if

$$a_n(X_1 + \cdots + X_n) + b_n \xrightarrow{d} Z, \tag{1.44}$$

as $n \to \infty$. In particular, for $\alpha = 2$ and finite second moments the classical central limit theorem is recovered which is proven considering the sum of n standardized random variables $Y = \sum_{i=1}^{n} Z_i$. The respective characteristic function is given by

$$\phi_Y(t) = (\phi_Z(t))^n \approx \left(1 - \frac{t^2 n}{2n}\right)^n. \tag{1.45}$$

In the limit $n \to \infty$ this becomes $\exp(-nt^2/2)$ which is the characteristic function of $\mathcal{N}(0, n)$. Due to Lévy's convergence [36] theorem pointwise convergence of the characteristic function implies convergence in distribution. There exists a variant of the theorem due to Lyapunov [8] that relaxes the assumption of identical distribution.

Finally, a random variable Z is said to be infinite divisible with respect to random variables $X_1^{(n)}, \ldots, X_n^{(n)}$ if

$$Z \stackrel{d}{=} X_1^{(n)} + \cdots + X_n^{(n)}. \tag{1.46}$$

Other, not necessarily equivalent, definitions of infinite divisibility exist. A concise review is presented by Mittnik and Rachev [77].

1.4 INFORMATION THEORY

For two discrete random variables X and Y and corresponding distinct events $x_i \in \mathcal{F}_X$ and $y_i \in \mathcal{F}_Y$ the joint probability is denoted by $P(x_i, y_i)$. The

probability to observe x_i is given by $P(x_i) = \sum_{y_i \in \mathcal{F}_X} P(x_i, y_i)$. The amount of information is defined by $\log(1/P(x_i)) = -\log(P(x_i))$ [60]. Units of information are fixed with respect to the basis of the logarithm. For example one unit of information related to base 2 is called bit whereas nat corresponds to the natural logarithm. Moreover, the expected information of a random variable X is called self-entropy and is given by

$$\mathcal{H}(X) \equiv \mathbb{E}[-\log P(x)] = -\sum_{x_i \in \mathcal{F}_X} P(x_i) \log P(x_i). \tag{1.47}$$

Self-entropy is a measure of uncertainty that X will take some specific value. Naturally, it increases with the sample size and is large if distinct observations are equally likely.

The mutual information measures the diference between the information of the joint probability $P(x_i, y_i)$ and the information if the random variables were independent. Thus, it is defined as

$$M(x_i, y_i) \equiv \log \frac{P(x_i, y_i)}{P(x_i)P(y_i)}. \tag{1.48}$$

The expected mutual information, i. e., the reduction of uncertainty about X or Y when the other is known reads

$$\mathcal{H}(X, Y) \equiv \sum_{x_i \in \mathcal{F}_X} \sum_{y_i \in \mathcal{F}_Y} P(x_i, y_i) \log \frac{P(x_i, y_i)}{P(x_i)P(y_i)}) = \mathcal{H}(X) - \mathcal{H}(X|Y) \tag{1.49}$$

which is naturally related to the Kullback-Leibler divergence D_{KL}. For example, the information gain with respect to Y due to the observation of some event x_i is

$$D_{KL}(P(Y|x_i)\|P(Y)) \equiv \sum_{y_i \in \mathcal{F}_Y} P(y_i|x_i) \log \frac{P(y_i|x_i)}{P(y_i)}. \tag{1.50}$$

Thus, the relation between expected mutual information (1.49) and Kullback-Leibler divergence (1.50) is

$$\mathcal{H}(X, Y) = \sum_{x_i \in \mathcal{F}_X} P(x_i) D_{KL}(P(Y|x_i)\|P(Y)). \tag{1.51}$$

1.5 ESTIMATION

Estimation theory deals with the problem to abstract unknown parameters as precise as possible from a function of the data. In principle different paradigms have to be considered. While parametric estimation relies on

the fact that observations are drawn from a probability distribution non-parametric estimators do not need this assumption. Both paradigms can be cast in a Bayesian or Frequentist framework.

In order to estimate parameters $\Theta \in \mathbb{R}^n$ one defines a loss-function generally given by

$$L_\gamma[\Theta] \equiv \mathbb{E}[|\Theta - \hat{\Theta}|^\gamma], \tag{1.52}$$

with the estimate $\hat{\Theta} \in \mathbb{R}^n$. The quadratic loss-function L_2 is a conveniant choice which can be decomposed into the squared bias and the variance

$$L_2[\Theta] = \underbrace{(\mathbb{E}[\hat{\Theta}] - \Theta)^2}_{\text{Bias}^2[\hat{\Theta}]} + \text{Var}[\hat{\Theta}]. \tag{1.53}$$

equation (1.53) has important implications. Firstly, for unbiased estimators, that is $E[\hat{\Theta}] = \Theta$, L_2 equals the variance of the estimator $\hat{\Theta}$. Secondly, one can trade a little bias for reduced variance. Within the context of model estimation (section 1.6) it is exactly this tradeoff that has to be optimized with respect to the minimization of the out-of-sample error.

However, given an unbiased estimator $\hat{\Theta}$ and the regularity condition

$$\mathbb{E}[\partial_{\Theta_i} \log(f(x|\Theta)] = 0 \tag{1.54}$$

it can be shown [79] that a minimum lower bound for the variance exists[5]. To this end the Fisher information matrix $\mathcal{I}$ with elements

$$\mathcal{I}_{ij}(\Theta) = -\mathbb{E}\left[\partial_{\Theta_i}\partial_{\Theta_j} \log f(x|\Theta)\right] \tag{1.55}$$

is introduced. If it is possible to rewrite the regularity condition (1.54) in the form

$$\partial_\Theta \log f(x|\Theta) = \mathcal{I}(\Theta)(h(x) - \Theta), \tag{1.56}$$

then $\hat{\Theta} = h(x)$ is an efficient estimator and the variance of its components is given by the Cramér Rao bound (CRB)

$$\text{Var}[\hat{\Theta}_i] = (\mathcal{I}(\Theta))^{-1}_{ii}. \tag{1.57}$$

1.5.1 *Maximum Likelihood*

Due to its computational simplicity and its asymptotic properties the maximum likelihood estimator (MLE) is often used in practice. Given the joint density distribution $f(x|\Theta)$ with the random vector of observations x and Bayes' formula (1.9) one considers the probability of an infinite small interval around Θ given observations x $f(\Theta|x)$ and varies Θ a way that make the occurrence of x most probable. Thus, MLE are given by

$$\hat{\Theta}_{ML} = \text{argmax}_\Theta f(\Theta|x). \tag{1.58}$$

5 This is the Cramér Rao bound (CRB).

The estimator $\hat{\Theta}_{ML}$ is consistent $\lim_{n\to\infty} \hat{\Theta}_{ML} \overset{d}{=} \Theta$. Furthermore, it can be shown that equation (1.58) yields an efficient estimator such that its variance is given by the CRB (1.57) provided an efficient estimator exists. Otherwise it is at least asymptotically efficient as

$$\hat{\Theta}_{ML} \xrightarrow{d} \mathcal{N}(\Theta, \mathcal{I}^{-1}(\Theta)), \tag{1.59}$$

as $n \to \infty$. Additionally, MLE are invariant under transformations, i. e., if $\Phi = \Psi(\Theta)$ it holds $\hat{\Phi} = \Psi(\hat{\Theta})$.

If observations x_i are iid the likelihood factorizes $f(x|\Theta) = \prod_i f(x_i|\Theta)$. Then it is conveniant to minimize the negative log-likelihood

$$\mathcal{L} \equiv -\sum_i \log f(x_i|\Theta) \tag{1.60}$$

which is equivalent to (1.58) due to the monotonicity of the logarithm.

1.6 SUPERVISED STATISTICAL LEARNING THEORY

In general the task in supervised learning is to approximate the functional relationship between independent and dependent variables. Within the context of machine learning independent variables are called inputs or features and dependent variables are refered to as outputs. To this end a set of training examples $\mathcal{T} \equiv \{(x_i, y_i)\}$ is presenetd to an appropriate learning algorithm.

Regression, i. e., the prediction of a quantitative output, seeks an approximation $\hat{f}(X)$ to the underlying relationship $Y = f(X)$ given a random variable $X \in \mathbb{R}^m$ of features and a quantitative output $Y \in \mathbb{R}$.

For a family of possible solutions the relationship is determined minimizing the expected prediction error

$$E[L_\gamma] = \mathbb{E}[|Y - f(X)|^\gamma] = \int (y - f(x))^\gamma f_{X,Y}(x, y) dx dy. \tag{1.61}$$

Commonly used L_2 errors and variational calculus, i. e.,

$$\frac{\delta \mathbb{E}[L_2]}{\delta f(x)} = -2 \int (y - f(x)) f_{X,Y}(x, y) dy \tag{1.62}$$

yield the conditional expectation

$$f(x) = \mathbb{E}[Y|X = x]. \tag{1.63}$$

Thus, the best estimator for Y is the conditional mean at any point $X = x$. For L_1 errors this has to be replaced by the conditional median at any point.

For k categorical outputs the loss function (1.61) is replaced by a loss matrix $\mathcal{L} \in \mathbb{R}^{k \times k}$. Elements $\mathcal{L}_{ij}$ indicate the cost induced by a misclassification of an example classified as class i when it belongs to class j. The expected test error is calculated with respect to the joint distribution of the categorical variable and features X.

1.6.1 *Model Selection*

Models that are obtained from the training of machine learning algorithms usually are used to predict outcomes of unknown data. Thus, one is interested in the generalization ability on independent test data, instead of the minimization of the error on the training sample. This is inherently linked to the choice of an appropriate learning algorithm.

The generalization error is given by

$$\mathrm{Err}_\gamma \equiv \mathbb{E}[L_\gamma(Y, \hat{f}(X))|\mathcal{T}] \tag{1.64}$$

that is the error of a learning algorithm that produced $\hat{f}$ trained on a set of training examples $\mathcal{T}$ applied on an independent test sample using the loss-function L_γ.

The training error is the mean loss over N training examples

$$\epsilon_\gamma \equiv \frac{1}{N} \sum_{i=1}^{N} L_\gamma(y_i, \hat{f}(x_i)). \tag{1.65}$$

In general, the training error ϵ_γ decreases as the model complexity increases. It vanishes as the number of parameters equals the number of training examples which accounts to a re-parametrization of the training sample. However, very complex models usually will not lead to the lowest generalization error (1.64).

In particular, it can be shown [41] that more complex models yield lower bias but higher variance with respect to the expected prediction error $\mathbb{E}[\mathrm{Err}_\gamma] = \mathbb{E}[L_\gamma(Y, \hat{f}(X))]$. The goal is to balance the model complexity with respect to an optimal trade between bias and variance that minimizes the generalization error.

In this thesis cross validation is used to estimate the expected out-of-sample error, i. e., the data is divided into k equal-sized sub-samples. Then, the prediction error is calculated for the i-th sub-sample using a model that is fitted to the remaining data. The estimated prediction error due to cross-validation is the average prediction error.

A common choice is 5-fold cross-validation. For example, if k is chosen too large the training samples are quite similar such that one gets an unbiased estimator for the test error with high variance. Thus, one

estimates $\mathbb{E}[\mathrm{Err}_\gamma]$ but not (1.64). However, for small k considerably high bias is introduced, especially for small training samples.

1.6.2 *Vapnik-Chervonenkis Dimension*

The Vapnik-Chervonenkis (VC) dimension formalizes the concept of generalizability by introducing a measure of capacity, that is how wiggly a set of functions can be. This stems from the fact one can demonstrate that the number of free parameters is not a good measure of capacity. For example, $\cos(fx)$ has only one free parameter. However, it can separate arbitrarily points as one increases its frequency f.

The VC dimension is given by the largest number of points $x \in \mathbb{R}^n$ that can perfectly be separated by a member of the function-class with respect to an *arbitrary* assignment of binary outputs. For example, the VC dimension of a hyperplane is 3. There are 2^3 possibilities to assign binary outputs to 3 points that all can be separated when the class boundary is linear. However, this does no longer work for 4 points.

Its importance for practical applications results from the fact that a given VC dimension h allows to state upper bounds of the expected out-of-sample error for regression and classification, i. e., for binary classification Hastie et al. [41] state

$$\mathrm{Err}_{\mathcal{T},2} = \epsilon_2 + \frac{\kappa}{2}\left(1 + \sqrt{1 + \frac{4\epsilon_2}{\kappa}}\right),\ \kappa = a_1 \frac{h(\log(\frac{a_2 N}{h}) + 1) - \log(\frac{\eta}{4})}{N}, \quad (1.66)$$

with probability $1 - \eta$ and parameters $a_1 = 4$ and $a_2 = 1$ corresponding to the worst-case. Thus, the correction to the training error[6] increases with the VC dimension h and decreases with the the size of the training sample N.

1.6.3 *Support Vector Classification*

Support Vector Classifiers (SVC) induce a class boundary on an input-space that maximizes the margin between members of different classes. For example, given the training set $\mathcal{T} = \{x_i, y_i\}$ with N members $x_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$, a hyperplane

$$x^T w + b = 0 \quad (1.67)$$

6 Also called optimism

is used to separate classes. If the problem is linearly separable, i. e., members of different classes do not overlap in the input-space, the maximization margin is equivalent to the convex optimization problem

$$\begin{aligned} &\min_{w,b} \; w^T w \\ &\text{subject to } y_i(x_i^T w + b) \geqslant 1, i = 1, \ldots, N. \end{aligned} \tag{1.68}$$

The margin that separates the two classes has the width of $2w^T w$. Most often it is convenient to allow some training examples to lie on the wrong side of the class boundary. The introduction of slack variables $\xi \in \mathbb{R}^n$ allows a certain degree of misclassification, i. e., ξ_i measures the relative distance by which x_i is misclassified in terms of $w^T w$. The equivalent convex optimization problem reads

$$\begin{aligned} &\min_{w,b,\xi} \; \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \\ &\text{subject to } y_i(x_i^T w + b) \geqslant 1 - \xi_i, \end{aligned} \tag{1.69}$$

with $\xi_i \geqslant 0$ and where the cost parameter C has to be calibrated such that the cross-validation error is minimal. Observations x_i for which the constraint in (1.69) becomes $y_i(x_i^T w + b) = 1 - \xi_i$ are called support vectors. This means that observations that are located neither on the frontier nor within the margin do not shape the decision boundary. Larger values of C lead to smaller margins. Hence, they stress the importance of values near the decision boundary.

The primal problem (1.69) can be transformed to the so-called dual problem

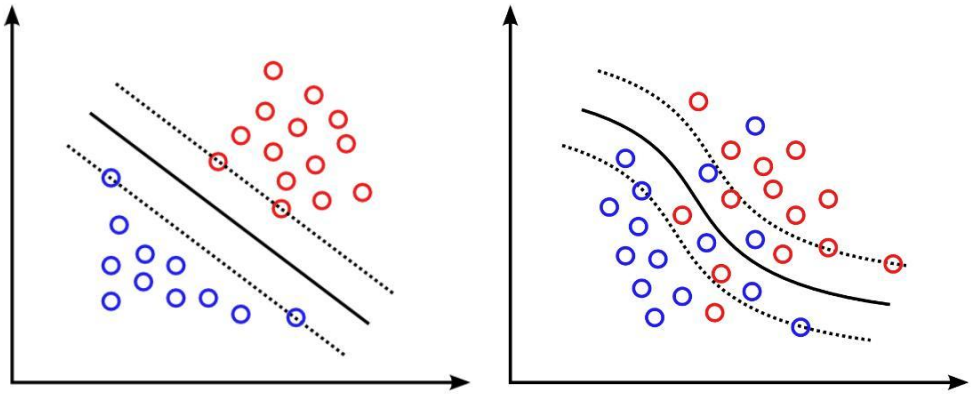
$$\begin{aligned} &\min_{\alpha} \; \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ &\text{subject to } y^T \alpha = 0, \end{aligned} \tag{1.70}$$

with $0 \leqslant \alpha_i \leqslant C$, $e = (1, \ldots, 1)^T$ and a positive semi-definite matrix Q with elements $Q_{ij} = y_i y_j x_i^T x_j$. Note that the slack variables vanished within the dual form. Solutions $\hat{\alpha}$ to equations (1.70) are transformed in terms of the original problem by

$$\hat{w} = \sum_{i=1}^{N} y_i \hat{\alpha}_i x_i. \tag{1.71}$$

A class label is assigned due to the classification rule

$$\hat{y} = \operatorname{sgn}\left(\hat{w}^T x + b\right) = \operatorname{sgn}\left(\sum_{i=1}^{N} y_i \hat{\alpha}_i x_i^T x + b\right). \tag{1.72}$$



(a) Linear and separable SVC (b) Nonlinear and non-separable SVC

Figure 1: Support vectors in the separable case are given by all points that lie exactly on the decision boundary. Given the non-separable case and the introduction of slack variables support vectors can also lie within the margin and on the wrong side of the decision boundary.

So far only linear decision boundaries are considered. Mapping the inputs x to a space of higher dimension $x \mapsto h(x) \in \mathbb{R}^m$ and applying the linear SVC induces nonlinear decision boundaries in the original feature space. To this end inner products $x_i^T x_j$ are replaced by kernel functions $K(x_i, x_j) = h(x_i)^T h(x_j)$. Common choices are

- linear $K(x_i, x_j) = x_i^T x_j$
- polynomials $K(x_i, x_j) = (\gamma x_i^T x + b)^d$, $\gamma > 0$
- radial $K(x_i, x_j) = \exp(-\gamma |x_i - x_j|_2)$
- sigmoid $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + b)$

where free parameters have to be calibrated within the model selection phase. In particular, small margins due to large C may lead to serious overfitting problems due increased flexibility of the class boundary.

Figure 1 is a sketch illustrating both (non)-separable and (non)-linear decision boundaries.

Whenever, we train multiclass SVC with a one-against all approach [52]. Thus, given k classes $k(k+1)/2$ classifiers are trained. For the final classification a majority vote is used, i. e., every classifier votes for one class and the one that aggregates most votes wins.

Furthermore, we can state upper bounds for the out-of-sample error due to the existence of upper bounds of the VC dimension i. e., given that the training set $\mathcal{T}$ fits into a sphere with radius r and $|w|_2 = \omega$ it is given by

$h = (\omega r)^2$. Thus, upper bounds of the test error are given with probability $1-\eta$ by

$$\mathrm{Err}_{\mathcal{T},2} = 4\frac{h}{N}(\log(2N/h+1) - \log(\eta/4)). \tag{1.73}$$

2

ECONOMETRICS OF FINANCIAL MARKETS

The econometrics of financial markets involves modelling of time series like prices, returns and interest rates. It is a quantitative field that relates topics of economy to mathematics. In particular, financial modelling is set up in a probabilistic world where the system in question takes a certain state in the equivalent probability space.

There are various reasons for an increased demand in financial modelling on an academic level as well as within finance industry. For example, sophistication of modelling resulted in milestones such as Black and Scholes famous work on option pricing [10], Merton's equilibrium model [67], heavy-tailed return distributions [63], the efficient market hypothesis [33], Markowitz's [65] efficient frontier and the capital asset pricing model due to Sharpe [84] and others to name only a few.

This was accompanied by advancements of the information infrastructure which had without any doubt a sustainable impact on the development of financial econometrics. For example, the possibility to collect and store vast amounts of data, increasingly cheap computer systems powerful enough to perform complex calculations even on the consumer level and finally low-latency, broadband signal-processing.

The vast amount of work conducted over the last decades both on a theoretical and empirical level renders it virtually impossible to give a concise overview of the field. Thus, we rather focus on the most important aspects as used in this thesis. For a comprehensive introduction to the field we refer to Rachev et al. [78] and Campbell et al. [15].

The remainder of this chapter introduces the most basic definitions of financial time series related to prices and returns after which we discuss empirical characteristics of unconditional return distributions, in particular heavy tails. This is followed by sections on modelling of conditional volatility, in particular, GARCH models, their extensions and parameter estimation.

2.1 PRICES AND RETURNS

Given the time series of prices of any asset k that pays no dividends $\{s_t^{(k)}\}$, $t \in \mathbb{Z}$ the rate of return, for simplicity also referred to as return, is defined as

$$r_t^{(k)} \equiv \frac{s_t^{(k)} - s_{t-1}^{(k)}}{s_{t-1}^{(k)}}. \tag{2.1}$$

There are at least two reasons to prefer the evolution of returns rather than that of prices. The return is scale-free as long as one does not consider actual investing where the size of the position impacts the price. Even more important, time series of returns often are stationary at least on local time horizons.

Furthermore, we consider the rate of return for a portfolio $r^{(p)}$ of n assets which is given by

$$r_t^{(p)} = \sum_{k=1}^{n} w_i r_t^{(k)} \tag{2.2}$$

where the weights are subject to an appropriate normalization. Another frequently used quantity is the log return also called continuously compound rate of return defined by

$$y_t^{(k)} \equiv \log\left(\frac{s_t^{(k)}}{s_{t-1}^{(k)}}\right). \tag{2.3}$$

It follows from the basic properties of logarithms that log returns on longer time horizons are just the simple sum of log returns of their constituent, shorter time horizons. However, due to the same property equation (2.2) is no longer valid for continuously compounded returns.

Expressions (2.1) and (2.3) can be modified to incorporate the payment of dividends d_t in the interval between $t-1$ and t, i. e.,

$$s_t' \to s_t - d_t. \tag{2.4}$$

Additionally, it is often desired to compare the performance of an asset or portfolio to the one containing low-risk securities or certain indices. Thus, the excess return $z_t^{(k)}$ is defined by

$$z_t^{(k)} \equiv x_t^{(k)} - x_t^{(0)} \tag{2.5}$$

where $x_t^{(k)}$ can be any of the above returns and $x_t^{(0)}$ is the corresponding return of the portfolio of reference securities.

2.1.1 *Distribution of Returns*

Given the time series $\{x_t^{(k)}\}_{t=1\ldots T}$ of an asset k and any of the above returns (2.1) and (2.3) $f(x^{(k)}|\Theta)$ denotes the unconditional, joint distribution of returns and Θ uniquely determines the statistical distribution.

Historically important is Bachelier's [6] introduction of the normal distribution into finance. Assuming that price changes on the transaction level are IID with finite variance the normal distribution follows from the central limit theorem. Ever since, the normal distribution has played a crucial role in many areas of modern finance including the classical formulation of the mean-variance portfolio theory by Markowitz [65] or the celebrated Black-Scholes formula [10].

However, as noticed by Mandelbrot [63] and subsequently promoted by Fama [32] empirical return distributions exhibit rather strong excess kurtoses. Not surprisingly the observation of volatility clustering, that is large returns are most likely followed by large returns whereas small returns are followed by small returns, violates the assumption of IID. Moreover, the occurrence of extreme events as corresponding to rather abnormal returns is much more likely as implicated by the normal distribution. For example, in the context of GARCH modelling Kim et al. [48] conclude that applying normal innovations a crash like the Black Monday 1987 would be expected every $2.554 \cdot 10^{39}$ years whereas such an event would be expected every 37.26 years using the corresponding α-stable GARCH model. It is worth mentioning that leptokurtic return behaviour is not a characteristic of modern financial markets and, thus, could be linked to the arrival of information technology but can also be found in financial markets of the eighteenth century [40]. In his seminal paper Mandelbrot introduced α-stable Paretian distributions into finance which are capable to model both excess kurtosis and skewness. For an review see, for example, Rachev et al. [77]. While α-stable distributions have nice theoretical properties among them domains of attraction and stability, financial modelling often requires finite second and higher moments which they do not exhibit due to polynomial decaying Lévy densities [49].

Only recently this was overcome due to the introduction of tempered stable distributions [81], also called Carr-Geman-Madan-Yor (CGMY) model [18] or truncated Lévy flights [54] and successfully applied to finance, in particular, GARCH modelling and option pricing by Kim et al. [49].

Figure 2 illustrates the z-transformed, unconditional return distribution of the S&P 500 on 10 minute time intervals from 3 March 2010 to 20 April 2011. While the normal distribution underestimates the probability of extreme events in the tails both normal and t-distribution underestimate

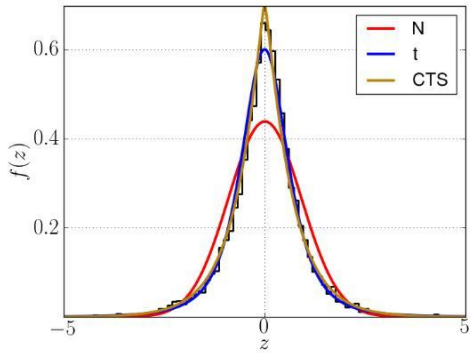


Figure 2: Empiral and theoretical unconditional return distributions for z-transformed log returns of the S&P 500 on 10 minute time intervals. The normal, the Student t and the classical tempered stable distribution are fitted to the data. The normal distribution underestimates risk corresponding to the central and tail region of the distribution. The classical tempered stable distribution represent the entire empirical distribution remarkably well.

the central region of the distribution, whereas the classical tempered stable (CTS) achieves both.

Appendix A reviews some heavy- and semi-heavy-tailed distributions which are used within this thesis.

2.1.2 *Heavy Tails*

In order to define the notion of fat tails more precisely we consider the probability that some random variable X takes values larger than x, i. e., we consider the upper tail. A distribution is said to have Pareto tails if it holds

$$P(X > x) = \frac{L(x)}{x^{\alpha}}, \tag{2.6}$$

with the tail index $\alpha > 0$ and a slowly varying function $L(\cdot) > 0$ such that

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1, \; c \in \mathbb{R}_+. \tag{2.7}$$

Thus, Pareto type distributions decay like a power law, i. e., $\lim_{x\to\infty} L(x)/x^{\alpha} = 0$. In particular, this means that moments of the respective distribution exists only up to the tail index α with possible modifications of the statement due to an appropriate choice of $L(\cdot)$.

Under the assumption that the tail behaviour is governed by Pareto-like distributions the tail index can be estimated via the Hill estimator [78]

$$\hat{\alpha}_k = \left(\frac{1}{k-1} \sum_{i=1}^{k} \log X_{(n+1-i)} - \log X_{(n-k)} \right)^{-1} \tag{2.8}$$

with the order statistics $X_{(i)}, \dots X_{(n)}$. In particular, α-stable Pareto distributions and the Student t-distribution show asymptotic Pareto type tail behaviour.

While it is generally accepted that the normal distribution cannot account for empirical return distributions as it has too light tails, i. e., it drops faster than any exponential, there is no consensus about the correct tail behaviour. In particular, the term semi-heavy tails was used in the context of generalised hyperbolic distributions, e. g., Variance Gamma (VG) distributions, in order to account for a tail behaviour that drops slower than a normal distributions but where all moments may exist.

2.2 MODELING CONDITIONAL VOLATILITY

Ross [82] showed that volatility measures the information flow within financial markets. Hence, there are many areas of modern finance including portfolio theory and option pricing that depend on the concise estimation of volatility.

Generally, it is accepted that volatility consists of a predictable and an unpredictable part, with research typically focusing on one of these aspects. Given the return y_t from time $t-1$ to t of a certain financial instrument the modelling of conditional volatility $\sigma_t^2 = \mathrm{Var}[y_t|\mathcal{I}_{t-1}]$ parametrizes the relation between the volatility and the filtration available $\mathcal{I}_{t-1}$. Time series of the unexpected return $\epsilon_t \equiv y_t - \mathbb{E}[y_t|\mathcal{I}_{t-1}]$ are also called news process or just noise.

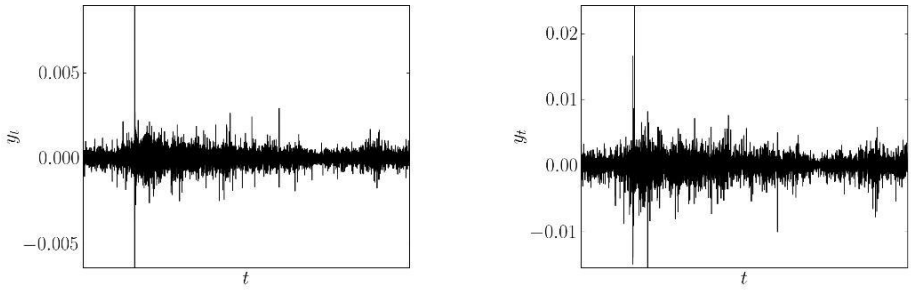
2.2.1 *Empirical Evidence*

Financial time series exhibit a number of stylized facts that a certain model parametrization has to acknowledge. Figure 3 depicts returns of the Standard & Poors 500 (S&P 500) from 3 March 2010 to 20 April 2011 on diferent time horizons.

As discussed in section 2.1.1 unconditional return distributions are more heavy-tailed then expected for a normal distribution.

Additionally, the realized stock return volatility exhibits a hyperbolic decay within its autocorrelation function [3], that is

$$\lim_{k\to\infty} \rho(k) \sim L(k) k^{2d-1}, \; d \in (0, 1/2), \tag{2.9}$$



(a) $\Delta t = 15s$ (b) $\Delta t = 600s$

Figure 3: Realized intraday returns of the S&P500 from 3 March 2010 to 20 April 2011 for time horizons of 15s and 600s. Volatility clustering is observed on both time horizons.

with k the lag number and and slowly varying function $L(\cdot)$. The corresponding long memory effect of volatility was refered to as volatility clustering[63]. Thus, the assumption of homoskedasticity, i. e., constant variance, is inadequate. To this end, Engle [31] introduced autoregressive conditional heteroskedastic models (ARCH) originally to model inflation rates of the UK. However, the clustering of volatility appears to be generic and, subsequently, was confirmed over many asset classes and time horizons.

Indeed, heavy tails of the unconditional return distribution can in part be explained due to volatility clustering. However, with increasing frequency of the time series the innovation process becomes more leptokurtic such that volatility clustering cannot be the only source of heavy tails.

Furthermore, among others Black [9] provided evidence that financial time series are asymmetric with respect to the response of volatility to good and bad news, i. e., the increase in volatility is larger if the unexpected return was negative. Black also discussed that the effect is too large to explained by pure leverage effects, i. e., a firm with outstanding debt becomes more leveraged if its value falls while stock volatility rises and the total return stays constant. Nevertheless the term *leverage* is commonly used in literature when referring to the asymmetry of volatility with respect to past returns. To account for the notion, various ideas have been proposed to model asymmetric dependencies of the news process ϵ_t (section 2.2.4).

Moreover, the information flow is not uniform when using intraday time series. For example, increased volatility after the open reflects the information aggregated over the corresponding non-trading period. The variance following weekends is not as high as one would expect if news

arrived at a constant rate in comparison to two succeeding weekdays. Ballie et al. [7] find that volatility is on average much higher during continuous trading than during non-trading periods. However, the assumption of constant arrival rates can be refuted as demonstrated in section 4.2. Moreover, concerning high-frequent data one has to consider enhanced volatility due to macroeconomic events and cross-market integration. This will be discussed in detail in chapter 3.

2.2.2 *ARCH Models*

We consider a stochastic process $\{\epsilon_t\}$ in discrete time and the autoregressive conditional heteroskedastic model (ARCH)

$$\epsilon_t = \sigma_t z_t,\ \mathbb{E}[z_t] = 0,\ \mathrm{Var}[z_t] = 1,\ \sigma_t > 0, \tag{2.10}$$

where the conditional volatility σ_t needs a reasonable parametrization due to the available filtration and z_t is the independent and identical distributed innovation term. If not stated otherwise and without loss of generality we assume $y_t = a_0 + \epsilon_t$ given the rate of return y_t i. e., the expected value $\mathbb{E}[y_t|\mathcal{I}_{t-1}] = a_0$ is constant.

Often, it is conveniant to assume a normal distribution $\mathcal{N}(0,1)$. In particular, using Jensen's inequality it can be shown that normality of the innovation term implies a leptokurtic news process ϵ_t, i. e.,

$$\mathbb{E}[\epsilon_t^2] > \mathbb{E}^2[\sigma_t^2]\mathbb{E}[z_t^4]. \tag{2.11}$$

However, in section 3.6 we show that for high-frequency data empirical innovation distributions still exhibit leptokurtic behavior such that the assumption of normality has to be rejected.

Furthermore, Engle [31] considers the following parametrization called linear ARCH model of order p (ARCH(p))

$$\epsilon_t = \sigma_t z_t,\ \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \tag{2.12}$$

with constant parameters $\alpha_i > 0,\ \omega > 0$. The simple equation (2.12) implies that conditional variance depends on past errors, that is past information. From this it can be interfered that $\alpha_{i+1} < \alpha_i$.

Some properties of ARCH(1) are in order. Due to $\mathbb{E}[z_t] = 0$ it follows that the unconditional expectation value of ϵ_t vanishes. Furthermore, using the geometric series the unconditional variance is

$$\mathrm{Var}[\epsilon_t] = \mathbb{E}[\epsilon_t^2] = \mathbb{E}[\omega + \alpha_1 \epsilon_{t-1}^2] = \frac{\omega}{1-\alpha_1}. \tag{2.13}$$

Given finite forth moments in ϵ_t, it can be shown [78] that the kurtosis is given by

$$\gamma_1[\epsilon] = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3. \tag{2.14}$$

An equivalent representation of equation (2.12) exists due to the time-dependent p-th order moving average (MA(p)) model

$$\epsilon_t = \tilde{\omega}_t + \sum_{i=1}^{p} \tilde{\alpha}_{i,t}\tilde{\epsilon}_{t-i}^2, \tag{2.15}$$

where former coefficients are now IID stochastic processes with vanishing mean and variance given by coefficients of the corresponding ARCH(p) model.

Moreover, equation (2.17) can be written in terms of the lag-operator L defined by $L^i y_t \equiv y_{t-1}$ such that it is an AR(p) process

$$\epsilon_t^2 = \omega + \alpha(L)\epsilon_{t-1}^2 + \nu_t, \tag{2.16}$$

where $\nu_t = \epsilon_t^2 - \sigma_t^2$ and $\alpha(L)$ is a p-th order polynomial in L which is covariance stationary if all of its roots lie outside the unit circle [78].

2.2.3 *Generalized ARCH Models*

By adding a moving average term Bollerslev [12] introduced the generalized ARCH model GARCH(p, q)

$$\epsilon_t = \sigma_t z_t,\ \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \tag{2.17}$$

where z_t is a IID, standartized innovation. Compared to the linear ARCH(p) the conditional variance depends on past realizations of itself, which renders it more flexible. In particular, a high order ARCH(p) model might be replaced by a relatively low order GARCH(p, q) version. Moreover, in practice it turns out that GARCH(1, 1) is a suitable parametrization of the data. It can be shown that equation (2.17) is a well-defined process if parameters $\alpha_i,\ \beta_j$ are nonnegative and ω is positive.

Again, using the lag-operator L equation (2.17) can be made equivalent to the ARMA(max(p, q), q) model

$$\epsilon_t^2 = \omega + (\alpha(L) + \beta(L))\epsilon_{t-1}^2 - \beta(L)\nu_{t-1} + \nu_t, \tag{2.18}$$

with $\nu_t = \epsilon_t^2 - \sigma_t^2$ and the q-th order moving average polynomial $\beta(L)$. Ergodicity and strict covariance stationarity follows if all roots of $\alpha(L) +$

$\beta(L) = 1$ lie outside the unit circle. Bollerslev showed [12] that this is equivalent to

$$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i \leqslant 1. \tag{2.19}$$

For the integrated GARCH (IGARCH) model due to Engle and Bollerslev [30] inequality (2.19) becomes an equation. Thus, past noise persists for all future time-horizons.

Given a strictly covariance stationary process $\{\epsilon_t\}$ its unconditional variance reads

$$\mathbb{E}[\epsilon_t^2] = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{q} \beta_i}. \tag{2.20}$$

For the simple GARCH$(1,1)$ it can be shown [78] that

$$2\alpha_1^2 + (\alpha_1 + \beta_1)^2 \geqslant 1 \tag{2.21}$$

yields finite fourths moments of ϵ_t and $\sqrt{\sigma_t^2}$ and leptokurtic hehaviour of the corresponding unconditional distributions. Letting ω be a function of time other stylized facts such as forecastable events or non-trading periods can be incorporated.

2.2.4 *Asymmetric GARCH*

While the most prominent motivation for asymmetric models with respect to the response of volatility on past shocks is rooted within the empirical work of Black [9], Nelson [70] lists further arguments that deem to make modifications of the symmetric model necessary. For example, the nonnegative constraints on the parameters within equation (2.17) render random oscillations impossible. Moreover, concerning the question how long shocks persist within the time series the choice of the criterion for convergence in probability may have a suitable influence on the result.

Thus, given a model of the form

$$\log \sigma_t^2 = \alpha_t + \sum_{i=1}^{q} \beta_i g(z_{t-i}) \tag{2.22}$$

with a suitable function $g(\cdot)$, σ_t^2 remains positive due to the use of the logarithm, i. e., the constraint that parameters α_t, β_i have to be nonnegative can be relaxed. One arrives at Nelson's exponential GACRH (EGARCH) model by introducing

$$g(z_t) = \Theta z_t + \gamma(|z_t| + \mathbb{E}[|z_t|]). \tag{2.23}$$

Equation (2.23) is linear with slope $\Theta + \gamma$ on $\mathbb{R}_+$ and linear with slope $\Theta - \gamma$ on $\mathbb{R}_-$ which allows to account for the asymmetry effect. Typically,

the coefficient γ is negative. Thus, negative news have a stronger impact on the conditional volatility. For normally distributed innovations z_t, the expected value in (2.23) becomes $\sqrt{2/\pi}$ whereas it reads

$$\mathbb{E}[|z_t|] = \sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \tag{2.24}$$

for a Student t-distribution with $\nu > 2$ degrees of freedom.

Moreover, EGARCH addresses the above stated points of criticism. For example, vanishing constraints such as the nonnegativity of the coefficients allows oscillatory behaviour in σ_t. Additionally, it is a linear model in $\log\sigma_t^2$ such that stationarity does not depend on the criterion of convergence in probability. For example, Nelson [70] proves strict stationarity and ergodicity by noting that $\log\sigma_t^2$ can be written as an ARMA(p,q) process where the corresponding theorems on stationarity hold. Within this thesis we abandon the original formulation and use

$$\log\sigma_t^2 = \omega + \sum_{i=1}^{q}\alpha_i(|z_t| + \mathbb{E}[|z_t|]) + \gamma_i z_{t-1} + \sum_{i=1}^{p}\beta_i \log\sigma_{t-1}^2 \tag{2.25}$$

instead. Another way to introduce asymmetries is the model

$$\sigma_t^\gamma = \omega + \sum_{i=1}^{q}(\alpha_i|\epsilon_{t-i}^\gamma| + \gamma_i|\epsilon_{t-i}^\gamma|\mathbb{1}_{\epsilon_{t-i}<0}) + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^\gamma, \tag{2.26}$$

with the indicator function $\mathbb{1}$. For $\gamma = 2$ equation (2.26) is known as the Glosten Jagannathan Runke (GJR) model [39] which is well defined if coefficients are nonnegative and $\alpha_i + \gamma_i \geqslant 0$. The symmetric GARCH model (2.17) can essentially be interpreted as a GJR model with $\gamma = 0$. Moreover, stationarity follows for

$$\sum_{i=1}^{q}\left(\alpha_i + \frac{1}{2}\gamma_i\right) + \sum_{i=1}^{p}\beta_i < 1. \tag{2.27}$$

2.2.5 *Parameter Estimation in GARCH models*

In practice parameter estimation within GARCH models might be performed by applying a maximum likelihood approach. Let Θ be a vector of parameters to be estimated, $\{y_i\}_{t=1,\dots,T}$ the realized observations and $f(z_t|\Theta)$ the density distribution of $z_t = \epsilon_t/\sigma_t$. Then, one maximizes the log likelihod

$$\mathcal{L}(y_1,\dots,y_t|\Theta) = \sum_{t=1}^{T}\mathcal{L}_t(y_t|\Theta) = \sum_{t=1}^{T}\log f(z_t|\Theta) - \frac{1}{2}\log\sigma_t^2. \tag{2.28}$$

Under the assumption of normal innovations the log-likelihood becomes

$$\mathcal{L}(y_1, \ldots, y_t|\Theta) = -\frac{1}{2}\sum_{i=1}^{T}\left(\log 2\pi + \log \sigma_t^2(\Theta) + \frac{\epsilon_t^2(\Theta)}{\sigma_t^2(\Theta)}\right) \tag{2.29}$$

which is maximized with respect to Θ. However, as noted earlier most models cannot remove leptokurtic behaviour from the unconditional distribution such that the innovation process still exhibits fat tails, especially in the context of high-frequency data. This is acknowledged by applying other distributions such as Student-t, CTS or VG.

Nevertheless, one may use Gaussian maximum likelihood estimation to fit parameters of the GARCH model independent from the actual fitting of the innovation distribution which is called quasi-maximum likelihood estimation. The validity of the approach is justified by the notion that quasi maximum likelihood estimates $\hat{\Theta}_{QML}$ are Fisher consistent [12], i. e., it holds

$$\mathbb{E}\left[\partial_\Theta \mathcal{L}(y_1, \ldots, y_t|\Theta)\right] = 0. \tag{2.30}$$

From this it can be shown [89] that an estimate $\hat{\Theta}_{QML}$ is consistent and asymptotically normal

$$\lim_{T\to\infty} \sqrt{T}(\hat{\Theta}_{QML} - \Theta) \stackrel{d}{=} \mathcal{N}\left(0, \mathcal{J}^{-1}(\Theta)\mathcal{I}(\Theta)\mathcal{J}^{-1}(\Theta)\right), \tag{2.31}$$

with elements of matrices $\mathcal{I}(\Theta)$ and $\mathcal{J}(\Theta)$ given by $\mathcal{J}_{ij} = -\mathbb{E}[\partial\Theta_i\partial\Theta_j\mathcal{L}_t]$ and $\mathcal{I}_{ij} = \mathbb{E}[\partial\Theta_i\mathcal{L}_t\partial\Theta_j\mathcal{L}_t]$ that can be estimated consistently from the data.

Brooks et al. [14] give an overview of various implementations of GARCH estimation and and their accuracy with respect to a benchmark.

Part II

INTRADAY VOLATILITY

3

MODELING INTRADAY PATTERNS OF VOLATILITY

3.1 INTRODUCTION

This chapter examines the intraday pattern of dispersion both for individual stocks and the collective movement of a given universe in the US and Europe. Stocks listed on the London Stock Exchange, XETRA and Euronext are taken into account. Strong integration for the given exchanges is found apart from effects due to specific market mechanics, e. g., the midday auction at XETRA. Subsequently, a method is proposed to identify *ex-post* effects related to unpredictable events in the data and remove them in an iterative fashion from a generic rank-k approximation of the intraday return dispersion.

Locally increased volatility in the low-rank approximation of the intraday data can often be explained by the release of macroeconomic news during the continuous trading. In the context of the US Treasury market Fleming et al. [35] find a nearly instant reaction due to macroeconomic events. This is confirmed by Bollerslev et al. [13] who analyse the 5 minute return volatility of US Treasury bond futures. They observe enhanced volatility due to the release of announced macroeconomic news such as Humphrey-Hawkins testimony and the employment report. Ederington et al. [29] investigate the response of macroeconomic news releases on interest rate and FX markets. They conclude that there are several announcements that effect Treasury bond future prices, Eurodollar future prices, and the DM/USD exchange rate. Concerning market efficiency they constitute that the price adjustment due to news is largest within the minute of publication and that the direction of subsequent returns are independent from the immediate adjustment. However, returns are more volatile for the 15 minutes following the release. Almeida et al. [2] detects additional impact due to German policy decisions on DM/USD exchange rates.

Hence, this thesis the impact of macroeconomic news releases on the stock market is explored. Chen et al. [20] analyse the intraday response of market returns as measured by the Dow Jones Industrial Average (DJIA). They identify significant effects due to unexpected changes. Furthermore, their evidence supports Cook et al. [23] who argue that the market responds only after 1979 due to a modification of the Fed's policy. More general macroeconomic news releases are studied by Wasserfallen [88]

who spots only small influences due to macroeconomic news releases on the stock market in Great Britain, Germany and Switzerland. However, Nikkinen et al. [71] observe that the US employment report and Federal Open Market Committee (FOMC) have impact on the European stock market, whereas domestic news have no significance.

Results from the literature are supported. Moreover, it is found that 14 out of 15 event classes impact prices during the continuous trading within the US. Additionally, 10 out of 12 events have a significant impact on the DAX.

Subsequently, a generic low-rank profile is proposed that takes into account mean corrections of the volatility due to the release of macroeconomic news. The corresponding intraday estimates are merged with classical GARCH paradigms such as GARCH, EGARCH and GJR. The respective boosted models are called SVD-GARCH, SVD-EGARCH and SVD-GJR. The model performance is measured by backtesting Value-at-Risk utilizing Kupiec's proportion of failures test for one-step-ahead forecasts of 1 minute returns of DAX and S&P500 data. It is found that risk is properly assessed for boosted variants of classical GARCH whereas classical variants overestimate risk leading to very conservative estimates. Additionally, for SVD-GARCH and SVD-GJR the performance as measures by the mean absolute percentage error (MAPE) is slightly improved.

Thereby, the goodness-of-fit as of various distributions such as the normal, the Student t, classical tempered stable and variance gamma distribution is examined on various time horizons. In particular, for high frequency data classical tempered stable as well as variance gamma distributions describe the data much better. Moreover, the scaling behaviour of parameters of the variance gamma distribution are explored.

The remainder of this chapter is organised as follows: section 3.2 describes the data and define collective and single stock return dispersion. Section 3.3 analyses *ex-post* intraday return dispersion within the US and Europe. *Black swan* events that dominate the rank-3 approximation are identified. Section 3.4 introduces an iterative rank-k model to account for unpredictable events. Furthermore, the impact of macroeconomic news releases on intraday return dispersion are explored. Section 3.6 focuses on GARCH modeling. In particular, SVD variants of classical GARCH models are introduced in section 3.6.1. The goodness-of-fit of various distributions with respect to the innovation process is examined in section 3.6.2, the scaling-behaviour of the variance gamma distribution is assessed in section 3.6.3 and the forecasting quality as well as risk backtests are investigated in section 3.6.4. Finally, section 3.7 summarizes the findings.

3.2 THE DATA

Two sets of data are considered. The first encompasses high-frequency data for the S&P 500 and the DAX from March 3, 2010 to April 20, 2011 on time intervals from 15 seconds to 10 minutes. The second dataset contains individual stock data on 1 minute intervals from January 2, 2008 to 29 December 2010 for the US and Europe. The single stock dataset of the US consists of time series of 465 different equities with market capitalization ranging from approximately 1 to 348 billion Euros[1], whereas the European single equity dataset comprises of 170 companies with market capitalization between 2 and 162 billion Euros. The data is adjusted for dividend payments. Furthermore, all European equities are chosen such that they are traded on XETRA, London Stock Exchange (LSE) and Euronext which leads to homogeneous trading hours concerning the continuous trading among the chosen exchanges. In order to account for a different dynamic concerning the midday action of XETRA the European market is divided into additional subsets for equities traded at Euronext (63 stocks), LSE (70 stocks) and XETRA (37 stocks). While this significantly reduces the statistic compared to the US. It is remarked that the only criterion for picking individual stocks was their liquidity.

The continuously compounded return – henceforth referred to as the return – at a time t is utilized

$$y_t^{(k)} \equiv \log\left(\frac{s_t^{(k)}}{s_{t-1}^{(k)}}\right), \quad t = 1, \dots, T, \tag{3.1}$$

where $\{s_t^{(k)}\}$ denotes the time series of the price of security k, i. e., price changes and no returns on investments are considered. Concerning time series of individual stocks last prices to which a trade took place of the corresponding trading minute are used.

Participants in the financial markets want to know the risk they are exposed to and the risk they will face in the future in order to construct optimal portfolios or price derivatives. Volatility, i. e., the second moment of return, is a widely accepted measure of risk and directly related to the information flow available within the market [82]. However, other metrics exist [28]. In order to estimate risk in a robust fashion the mean absolute deviation of the return $y_t^{(k)}$ is used. In particular, the single stock representation reads

$$\mathcal{C} \equiv (\sigma_t^{(k)})_{k=1,\dots,m \; t=1,\dots,T} \quad \sigma_t^{(k)} \equiv d_1(y_{it}^{(k)}, 0) = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_{i,t}^{(k)}|}{\hat{\sigma}_i^{(k)}}\sqrt{T}, \tag{3.2}$$

1 As at 2 January 2008

with T time intervals of the trading day, securities k, intraday intervals t and days i. Here, the intraday measure of risk is scaled by an estimator of the daily volatility $\hat{\sigma}_i^{(k)}$ which is calculated as a 14-period exponential moving average $EMA_{14}(\omega)$ with

$$\omega \equiv \sup\left\{|\alpha-\beta|:\ \alpha,\beta\in\{\text{open}^{\pm},\text{close}^{\pm},\text{high}^{\pm},\text{low}^{\pm}\}\right\} \tag{3.3}$$

where $\pm$ accounts for realizations of succeeding trading days. In the face of non-Gaussian returns the scaling may not be exact, however, it approximately removes interday effects within an intraday volatility analysis. Furthermore, it transforms intraday volatility to the same magnitude for different dates and equities which is essential within the context of the subsequent singular value decomposition.

Moreover, the collective representation for a market l is stated by

$$\mathcal{D}_l \equiv (\sigma_{i,t})_{i=1,\dots,n\ t=1,\dots,T} \quad \sigma_{i,t} = \frac{1}{m}\sum_{k=1}^{m}\frac{|y_{i,t}^{(k)}|}{\hat{\sigma}_i^{(k)}}\sqrt{T}. \tag{3.4}$$

Index variants of expression (3.4) replace $\sigma_{i,t}$ by realizations of the corresponding index. Thus, matrices $\mathcal{C}$ and $\mathcal{D}$ are constructed such that rows of $\mathcal{C}$ correspond to different equities, i. e., elements average over different dates, whereas $\mathcal{D}$ measures risk for different indices k where rows corresponds to different days. Notice that overnight returns are removed explicitly from the analysis.

For the sake of a general intraday volatility profile linear independent projections of the data that show the highest variance compared to all other combinations have to be found. Indeed, it can be shown that the singular value decomposition of any $m \times n$ matrix $\mathcal{X} = \{x_1,\dots,x_n\}$

$$\mathcal{X} = \mathcal{U}\Sigma\mathcal{V}^T \tag{3.5}$$

serves the purpose. Here, $\mathcal{U}$ is an $m \times m$ orthonormal matrix, while $\mathcal{V}$ is an $n \times n$ orthonormal matrix. Columns u_i and v_i are called left and right singular vectors. In the context of matrices $\mathcal{C}$ and $\mathcal{D}$ one also refers to them as *eigenequites/dates* and *eigenminutes*, respectively. The $m \times n$ diagonal matrix Σ has elements $s_i \geqslant s_{i+1}$ called singular values. To specify our introductory statement projections $\mathcal{X}v_i$ show the highest variance of the data within the subspace orthogonal to

$$\mathcal{L} \equiv \text{span}\{v_1,\dots,v_{i-1}\}. \tag{3.6}$$

Furthermore, following a theorem due to Eckart and Young [27] the rank-k approximation

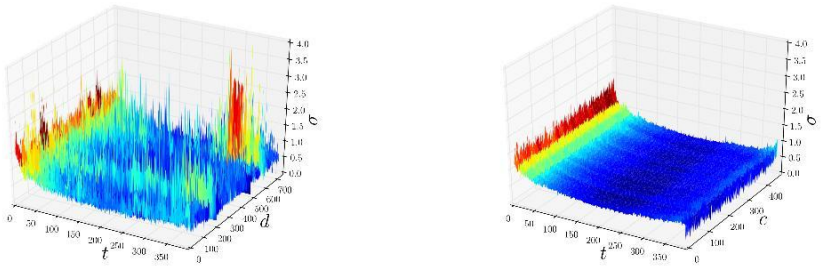
$$\mathcal{X}_k = \mathcal{U}\Sigma_k\mathcal{V}^T = \sum_{i=1}^{k} s_i u_i v_i^T \tag{3.7}$$

minimizes the Frobenius norm $\|\mathcal{X} - \mathcal{X}_k\|_F$, which is equivalent to the minimization of the average Euclidean error of columns x_i with respect to the chosen subspace.

3.3 HISTORICAL VOLATILITY

3.3.1 *Patterns in the US Stock's Volatility Profile*

Figure 4 depicts matrices $\mathcal{C}$ and $\mathcal{D}$ within the US. Figure 4a approximately represents the collective risk with respect to the chosen stock universe. Local and non-periodic enhanced volatility is most likely due to new information which eludes predictability within efficient markets. In order to estimate future market-risk one wants to identify these events reliably and remove them in an iterative fashion from a generic model for the intraday dispersion. The date averaged intraday dispersion for every stock as illustrated in figure 4b is much smoother.



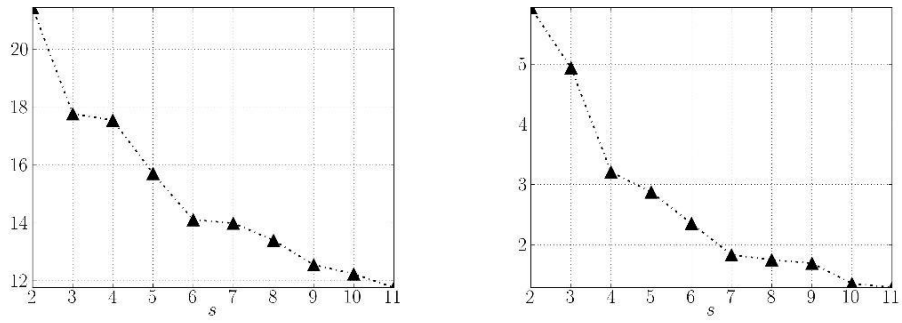
(a) Volatility for different dates averaged over 465 equities.

(b) Volatility for different equities averaged over 770 dates.

Figure 4: Matrices $\mathcal{C}$ and $\mathcal{D}$ for the US, time intervals of 1 minute for equities with market capitalization that differ by a factor 436 over 3 years. The color encoding is averaged over 10×10 entries.

A reduction of dimension by applying a singular value decomposition (3.5) to matrices $\mathcal{C}$ and $\mathcal{D}$ is performed.

The rank-1 approximation of the data $s_1 u_1 v_1^T$ equals the mean volatility with respect to different dates and stocks. However, in the case of market dispersion it accounts for only 14.2%, whereas mean risk as measured by the single stock matrix $\mathcal{C}$ represents 60.3% of the variation among the data. A more complex model such as a rank-3 approximation accounts for 16.7% and 64.3%, respectively. Figure 6 exhibits components of the respective rank-3 approximations. The mean intraday profile of volatility for both market indices and individual stocks decays quickly from its maximum at



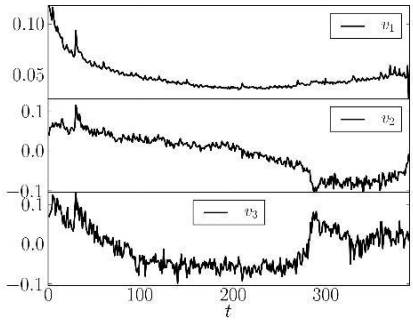
(a) $\Sigma_{\mathcal{D}}$, $s_1/s_2 = 9.67$ (b) $\Sigma_{\mathcal{C}}$, $s_1/s_2 = 22.87$

Figure 5: Singular values s_i, $i \in 2, \ldots 11$ for matrices $\mathcal{D}$ and $\mathcal{C}$. First singular values s_1 are omitted as they are much larger in magnitude. The mean represents 14.2% and 60.3% whereas the rank-3 model accounts for 16.7% and 64.3% of the variation among the data.

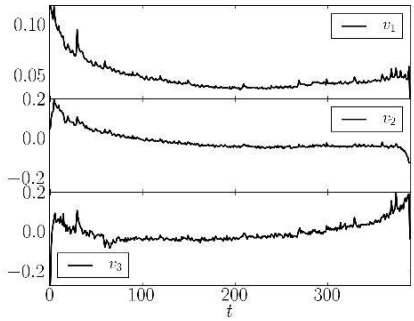
market open and reaches its minimum after roughly 200 minutes around midday. Enhanced volatility in the price finding phase 30 minutes after open are due to overnight news both for macroeconomic news (figure 6a) and corporate events (figure 6b). Additionally, a peak of volatility 30 minutes after trading begin due to the release of macroeconomic news such as the Institute for Supply Management's Manufacturing Index (ISM Mfg), the Housing-Market Index or Consumer Confidence is observed. After midday return dispersion rises again weakly.

The respective left singular vector (figure 6c) describing eigendates exhibits two dominating components, one at October 10, 2008 and another at May 7, 2010. The former corresponds to a market-situation governed by fear that the financial crisis slides into a global recession with the Dow Jones index loosing up to 5%. The latter refers to the day after the so-called *Flash-Crash*. Interestingly, the intraday volatility pattern that day resembles the average intraday volatility but with risk being much higher in magnitude. Furthermore, the second singular-vector u_2 is dominated by two occasions. One at May 6, 2010 the already mentioned *Flash-Crash*, the other is July 3, 2008 with markets closing early. Thus, the intraday profile as represented by v_2 mainly represents a superposition of both events. Finally, $s_3 u_3 v_3^T$ is a superposition of the aftermath of the *Flash-Crash* at May 7, 2010 and of September 18, 2008, where actually two things influenced the markets. To begin with, central banks agreed to infuse \$180 billion into global money markets. In addition, stocks rallied on the Paulson rumors that government would absorb bad dept.

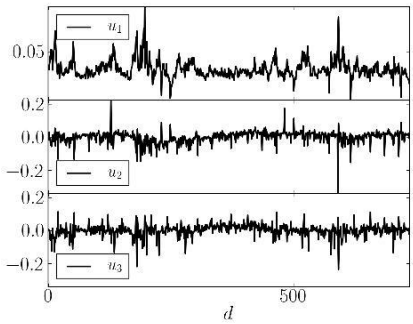
In contrast, the second and third eigenstocks correspond to a correction of the mean volatility profile for a number of stocks mainly around the



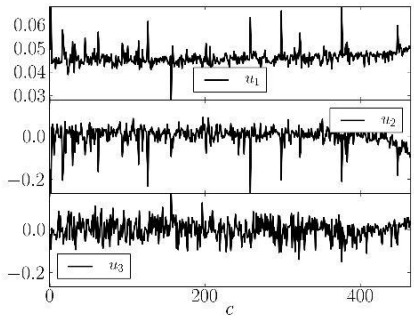
(a) eigenminutes of $\mathcal{D}$



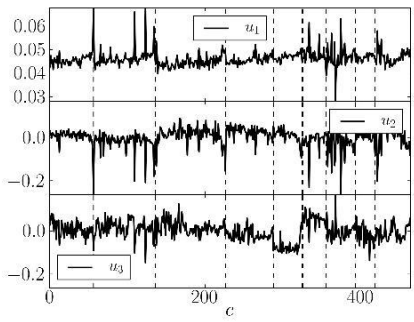
(b) eigenminutes of $\mathcal{C}$



(c) eigendates



(d) eigenstocks ordered by market cap.



(e) eigenstocks ordered by sector

Figure 6: First three left and right singular vectors of matrices $\mathcal{D}$ and $\mathcal{C}$ in the US. Eigenminutes exhibit enhanced volatility in the price finding phase due to overnight news. In the last minute volatility decreases. The announcement of macroeconomic news results in locally increased return dispersion. Eigendates reveal the well-known volatility clustering. The collective risk depends strongly on certain singular events such as the *Flash-Crash*. Components of the single stock rank-3 model feature neither a connection to market capitalization nor to sector affiliation.

first and last minutes of a trading day. In particular, some companies dominating u_2 exhibit lower return dispersion in the price-finding phase

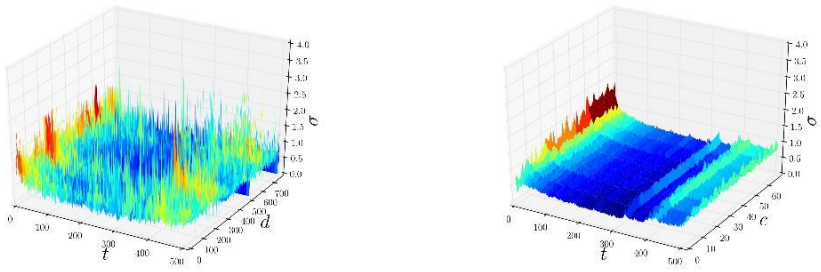
and increased volatility during the closing phase as compared to the mean profile, whereas certain stocks extraordinary corrections to the mean dispersion within the first and last trading minutes and increasing volatility in the last trading hour. However, there is neither a relation to market capitalization (figure 6d) nor to the sector affiliation (figure 6e). It is remarked that events changing the volatility for the individual stock are not detected due to the averaging over 770 dates.

3.3.2 *Patterns in the European Stock Market Data*

While some effects such as an increased return dispersion due to the price-finding phase after open are comparable to results of the US market analysis, a lot of features differ due to market mechanics and regulatory frameworks. Additionally, different trading hours result in different information both available and anticipated.

The data sample containing European stocks is split into equities listed on London Stock Exchange, Euronext and XETRA. A focus on these trading venues results in homogeneous trading hours from 08:00 to 16:30 GMT.

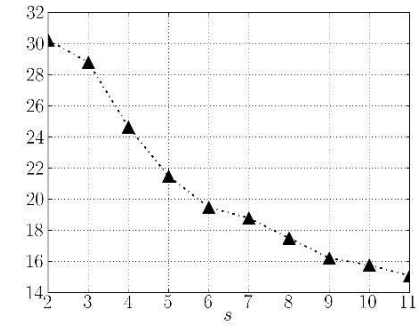
London Stock Exchange data for matrices $\mathcal{D}_{\mathrm{LSE}}$ and $\mathcal{C}_{\mathrm{LSE}}$ is depicted in figure 7. A singular value decomposition of the respective matrices is performed. The mean of $\mathcal{D}_{\mathrm{LSE}}$ represents 12.3% whereas the mean of $\mathcal{C}_{\mathrm{LSE}}$ constitutes 69.2% of the variation among the data. The rank-3 approximation accounts for 14.3% and 73.9%, respectively.



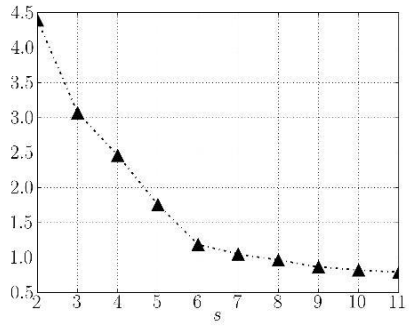
(a) $\mathcal{D}_{\mathrm{LSE}}$ (b) $\mathcal{C}_{\mathrm{LSE}}$

Figure 7: Matrices $\mathcal{C}$ and $\mathcal{D}$ for London Stock Exchange, time intervals of 1 minute for 70 stocks from January 2, 2008 to December 29, 2010. The color encoding is averaged over 10×10 entries.

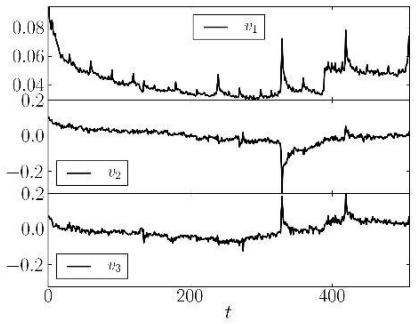
The mean dispersion profile for both matrices $\mathcal{D}_{\mathrm{LSE}}$ and $\mathcal{C}_{\mathrm{LSE}}$ starts with high volatility in the price finding phase lasting approximately 30 minutes after the open. This effect due to overnight news is comparable to the US.



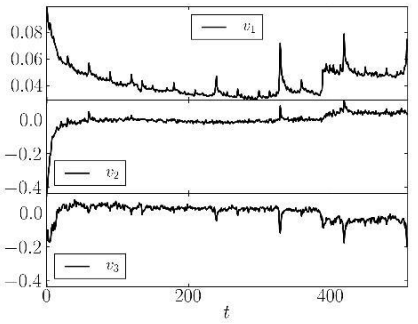
(a) $\Sigma_{\mathcal{D}_{\mathrm{LSE}}}$, $s_1/s_2 = 12.15$



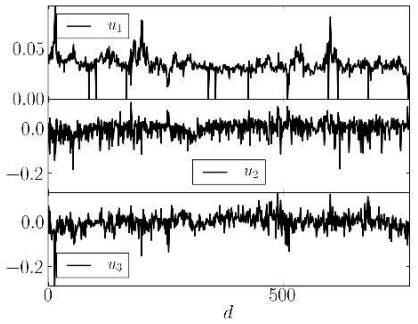
(b) $\Sigma_{\mathcal{C}_{\mathrm{LSE}}}$, $s_1/s_2 = 24.97$



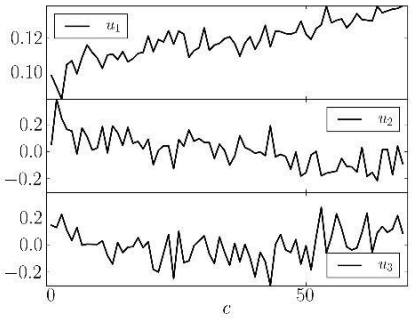
(c) eigenminutes of $\mathcal{D}_{\mathrm{LSE}}$



(d) eigenminutes of $\mathcal{C}_{\mathrm{LSE}}$



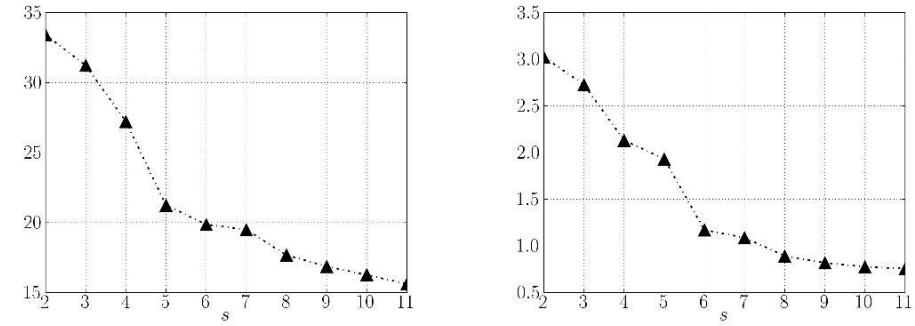
(e) eigendates



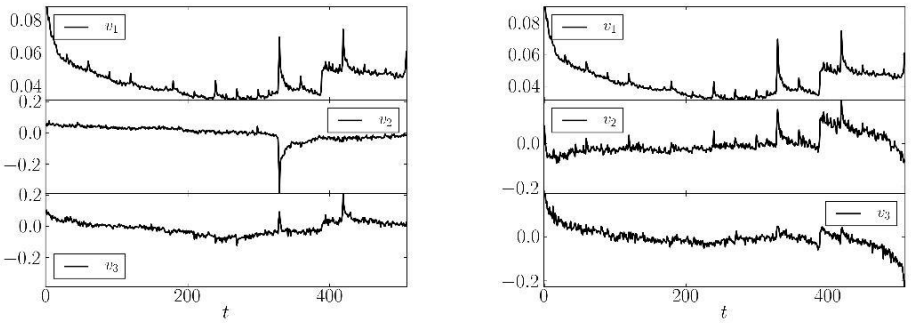
(f) eigenstocks

Figure 8: Rank-3 approximation for the London Stock Exchange (14.3% and 73.9% of the variation). Locally enhanced return dispersion is due to the regular announcement of macroeconomic events. A barrier of volatility is observed due to begin of continuous trading in the US. Volatility rises within the last minutes of continuous trading. Singular events dominating the decomposition can be linked to the financial crisis, the *Flash-Crash* and Hungary's debt problem. Return dispersion for the stock representation is correlated to market capitalization of single stocks.

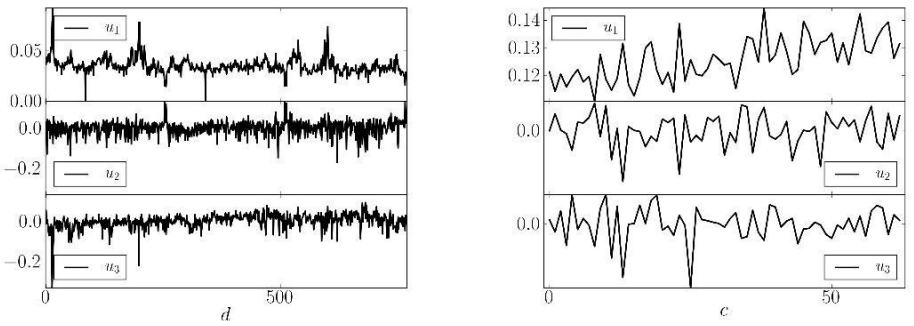
Then, enhanced volatility is observed periodically every 30 minutes due to the announced macroeconomic news within Europe such as European



(a) $\Sigma_{\mathcal{D}_{Eu}}$, $s_1/s_2 = 12.70$ (b) $\Sigma_{\mathcal{C}_{Eu}}$, $s_1/s_2 = 39.42$



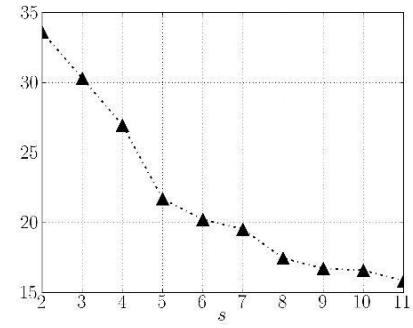
(c) eigenminutes of $\mathcal{D}_{Eu}$ (d) eigenminutes of $\mathcal{C}_{Eu}$



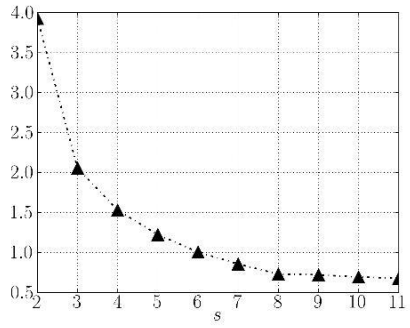
(e) eigendates (f) eigenstocks

Figure 9: Singular values s_i, $i \in 2, \ldots 11$ and the first three left and right singular vectors of intraday volatility at Euronext (77.78% and 15.47% of the variation). Both the mean profile of collective return dispersion and single stock representation are strongly correlated to corresponding results at LSE. This stays true for higher corrections at the collective dispersion level rendering strong integration of the European market. However, higher corrections to the single stock representation discriminate distinct stocks which results in different profiles.

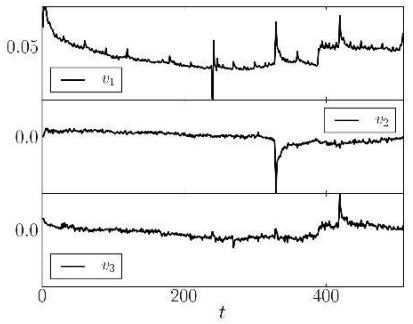
central banks rate decisions or consumer price indices. Locally enhanced volatility after midday is due to macroeconomic news releases within the



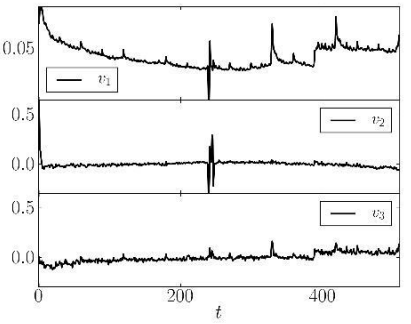
(a) $\Sigma_{\mathcal{D}_X}$, $s_1/s_2 = 12.34$



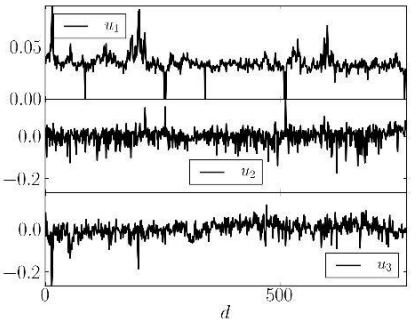
(b) $\Sigma_{\mathcal{C}_X}$, $s_1/s_2 = 22.81$



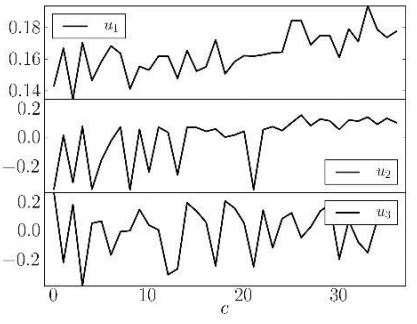
(c) eigenminutes of $\mathcal{D}_X$



(d) eigenminutes of $\mathcal{C}_X$



(e) eigendates



(f) eigenstocks

Figure 10: Singular values s_i, $i \in 2, \ldots 11$ and the first three left and right singular vectors intraday volatility at XETRA (82.30% and 14.54% of the variation). The mean profile for collective return dispersion, single stock representation and higher market corrections are strongly correlated to LSE and Euronext. Locally enhanced volatility around midday is due to the intraday auction which also dominates the second eigenminutes of the single stock representation.

pre-trade phase of the US and Canada. For example, the peak of volatility at 13:30 GMT (330 minutes after open) is linked to the Canadian quarterly GDP or the US unemployment rate. Interestingly, the impact of announced

news within the US on stocks traded at LSE is stronger than the effect of European news. Furthermore, a barrier of volatility at 14 : 30 GMT (390 minutes after open) is observed due to begin of continuous trading within the US, i. e., European investor's uncertainty about the impact US overnight news is released. Increased volatility at 15:00 GMT corresponds the release of macroeconomic news in the US such as consumer confidence or the ISM Mfg index. Interestingly, dispersion rises within the last minutes of continuous trading due to the succeeding closing auction.

There are a number of events that dominate the respective right singular vectors for $\mathcal{D}_{\text{LSE}}$, i. e., the mean dispersion profile roughly corresponds to January 21 and 22, 2008 and May 7, 2010. While the latter can be linked to the aftermath of the *Flash-Crash*, the former is connected to investors worry about global recession. The Fed's emergency cut on interest rates as announced on January 22 further increased volatility within the markets. The second right singular vector v_2 constitutes further increased volatility 330 minutes after the begin of continuous trading at June, 4, 2010 due to Hungary's dept problem, whereas v_3 again relates to January 21, 2008.

Interestingly, the average dispersion of returns $s_1 u_1 v_1^T$ depends linearly on the market capitalization of chosen companies (figure 8f and 8d), i. e., return dispersion of big-cap companies is larger in magnitude over the day as compared to small caps. Also, this means that the daily estimator of volatility σ_d cannot capture the influence of overnight effects on daily return dispersion.

Corrections of the mean dispersion profile due to second and third singular components result in reduced return dispersion of small caps in the price-finding and closing phase as compared to large caps.

Figure 9 and 10 illustrate components of the rank-3 approximation of matrices $\mathcal{C}$ and $\mathcal{D}$ for stocks traded on Euronext and XETRA. The respective rank-3 models account for 77.78% and 15.47% on Euronext as well as 82.30% and 14.54% on XETRA of the variation among the data. It is observed that the collective return dispersion both at Euronext and XETRA is strongly correlated to the intra-daily profile at LSE, i. e., the mean pattern as well as higher corrections of the rank-3 model resemble those at LSE. However, for the market profile at XETRA one has to account for the influence of the intraday auction around midday, i. e., during a pre-specified interval one has only one price with the assigned dispersion being larger due to concentrated liquidity. Additionally, the auction interval differs for distinct stocks, e. g., those constituting the DAX, MDAX, SDAX etc. Not surprisingly the single stock decomposition features varying intraday auctions within the first and second component of the SVD.

3.3.3 *Indices*

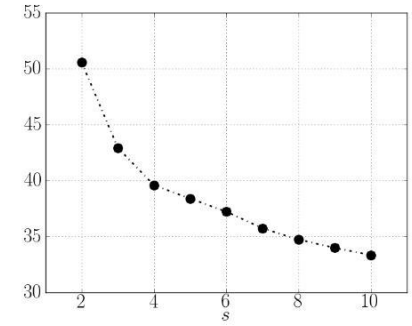
In order to verify if features of the analysis performed in section 3.3.1 and 3.3.2 merely depend on the choice of equities the investigation is repeated using the S&P 500 performance index and the DAX on 1 minute time intervals.

The 465 equities chosen in the US constitute more than 90% of the S&P500, whereas the 37 equities traded on XETRA account for the blue chips of the DAX and a few stocks from MDAX and TECDAX. Indices are observed from March 3, 2010 to April 20, 2011 while the single stock analysis is performed from January 2, 2008 to December 29, 2010 such that effects due to external events may differ. However, in section 3.4 a method is proposed to remove effects depending on singular events.

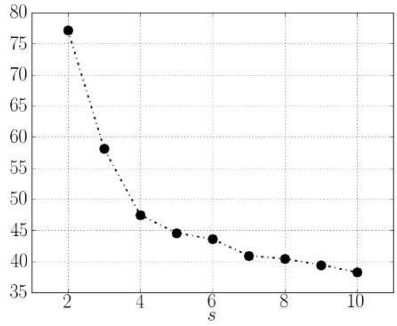
Figure 11 depicts the rank-3 approximation of the market representation (3.4) for the respective indices. It indicates that high-order corrections to the mean are more important compared to the simple average of the single stock time series which is expressed due to the fact that s_1/s_2 is much smaller. This might be due to the fact that many trading days of the single stock analysis are governed by the recent financial crisis where the correlation among different assets are stronger compared to non-crisis times [45]. Thus, the rank-1 representation carries more explanatory power. Concerning the market-representation of the S&P500 the first eigenminutes, depicted in figure 11c, resemble those of the single stock analysis. However, the decay of volatility in the price finding phase due to overnight news is much steeper. The corresponding eigendate profile illustrated in figure 11e is strongly influenced by the Greek dept crisis. In particular, May 19, 2010 to May 21, 2010 and May 25, 2010 effect the corresponding eigenminutes. Additionally, it is observed that the first correction due to the second eigenminutes $v^{(2)}$ increase volatility in the first minute of the continuous trading which can be traced back to May 10, 2010, May 25, 2010, May 27, 2010, June 10, 2010 and March 15 to 16, 2011. Most likely it is linked to a superposition of the aftermath of the Flash-Crash, the Greek dept crisis and the Japanese nuclear crisis.

The rank-1 approximation of the DAX indicates that features of the eigenminutes profile as found in the single stock analysis are conserved (figure 11d).

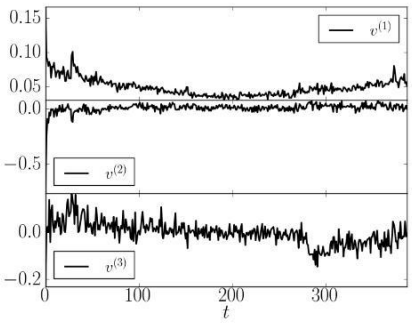
For example, increased volatility due in the price finding phase, the volatility barrier due to the open of the US market and peaks linked to macroeconomic news releases – in particular, to news published during the pre-trade phase of the US and Canada – are observed. The first eigenminutes are influenced by March 15 to 16, 2011, the Japanese nuclear crisis. Interestingly, the Greek crisis has a much weaker influence as compared to the S&P 500.



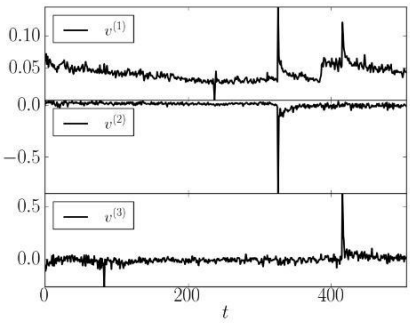
(a) $\Sigma_{\mathcal{D}_{S\&P}}$, $s_1/s_2 = 5.88$



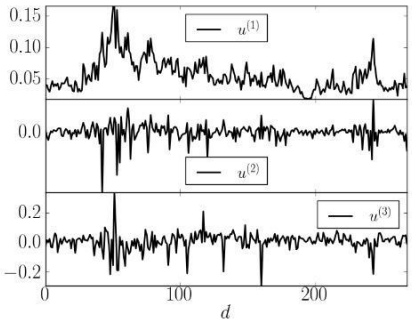
(b) $\Sigma_{\mathcal{D}_{DAX}}$, $s_1/s_2 = 4.10$



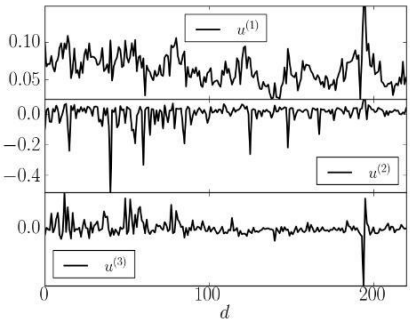
(c) eigenminutes of $\mathcal{D}_{S\&P}$



(d) eigenminutes of $\mathcal{D}_{DAX}$



(e) eigendates of $\mathcal{D}_{S\&P}$



(f) eigendates of $\mathcal{D}_{DAX}$

Figure 11: The rank-3 approximation for the S&P500 and DAX accounts for 10.6% and 10.4% of the variation. First eigenminutes exhibit qualitatively the same features as the corresponding single stock analysis perfomred in section 3.3.1 and section 3.3.2. However, stronger characteristics and varying corrections to the mean profile are due to different time horizons, i. e., the former analysis is influenced by the recent financial crisis whereas the index analysis is impacted by the Greek and Japanese nuclear crisis.

It is concluded that the corresponding index representation as opposed to the market-representation of single stocks exhibits qualitatively the

same features. Variations are induced to different time intervals, i. e., the single stock analysis is governed by the recent financial crisis whereas the index investigation is influenced by different singular events such as the Greek crisis, or the Japanese nuclear crisis.

3.4 ITERATIVE RANK-k MODELS

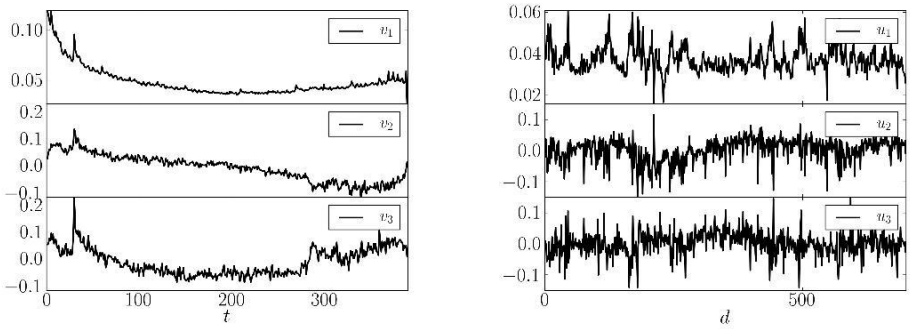
In order to improve the empirical model of intraday dispersion it is remarked that there are different kind of information shocks that drive the stock market. Firstly, announced events such as certain macroeconomic indicators, e. g., consumer confidence, and quarterly figures from companies are anticipated by investors. Secondly, it is rooted within the unique structure of unexpected events that without certain event-specific exogenous variables it is impossible to predict events such as the *Flash-Crash*. Thus, their impact on the stock market is fundamentally different from expected events. In order to account for this notion a framework is introduced to identify unexpected events which are characterized by an extraordinary high impact on the collective intraday return dispersion. Events that are market moving will not only exhibit different market volatility as compared to the mean intraday profile they will also result in large outliers within components of the eigendates. Thus, a filter is proposed that removes days for which it holds

$$|u_j^{(i)}| \geqslant mS_{u^{(i)}}, \quad i = 1, \dots, k, \; m \in \mathbb{N}, \tag{3.8}$$

with $S_{u^{(i)}}$ the sample standard deviation of components j of the given i^{th} eigendate. It is emphasized that the filter (3.8) removes events that have strong impact on the overall voltility profile but have nearly no significance for a typical intraday profile of volatility.

For example, for $m = 5$, i. e., 5 standard deviations, and the single stock analysis within the US as performed in section 3.3.1 it is identified days that can be linked to the financial crisis of 2008 and the so-called Flash-Crash. The virtue of the method lies in the fact that it identifies untypical volatiliy profiles without the need to tie it to individual events. In this sense volatility profiles prodcued by the filter (3.8) are robust.

Figure 12 depicts components of the iterative $\mathrm{rank}-3$ approximation filtering data outliers with 3 standard deviations. This amounts to 13 dates within to u^1, 14 in u^2 and 15 in u^3. While the mean profile of return dispersion is not affected due to singular events, their removal changes higher corrections significantly, i. e., essentially they resemble orthogonal polynomials.

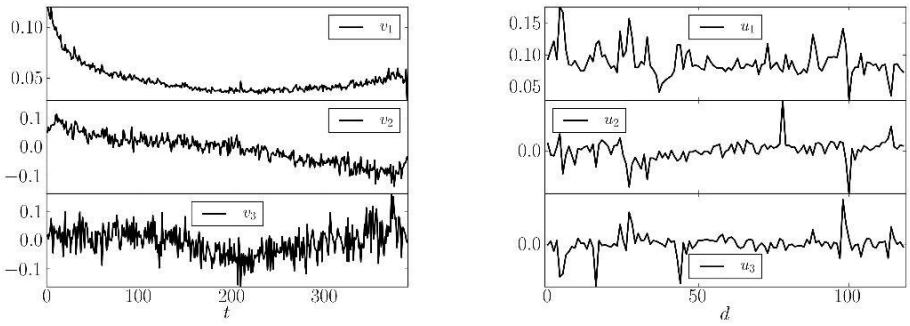


(a) eigenminutes (b) eigendates

Figure 12: Iterative rank-3 $\mathfrak{m} = 3$ approximation of the collective return dispersion with a total of 42 days filtered from the data. Corrections to the mean dispersion profile are essentially orthogonal polynomials. Deviations are caused by macroeconomic news releases.

3.5 IMPACT OF MACROECONOMIC NEWS

In order to analyse if the peak in volatility 30 minutes after open of the US market corresponds to the release of macroeconomic news those days are removed from $\mathcal{D}_{\mathrm{US}}$ for which news were announced. Figure 13 depicts components of rank-3 approximation for the corresponding singular value decomposition. One observes that the locally increased volatility is removed from eigenminutes of the collective return dispersion.



(a) eigenminutes (b) eigendates

Figure 13: Rank-3 approximation of the collective return dispersion with days filtered that exhibit the release of macroeconomic news releases in the US. Thus, locally enhanced volatility is removed from the eigenminute profile.

Thus, the impact of the release of macroeconomic news in the US on the S&P 500 and the DAX on time intervals of 15 seconds for release dates

from March 3, 2010 to April 20, 2011 is considered. In general, news are not released to the public earlier than the announced times due to the usage of lock-up rooms that allow accredited journalists only at pre-specified times to submit information to their agencies.

Events published during continuous trading are considered. Concerning the S&P 500 15 different events with a total of 230 instances are examined. Most of them are released at 10:00 EST. However, a number of events is distributed over the trading day. For example, Crude Inventories (10:30 EST), the Fed's Beige Book, Treasury Budget, Minutes of the FOMC (14:00 EST), and Consumer Credit (15:00 EST). The same analysis is repeated using the DAX on 15 second time intervals with 12 different events and a total of 168 instances. There, all events occuring after the end of the continuous trading are removed. However, one gains access to events that fall into the pre-trade phase of the US (Retail Sales 8:30 EST).

The market response due to the release of an announcement at date i by is calculated according to

$$\Delta\sigma_{it} = \sigma_{it} - \sum_{j=1}^{3} s_j u_{i-1,j} v_j^T, \tag{3.9}$$

where $u_{i-1,j}$ accounts to the component of the j^{th} left singular vector of the previous day which is obtained by an iterative rank-3 $m = 3$ approximation of $\mathcal{D}$. The rationale for doing so is rooted within the well-known volatility clustering as noted by Mandelbrot [63]. Here, $\mathcal{D}$ is calculated using intraday time series up to day i. However, time series that exhibit the announcement of macroeconomic news are filtered. Essentially, this means that background return dispersion is removed and the effect of news releases on the collective return dispersion is explored.

Figure 14 displays the average response and the corresponding statistical error of the S&P500 whereas figure 15 exhibits the reaction of the DAX due to the announcements.

Both the response in volatility based on time series of the S&P 500 and the DAX show similar qualitative behaviour. That is the immediate response is very strong and drops rapidly. No systematic drop in volatility is observed in the 5 minutes preceding the release of the news. Concerning the S&P 500 Treasury budget has no significant impact on the response volatility. With respect to the DAX the announcement of retail sales is insignificant. The latter one is published during the pre-trade phase of the US such that European investor's might await the opening of the continuous trading in the US for which we have evidence due to the volatility barrier within the general intraday profiles in Europe. Concerning the release of crude inventories with respect to the reaction of the DAX

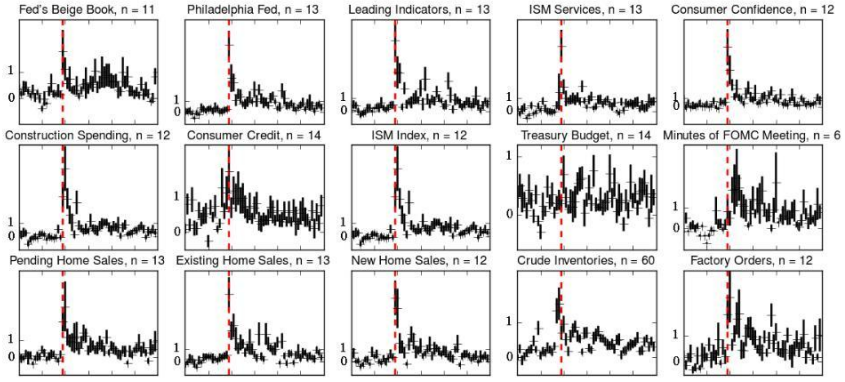


Figure 14: Mean market volatility and statistical error of the S&P500 on 15 second time intervals 5 minutes before and 10 minutes after an event takes place. A dashed line indicates the event in question. Most events show an immediate response due to the release. For some events such as Minutes of FOMC Meeting the response in volatility exhibits long term dependencies.

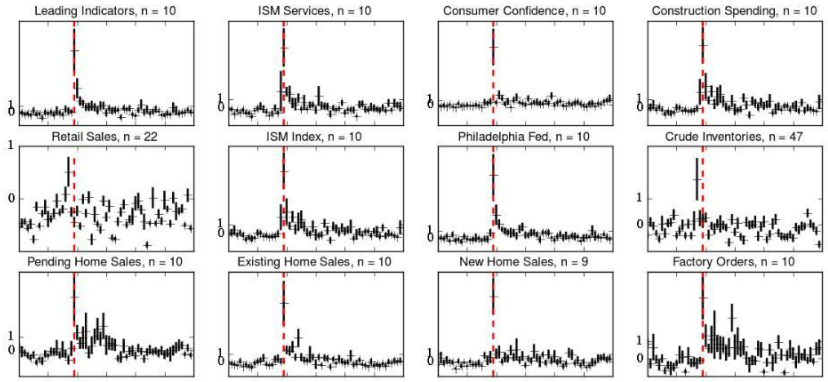


Figure 15: Mean market volatility and statistical error of the DAX on 15 second time intervals. The qualitative behaviour of the response volatility is comparable to that of the S&P 500 that is the immediate response is high and decays very quickly. In particular, response due to leading indicators, consumer confidence and new home sales ist very high. No significant reaction is observed for retail sales.

it is observed that the market reaction is highest 30 seconds prior to the offical relase.

In order to estimate the immediate response and lifetimes of certain events announced at times t_S a volatility pattern of the form

$$\Delta\sigma_{it} = 1_{t\geqslant\tau}\,(1 + \Delta\sigma_{t_S}\exp(-\lambda t)) \tag{3.10}$$

is assumed. Table 1 and table 2 list the instant market reaction due to the news release $\Delta\sigma_\tau$ that is the deviation of the background volatility given by the respective rank-3 approximation at times t_S, rate parameters λ, the corresponding estimation errors and t-statistics.

Returns are normed. Thus, concerning the S&P 500 it is found that the immediate response in terms of added volatility to the background model due to consumer confidence (683%) is biggest followed by the Philadelphia Fed (582%) and leading indicators (543%). The the smallest impact is related to Treasury budget (19%) and Minutes of the FOMC meeting (84%).

Lifetimes $\tau = \lambda^{-1}$ are calculated for events that are found to have impact on the intraday return dispersion. For example, leading indicators (27 seconds) and new home sales (23 seconds) exhibit rather short lifetimes, whereas the Minutes of the Federal Open Market Committee (FOMC) Meeting (401 seconds) and Factory Orders (258 seconds) have the biggest long term impact.

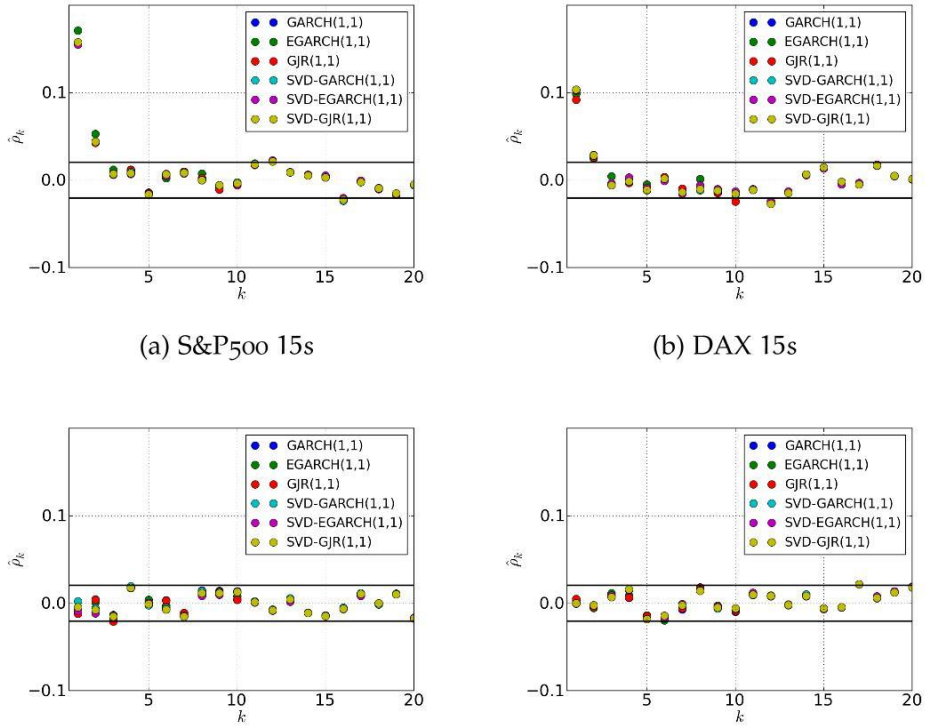
With respect to the DAX it is noticed that both the immediate response is stronger and lifetimes are shorter as compared to the respective events of the S&P 500. Again it is found that Consumer Confidence (1374%) has the strongest impact followed by leading indicators (945%). Very short lifetimes exhibit Consumer Confidence (5 seconds) and New Home Sales (8 seconds) whereas response volatility decays rather slow for Leading Indicators (22 seconds) and the Philadelphia Fed (24 seconds). However, they are in the same order of magnitude as the fastest adjustments of the S&P 500.

3.6 GARCH MODELING

In order to model and forecast the conditional volatility $\sigma_t^2 = \mathrm{Var}[y_t|\mathcal{I}_{t-1}]$ of a time series $\{y_t\}$ of log returns different GARCH paradigms utilized. The data corresponds to high frequency data of the S&P500 and the DAX from 3 March 2010 to 20 April 2011 on time horizons from 15 seconds to 10 minutes in steps of 15 seconds. In line with the previous analysis only intraday returns are considered. Thus, all overnight returns are removed.

Concerning the model specification GARCH(1, 1), EGACRH(1, 1), and GJR(1, 1) are applied as reviewed in section 2.2.3. Additionally, the SVD-GARCH(1, 1), SVD-EGARCH(1, 1) and SVD-GJR(1, 1) are introduced. Essentially, they take into account high-order, intraday patterns of volatility due to a rank-k approximation of the intraday-time series. Additionally, forecasts of the enhanced volatility due to macroeconomic news releases are included. Corresponding model are discussed in detail in section 3.6.1.

Figure 16 depicts the autocorrelogram for the standardized innovation processes on 15 seconds and 1 minute time intervals.



(a) S&P500 15s

(b) DAX 15s

(c) S&P500 60s

(d) DAX 60s

Figure 16: Autocorrelation of the standartized innovation processes of various models on 15 seconds and 1 minute time intervals for data of the S&P500 and the DAX. Horizontal lines correspond to the 95% confidence interval. For ultra-high frequency data the model cannot remove the prevailing autocorrelation.

It is noted that the use of the above models of conditional volatility yields a non-vanishing sample autocorrelation $\hat{\rho}$ over the first $k = 20$ lags ultra-high frequency data with respect to the 95% confidence interval. In particular, the result is consistent over different data sets, i. e., the DAX and the S&P500. However, as one proceeds to lower frequencies, e. g., 1 minute, the models remove the autocorrelation of the data much better. Thus, for high frequent data the model might not remove heterosekasticity from the data which in part may explain the extraordinary variation with respect to the shape paramter of the CTS distribution below certain time horizons. This will be discussed in detail in section 3.6.3.

3.6.1 *Introducing SVD-GARCH*

A number of stylized facts are known about financial time series. For example, the aggregation of information over the trading day is not uniform such that various intraday profiles may exist. Moreover, increased volatility due to trading on overnight news is observed after the open of the corresponding market. However, following weekends and holidays the increase of volatility is not as high as one would expect if information aggregated at an constant pace.

Additionally, different markets could be integrated, i. e., increased volatility is observed due to the open of the US market in Europe. Finally, the announcement of macroeconomic news might be anticipated which results in a different market reaction as compared to news that are released without prior knowledge of market participants.

Typically, these questions can be tackled relaxing the assumption that the unconditional variance ω in equation (2.17) does not depend on the time. The proposed approach is equivalent. However, intraday log returns y_{it} of day i and time t are transformed utilizing an estimator of the intraday volatility, i. e., $\tilde{y}_{it} = y_{it}/\hat{\sigma}_{it}$.

To this end the rank-3 approximation is calculated using all intraday time series up to day i with the constraint that no macroeconomic news are announced a certain day. Thus, the mean intraday profile as well as higher corrections are removed from the data. Concerning the S&P500 this essentially removes a volatility profile with increased volatility at the beginning and the end of the trading session. In Europe also effects due to the open of the US market and intraday auctions are taken into account, additionally.

If macroeconomic news are announced at day i the event-specific mean response curves are estimated. Thus, an estimator $\hat{\sigma}_{it}$ for the intraday volatility is calculated according to

$$\hat{\sigma}_{it} \equiv \sum_{j=1}^{3} s_j u_{i-1,j} v_j^T + \sum_{k=1}^{n} \mathbb{1}_{t \geqslant \tau_k} \left(1 + \Delta\sigma_{t_S} \exp(-\lambda t)\right), \tag{3.11}$$

with the indicator function $\mathbb{1}$, n events k released at time τ_k. Of course, the additional term due to the release of news becomes zero when there are no news released during the trading day. Equation (3.11) essentially preprocesses the actual GARCH modeling by removing parts of the herteroskedasticity within the conditional volatility that are related to high-order intraday profiles.

3.6.2 *Fitting the Innovation Process*

Describing the innovation process in terms of a parametric model might have various reasons ranging from applications in risk measurement to the theoretical understanding of microscopic processes that leads to the price formation. Thus, different distributions such as the normal distribution, the Student t- distribution, the standardized CTS distribution and the VG distribution are utilized (section A). From table 3 and table 4 it can be concluded that the normal distribution is inappropriate to describe the innovation process on all time horizons from 15 seconds to 10 minutes, i. e., both the Kolmogorow-Smirnov as well as the Anderson-Darling goodness-of-fit statistic is rejected on all confidence intervals.

Figure 17 and figure 18 depict maximum likelihood fits of the remaining distributions, i. e., Student t, standardized CTS and VG with respect to innovation processes of the GARCH(1, 1) and SVD-GARCH(1, 1) on various time horizons.

Most notably it is observed that the VG distribution overestimates risk related to the central region of the distribution for data on 15 second time intervals. However, as one proceeds to more low frequent data the VG distribution approaches the empirical distribution and describes it rather well. This appears to be consistent for various models. Both the Student t-distribution and the standardized CTS distribution describe the empirical distribution well in the central region. However, figure 17 and figure 18 allow no quantitative assessment of the goodness-of-fit with respect to the tails of the distribution.

To this end, the Anderson-Darling (AD) statistic as reported in table 5 and table 10 is investigated. Unfortunately, single outliers might have a severe impact on the numerical optimization of the log likelihood. Thus, no smooth dependency of the test statistic from time horizons is recovered. However, the AD test statistic is widely improved for the Student t, the CTS and the VG distribution with respect to the normal distribution. In general, the test statistic is accepted for more low-frequent time intervals with an increased probability. Especially, the DAX data as depicted in table 6 is described remarkably well except for the 30 seconds interval. However, no improvement of goodness-of fit with respect to the usage of the class of SVD-GARCH models can be reported.

Concerning the AD statistics of the CTS distribution on S&P500 data it is observed that the data is well described for time horizons above 30 seconds with a few exceptions. Especially, no remarkable improvement of the goodness-of-fit with respect to the SVD version of GARCH models are observed. However, the goodness-of-fit is enhanced due to the utilization of SVD-GARCH models on nearly all time horizons for the DAX data. In particular, it is noticed that the p-value of the SVD-GARCH model on

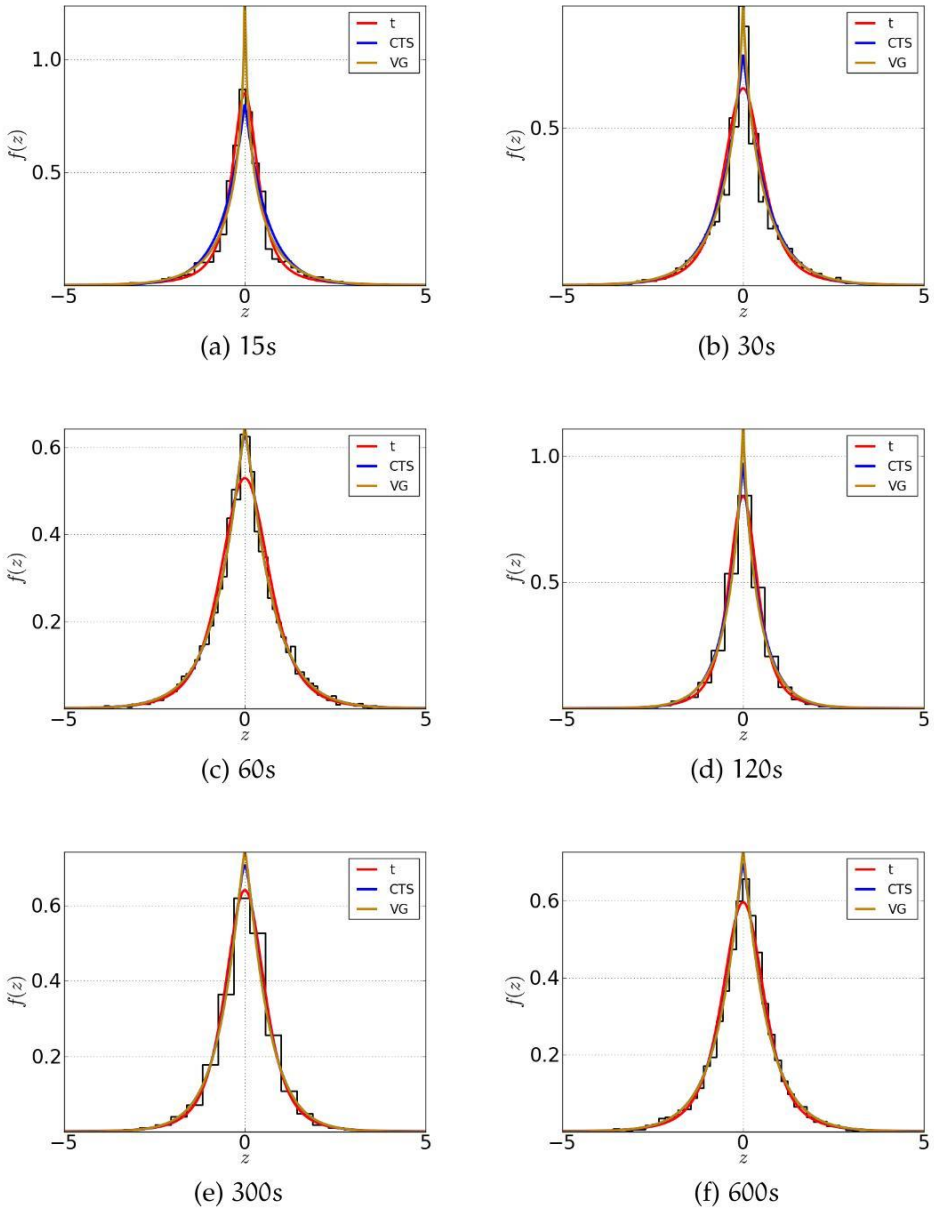


Figure 17: Maximum likelihood fits of various theoretical distributions to the innovation process of a GARCH(1, 1) model on different time horizons. As the time horizons get longer the pronounced peak of the VG distribution approaches the data. The Student t-distribution underestimates risk related to the central region slightly. Furthermore, the excess kurtosis is much more pronounced for very short timescales. The CTS distribution describes the data well on all time horizons above 30 seconds.

30 seconds time horizons is strongly improved compared to the classical GARCH model with a p-value being roughly three times smaller. The

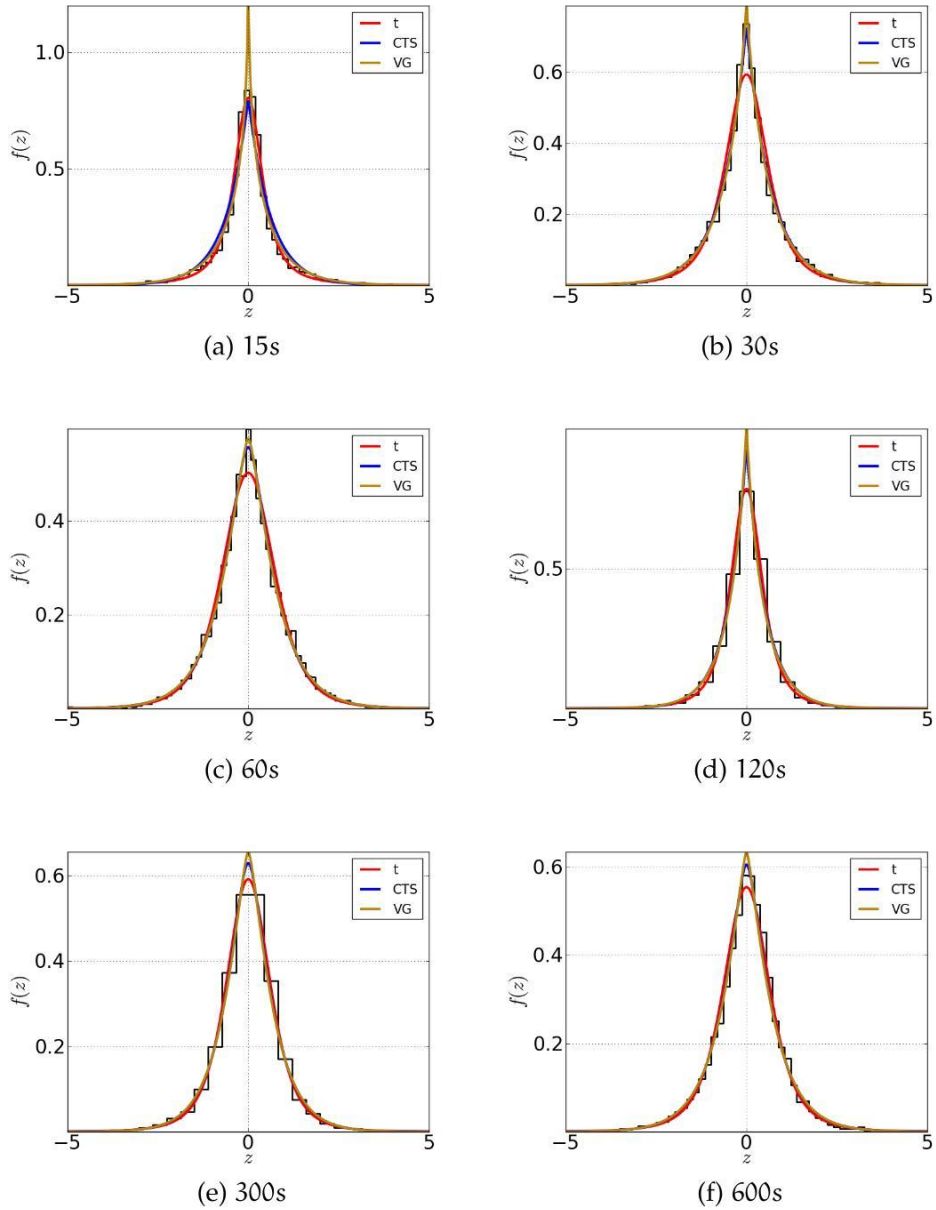


Figure 18: Maximum likelihood fits of the Student t, the CTS and the VG distribution on the innovation process of the SVD-GARCH(1, 1) model for various time horizons on data of the S&P500. Again, increased excess kurtosis of very high frequent data yields probability estimates too extreme in the central region of the VG distribution.

improvement is even better for the EGARCH model, where the p value increases by roughly one order in magnitude on the 15 seconds time interval which renders the CTS distribution possible to describe ultra-high frequency data.

Finally, the AD test statistics of the VG distribution is investigated. It is emphasized that weights in the test statistic stress the importance of the tails of the distribution. Thus, strong deviations of the VG distribution from of high frequency innovations in the center of the distribution are inhibited yielding very good results for the AD test indicating that the VG distribution is capable to model the tail behaviour of the distributions. However, if one takes into account the KS test statistic that considers the single greatest deviation of theoretical and empirical distribution large deviations around zero explain a bad test statistic.

In detail it is observed that the goodness-of-fit with respect to the AD test statistic and S&P500 data is very well above the 15 seconds time interval and improves for SVD-GARCH models with exception of the SVD-GJR model. Concerning the DAX data goodness-of-fit improves which is consistent with the behaviour of the t and CTS distribution.

With respect to the SVD-GARCH paradigm it is found that the goodness-of-fit improves significantly even on the 15 seconds time interval compared to the classical GARCH formulation.

Figure 19 indicates the goodness-of-fit for various models and innovations, both for the S&P500 and the DAX data on 1 minute time horizons.

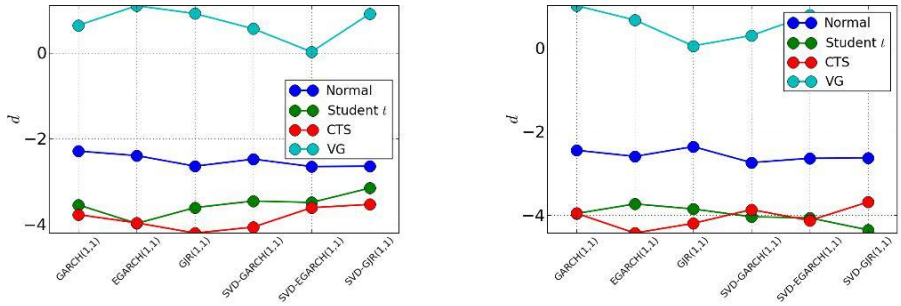
3.6.3 *Random Scaling Behaviour*

In this section the scaling behaviour of parameters of the VG distribution fitted against the GARCH$(1,1)$ innovation and the corresponding SVD-boosted model over varying time horizons are investigated.

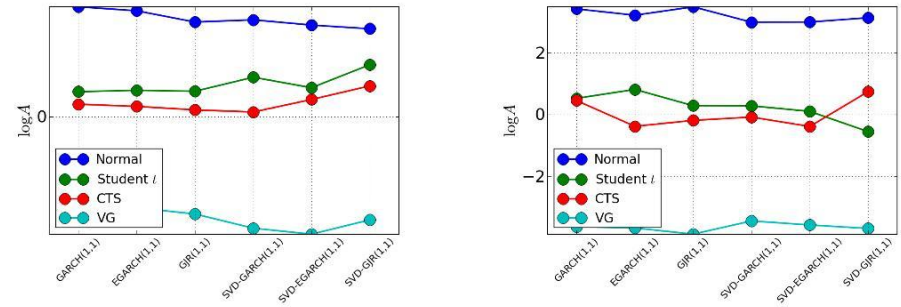
It is found that ν describing the relative excess kurtosis of the VG distribution drops very fast for time scales below 2 minutes. Moreover, innovations of SVD boosted GARCH models exhibits less excess kurtosis as innovations related to the classical GARCH paradigms. Parameters σ and β corresponding to scale and skewness of the distribution show a nonlinear dependence of time. Moreover, the scale of innovations of the SVD-GARCH model diverges compared to simple GARCH formulations, i. e., on longer time horizons innovations of SVD-GARCH exhibit a larger scale than the classical GARCH equivalent. Finally, the skewness of the distribution increases on lower frequent time horizons. However, no distinction with respect to SVD-GARCH and GARCH models is possible.

3.6.4 *Forecasting and Backtesting Value-at-Risk*

In order to assess the forecasting quality of classical GARCH paradigms and their corresponding SVD variants the models are fitted to 5000 observations of DAX and S&P500 data on 1 minute time horizons starting at 3



(a) KS S&P500 (b) KS DAX



(c) AD S&P500 (d) AD DAX

Figure 19: KS and AD test statistics for various models, distributions and data to the 1 minute time horizon.

March 2010. The one-step-ahead forecast $\hat{\sigma}_\tau$ is calculated for the next 5000 observations, i. e., the model is fitted to a moving frame where the oldest observations gradually drop out.

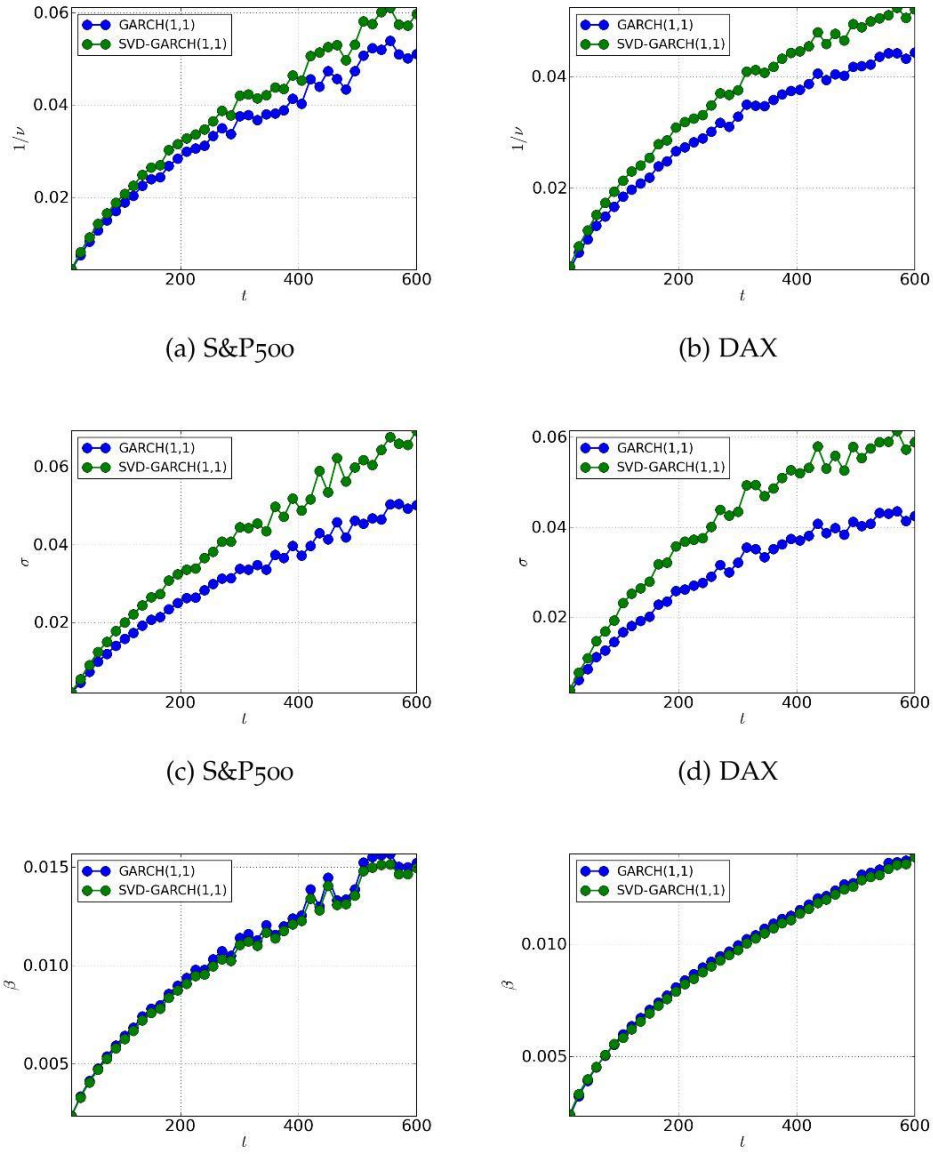
The mean absolute percentage error (MAPE) is scale-free and given by

$$\text{MAPE} = \frac{1}{n} \sum_{\tau=t+1}^{t+n} \left| \frac{\hat{\sigma}_\tau - \sigma_\tau}{\sigma_\tau} \right|, \tag{3.12}$$

with n the number of forecasts. Figure 21 depicts results for the DAX and the S&P500 data. The overall performance as measured by equation (3.12) is slightly better for the DAX data.

One observes that SVD-boosted variants of GARCH$(1,1)$ and GJR$(1,1)$ have lower mean absolute percentage error than their classical variants. However, such a behaviour is not found for the EGARCH$(1,1)$ model. In particular, for S&P500 data classical EGARCH performes competitive to boosted versions of GARCH and GJR and even better than its SVD variant.

The second paradigm to assess a model's quality applied here is back-testing Value-at-Risk (VaR) which is one of the most prominent methods



(a) S&P500 (b) DAX

(c) S&P500 (d) DAX

(e) S&P500 (f) DAX

Figure 20: Random scaling behaviour of parameters ν, β, σ of the the VG distribution. For short time scales ν representing the relative excess kurtosis drops very fast. In particular, Parameters σ and β have a nonlinear dependence from time. On longer time horizons

to measure risk and became an industry standard following J.P. Morgan's introduction of RiskMetrics [26].

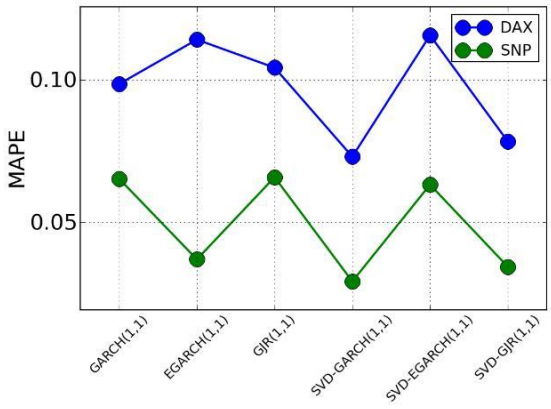


Figure 21: MAPE of various GARCH models with respect to one-step-ahead forecasts. In general, the accuracy with respect to DAX data is slightly better than with respect to S&P500 data. The performance of SVD variants of GARCH and GJR improves with respect to their classical versions.

Value-at-Risk is a lower bound for the loss that hat to be expected with a given probability η, i. e., for a random variable X VaR at the significance level η is given by

$$\mathrm{VaR}_\eta(X) \equiv -\inf_{x\in\mathbb{R}}\{P(X \leqslant x) > \eta\} \tag{3.13}$$

which is just the quantile-function (1.13).

Despite its well-known points of criticism among them a lack of subadditivity[2], it is central to the current regulatory framework, Basel II, which binds banks to hold market risk capital corresponding to the loss expected in 1% of the time on a 10 day time horizon. Specifically, banks have to calculate VaR estimates with respect to the previous 250 trading days and the required market risk capital depends on $\max(\mathrm{VaR}_{t,0.01}, \bar{\rho}_t)$, with

$$\bar{\rho}_t = S_t \frac{1}{60}\sum_{i=0}^{59} \mathrm{VaR}_{t-i,0.01}, \tag{3.14}$$

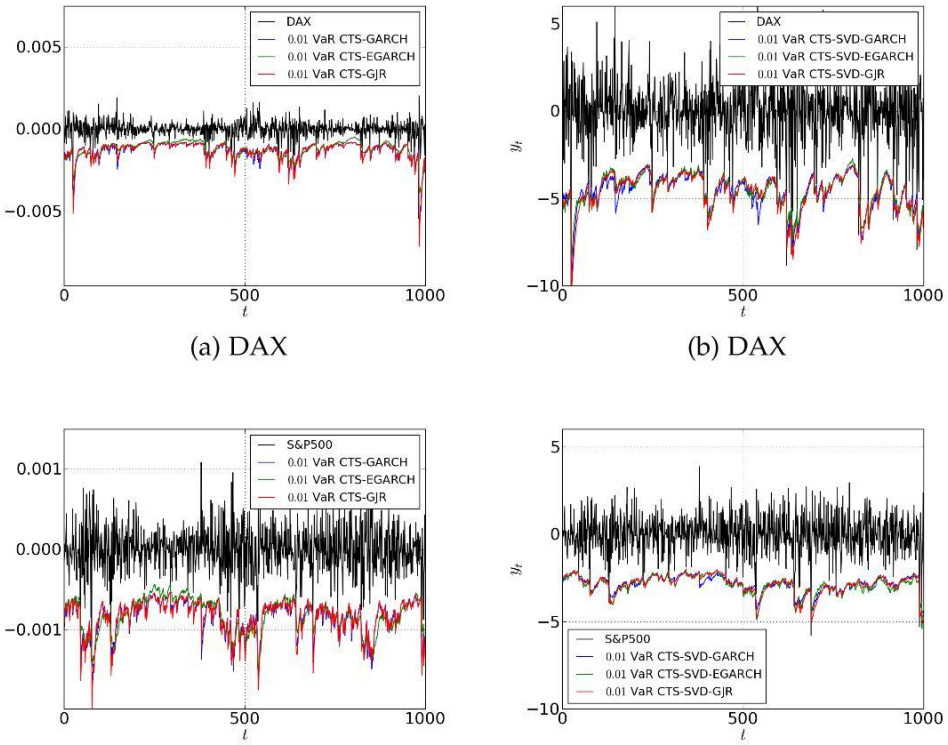
that is the average VaR over the last 60 trading days and S_t depends on the number of VaR exceptions k in the previous 250 trading days. The risk factor S_t is given by the so-called traffic-light approach

$$S_t = \begin{cases} 3, & k \leqslant 4 \quad \text{green} \\ 3 + 0.2(k-4), & 5 \leqslant k \leqslant 0 \quad \text{yellow} \\ 4, & 10 < k \quad \text{red} \end{cases} \tag{3.15}$$

2 VaR is not a coherent risk measure [5].

which is motivated on basis of Kupiec's proportion of failures (POF) test (section B.4). Having sketched the prevailing framework one-step-ahead forecasts of VaR estimates are calculated for the different models. Presented results have a stronger statistical significance as compared to the regulatory requirements. GARCH models are fitted on the recent 5000 historical observations. Moreover, the 5000 one-step-ahead forecasts are calculated. Again, results are obtained for DAX and S&P500 data.

Figure 22 exhibits VaR estimates at the 0.01 significance level utilizing CTS innovations.



(a) DAX (b) DAX

(c) S&P500 (d) S&P500

Figure 22: VaR estimates for various formulations for CTS-GARCH(1, 1), CTS-EGARCH(1, 1), CTS-GJR(1, 1) and their SVD boosted variants on DAX and S&P500 data and 1 minute time horizons starting at 3 March 2010. Only the first 1000 observations are depicted.

The individual model quality is assessed using Kupiec's proportion of failures test. Table 12 and table 11 summarize our findings, with α corresponding to the critical level of Value-at-Risk, k the number of exceptions, Λ the test statistic and the chi-squared distribution with one degree of freedom χ^2_1 evaluated at the 95% confidence interval.

Concerning DAX data results of CTS and VG fitted distributions coincide. Classical versions of the respective GARCH paradigms are generally

rejected except for CTS and VG variants at a critical level of 0.01. Moreover, t-EGARCH is accepted for all critical levels, i. e., 0.01, 0.05 and 0.1. SVD-boosted variants of GARCH, EGARCH and GJR are found to be rejected with repect to a critical value of 0.01 but accepted for the remaining critical values except for the CTS/VG-SVD-EGARCH model.

With regard to S&P500 data it is found that classical GARCH, EGARCH and GJR are rejected for all critical values except for the CTS/VG-EGARCH model with critical values of 0.05 and 0.1 and the t-GARCH and t-GJR model and critical values of 0.1. Acceptance rates of SVD variants are improved compared to classical GARCH paradigms. In particular, all models pass Kupiec's proportion of failures test at the 0.01 VaR level. On the 0.05 level all models except for the VG/CTS versions of GARCH and GJR are accepted. On the 0.1 level only the t-SVD-GJR model is rejected.

Again, CTS and VG lead to a quite similar number of exceptions. Moreover, both for DAX and S&P500 data as well as SVD boosted and classical versions of GARCH modeling a rejection is caused by a too conservative VaR estimate, i. e., the number of exceptions are too small.

3.7 CONCLUSION

This chapter analysed intraday volatility of single stocks as well as different market representations in the US and Europe. Concerning the market representation two approaches were utilized. The first ansatz was based on ultra-high frequency realizations of broad market indices such as the S&P500 and the DAX. The seconds ansatz averages over the single stock universe on 1 minute time horizons. Moreover, sub-samples of the single stock representation are investigated for equities listed on Euronext, LSE and XETRA. To that end a low-rank approximation of the data is performed.

It is found that volatility drops quickly in the price finding phase corresponding to the first 30 minutes of the continuous trading and reaches its minimum around midday. In the second half of the trading session volatility rises only weakly. Locally enhanced volatility is due to the release of macroeconomic news. European volatility profiles are modified to to the open of the US market resulting in a barrier of volatility and various intraday auctions. Unlike the single stock universe in the US mean volatility profiles in Europe depend linearly on the market capitalization. However, no such relation can be reported for the sector affiliation.

Higher corrections to the low-rank approximation of the market representation are impacted by *black swan* events that elude predictability in efficient markets. A procedure is introduced to identify effects that are related to external events which allows to improve generic intraday ap-

proximations. As a result higher corrections to the mean profile resemble orthogonal polynomials. For the US from 2 January 2008 to 29 December 2010 a number of events are identified, e. g., May 6 and 7, 2010 corresponding to the so-called *Flash-Crash* and its aftermath as well as September 18, 2008 and October 10, 2008. The former corresponds to a market-situation governed by the fear that the financial crisis slides into a global recession and the latter is caused by central banks which agreed to infuse $180 billion into global money markets. Concerning Europe additional events are identified. Fox example, June 4, 2010 which is related to Hungary's debt problem.

In order to quantify the impact of macroeconomic news releases in the US on the S&P500 and the DAX a response model and corresponding immediate corrections as well as lifetimes of various event classes are estimated. It is found that 14 out of 15 events impact the S&P500 whereas 10 out of 12 events affect volatility of the DAX. More specifically, Consumer Confidence as well as Leading Indicators have eminent impact. With respect to the S&P500 response-volatility corresponding to the release of macroeconomic news exhibit rather short lifetimes, e. g., , Leading indicators (27 seconds) and New Home Sales (23 seconds). Minutes of the FOMC meeting (401 seconds) has the biggest long-term impact. For the DAX data it is found that response-volatility decays even faster. For example, mean lifetimes of the quickest adjustments of the DAX are Consumer Confidence (5 seconds) and New Home Sales (8 seconds).

Subsequently, the low-rank approximations of the intraday data and models of response-volatility due to the release of macroeconomic news are merged with classical GARCH paradigms. To this end, classical GARCH models such as GARCH$(1,1)$, EGARCH$(1,1)$ and GJR$(1,1)$ are fitted to S&P500 and DAX data on time horizons from 15 seconds to 10 minutes. Thus, boosted GARCH variants called SVD-GARCH, SVD-EGARCH and SVD-GJR that take into account low-rank approximations of intraday volatility as well as response-volatility due to the release of macroeconomic news are introduced.

It is found that both classical as well as boosted GARCH models exhibit non-vanishing autocorrelation of the innovation process for ultra-high-frequency data on time horizons below 1 minute. To this end the goodness-of-fit of various distributions such as the normal, the Student t, the classical tempered stable and the variance gamma distribution to the innovation process of the respective GARCH model are examined. It is found that only Student-t, classical tempered stable and variance gamma distribution describe the empirical data. In particular, classical tempered stable and variance gamma distribution characterize the ultra-high-frequency data as assessed by the Anderson-Darling statistic. In general, SVD-GARCH models lead to an improved goodness-of-fit with

respect to the high-frequency time-scales. Moreover, the random scaling behaviour of the variance gamma distribution is investigated. It is found that for time horizons below 2 minutes relative excess kurtosis increases in an extraordinary manner. In part, this may be associated to the non-vanishing autocorrelation of the innovation process.

Finally, the forecasting quality both of classical and SVD-GARCH models for different distributions and time horizons of 1 minute is analysed. To this end the one-step-ahead forecasting performance of the various models is measured by the mean absolute percentage error. It is found that both for the DAX and the S&P500 the SVD-GARCH$(1,1)$ and the SVD-GJR$(1,1)$ model perform better than their classical analogues. However, no improvement can be reported for the EGARCH model. Additionally, a Value-at-Risk backtest using Kupiec's proportion of failures is conducted. It is found that acceptance rates of Kupiec's proportion of failures test is improved for SVD-boosted variants of GARCH paradigms. In particular, most Value-at-Risk estimates are too conservative with respect to classical GARCH models whereas SVD-GARCH models assess risk in an appropriate manner.

EVENT	$\Delta\sigma_{t_s}$	$s_{\Delta\sigma_{t_s}}$	$t_{\Delta\sigma_{t_s}}$	λ	s_λ	t_λ
Fed's Beige Book	1.7825	0.8160	2.1843	0.3517	0.0588	5.9795
Philadelphia Fed	5.8146	1.2149	4.7861	0.2427	0.0255	9.5242
Leading Indicators	5.4285	1.2988	4.1797	0.5538	0.0747	7.4171
ISM Services	4.9875	1.2211	4.0844	0.4325	0.0746	5.7941
Consumer Confidence	6.8256	2.0708	3.2960	0.3892	0.0471	8.2657
Construction Spending	3.0834	0.9235	3.3390	0.1186	0.0128	9.2479
Consumer Credit	1.7388	0.6314	2.7539	0.0825	0.0075	11.0081
ISM Index	3.0834	0.9235	3.3390	0.1186	0.0128	9.2479
Treasury Budget	0.1919	0.1325	1.4486	N/A	N/A	N/A
Minutes of FOMC Meeting	0.8411	0.6020	1.3971	0.0374	0.0047	7.9017
Pending Home Sales	2.0497	0.5509	3.7207	0.0937	0.0086	10.8836
Existing Home Sales	5.1869	1.2833	4.0418	0.2122	0.0190	11.1893
New Home Sales	4.8963	1.3222	3.7032	0.6517	0.1054	6.1843
Crude Inventories	N/A	N/A	N/A	N/A	N/A	N/A
Factory Orders	1.8115	0.5423	3.3406	0.0582	0.0048	12.0176

Table 1: Immediate impact $\Delta\sigma_{t_s}$, characteristic lifetime $\tau = \lambda^{-1}$ and corresponding estimation errors for certain macroeconomic events. Events with the highest immediate impact are Consumer Confidence and the Philadelphia Fed, whereas Treasury Budget and Minutes of the FOMC Meeting have only weak impact. The shortest lifetimes exhibit Leading Indicators and New Home Sales. Comparatively long range dependencies show Factory Orders and Minutes of the FOMC meeting.

EVENT	$\Delta\sigma_{t_s}$	$s_{\Delta\sigma_{t_s}}$	$t_{\Delta\sigma_{t_s}}$	λ	s_λ	t_λ
Leading Indicators	9.4523	3.3523	2.8197	0.6686	0.0894	7.4797
ISM Services	6.9361	2.2009	3.1515	1.1307	0.2548	4.4380
Consumer Confidence	13.7393	4.3946	3.1264	3.2409	0.8792	3.6862
Construction Spending	7.1790	2.1069	3.4073	0.8323	0.1592	5.2278
Retail Sales	N/A	N/A	N/A	N/A	N/A	N/A
ISM Index	7.1790	2.1069	3.4073	0.8323	0.1592	5.2278
Philadelphia Fed	9.6333	3.1834	3.0261	0.6384	0.0654	9.7629
Crude Inventories	N/A	N/A	N/A	N/A	N/A	N/A
Pending Home Sales	4.0898	1.9099	2.1414	0.9889	0.2951	3.3515
Existing Home Sales	7.7962	2.9572	2.6363	1.2718	0.2956	4.3028
New Home Sales	8.6760	2.4703	3.5121	1.9815	0.6219	3.1863
Factory Orders	3.4924	1.2343	2.8293	0.7790	0.1884	4.1350

Table 2: Compared to the S&P 500 the DAX responds stronger but with shorter lifetimes to the release of macroeconomic news. In particular, Consumer Confidence and Leading Indicators have strong impact. Short lifetimes are related to Consumer Confidence and New Home Sales. Long lifetimes of the response volatility can be linked to Leading Indicators and Philadelphia Fed.

MODEL	Δt	c	k	α_1	β_1	l	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.65710 (0.00000)	0.26673 (0.00000)		0.14508	0.00000	88.73117	0.00000
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.64771 (0.00056)	0.26741 (0.00044)		0.11016	0.00000	55.20060	0.00000
	60	0.00000 (0.00000)	0.00000 (0.00000)	0.65148 (0.00085)	0.25683 (0.00074)		0.10444	0.00000	44.04357	0.00000
	120	0.00001 (0.00000)	0.00000 (0.00000)	0.65405 (0.00385)	0.25624 (0.00416)		0.09433	0.00000	34.49353	0.00000
	300	0.00002 (0.00001)	0.00000 (0.00000)	0.61991 (0.00477)	0.28817 (0.00576)		0.10080	0.00000	46.34471	0.00000
	600	0.00003 (0.00001)	0.00000 (0.00000)	0.65204 (0.01021)	0.26780 (0.00917)		0.10162	0.00000	44.58840	0.00000
EGARCH(1,1)	15	−0.00000 (0.00000)	−0.08398 (0.00069)	0.99461 (0.00004)	0.09370 (0.00029)	−0.01244 (0.00019)	0.14901	0.00000	92.53985	0.00000
	30	−0.00000 (0.00000)	−0.10371 (0.00138)	0.99330 (0.00008)	0.11834 (0.00067)	−0.01812 (0.00042)	0.11586	0.00000	63.13461	0.00000
	60	−0.00000 (0.00000)	−0.14098 (0.00258)	0.99070 (0.00016)	0.15342 (0.00112)	−0.02543 (0.00084)	0.09542	0.00000	39.37244	0.00000
	120	0.00000 (0.00000)	−0.19087 (0.00545)	0.98689 (0.00036)	0.19287 (0.00224)	−0.03820 (0.00149)	0.10015	0.00000	44.00580	0.00000
	300	0.00001 (0.00000)	−0.47880 (0.01517)	0.96542 (0.00105)	0.30200 (0.00409)	−0.06677 (0.00289)	0.09533	0.00000	47.69959	0.00000
	600	0.00001 (0.00001)	−0.78695 (0.03277)	0.94020 (0.00243)	0.34115 (0.00712)	−0.05403 (0.00517)	0.09149	0.00000	38.88375	0.00000
GJR(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.65754 (0.00000)	0.24211 (0.00000)	0.04802 (0.00000)	0.14258	0.00000	87.63072	0.00000
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.64873 (0.00051)	0.23326 (0.00041)	0.06573 (0.00022)	0.10905	0.00000	55.14769	0.00000
	60	0.00000 (0.00000)	0.00000 (0.00000)	0.65273 (0.00076)	0.21159 (0.00067)	0.08787 (0.00032)	0.10222	0.00000	45.51421	0.00000
	120	0.00000 (0.00000)	0.00000 (0.00000)	0.65542 (0.00388)	0.19530 (0.00444)	0.11926 (0.00612)	0.10977	0.00000	45.01820	0.00000
	300	0.00001 (0.00001)	0.00000 (0.00000)	0.62580 (0.00460)	0.18811 (0.00653)	0.18508 (0.01063)	0.10187	0.00000	49.51418	0.00000
	600	0.00001 (0.00001)	0.00000 (0.00000)	0.65737 (0.00973)	0.18952 (0.00928)	0.14032 (0.01540)	0.07150	0.00000	26.34175	0.00000
SVD-GARCH(1,1)	15	0.01051 (0.00156)	0.00161 (0.00003)	0.97668 (0.00010)	0.02332 (0.00011)		0.15230	0.00000	86.84747	0.00000
	30	0.04795 (0.00136)	0.00000 (0.00002)	0.95908 (0.00024)	0.04092 (0.00027)		0.10246	0.00000	45.94194	0.00000
	60	0.01574 (0.00307)	0.00290 (0.00012)	0.95723 (0.00050)	0.04277 (0.00052)		0.08704	0.00000	39.84155	0.00000
	120	0.02093 (0.00442)	0.00398 (0.00024)	0.94998 (0.00086)	0.04968 (0.00087)		0.09275	0.00000	36.22863	0.00000
	300	0.03177 (0.00708)	0.00909 (0.00073)	0.92501 (0.00164)	0.07349 (0.00144)		0.08271	0.00000	33.78079	0.00000
	600	0.03182 (0.01055)	0.01238 (0.00127)	0.92317 (0.00218)	0.07370 (0.00197)		0.08430	0.00000	28.44704	0.00000
SVD-EGARCH(1,1)	15	−0.03873 (0.00018)	0.01189 (0.00006)	0.99609 (0.00004)	0.07594 (0.00027)	−0.01199 (0.00017)	0.14261	0.00000	81.26330	0.00000
	30	0.00903 (0.00208)	0.07355 (0.00069)	0.91749 (0.00057)	0.36145 (0.00200)	−0.00026 (0.00110)	0.09701	0.00000	43.50730	0.00000
	60	0.00073 (0.00248)	0.00728 (0.00016)	0.99433 (0.00015)	0.10109 (0.00099)	−0.02581 (0.00072)	0.09501	0.00000	44.04592	0.00000
	120	0.00764 (0.00376)	0.00635 (0.00028)	0.99403 (0.00027)	0.10545 (0.00171)	−0.03630 (0.00120)	0.08395	0.00000	33.57375	0.00000
	300	0.02078 (0.00747)	0.03652 (0.00197)	0.91608 (0.00264)	0.30682 (0.00671)	−0.04080 (0.00395)	0.08064	0.00000	31.55852	0.00000
	600	−0.09899 (0.00935)	0.01201 (0.00134)	0.95565 (0.00195)	0.20826 (0.00475)	−0.05724 (0.00357)	0.07085	0.00000	23.71491	0.00000
SVD-GJR(1,1)	15	0.00798 (0.00159)	0.00158 (0.00003)	0.97762 (0.00010)	0.01744 (0.00013)	0.00989 (0.00016)	0.14788	0.00000	91.58995	0.00000
	30	0.00782 (0.00219)	0.00206 (0.00006)	0.96843 (0.00026)	0.02351 (0.00032)	0.01612 (0.00041)	0.10955	0.00000	55.57181	0.00000
	60	0.00927 (0.00310)	0.00299 (0.00012)	0.95994 (0.00047)	0.02768 (0.00057)	0.02386 (0.00081)	0.10527	0.00000	42.16724	0.00000
	120	0.01159 (0.00443)	0.00406 (0.00024)	0.95382 (0.00084)	0.02806 (0.00108)	0.03416 (0.00147)	0.08406	0.00000	29.65738	0.00000
	300	0.01672 (0.00705)	0.00899 (0.00068)	0.93251 (0.00173)	0.03513 (0.00236)	0.05849 (0.00286)	0.08390	0.00000	36.75187	0.00000
	600	0.01893 (0.01076)	0.01297 (0.00124)	0.92626 (0.00222)	0.04271 (0.00145)	0.05184 (0.00399)	0.07150	0.00000	20.90336	0.00000

Table 3: Parameter Estimates for normal-GARCH and S&P 500 data. Numbers in parentheses indicate estimation error. Only 6 different time horizons including the horizons of highest and lowest frequency are reported. The sum of GARCH parameters is very close to unity indicating strong persistence of the noise process. However, for normal innovations goodness-of-fit with respect to the Anderson-Darling and Kolmogorow-Smirnov test statistic is rejected at all confidence levels.

MODEL	Δt	c	k	α_1	β_1	l	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	−0.00000 (0.00000)	0.00000 (0.00000)	0.69167 (0.00000)	0.30833 (0.00000)		0.13015	0.00000	72.72874	0.00000
	30	−0.00000 (0.00000)	0.00000 (0.00000)	0.70685 (0.00024)	0.22335 (0.00033)		0.12901	0.00000	55.90930	0.00000
	60	−0.00000 (0.00000)	0.00000 (0.00000)	0.71228 (0.00043)	0.21450 (0.00047)		0.10619	0.00000	46.33243	0.00000
	120	−0.00000 (0.00000)	0.00000 (0.00000)	0.71159 (0.00219)	0.21487 (0.00176)		0.07635	0.00000	27.07157	0.00000
	300	0.00001 (0.00000)	0.00000 (0.00000)	0.72692 (0.00507)	0.19023 (0.00403)		0.08377	0.00000	29.72326	0.00000
	600	0.00002 (0.00001)	0.00000 (0.00000)	0.75636 (0.00811)	0.15513 (0.00564)		0.08699	0.00000	30.60599	0.00000
EGARCH(1,1)	15	−0.00000 (0.00000)	−0.30715 (0.00073)	0.98096 (0.00004)	0.19253 (0.00012)	0.00044 (0.00017)	0.12765	0.00000	75.19521	0.00000
	30	−0.00001 (0.00000)	−0.37527 (0.00170)	0.97611 (0.00010)	0.22614 (0.00040)	0.00480 (0.00027)	0.11333	0.00000	56.02438	0.00000
	60	−0.00000 (0.00000)	−0.62522 (0.00335)	0.95925 (0.00020)	0.25699 (0.00098)	0.00203 (0.00067)	0.10359	0.00000	46.65844	0.00000
	120	−0.00000 (0.00000)	−0.76222 (0.00644)	0.94814 (0.00041)	0.27890 (0.00149)	−0.00869 (0.00119)	0.08767	0.00000	37.61912	0.00000
	300	−0.00001 (0.00000)	−1.09683 (0.02517)	0.92133 (0.00176)	0.26666 (0.00403)	−0.03404 (0.00296)	0.09133	0.00000	28.97623	0.00000
	600	0.00001 (0.00001)	−1.55059 (0.06666)	0.88340 (0.00495)	0.27276 (0.00816)	−0.06327 (0.00526)	0.07516	0.00000	24.94184	0.00000
GJR(1,1)	15	−0.00000 (0.00000)	0.00000 (0.00000)	0.69813 (0.00000)	0.22065 (0.00000)	0.03134 (0.00000)	0.13466	0.00000	71.00460	0.00000
	30	−0.00000 (0.00000)	0.00000 (0.00000)	0.70678 (0.00021)	0.22522 (0.00029)	−0.00353 (0.00013)	0.11402	0.00000	58.77693	0.00000
	60	−0.00000 (0.00000)	0.00000 (0.00000)	0.71226 (0.00039)	0.21486 (0.00042)	−0.00067 (0.00019)	0.10672	0.00000	39.21116	0.00000
	120	−0.00000 (0.00000)	0.00000 (0.00000)	0.71294 (0.00190)	0.19649 (0.00169)	0.03305 (0.00065)	0.07719	0.00000	29.96902	0.00000
	300	−0.00000 (0.00000)	0.00000 (0.00000)	0.73051 (0.00515)	0.15196 (0.00396)	0.06786 (0.00680)	0.07440	0.00000	27.15058	0.00000
	600	0.00001 (0.00001)	0.00000 (0.00000)	0.75980 (0.00826)	0.10327 (0.00594)	0.09498 (0.00939)	0.09494	0.00000	32.76157	0.00000
SVD-GARCH(1,1)	15	−0.00463 (0.00138)	0.01633 (0.00006)	0.93738 (0.00009)	0.06262 (0.00009)		0.12545	0.00000	68.09171	0.00000
	30	−0.00671 (0.00194)	0.02540 (0.00014)	0.91622 (0.00020)	0.08378 (0.00021)		0.09713	0.00000	45.09571	0.00000
	60	−0.01407 (0.00292)	0.04784 (0.00018)	0.89957 (0.00036)	0.08910 (0.00052)		0.08233	0.00000	29.93850	0.00000
	120	0.00651 (0.00519)	0.02941 (0.00043)	0.91777 (0.00097)	0.07156 (0.00096)		0.08498	0.00000	26.99600	0.00000
	300	0.01545 (0.00859)	0.03598 (0.00143)	0.91934 (0.00211)	0.06478 (0.00164)		0.06695	0.00000	18.88508	0.00000
	600	0.02091 (0.01252)	0.05675 (0.00312)	0.90477 (0.00377)	0.06948 (0.00290)		0.06478	0.00000	19.89987	0.00000
SVD-EGARCH(1,1)	15	−0.01962 (0.00010)	0.03297 (0.00005)	0.98599 (0.00002)	0.16357 (0.00014)	0.00457 (0.00014)	0.11578	0.00000	55.41803	0.00000
	30	0.00775 (0.00184)	0.10596 (0.00028)	0.91138 (0.00027)	0.27981 (0.00050)	0.01728 (0.00058)	0.09323	0.00000	37.63514	0.00000
	60	−0.01097 (0.00077)	0.05799 (0.00029)	0.95269 (0.00016)	0.25785 (0.00100)	0.02809 (0.00058)	0.08839	0.00000	36.66572	0.00000
	120	−0.01432 (0.00126)	0.03216 (0.00042)	0.96876 (0.00034)	0.19369 (0.00185)	−0.01886 (0.00115)	0.06821	0.00000	22.62421	0.00000
	300	−0.00534 (0.00656)	0.03046 (0.00098)	0.96624 (0.00119)	0.15708 (0.00316)	−0.02558 (0.00207)	0.07149	0.00000	20.93988	0.00000
	600	0.01779 (0.01101)	0.03248 (0.00161)	0.96438 (0.00183)	0.15186 (0.00527)	−0.03420 (0.00338)	0.07185	0.00000	20.01490	0.00000
SVD-GJR(1,1)	15	−0.00453 (0.00146)	0.01632 (0.00006)	0.93739 (0.00009)	0.06283 (0.00013)	−0.00045 (0.00018)	0.11746	0.00000	59.55608	0.00000
	30	−0.00495 (0.00202)	0.02514 (0.00015)	0.91659 (0.00022)	0.08753 (0.00022)	−0.00823 (0.00024)	0.09529	0.00000	44.90169	0.00000
	60	−0.00908 (0.00364)	0.04739 (0.00021)	0.89796 (0.00042)	0.09998 (0.00057)	−0.01693 (0.00084)	0.08703	0.00000	30.97288	0.00000
	120	0.00023 (0.00527)	0.02887 (0.00044)	0.92100 (0.00097)	0.05650 (0.00107)	0.02347 (0.00137)	0.08084	0.00000	28.76359	0.00000
	300	0.00809 (0.00859)	0.03457 (0.00145)	0.92301 (0.00206)	0.04819 (0.00162)	0.02705 (0.00251)	0.06804	0.00000	21.34484	0.00000
	600	0.01332 (0.01260)	0.05495 (0.00316)	0.90789 (0.00377)	0.04865 (0.00319)	0.03696 (0.00446)	0.07255	0.00000	22.99602	0.00000

Table 4: Parameter Estimates for normal GARCH on DAX data. Results are similar to the findings on S&P500 data. High persistence of the noise process, i. e., the sum of GARCH coefficients being close to unity, is observed. Moreover, fits are rejected with respect to the KS and AD test statistic for all models and time horizons.

model	Δt	c	k	α_1	β_1	l	ν	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.68512 (0.00000)	0.31488 (0.00000)		2.57126 (0.00000)	0.03253	0.00341	5.36659	0.00192
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.69011 (0.00141)	0.30989 (0.00104)		2.87638 (0.00020)	0.02894	0.01290	3.38722	0.01748
	60	0.00000 (0.00000)	0.00000 (0.00000)	0.69010 (0.00207)	0.30990 (0.00155)		3.17353 (0.00019)	0.02809	0.01727	3.10316	0.02427
	120	0.00001 (0.00000)	0.00000 (0.00000)	0.69332 (0.00895)	0.30668 (0.01123)		3.28659 (0.00165)	0.04565	0.00001	6.95573	0.00035
	300	0.00002 (0.00000)	0.00000 (0.00000)	0.66213 (0.01492)	0.33787 (0.01895)		3.46311 (0.00222)	0.03150	0.00508	3.18160	0.02216
	600	0.00005 (0.00001)	0.00000 (0.00000)	0.66203 (0.01943)	0.31434 (0.02619)		3.79911 (0.18582)	0.02901	0.01255	2.37877	0.05739
EGARCH(1,1)	15	−0.00000 (0.00000)	−0.01051 (0.00167)	0.99920 (0.00010)	0.14245 (0.00209)	−0.02554 (0.00091)	2.46872 (0.01406)	0.03067	0.00693	3.01084	0.02702
	30	−0.00001 (0.00000)	−0.03761 (0.00301)	0.99763 (0.00018)	0.13029 (0.00176)	−0.02775 (0.00111)	3.71759 (0.03682)	0.02930	0.01134	3.15391	0.02288
	60	−0.00000 (0.00000)	−0.07146 (0.00536)	0.99553 (0.00033)	0.14913 (0.00276)	−0.02850 (0.00174)	5.22171 (0.08516)	0.01746	0.31592	1.40394	0.20117
	120	0.00000 (0.00000)	−0.13121 (0.01020)	0.99145 (0.00066)	0.19054 (0.00455)	−0.03776 (0.00298)	5.65189 (0.13167)	0.03166	0.00477	2.44211	0.05313
	300	0.00001 (0.00000)	−0.33139 (0.02794)	0.97695 (0.00194)	0.27718 (0.00940)	−0.05672 (0.00624)	5.59500 (0.19934)	0.02644	0.02965	2.45194	0.05250
	600	0.00003 (0.00001)	−0.51104 (0.05345)	0.96238 (0.00392)	0.31018 (0.01567)	−0.05995 (0.01005)	4.69090 (0.21727)	0.01881	0.23578	2.50046	0.04951
GJR(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.68600 (0.00000)	0.29142 (0.00000)	0.04514 (0.00000)	2.57041 (0.00000)	0.03591	0.00085	3.91139	0.00964
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.69177 (0.00128)	0.27310 (0.00096)	0.07025 (0.00049)	2.87342 (0.00018)	0.02541	0.04079	3.17520	0.02232
	60	0.00000 (0.00000)	0.00000 (0.00000)	0.69234 (0.00188)	0.25484 (0.00143)	0.10564 (0.00071)	3.16879 (0.00017)	0.02138	0.12704	1.69827	0.13549
	120	0.00001 (0.00000)	0.00000 (0.00000)	0.69515 (0.00886)	0.23726 (0.01202)	0.13517 (0.01556)	3.28543 (0.00133)	0.03163	0.00483	2.33320	0.06068
	300	0.00001 (0.00000)	0.00000 (0.00000)	0.66275 (0.01455)	0.23335 (0.01992)	0.20780 (0.02770)	3.47821 (0.00193)	0.02879	0.01356	2.14655	0.07641
	600	0.00003 (0.00001)	0.00000 (0.00000)	0.66492 (0.01909)	0.21686 (0.02562)	0.17555 (0.03423)	3.84342 (0.18826)	0.02723	0.02298	2.41533	0.05489
SVD-GARCH(1,1)	15	−0.01186 (0.00100)	0.00009 (0.00003)	0.97090 (0.00031)	0.02910 (0.00038)		3.32413 (0.01843)	0.03912	0.00020	3.96027	0.00912
	30	0.00011 (0.00174)	0.00025 (0.00007)	0.96640 (0.00052)	0.03360 (0.00058)		4.55546 (0.04648)	0.03284	0.00303	3.07765	0.02500
	60	0.01343 (0.00276)	0.00127 (0.00017)	0.96111 (0.00088)	0.03889 (0.00094)		6.12622 (0.10912)	0.02634	0.03054	2.13485	0.07754
	120	0.02230 (0.00407)	0.00246 (0.00035)	0.95322 (0.00145)	0.04678 (0.00155)		6.84321 (0.18762)	0.98896	0.00000	10.65492	0.00000
	300	0.03617 (0.00662)	0.00489 (0.00088)	0.93822 (0.00288)	0.06178 (0.00307)		7.15999 (0.30870)	0.01628	0.39971	0.93929	0.39076
	600	0.04658 (0.00992)	0.00657 (0.00149)	0.93634 (0.00410)	0.06140 (0.00424)		7.70805 (0.43608)	0.03174	0.00463	3.92165	0.00952
SVD-EGARCH(1,1)	15	−0.01490 (0.00081)	0.01231 (0.00047)	0.99486 (0.00022)	0.24810 (0.00333)	−0.01517 (0.00121)	2.57024 (0.01519)	0.99115	0.00000	13.03305	0.00000
	30	−0.00714 (0.00194)	0.04917 (0.00134)	0.96490 (0.00079)	0.33860 (0.00490)	−0.08860 (0.00247)	3.57125 (0.04131)	0.03440	0.00161	3.69261	0.01234
	60	0.03040 (0.00270)	0.00645 (0.00091)	0.96616 (0.00106)	0.26410 (0.00514)	−0.05322 (0.00262)	4.80387 (0.08059)	0.02356	0.07039	1.48992	0.17890
	120	0.05777 (0.00402)	0.00180 (0.00101)	0.96714 (0.00137)	0.26281 (0.00635)	−0.01664 (0.00336)	7.88219 (0.21312)	0.03421	0.00174	3.14046	0.02324
	300	0.02767 (0.00643)	0.00409 (0.00193)	0.96037 (0.00263)	0.29914 (0.01127)	−0.05099 (0.00632)	7.28341 (0.31905)	0.01914	0.21913	1.35432	0.21543
	600	0.02993 (0.00967)	0.00130 (0.00107)	0.99203 (0.00123)	0.11188 (0.00827)	−0.05555 (0.00591)	6.84382 (0.38015)	0.03070	0.00687	2.72947	0.03763
SVD-GJR(1,1)	15	−0.00119 (0.00103)	0.00000 (0.00001)	0.96912 (0.00032)	0.02753 (0.00048)	0.00669 (0.00053)	3.13801 (0.01665)	0.03861	0.00025	4.96976	0.00296
	30	0.00281 (0.00179)	0.00034 (0.00007)	0.96786 (0.00051)	0.02356 (0.00064)	0.01717 (0.00084)	4.55887 (0.04664)	0.02507	0.04522	2.89002	0.03113
	60	0.01041 (0.00278)	0.00133 (0.00017)	0.96288 (0.00085)	0.02535 (0.00108)	0.02355 (0.00143)	6.12764 (0.10992)	0.02574	0.03685	1.69030	0.13692
	120	0.01561 (0.00408)	0.00234 (0.00033)	0.95607 (0.00139)	0.02784 (0.00180)	0.03217 (0.00241)	6.86589 (0.19121)	0.02338	0.07400	1.55410	0.16407
	300	0.02555 (0.00665)	0.00506 (0.00083)	0.94195 (0.00274)	0.03295 (0.00358)	0.04783 (0.00485)	7.32287 (0.32495)	0.02981	0.00946	2.32525	0.06127
	600	0.03714 (0.00982)	0.00759 (0.00152)	0.93917 (0.00413)	0.03090 (0.00548)	0.05275 (0.00736)	6.68039 (0.37437)	0.04294	0.00003	6.04424	0.00092

Table 5: Parameter estimates for t-GARCH and S&P 500 data. The sum of GARCH coefficients is close to unity. The Student t-distribution describes the innovation process much better compared to the normal distribution. However, for ultra-high-frequency time intervals the distribution is rejected by the Anderson-Darling test at the 95% confidence interval.

MODEL	Δt	c	k	α_1	β_1	l	ν	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.73525 (0.00000)	0.26475 (0.00000)		2.64086 (0.00000)	0.02385	0.06485	2.48991	0.05014
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.73585 (0.00127)	0.26415 (0.00097)		2.84544 (0.00017)	0.02130	0.12953	2.70899	0.03856
	60	0.00000 (0.00000)	0.00000 (0.00000)	0.72383 (0.00188)	0.27617 (0.00143)		3.06328 (0.00018)	0.02415	0.05952	1.83089	0.11401
	120	0.00000 (0.00000)	0.00000 (0.00000)	0.73009 (0.00743)	0.26991 (0.00914)		3.12459 (0.00123)	0.01420	0.57557	0.67277	0.58199
	300	0.00000 (0.00000)	0.00000 (0.00000)	0.71233 (0.01164)	0.28356 (0.01413)		3.31116 (0.00148)	0.01442	0.55626	0.67565	0.57950
	600	0.00001 (0.00001)	0.00000 (0.00000)	0.69527 (0.01632)	0.30038 (0.02422)		3.34357 (0.12772)	0.01912	0.21984	1.69197	0.13662
EGARCH(1,1)	15	−0.00000 (0.00000)	−0.15212 (0.00162)	0.99015 (0.00010)	0.35275 (0.00162)	−0.05663 (0.00157)	2.23158 (0.00099)	0.01917	0.21728	1.59889	0.15453
	30	−0.00000 (0.00000)	−0.13460 (0.00370)	0.99195 (0.00022)	0.18813 (0.00200)	−0.01929 (0.00134)	3.31556 (0.02655)	0.99286	0.00000	14.95160	0.00000
	60	−0.00000 (0.00000)	−0.14773 (0.00698)	0.99079 (0.00043)	0.21756 (0.00290)	−0.01942 (0.00207)	3.81422 (0.04384)	0.02361	0.06937	1.36082	0.21350
	120	−0.00000 (0.00000)	−0.15524 (0.00931)	0.98985 (0.00061)	0.24735 (0.00452)	−0.02640 (0.00316)	3.88905 (0.06207)	0.02795	0.01807	1.81235	0.11677
	300	−0.00001 (0.00000)	−0.47901 (0.03459)	0.96639 (0.00241)	0.28863 (0.00904)	−0.03890 (0.00631)	3.87213 (0.10355)	0.02054	0.15681	1.76257	0.12457
	600	−0.00000 (0.00001)	−0.95224 (0.08580)	0.92940 (0.00633)	0.36447 (0.01777)	−0.04988 (0.01135)	3.60493 (0.13483)	0.02407	0.06083	2.25958	0.06643
GJR(1,1)	15	0.00000 (0.00000)	0.00000 (0.00000)	0.73556 (0.00000)	0.25517 (0.00000)	0.01853 (0.00000)	2.64023 (0.00000)	0.02649	0.02918	2.43399	0.05366
	30	0.00000 (0.00000)	0.00000 (0.00000)	0.73650 (0.00115)	0.25130 (0.00089)	0.02439 (0.00047)	2.84361 (0.00016)	0.03328	0.00254	3.74972	0.01157
	60	−0.00000 (0.00000)	0.00000 (0.00000)	0.72511 (0.00170)	0.25479 (0.00131)	0.04020 (0.00068)	3.05938 (0.00017)	0.02597	0.03433	2.34619	0.05972
	120	−0.00000 (0.00000)	0.00000 (0.00000)	0.73263 (0.00740)	0.22054 (0.01010)	0.09365 (0.01340)	3.12109 (0.00099)	0.02302	0.08196	1.31416	0.22780
	300	−0.00000 (0.00000)	0.00000 (0.00000)	0.71650 (0.01152)	0.21108 (0.01488)	0.13779 (0.02161)	3.30795 (0.00127)	0.02100	0.13998	1.76850	0.12361
	600	−0.00000 (0.00001)	0.00000 (0.00000)	0.69793 (0.01609)	0.23158 (0.02412)	0.13093 (0.03158)	3.34913 (0.12837)	0.02131	0.12928	1.33361	0.22171
SVD-GARCH(1,1)	15	0.01212 (0.00046)	0.00000 (0.00000)	0.93966 (0.00039)	0.06034 (0.00055)		3.38596 (0.01640)	0.02320	0.07788	1.93692	0.09951
	30	−0.00026 (0.00026)	0.00000 (0.00000)	0.92134 (0.00057)	0.07866 (0.00077)		4.37279 (0.03254)	0.01952	0.20059	1.66615	0.14135
	60	−0.00492 (0.00024)	0.00000 (0.00000)	0.93007 (0.00081)	0.06993 (0.00102)		4.87973 (0.05303)	0.01744	0.31715	0.89871	0.41504
	120	0.00414 (0.00429)	0.01252 (0.00104)	0.92365 (0.00206)	0.07635 (0.00234)		4.67410 (0.08851)	0.01944	0.20406	0.72998	0.53429
	300	0.00811 (0.00698)	0.01967 (0.00236)	0.91613 (0.00377)	0.08158 (0.00414)		4.57166 (0.13913)	0.01346	0.64372	0.33084	0.91332
	600	0.00758 (0.01011)	0.04077 (0.00587)	0.88649 (0.00731)	0.10620 (0.00800)		4.20940 (0.17355)	0.01770	0.30075	1.32834	0.22334
SVD-EGARCH(1,1)	15	−0.00550 (0.00006)	0.00798 (0.00032)	0.99336 (0.00011)	0.19550 (0.00181)	−0.00750 (0.00096)	2.73530 (0.01477)	0.02232	0.09922	2.41272	0.05507
	30	−0.00584 (0.00007)	0.00447 (0.00034)	0.99421 (0.00019)	0.17484 (0.00177)	−0.01271 (0.00114)	3.44472 (0.02869)	0.02932	0.01126	2.49220	0.05001
	60	−0.00667 (0.00023)	0.00311 (0.00044)	0.99465 (0.00024)	0.18321 (0.00243)	−0.02055 (0.00184)	4.07587 (0.04865)	0.02079	0.14743	1.09279	0.31205
	120	−0.00990 (0.00065)	0.00408 (0.00078)	0.99261 (0.00067)	0.21698 (0.00432)	−0.02492 (0.00310)	4.38526 (0.07751)	0.02532	0.04197	2.65345	0.04120
	300	−0.01049 (0.00633)	0.01131 (0.00161)	0.98005 (0.00190)	0.20008 (0.00738)	−0.03303 (0.00514)	4.52306 (0.13488)	0.01695	0.35089	0.65165	0.60055
	600	−0.00185 (0.00998)	0.01659 (0.00297)	0.97232 (0.00363)	0.21530 (0.01236)	−0.03047 (0.00799)	4.25281 (0.17488)	0.01722	0.33199	1.10967	0.30452
SVD-GJR(1,1)	15	−0.00450 (0.00006)	0.00000 (0.00000)	0.94157 (0.00034)	0.05556 (0.00065)	0.00573 (0.00083)	3.28023 (0.01561)	0.02935	0.01116	3.17355	0.02237
	30	−0.00140 (0.00012)	0.00000 (0.00000)	0.92680 (0.00056)	0.06796 (0.00098)	0.01047 (0.00138)	4.39789 (0.03338)	0.02132	0.12899	1.57358	0.15984
	60	−0.00642 (0.00025)	0.00000 (0.00000)	0.92214 (0.00089)	0.06883 (0.00143)	0.01805 (0.00208)	4.99454 (0.05332)	0.02743	0.02147	2.45716	0.05217
	120	−0.00100 (0.00430)	0.01345 (0.00109)	0.92507 (0.00206)	0.05729 (0.00256)	0.03528 (0.00363)	4.56503 (0.08624)	0.02165	0.11846	1.06498	0.32490
	300	0.00062 (0.00701)	0.02066 (0.00241)	0.91597 (0.00376)	0.05977 (0.00435)	0.04408 (0.00643)	4.57802 (0.13892)	0.01583	0.43534	0.62229	0.62717
	600	0.00186 (0.01016)	0.04116 (0.00585)	0.88735 (0.00723)	0.08306 (0.00856)	0.04451 (0.01159)	4.21907 (0.17425)	0.01292	0.69383	0.57665	0.67033

Table 6: Parameter estimates for t-GARCH and DAX data. Results are compatible with the ones achieved on the S&P500. In particular, goodness-of-fit as indicated by the Kolmogorow-Smirnov and Anderson-Darling test is improved compared to the normal model.

MODEL	Δt	α	λ_+	λ_-	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	0.00010 (0.00002)	1.73885 (0.00472)	1.73837 (0.00472)	0.05656	0.00000	13.18993	0.00000
	30	0.00010 (0.00007)	1.63626 (0.00625)	1.63650 (0.00625)	0.03866	0.00025	5.22969	0.00223
	60	0.00010 (0.00038)	1.58636 (0.00856)	1.58527 (0.00854)	0.02349	0.07184	1.29152	0.23513
	120	0.00010 (0.00302)	1.57614 (0.01202)	1.57427 (0.01205)	0.01306	0.68084	0.65682	0.59595
	300	0.04528 (0.02229)	1.63868 (0.02130)	1.63770 (0.02139)	0.01588	0.43173	1.26007	0.24574
	600	0.05655 (0.03413)	1.52347 (0.02791)	1.52551 (0.02816)	0.02310	0.08017	1.54538	0.16600
EGARCH(1,1)	15	0.00010 (0.00002)	1.73885 (0.00472)	1.73837 (0.00472)	0.04895	0.00000	14.18979	0.00000
	30	0.00010 (0.00007)	1.63626 (0.00625)	1.63650 (0.00625)	0.03117	0.00575	5.36308	0.00193
	60	0.00010 (0.00038)	1.58636 (0.00856)	1.58527 (0.00854)	0.02440	0.05525	2.59871	0.04399
	120	0.00010 (0.00302)	1.57614 (0.01202)	1.57427 (0.01205)	0.02907	0.01232	4.23511	0.00670
	300	0.04528 (0.02229)	1.63868 (0.02130)	1.63770 (0.02139)	0.02007	0.17584	1.53089	0.16927
	600	0.05655 (0.03413)	1.52347 (0.02791)	1.52551 (0.02816)	0.01909	0.22158	1.43375	0.19312
GJR(1,1)	15	0.00010 (0.00002)	1.73886 (0.00472)	1.73837 (0.00472)	0.05746	0.00000	17.67640	0.00000
	30	0.00010 (0.00007)	1.63626 (0.00625)	1.63650 (0.00625)	0.03239	0.00362	4.63016	0.00431
	60	0.00010 (0.00038)	1.58636 (0.00856)	1.58527 (0.00854)	0.02387	0.06440	2.53694	0.04738
	120	0.00010 (0.00302)	1.57614 (0.01202)	1.57428 (0.01205)	0.03281	0.00306	4.14856	0.00738
	300	0.04528 (0.02229)	1.63868 (0.02130)	1.63770 (0.02139)	0.02450	0.05364	1.30649	0.23025
	600	0.05655 (0.03413)	1.52347 (0.02791)	1.52551 (0.02816)	0.01505	0.50076	1.26997	0.24234
SVD-GARCH(1,1)	15	0.00010 (0.00002)	1.70447 (0.00459)	1.70331 (0.00461)	0.04686	0.00030	7.23498	0.00026
	30	0.00010 (0.00010)	1.59391 (0.00605)	1.59338 (0.00607)	0.04374	0.00092	5.12827	0.00249
	60	0.00010 (0.00122)	1.54010 (0.00826)	1.53879 (0.00828)	0.01434	0.79943	0.58681	0.66054
	120	0.05083 (0.01523)	1.52858 (0.01227)	1.52763 (0.01233)	0.02385	0.20211	2.07366	0.08371
	300	0.14714 (0.02617)	1.59301 (0.02172)	1.59007 (0.02172)	0.01777	0.54689	0.90870	0.40891
	600	0.18132 (0.04003)	1.49871 (0.02923)	1.49971 (0.02953)	0.01729	0.58245	1.18045	0.27510
SVD-EGARCH(1,1)	15	0.00010 (0.00002)	1.70448 (0.00459)	1.70331 (0.00461)	0.05457	0.00001	10.59433	0.00000
	30	0.00010 (0.00010)	1.59391 (0.00605)	1.59338 (0.00607)	0.02529	0.15211	2.45835	0.05210
	60	0.00010 (0.00122)	1.54010 (0.00826)	1.53879 (0.00828)	0.01975	0.41126	0.86121	0.43891
	120	0.05083 (0.01523)	1.52858 (0.01227)	1.52763 (0.01233)	0.01362	0.84712	0.53296	0.71342
	300	0.14714 (0.02617)	1.59301 (0.02172)	1.59007 (0.02172)	0.03858	0.00506	3.66076	0.01280
	600	0.18132 (0.04003)	1.49871 (0.02923)	1.49971 (0.02953)	0.02717	0.10257	1.81894	0.11578
SVD-GJR(1,1)	15	0.00010 (0.00002)	1.70447 (0.00459)	1.70331 (0.00461)	0.04145	0.00201	6.65531	0.00048
	30	0.00010 (0.00010)	1.59392 (0.00605)	1.59338 (0.00607)	0.04738	0.00024	6.27828	0.00072
	60	0.00010 (0.00122)	1.54010 (0.00826)	1.53879 (0.00828)	0.02317	0.22978	1.27028	0.24224
	120	0.05083 (0.01523)	1.52858 (0.01227)	1.52763 (0.01233)	0.02235	0.26653	1.88179	0.10678
	300	0.14714 (0.02617)	1.59301 (0.02172)	1.59007 (0.02172)	0.01239	0.91495	0.37778	0.87045
	600	0.18132 (0.04003)	1.49871 (0.02923)	1.49971 (0.02953)	0.02935	0.06249	2.90791	0.03049

Table 7: Parameter estimates for CTS-GARCH and S&P500 data. GARCH parameters were estimated using the Student t-distributions. The CTS distribution is fitted to the final innovation process.

MODEL	Δt	α	λ_+	λ_-	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	0.00010 (0.00003)	1.67606 (0.00439)	1.67953 (0.00441)	0.03921	0.00019	7.08431	0.00030
	30	0.00010 (0.00012)	1.58466 (0.00585)	1.58903 (0.00589)	0.01941	0.20550	1.88929	0.10576
	60	0.00010 (0.00053)	1.53311 (0.00799)	1.53739 (0.00805)	0.02115	0.13469	1.16554	0.28102
	120	0.00010 (0.00473)	1.50362 (0.01109)	1.50761 (0.01119)	0.01707	0.34271	1.11313	0.30300
	300	0.01960 (0.02461)	1.47483 (0.01797)	1.47784 (0.01814)	0.02156	0.12122	1.15348	0.28592
	600	0.00010 (0.04290)	1.46225 (0.02442)	1.46441 (0.02460)	0.01919	0.21657	1.57266	0.16004
EGARCH(1,1)	15	0.00010 (0.00003)	1.67606 (0.00439)	1.67953 (0.00441)	0.02993	0.00908	5.59765	0.00149
	30	0.00010 (0.00012)	1.58466 (0.00585)	1.58903 (0.00589)	0.02776	0.01925	2.89958	0.03078
	60	0.00010 (0.00053)	1.53311 (0.00799)	1.53739 (0.00805)	0.01184	0.78960	0.66620	0.58771
	120	0.00010 (0.00473)	1.50362 (0.01109)	1.50761 (0.01119)	0.01743	0.31818	1.73567	0.12902
	300	0.01960 (0.02461)	1.47483 (0.01797)	1.47784 (0.01814)	0.01573	0.44313	1.19556	0.26924
	600	0.00010 (0.04290)	1.46225 (0.02442)	1.46441 (0.02460)	0.01198	0.77772	0.68389	0.57243
GJR(1,1)	15	0.00010 (0.00003)	1.67606 (0.00439)	1.67953 (0.00441)	0.03447	0.00157	5.80449	0.00119
	30	0.00010 (0.00012)	1.58466 (0.00585)	1.58903 (0.00589)	0.02435	0.05615	2.65331	0.04121
	60	0.00010 (0.00053)	1.53311 (0.00799)	1.53739 (0.00805)	0.01469	0.53210	1.01160	0.35125
	120	0.00010 (0.00473)	1.50362 (0.01109)	1.50761 (0.01119)	0.02521	0.04343	2.85757	0.03234
	300	0.01960 (0.02461)	1.47483 (0.01797)	1.47784 (0.01814)	0.01189	0.78568	0.74493	0.52245
	600	0.00010 (0.04290)	1.46225 (0.02443)	1.46441 (0.02460)	0.01510	0.49588	0.82939	0.46030
SVD-GARCH(1,1)	15	0.00010 (0.00005)	1.69173 (0.00441)	1.69180 (0.00443)	0.03601	0.01091	3.45261	0.01622
	30	0.00010 (0.00024)	1.59948 (0.00588)	1.59916 (0.00591)	0.02211	0.27797	1.11865	0.30060
	60	0.00968 (0.01189)	1.51222 (0.00826)	1.51140 (0.00830)	0.01965	0.41760	1.90694	0.10340
	120	0.10726 (0.01738)	1.46249 (0.01160)	1.46264 (0.01168)	0.01966	0.41659	0.75024	0.51830
	300	0.14972 (0.02794)	1.44527 (0.01850)	1.44539 (0.01851)	0.01548	0.71785	0.50346	0.74327
	600	0.12023 (0.04016)	1.44046 (0.02593)	1.44076 (0.02597)	0.02099	0.33675	0.92253	0.40059
SVD-EGARCH(1,1)	15	0.00010 (0.00005)	1.69173 (0.00441)	1.69180 (0.00443)	0.02836	0.07867	1.81894	0.11578
	30	0.00010 (0.00024)	1.59948 (0.00588)	1.59916 (0.00591)	0.02023	0.38148	1.08623	0.31503
	60	0.00968 (0.01189)	1.51222 (0.00826)	1.51140 (0.00830)	0.02638	0.12135	0.75970	0.51100
	120	0.10726 (0.01738)	1.46249 (0.01160)	1.46264 (0.01168)	0.01487	0.76225	0.73417	0.53095
	300	0.14972 (0.02794)	1.44527 (0.01850)	1.44539 (0.01851)	0.01996	0.39790	0.58516	0.66212
	600	0.12023 (0.04016)	1.44046 (0.02593)	1.44076 (0.02597)	0.01618	0.66533	0.68110	0.57481
SVD-GJR(1,1)	15	0.00010 (0.00005)	1.69173 (0.00441)	1.69180 (0.00443)	0.03933	0.00399	4.51741	0.00489
	30	0.00010 (0.00024)	1.59948 (0.00588)	1.59916 (0.00591)	0.02414	0.19120	2.20050	0.07146
	60	0.00968 (0.01069)	1.51220 (0.00543)	1.51139 (0.00620)	0.01425	0.80585	0.47502	0.77238
	120	0.10726 (0.01738)	1.46249 (0.01160)	1.46264 (0.01168)	0.02633	0.12262	2.83988	0.03303
	300	0.14972 (0.02794)	1.44527 (0.01850)	1.44539 (0.01851)	0.01183	0.93896	0.34442	0.90144
	600	0.12023 (0.04016)	1.44046 (0.02593)	1.44076 (0.02597)	0.02519	0.15542	2.10154	0.08083

Table 8: Parameter estimates for CTS GARCH and DAX data. The goodness-of-fit with respect to the Anderson-Darling statistic is improved for SVD-GARCH, SVD-EGARCH and SVD-GJR and high-frequency time intervals compared to classical GARCH paradigms.

model	Δt	σ	β	λ	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	1.02052 (0.00422)	−0.00230 (0.00236)	0.44953 (0.00184)	5.80621	0.00119	0.05918	0.00000
	30	1.19238 (0.00743)	−0.00128 (0.00334)	0.61151 (0.00452)	1.01527	0.34936	0.02925	0.06393
	60	1.37356 (0.01284)	−0.00145 (0.00477)	0.79510 (0.00992)	3.55669	0.01441	0.03421	0.01809
	120	1.47578 (0.02034)	−0.00161 (0.00684)	0.90024 (0.01735)	1.00464	0.35485	0.01923	0.44452
	300	1.63527 (0.03754)	−0.00058 (0.01140)	1.01086 (0.03382)	2.08058	0.08298	0.02731	0.09935
	600	1.53705 (0.05106)	0.00096 (0.01523)	1.02678 (0.05012)	1.90452	0.10372	0.03251	0.02852
EGARCH(1,1)	15	1.02052 (0.00422)	−0.00230 (0.00236)	0.44953 (0.00184)	2.95666	0.02879	0.05293	0.00003
	30	1.19238 (0.00743)	−0.00128 (0.00334)	0.61151 (0.00452)	0.72376	0.53929	0.02927	0.06373
	60	1.37356 (0.01284)	−0.00145 (0.00477)	0.79510 (0.00992)	4.90698	0.00318	0.05172	0.00004
	120	1.47578 (0.02034)	−0.00161 (0.00684)	0.90024 (0.01735)	1.42297	0.19599	0.02277	0.24729
	300	1.63527 (0.03754)	−0.00058 (0.01140)	1.01086 (0.03382)	2.92269	0.02996	0.03299	0.02517
	600	1.53705 (0.05106)	0.00096 (0.01523)	1.02678 (0.05012)	3.01155	0.02700	0.04332	0.00106
GJR(1,1)	15	1.02052 (0.00422)	−0.00230 (0.00236)	0.44953 (0.00184)	3.25241	0.02042	0.03666	0.00902
	30	1.19238 (0.00743)	−0.00128 (0.00334)	0.61151 (0.00452)	0.61639	0.63262	0.02375	0.20579
	60	1.37356 (0.01284)	−0.00145 (0.00477)	0.79510 (0.00992)	5.76175	0.00125	0.04176	0.00181
	120	1.47578 (0.02034)	−0.00161 (0.00684)	0.90024 (0.01735)	1.72155	0.13142	0.02204	0.28159
	300	1.63527 (0.03754)	−0.00058 (0.01140)	1.01086 (0.03382)	1.06417	0.32528	0.02437	0.18287
	600	1.53705 (0.05106)	0.00096 (0.01523)	1.02678 (0.05012)	2.50645	0.04915	0.03496	0.01472
SVD-GARCH(1,1)	15	1.03761 (0.00436)	−0.00206 (0.00233)	0.47350 (0.00205)	1.78126	0.12158	0.03040	0.04854
	30	1.22361 (0.00806)	−0.00092 (0.00329)	0.66178 (0.00546)	0.88815	0.42161	0.02092	0.34069
	60	1.43642 (0.01421)	−0.00081 (0.00470)	0.89481 (0.01246)	1.54591	0.16588	0.02971	0.05742
	120	1.53967 (0.02254)	−0.00051 (0.00668)	1.02434 (0.02209)	3.41363	0.01696	0.03531	0.01332
	300	1.71188 (0.04203)	−0.00141 (0.01104)	1.17705 (0.04445)	1.83754	0.11304	0.02195	0.28601
	600	1.65416 (0.05975)	0.00071 (0.01497)	1.22896 (0.06922)	1.76095	0.12483	0.02140	0.31453
SVD-EGARCH(1,1)	15	1.03761 (0.00436)	−0.00206 (0.00233)	0.47350 (0.00205)	1.98319	0.09382	0.02897	0.06831
	30	1.22361 (0.00806)	−0.00092 (0.00329)	0.66178 (0.00546)	1.56983	0.16065	0.02964	0.05833
	60	1.43642 (0.01421)	−0.00081 (0.00470)	0.89481 (0.01246)	4.50099	0.00498	0.03501	0.01449
	120	1.53967 (0.02254)	−0.00051 (0.00668)	1.02434 (0.02209)	1.10730	0.30556	0.02008	0.39018
	300	1.71188 (0.04203)	−0.00141 (0.01104)	1.17705 (0.04445)	1.30082	0.23208	0.01698	0.60552
	600	1.65416 (0.05975)	0.00071 (0.01497)	1.22896 (0.06922)	1.02805	0.34288	0.01743	0.57204
SVD-GJR(1,1)	15	1.03761 (0.00436)	−0.00206 (0.00233)	0.47350 (0.00205)	1.84971	0.11128	0.04640	0.00035
	30	1.22361 (0.00806)	−0.00092 (0.00329)	0.66178 (0.00546)	4.08169	0.00796	0.03915	0.00423
	60	1.43642 (0.01421)	−0.00081 (0.00470)	0.89481 (0.01246)	0.69670	0.56157	0.01212	0.92711
	120	1.53967 (0.02254)	−0.00051 (0.00668)	1.02434 (0.02209)	12.76188	0.00000	0.99147	0.00000
	300	1.71188 (0.04203)	−0.00141 (0.01104)	1.17705 (0.04445)	4.64243	0.00426	0.03794	0.00616
	600	1.65416 (0.05975)	0.00071 (0.01497)	1.22896 (0.06922)	2.49454	0.04987	0.02853	0.07568

Table 9: Parameter estimates for VG-GARCH and S&P500 data. A quasi maximum likelihood approach was pursued, i. e., GARCH parameters were fitted using a Student t-distribution. The VG distribution was fitted to the final innovation process. The KS test statistic is deteriorated compared to the CTS-GARCH model due to a peak around zero. However, the AD test is slightly improved.

MODEL	Δt	σ	β	λ	d	p_{ks}	a^2	p_{ad}
GARCH(1,1)	15	1.19703 (0.00527)	−0.00047 (0.00233)	0.59404 (0.00310)	2.09727	0.08127	0.03306	0.02468
	30	1.28738 (0.00839)	0.00075 (0.00323)	0.71935 (0.00607)	0.79503	0.48462	0.01645	0.64516
	60	1.37561 (0.01323)	0.00125 (0.00452)	0.83893 (0.01113)	1.27733	0.23985	0.02055	0.36240
	120	1.43099 (0.01974)	0.00137 (0.00635)	0.92105 (0.01806)	3.16146	0.02268	0.03165	0.03565
	300	1.45686 (0.03284)	0.00075 (0.00996)	0.98132 (0.03213)	1.07280	0.32122	0.02225	0.27117
	600	1.40877 (0.04437)	0.00095 (0.01403)	0.94068 (0.04247)	2.77632	0.03560	0.02649	0.11856
EGARCH(1,1)	15	1.19703 (0.00527)	−0.00047 (0.00233)	0.59404 (0.00310)	2.57698	0.04515	0.03966	0.00361
	30	1.28738 (0.00839)	0.00075 (0.00323)	0.71935 (0.00607)	0.48757	0.75951	0.01719	0.58950
	60	1.37561 (0.01323)	0.00125 (0.00452)	0.83893 (0.01113)	1.22304	0.25893	0.02065	0.35629
	120	1.43099 (0.01974)	0.00137 (0.00635)	0.92105 (0.01806)	0.97192	0.37236	0.01863	0.48563
	300	1.45686 (0.03284)	0.00075 (0.00996)	0.98132 (0.03213)	0.50173	0.74503	0.01445	0.79222
	600	1.40877 (0.04437)	0.00095 (0.01403)	0.94068 (0.04247)	1.96972	0.09544	0.02549	0.14617
GJR(1,1)	15	1.19703 (0.00527)	−0.00047 (0.00233)	0.59404 (0.00310)	14.39435	0.00000	0.99675	0.00000
	30	1.28738 (0.00839)	0.00075 (0.00323)	0.71935 (0.00607)	0.81692	0.46898	0.02077	0.34933
	60	1.37561 (0.01323)	0.00125 (0.00452)	0.83893 (0.01113)	1.14206	0.29064	0.02348	0.21675
	120	1.43099 (0.01974)	0.00137 (0.00635)	0.92105 (0.01806)	0.99843	0.35810	0.01683	0.61684
	300	1.45686 (0.03284)	0.00075 (0.00996)	0.98132 (0.03213)	4.07062	0.00806	0.03949	0.00381
	600	1.40877 (0.04437)	0.00095 (0.01403)	0.94068 (0.04247)	1.05921	0.32763	0.02085	0.34465
SVD-GARCH(1,1)	15	1.33493 (0.00588)	−0.00095 (0.00241)	0.69050 (0.00381)	2.11053	0.07993	0.02488	0.16526
	30	1.42639 (0.00954)	−0.00056 (0.00332)	0.83328 (0.00764)	1.16472	0.28135	0.02392	0.19929
	60	1.49503 (0.01518)	−0.00046 (0.00453)	0.98414 (0.01463)	1.19785	0.26836	0.01953	0.42521
	120	1.53471 (0.02302)	0.00013 (0.00622)	1.10526 (0.02519)	0.93125	0.39544	0.02186	0.29037
	300	1.56078 (0.03755)	0.00008 (0.00975)	1.16885 (0.04347)	0.77644	0.49833	0.01860	0.48732
	600	1.53005 (0.05207)	0.00017 (0.01388)	1.12899 (0.05891)	1.36064	0.21355	0.03197	0.03283
SVD-EGARCH(1,1)	15	1.33493 (0.00588)	−0.00095 (0.00241)	0.69050 (0.00381)	1.40825	0.19998	0.02347	0.21732
	30	1.42639 (0.00954)	−0.00056 (0.00332)	0.83328 (0.00764)	1.41401	0.19841	0.02201	0.28282
	60	1.49503 (0.01518)	−0.00046 (0.00453)	0.98414 (0.01463)	1.32605	0.22406	0.02131	0.31927
	120	1.53471 (0.02302)	0.00013 (0.00622)	1.10526 (0.02519)	0.74623	0.52143	0.01719	0.58969
	300	1.56078 (0.03755)	0.00008 (0.00975)	1.16885 (0.04347)	0.84871	0.44718	0.01460	0.78181
	600	1.53005 (0.05207)	0.00017 (0.01388)	1.12899 (0.05891)	2.22326	0.06947	0.02806	0.08405
SVD-GJR(1,1)	15	1.33493 (0.00588)	−0.00095 (0.00241)	0.69050 (0.00381)	0.93925	0.39078	0.01812	0.52170
	30	1.42639 (0.00954)	−0.00056 (0.00332)	0.83328 (0.00764)	1.31417	0.22780	0.02301	0.23660
	60	1.49503 (0.01518)	−0.00046 (0.00453)	0.98414 (0.01463)	0.77922	0.49626	0.01677	0.62098
	120	1.53471 (0.02302)	0.00013 (0.00622)	1.10526 (0.02519)	2.83223	0.03333	0.03425	0.01792
	300	1.56078 (0.03755)	0.00008 (0.00975)	1.16885 (0.04347)	0.49196	0.75501	0.01624	0.66060
	600	1.53005 (0.05207)	0.00017 (0.01388)	1.12899 (0.05891)	1.22024	0.25996	0.02511	0.15772

Table 10: Parameter estimates for VG-GARCH and DAX data. Nearly all parameter fits are accepted for the SVD-GARCH, SVD-EGARCH and SVD-GJR model at the 95% confidence interval. Notice how the excess kurtosis measured by $\nu = 1/\lambda$ is much higher for the high-frequency data compared to returns on several minute horizons. However, this might be explained by a non-vanishing autocorrelation of the innovation process.

MODEL	α	k	Λ	χ^2_1	OUTCOME	α	k	Λ	χ^2_1	OUTCOME	α	k	Λ	χ^2_1	OUTCOME
t-GARCH	0.01	40	2.17	1.74	Rejected	0.05	193	14.80	1.74	Rejected	0.10	395	26.21	1.74	Rejected
CTS-GARCH	0.01	45	0.52	1.74	Accepted	0.05	176	25.60	1.74	Rejected	0.10	365	44.27	1.74	Rejected
VG-GARCH	0.01	45	0.52	1.74	Accepted	0.05	176	25.60	1.74	Rejected	0.10	365	44.27	1.74	Rejected
t-EGARCH	0.01	53	0.18	1.74	Accepted	0.05	247	0.04	1.74	Accepted	0.10	476	1.30	1.74	Accepted
CTS-EGARCH	0.01	58	1.23	1.74	Accepted	0.05	218	4.50	1.74	Rejected	0.10	449	5.96	1.74	Rejected
VG-EGARCH	0.01	58	1.23	1.74	Accepted	0.05	218	4.50	1.74	Rejected	0.10	449	5.96	1.74	Rejected
t-GJR	0.01	37	3.75	1.74	Rejected	0.05	190	16.47	1.74	Rejected	0.10	403	22.24	1.74	Rejected
CTS-GJR	0.01	42	1.37	1.74	Accepted	0.05	168	31.85	1.74	Rejected	0.10	366	43.59	1.74	Rejected
VG-GJR	0.01	42	1.37	1.74	Accepted	0.05	168	31.85	1.74	Rejected	0.10	366	43.59	1.74	Rejected
t-SVD-GARCH	0.01	62	2.70	1.74	Rejected	0.05	264	0.81	1.74	Accepted	0.10	509	0.18	1.74	Accepted
CTS-SVD-GARCH	0.01	64	3.64	1.74	Rejected	0.05	243	0.21	1.74	Accepted	0.10	475	1.41	1.74	Accepted
VG-SVD-GARCH	0.01	66	4.70	1.74	Rejected	0.05	243	0.21	1.74	Accepted	0.10	475	1.41	1.74	Accepted
t-SVD-EGARCH	0.01	60	1.90	1.74	Rejected	0.05	246	0.07	1.74	Accepted	0.10	485	0.50	1.74	Accepted
CTS-SVD-EGARCH	0.01	60	1.90	1.74	Rejected	0.05	235	0.97	1.74	Accepted	0.10	447	6.45	1.74	Rejected
VG-SVD-EGARCH	0.01	61	2.28	1.74	Rejected	0.05	235	0.97	1.74	Accepted	0.10	447	6.45	1.74	Rejected
t-SVD-GJR	0.01	62	2.70	1.74	Rejected	0.05	264	0.81	1.74	Accepted	0.10	508	0.14	1.74	Accepted
CTS-SVD-GJR	0.01	63	3.15	1.74	Rejected	0.05	245	0.11	1.74	Accepted	0.10	475	1.41	1.74	Accepted
VG-SVD-GJR	0.01	65	4.15	1.74	Rejected	0.05	245	0.11	1.74	Accepted	0.10	475	1.41	1.74	Accepted

Table 11: Results of Kupiec's proportion of failures test and DAX data on 1 minute time horizons. The test is based on $n = 5000$ observations. Moreover, the critical level α corresponds to the VaR level, k are the number of violations, Λ is the test statistic and χ^2_1 is the chi-sqared distribution with one degree of freedom evaluated at a confidence level of 95%. Except for the CTS and VG versions of SVD-EGARCH boosting improves the acceptance rate of different models apart from the 0.01 level.

MODEL	α	k	Λ	χ^2_1	OUTCOME	α	k	Λ	χ^2_1	OUTCOME	α	k	Λ	χ^2_1	OUTCOME
t-GARCH	0.01	27	12.83	1.74	Rejected	0.05	217	4.79	1.74	Rejected	0.10	489	0.27	1.74	Accepted
CTS-GARCH	0.01	37	3.75	1.74	Rejected	0.05	196	13.22	1.74	Rejected	0.10	421	14.58	1.74	Rejected
VG-GARCH	0.01	37	3.75	1.74	Rejected	0.05	186	18.85	1.74	Rejected	0.10	421	14.58	1.74	Rejected
t-EGARCH	0.01	33	6.63	1.74	Rejected	0.05	275	2.55	1.74	Rejected	0.10	554	6.28	1.74	Rejected
CTS-EGARCH	0.01	40	2.17	1.74	Rejected	0.05	240	0.43	1.74	Accepted	0.10	495	0.06	1.74	Accepted
VG-EGARCH	0.01	40	2.17	1.74	Rejected	0.05	233	1.24	1.74	Accepted	0.10	495	0.06	1.74	Accepted
t-GJR	0.01	26	14.11	1.74	Rejected	0.05	214	5.72	1.74	Rejected	0.10	489	0.27	1.74	Accepted
CTS-GJR	0.01	36	4.39	1.74	Rejected	0.05	190	16.47	1.74	Rejected	0.10	427	12.40	1.74	Rejected
VG-GJR	0.01	36	4.39	1.74	Rejected	0.05	183	20.76	1.74	Rejected	0.10	427	12.40	1.74	Rejected
t-SVD-GARCH	0.01	46	0.33	1.74	Accepted	0.05	270	1.64	1.74	Accepted	0.10	532	2.23	1.74	Rejected
CTS-SVD-GARCH	0.01	51	0.02	1.74	Accepted	0.05	233	1.24	1.74	Accepted	0.10	512	0.32	1.74	Accepted
VG-SVD-GARCH	0.01	52	0.08	1.74	Accepted	0.05	233	1.24	1.74	Accepted	0.10	490	0.22	1.74	Accepted
t-SVD-EGARCH	0.01	43	1.04	1.74	Accepted	0.05	251	0.00	1.74	Accepted	0.10	515	0.50	1.74	Accepted
CTS-SVD-EGARCH	0.01	47	0.19	1.74	Accepted	0.05	225	2.72	1.74	Rejected	0.10	497	0.02	1.74	Accepted
VG-SVD-EGARCH	0.01	49	0.02	1.74	Accepted	0.05	225	2.72	1.74	Rejected	0.10	478	1.09	1.74	Accepted
t-SVD-GJR	0.01	42	1.37	1.74	Accepted	0.05	259	0.34	1.74	Accepted	0.10	529	1.84	1.74	Rejected
CTS-SVD-GJR	0.01	45	0.52	1.74	Accepted	0.05	227	2.30	1.74	Rejected	0.10	506	0.08	1.74	Accepted
VG-SVD-GJR	0.01	45	0.52	1.74	Accepted	0.05	227	2.30	1.74	Rejected	0.10	482	0.73	1.74	Accepted

Table 12: Results are based on $n = 5000$ observations, with k the number of exceptions, $alpha$ the critical level of VaR estimates, Λ the result of Kupiec's POF. The chi-sqared distributions with one degree of freedom is evaluated at the 95% confidence level. In general, SVD-boosted variants of GARCH, EGARCH and GJR improves the acceptance rate except for the CTS/VG-EGARCH model. Moreover, rejecting a model is based exclusively on a too conservative estimation of risk.

Part III

FINANCIAL NEWS RECOMMENDATION

4

TEXT MINING OF FINANCIAL NEWS

4.1 INTRODUCTION

The retrieval of information from financial news is a vibrant topic both for academia and professionals involved with financial markets. It touches on questions regarding market efficiency and makes use of concepts from various fields of research, e. g., computational linguistics, behavioural economics and statistical decision theory.

Within the last decade various prototypes for financial news recommendation aiming to predict short-term price movements have been proposed. For example Wüthrich et al. [90] trained a number of machine learning algorithms using 423 expert-compiled features to predict whether or not certain equity indices would increase or decrease by at least 0.5% or stay below the given threshold. Input variables consisted of word tuples. Overnight news were collected from web sites of the Financial Times or Reuters. However, opening prices of the trading day following the release of a given news article were replaced by the previous closing price. This led to the use of future information that would not be available in real-time applications.

Lavrenko et al. [57, 58] used a Naïve Bayes Classifier based on appropriately weighted term-frequencies to categorize intraday trends of certain US equities. They report a performance of 23bps per roundtrip on average. However, only those stocks were considered showing the largest impact to the release of news. Again the model is adjusted using future information that is not available in realtime applications.

Mittermayer and Knolmayer [69] trained various machine learning algorithms including Support Vector Machines with polynomial kernels trained on words, tuple of words and phrases whether equities in the US increase or decrease by 3% or stay below the respective threshold following the 15 minutes after the release. They discard all news with solely editorial character. Neglecting all implicit and explicit transaction costs their strategy yielded a return of 27bps per roundtrip on average.

Emotions drive human decisions [24]. Moreover, behaviour finance shows that collective emotions and feedback decisions influence the markets [72].

Only recently Bollen et al. [11] demonstrated that the public mood as measured by the information content of Twitter messages can be used to

improve the prediction accuracy of closing values of the Dow Jones Industrial Average (DJIA) training a Self-Organizing Fuzzy Neural Network. To this end they constructed features that indicate public mood levels such as *calm*, *alert*, *sure*, *vital*, *kind* and *happy* using a large-scale corpus of tweets; 140 character messages from the social web. It is shown that the prediction accuracy of the learning algorithm is improved compared with a training that makes use of only the latest closing values of the DJIA. However, no trading strategy is proposed implying that it remains ambiguous whether or not actual profit can be deduced.

Further prototypes were proposed [86, 38, 73, 76]. A survey of the respective text mining systems was published by Mittermayer and Knolmayer [68].

To the author's knowledge existing prototypes generally lack one of the following aspects. Usually the statistical significance of expert-compiled features is unkwnown. Moreover, the use of phrases or tuples of words is difficult to justify on a statistical basis. Some learning algorithms are trained with respect to targets that have at best a very poor signal to noise ratio. Additionally, the classification accuracy is usually not compared to random forecasts that take only into account the target's global probabilities. Finally, some approaches lack a trading strategy. The following approach handles all of the above points of criticism.

The remainder of this chapter is structured as follows: Section 4.2 introduces the dataset. This is followed by an analysis of response patterns of stocks that can be linked to individual news non-ambiguously (section 4.3). In particular, intraday as well as nightly news are taken into account. Section 4.4 examines the transformation of unstructured data. In section 4.5 various Support Vector machines with linear as well as non-linear kernels are trained. The out-of-sample accuracy is compared to a random classification. Moreover, it is reported that a simple strategy yields 40bps per roundtrip on average.

4.2 THE FINANCIAL NEWS CORPUS

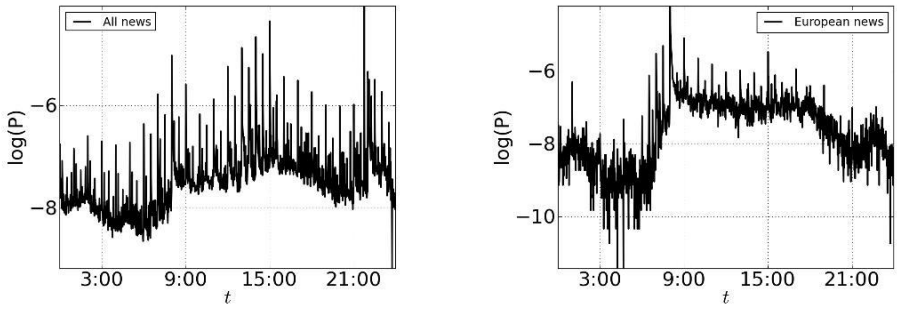
The data consists of 1.4 million realtime news from August 13, 2010 to April 4, 2011 as received via Dow Jones Financial Wire. It covers over 13600 companies listed at global stock markets and includes ad-hoc news, commentaries and analysis as well as economic indicators and political events.

The data is received in a semi-structured way. Thus, the actual information is encoded in a natural language. We receive messages in English. However, some news are tagged with respect to the company they may refer to by reporting the International Securities Identification Number

(ISIN). Moreover, news are tagged with respect to various categories they belong to. For example, some news report corporate actions such as dividends, earnings or stock splits whereas others contain calendars or stock comments.

Within the subsequent analysis we skip all news that cannot be linked to a certain stock non-ambiguously. Thus, all news are discarded that carry information about the economy as a whole or that refer to more than one stock, This affects most acquisitions, for example. Moreover, the analysis is limited to the European market.

Figure 23 depicts the logarithmic probability that a certain news is published over the day. Times refer to the Central European time. One observes peaks where an extraordinary number of news is published. However, it cannot be resolved whether or not this is related to regularly published news such as certain economic events or to the publication policy of the service provider, e. g., for non-time-critical information such as commentaries.



(a) Global coverage

(b) European coverage

Figure 23: Probability that a certain news is published during the day. We report logarithmic probabilities that a news corresponding to global and European coverage is published.

Figure 23a depicts the probability to receive news concerning global coverage. The probability to receive news is relatively low during the night and increases prior to the opening of European markets. It rises around 3pm and briefly after the close of the US market. The former time corresponds to the pre-trade phase of the US market.

Figure 23b exhibits the logarithmic probability that a news concerning a stock primarily listed at an European stock exchange is published. In particular, the rate of news publications increases from roughly 6am to 8:30am prior to the open of the continuous trading. This is followed by a constant news rate during the continuous trading. The publication rate drops again exponentially after the closing.

4.3 RESPONSE PATTERNS

In order to design a text mining algorithm capable of predicting abnormal returns one has to analyse the response of various equities due to the release of new information. To this end we construct the matrices $\mathcal{R}$ and $\mathcal{V}$ with the elements

$$\mathcal{R}_{it} = \frac{y_t^{(i)}}{\hat{\sigma}^i}\sqrt{T}, \quad \mathcal{V}_{it} = |\mathcal{R}_{it}|, \tag{4.1}$$

with T minutes per trading day and $t \in [-\tau, \tau - 1 \ldots, t_0, 1, \ldots \tau]$. Log-returns are calculated for the midprice on 1 minute time horizons as defined by equation (2.3). Moreover, $\hat{\sigma}^i$ is an estimator of the daily volatility of a certain equity at day i. It is calculated as a 14-day exponential moving average of equation (3.3).

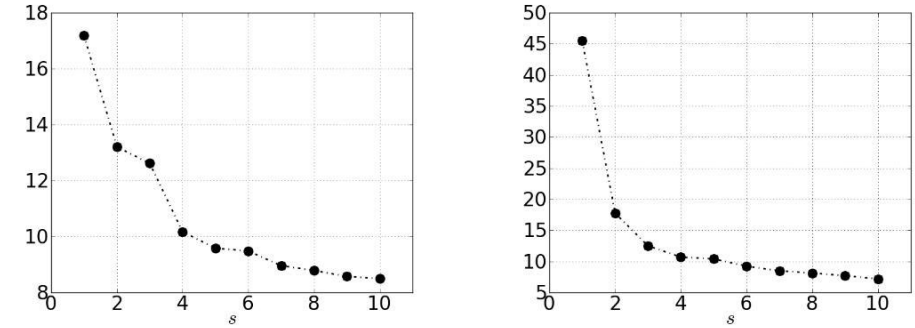
Eventually, rows of $\mathcal{R}$ correspond to the appropriately normalized 1 minute log-returns for the $2\tau + 1$ minutes around a the release of certain news at time t_0. Furthermore, $\mathcal{V}$ is a metric for the response volatility also referred to as the return dispersion.

Concerning the 1.4 million news from the dataset only those are chosen that are published $\tau = 60$ minutes after the opening and before the closing. Additionally, news are selected that can be mapped non-ambiguously to listed stocks. Moreover, a subset of 20000 news is picked randomly. Figure 24 displays the singular value decomposition of matrices $\mathcal{R}$ and $\mathcal{V}$. Right singular vectors are also called the eigenresponse whereas left singular vectors are the eigennews.

The first singular values of $\mathcal{R}$ drop much slower than the singular values of $\mathcal{V}$. In particular the rank-3 approximation of $\mathcal{R}$ accounts for 5.23% whereas the latter accounts for 11.67% of the variation among the data. Eigenresponses of the return are strongly impacted by a few outliers of the respective eigennews. For example, there are 153, 333 and 497 news which deviate more than 3 standard deviations from the mean within the corresponding eigennews profile of the return.

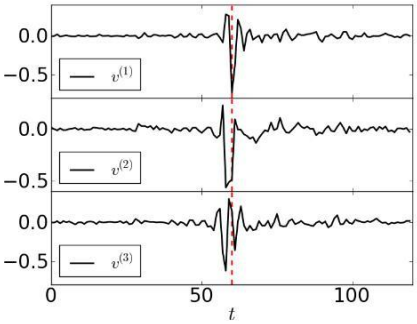
The first corrections to the rank-1 dispersion profile allow to distinguish news that actually impact the return dispersion and those that do not. Components of the third eigennews have both positive and negative components which allows to adjust for the skewness of the response dispersion, i. e., there are news for which the return dispersion is higher after the official release with only a little response prior to the actual release time and vice versa.

The eigenresponse of return and return dispersion as illustrated in figure 24c and 24d indicate that the reaction is strongest within the minute of the release of new information which is in agreement with the efficient market hypothesis. However, concerning some news we have evidence for

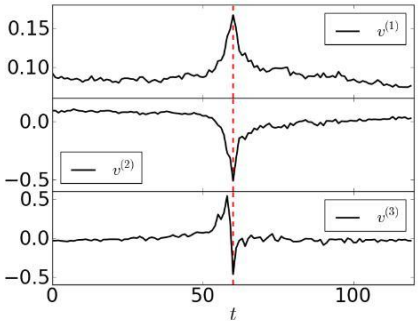


(a) $\Sigma_{\mathcal{R}}$, $s_1/s_2 = 1.23$

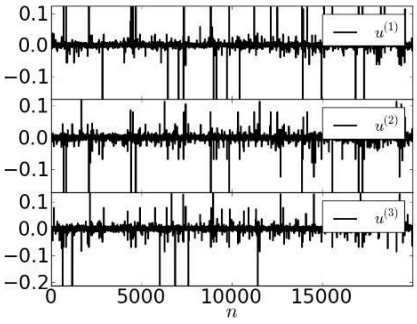
(b) $\Sigma_{\mathcal{V}}$, $s_1/s_2 = 2.55$



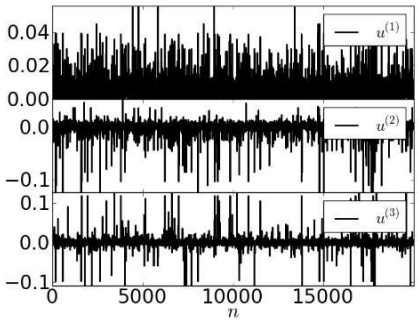
(c) eigenresponse of $\mathcal{R}$



(d) eigenresponse of $\mathcal{V}$



(e) eigennews of $\mathcal{R}$



(f) eigennews of $\mathcal{V}$

Figure 24: Rank-3 approximation of matrices $\mathcal{R}$ and $\mathcal{V}$. The dotted red line corresponds to the release of the corresponding news at time t_0. The response is strongest within the minute of the release. The eigenresponse of the return is strongly influenced by a few outliers whereas the eigenresponse of the dispersion indicates increased volatility for some news preceding its actual release. Volatility corrections due to higher rank approximations account for impact and skewness.

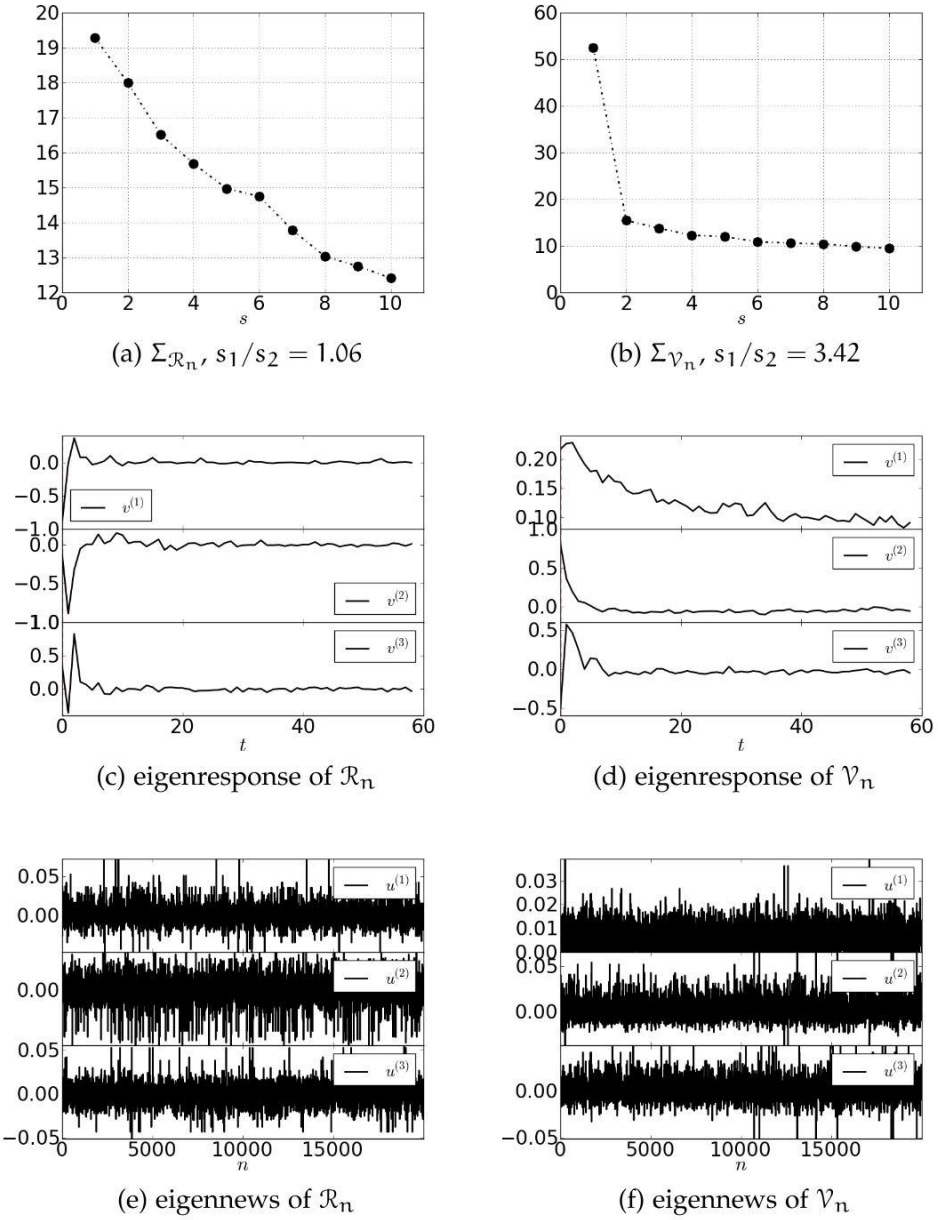
increased return dispersion with respect to minutes following/preceding their release.

For example, the first return dispersion profile increases 10 minutes prior to the release of the news. Even though one has to consider the appropriately weighted superposition of eigenprofiles to obtain the actual response patterns there are news for which the weights of high-order corrections to the rank-1 approximation are negligible. Thus, the existence of the corresponding profile could be linked to the trading of non-public information. Besides the effect cannot be explained due to inaccurate time stamps, as the effect is largest within the minute of the release.

Additionally, the matrices $\mathcal{R}_n$ and $\mathcal{V}_n$ are constructed from news published between the closing and the opening of two succeeding trading days. In this context one refers to them as nightly news. Again 20000 news are picked randomly. A natural choice for the respective time series selects the τ minutes following the opening of the continuous trading, i. e., minutes are discarded that precede the release of the news which had to be mapped to the previous trading day. Figure 25 depicts the singular value decomposition of $\mathcal{R}_n$ and $\mathcal{V}_n$.

The rank-3 approximation accounts for 9.52% and 17.84% of the variation among the data. Complications arise due to the fact that responses no longer can be mapped to the information content of single news but are rather a measure of the collective information accumulated over the given period. In particular, the rank-1 approximation of the return dispersion accounts for 11.46% of the data. Singular values of eigenreturns drop much slower compared to singular values of the eigendisperson. However, the rank-1 approximation of $\mathcal{V}_n$ depicts the typical mean intraday profile as discussed in section 3.3, i. e., the intraday volatility drops in the 30 minutes after the opening. Moreover, higher corrections to the intraday profile can be associated with the collective news published about a certain equity over the night. The corresponding response decays much faster. For example, first-order corrections to the eigendispersion can be linked to increased volatility within the first 5 minutes of the trading day. As eigennews of the second corrections are distributed around zero the corresponding dispersion profile accounts for two aspects. Either the price adjustment due to the news is immediate implying that volatility drops sharply or the adjustment due to overnight news is delayed.

We stress that we have evidence for volatility adjustments due to news only for the minutes around the release. Thus, any algorithms that aims to predict price movements has to take into account the corresponding time scales.



(a) $\Sigma_{\mathcal{R}_n}$, $s_1/s_2 = 1.06$

(b) $\Sigma_{\mathcal{V}_n}$, $s_1/s_2 = 3.42$

(c) eigenresponse of $\mathcal{R}_n$

(d) eigenresponse of $\mathcal{V}_n$

(e) eigennews of $\mathcal{R}_n$

(f) eigennews of $\mathcal{V}_n$

Figure 25: SVD of $\mathcal{R}_n$ and $\mathcal{V}_n$. News are published between the closing and opening of two succeeding trading days. Minute 0 corresponds to the opening of the following trading day. Concerning the return dispersion the rank-1 approximation accounts for the well-known mean intraday-profile following the minutes after the opening. High-order corrections measure the corrections due to overnight news related to an individual stock.

4.4 INFORMATION RETRIEVAL FROM TEXTUAL DATA

The problem to find a suitable representation of unstructured data such as large corpora of text encoded in some natural language is common to many

areas including computer science, mathematics, linguistics, psychology and economics. A very popular approach reduces the textual information of a document to a vector of real numbers. To this end a document is tokenized into single words which will be discussed in section 4.4.1. Further pre-processing steps include the transformation to word stems as described in section 4.4.2.

The assumption that the order of the words within a document can be neglected is also called *bag-of-words* hypothesis and induces exchangeability of the corresponding random variables, i. e., words are assumed to be conditionally IID. Of course removing the ordering of words destroys relevant information. However, one of the major problems involving corpora of documents is that the respective representation even with respect to the *bag-of-words* model is very sparse.

There are a number of different approaches to measure the relevance of an individual word. One of the most common approaches is the so-called term frequency-inverse document frequency (tf-idf) measure that can be motivated on empirical and information theoretic grounds [1]. Other measures include simple term frequencies [59], signal-to-noise ratios [83] or relevance weighted metrics [80].

Given a corpus of N documents $D = \{d_1, \ldots, d_N\}$, a vocabulary $W = \{w_1, \ldots, w_M\}$ of distinct terms contained in D and the frequency f_{ij} that a word w_i occurs in document d_j the tf-idf measure is defined as

$$h_{ij} = f_{ij} \log\left(\frac{N}{N_i}\right), \tag{4.2}$$

with N_i the number of documents that contain the word w_i There exist variants that normalize the term frequency by the total number of words in a document and, thus emphasize the importance of single words in shorter documents. The term related to the inverse document frequency can be interpreted as an information measure, i. e., the logarithm of an inverse probability. The use of equation (4.2) may also be motivated by the fact that the simple term-frequency overemphasizes words that occur very frequently but contain little information. However, measures that use only the entropy over documents prioritizes terms that occur with a very low frequency. The tf-idf measure, on the other hand, is a trade-off between the two extremes. The above statement is formalized following the motivation originally given by Aizawa [1]. Given random variables $\mathcal{W}$ and $\mathcal{D}$ over D and W, the expected mutual information between $\mathcal{W}$ and $\mathcal{D}$ is calculated. Thereby, it is found how well a document is described by words w_i. The self-entropy of $\mathcal{D}$ is given by

$$\mathcal{H}(\mathcal{D}) = -\sum_{d_j \in D} P(d_j) \log(d_j) = -N\frac{1}{N}\log\frac{1}{N} = -\log\frac{1}{N} \tag{4.3}$$

where it is assumed that it is equally likely to select any of the documents d_j. The self-entropy of a subset $\mathcal{D}$ with N_i documents that contain the term w_i reads

$$\mathcal{H}(\mathcal{D}|w_i) = -\sum_{d_j \in D} P(d_j|w_i) \log(d_j|w_i) = -\log \frac{1}{N_i}. \tag{4.4}$$

Given the frequency $f_{w_i} = \sum_{d_j \in D} f_{ij}$ that the term occurs in the whole corpus D and the total frequency $F = \sum_{w_i \in W} f_{w_i}$ of all terms in the corpus the expected mutual information is

$$\begin{aligned}\mathcal{H}(\mathcal{D}, \mathcal{W}) &= \sum_{w_i \in W} P(w_i) \left(\mathcal{H}(\mathcal{D}) - \mathcal{H}(\mathcal{D}|w_i)\right) \\ &= \sum_{w_i \in W} \frac{f_{w_i}}{F} \log \frac{N}{N_i} = \sum_{w_i \in W} \sum_{d_j \in D} \frac{f_{ij}}{F} \log \frac{N}{N_i} \\ &= \frac{1}{F} \sum_{w_i \in W} \sum_{d_j \in D} h_{ij}. \end{aligned} \tag{4.5}$$

Thus, the expected mutual information $\mathcal{H}(\mathcal{D}, \mathcal{W})$ is the sum over the tf-idf measure. This leads to the following interpretation: the inverse document frequency measures the change of information observing the term w_i whereas the word frequency term estimates the probability to do so.

4.4.1 *Tokenization*

Tokenization refers to the segmentation of texts into subunits such as words, phrases or sentences. More generally, a token is a sequence of characters that carries information.

While tokenization with respect to white spaces and related control characters as well as the removal of all punctuations might appear to be a good idea a closer look reveals its difficulties.

For example, considering Asian languages such as Chinese or Korean no whitespace boundaries between words or sentences may exist [64]. Other languages introduce different subtleties, e. g., the German language uses compound words that have to be split up whereas the French language introduces reduced definite articles for words starting with a vowel.

Having established that algorithms the tokenization of text are strongly dependant on the language to be processed we focus on English texts.

The English language makes use of apostrophes in order to contradict or to indicate possession [64]. For example, within the sentence

```
Bernanke's first name isn't Benjamin.
```

tokenization could be done with respect to whitespaces and apostrophes. While this may work out for *Benjamin | s* it turns out to be a bad idea for *isn | t*. Certainly, a better choice would be *is | n't* or even *is | not*.

Moreover, hyphenation is used in miscellaneous ways, e. g., prefixes and suffixes, splitting up vowels like *anti-intellectual*, word grouping *hundred-year-old* or just compound names *Anheuser-Busch*. These cases have to be handled differently, i. e., for prefixes one would remove hyphenations, word groupings might be split up and compound names just stay the way they are.

Additionally, tokenization with respect to whitespaces or non-alphanumeric characters is complicated by names, dates and other cases where a certain collocation of non-alphanumeric characters carries information that would be reduced by a wrong tokenization. For instance, it is preferable to group *New York* together as well as *Aug 27, 2011* and *someone@somewhere.com*.

The proposed approach uses a two-step procedure. First text documents are split into distinct sentences using the the unsupervised multilingual sentence boundary detection algorithm due to Kiss and Strunk [50], i. e., much uncertainty about the boundary detection of sentences is resolved if abbreviations within the sentence are identified reliably. To this end three criteria are introduced that are independent of the respective language. Firstly, abbreviations may be truncated words that end with a period. Secondly, abbreviations are short. And finally, abbreviations may contain internal periods. Moreover, it is important to identify initials and ordinal numbers. Using the above heuristics the sentence boundary detections outperforms other algorithms proposed in the literature.

After the identification of the sentence boundaries the Treebank tokenizer[1] is applied.

In some applications it is common to remove a number of terms that carry little information with respect to the target. Usually, one refers to these terms as stopwords. Stopwords can be characterized by their frequency of occurrence over the corpus of documents, i. e., common stopwords like *a*, *have* or *the* lie at the upper end of the spectrum.

However, we note that within this framework the importance of common stopwords is decreased due to the use of weights that measure the information gain.

The two-step procedure is exemplified with respect to the following sentence[2]

```
The bank postponed its IPO in November because of market volatility,
particularly in financials, at the time. The company had planned
to sell 15 million shares at $14 to $16 each.
```

1 The heuristic is described in http://www.cis.upenn.edu/~treebank/tokenization.html

2 The example sentence is quoted from the dataset.

leading to the tokenization as given by table 13.

the	bank	postponed	its	ipo	in	november
because	of	market	volatility	,	particularly	in
financials	,	at	the	time	.	the
company	had	planned	to	sell	15	million
shares	at	$14	to	$16	each	.

Table 13: Example of word tokenization using the Treebank algorithm. Note how monetary units are grouped together.

4.4.2 *Stemming*

Stemming refers to the removal of suffices from words using an iterative procedure. In particular, stemming maps words of a similar meaning to the same term, not necessarily being the morphological stem of the word. Thus, its use reduces the size of the vocabulary while increasing the frequency of individual terms. The approach generally improves the performance of an information retrieval (IR) task.

In many languages different forms of a word exist depending on its grammatical use. For example, the English verb *to be* might come along as *am*, *was* or *is*. Stemming builds on a number of heuristics that removes endings of a word such as suffixes. However, unlike the so-caleld lemmatization it does not try to find the morphological stem of a word, i. e., the stemming of the word *saw* might lead to *s* whereas its lemmatization would produce *see* or *saw* depending on its use as a verb or as a noun [64].

Stemming destroys a certain amount of information about the context of the word. However, a major problem of text-mining is the statistical weight related to features that are used in a machine learning algorithm, i. e., most words occur only once or twice. Stemming tries to enhance the statistical weight of certain terms.

Within this approach Snowball [75] is used, an open-source software package with implementations of a number of stemming algorithms for different languages. Additionally it presents a language to implement customized stemming algorithms. We use the well known Porter algorithm [74].

Any word may be written as a sequence of consonants and vowels with the following ordering

$$[C](VC)^m[V] \tag{4.6}$$

with the measure $m \in \mathbb{N}_0$, and an arbitrary consecution of consonants C and vowels V. Rules of the stemming algorithm are stated in the form

$$(\text{condition}) : \; S_1 \to S_2, \tag{4.7}$$

where a suffix S_1 is replaced by S_2 if a certain condition involving, for example, the measure m is met. While this approach is relatively easy a large number of explicit rules are formulated that we do not repeat here. However, their character is illustrated by a few examples

$$(\;) : \; \text{IES} \to \text{I} \tag{4.8}$$

$$(m > 0) : \; \text{EED} \to \text{EE} \tag{4.9}$$

$$(*v*) : \; \text{ING} \to \tag{4.10}$$

where the empty condition $(\;)$ is always true and $(*v*)$ indicates that the stem includes vowels.

the	bank	postpon	it	ipo	in	novemb	becaus
of	market	volatil	,	particular	in	financi	,
at	the	time	.	the	compani	had	plan
to	sell	15	million	share	at	$14	to
$16	each	.					

Table 14: Stemming of the example sentence. Terms such as *postpon* or *compani* are no morphological stems.

The result of an application of the full Porter stemming algorithm to the example sentence is summarized in table 14.

4.4.3 *Feature Selection*

The complexity of the approach is justified on the basis that the feature space is automatically compiled trying to preserve as much information about the original document as possible while acknowledging a certain amount of statistical weight. Unlike other approaches we neither use handcrafted dictionaries nor tuples of words of phrases.

In order to analyse word distributions in financial news we choose 40205 messages that can be linked non-ambiguously to European stocks. In total the news contain roughly 7.2 million tokens with a vocabulary size of 124.676 different terms.

For demonstrative purposes we removed the list of stopwords that are stated in table 15. Some of these are classical stopwords that occur very

frequently while carrying no information. Others beglong to the footing of the respective news provider. However, this was only done for illustration purposes with respect to term frequencies. In particular, the application of statistical learning algorithms is done with the full vocabulary.

The empirical word frequency diagram indicates that 48% of the words occur only once or twice over the text corpus. The probability that words occur exactly 3 times is already smaller than 10%.

-	have	for	that	on	2010	is	are	2011
has	it	or	as	from	at	by	this	compani
dow	jone	an	the	-	newsplus	click	newswir	plc

Table 15: An demonstrative list of stopwords that are removed in order to illustrate he distribution of word frequencies among financial text news. The list contains classical stopwords such as *the* as well as terms belonging to the news footer added by the service provider.

Moreover, figure 26 indicates the most frequent words. Disregarding stopwords the most frequent word is *said* with a probability of 0.008. This is followed by words like *market*, *rate* and *busi*. This justifies the prior notion on tuples of words. For example, an order of magnitude estimation under the assumption of conditional independent word occurrences yields a lower bound of $6.4 \cdot 10^{-5}$ for the probability to observe a certain pair of words .

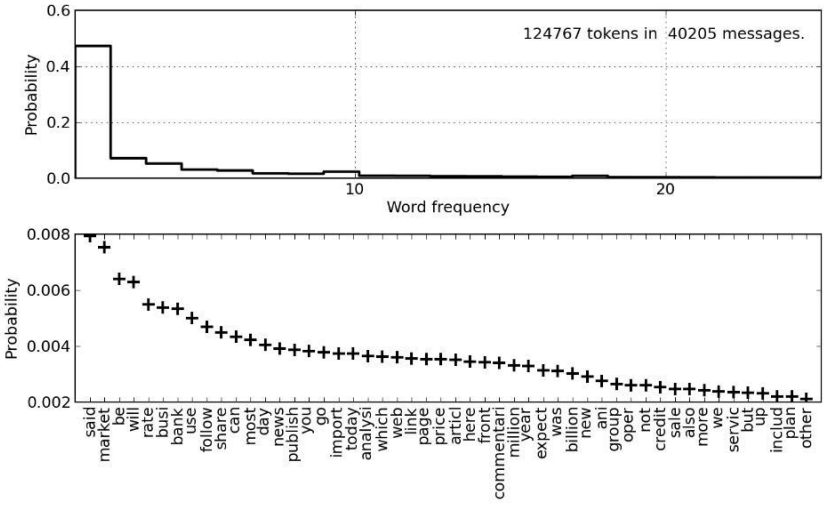


Figure 26: Probabilites that a term occurs n times as well the most frequent individual terms. The 40.205 news contained 124.767 different words with a total of 7.208.220 terms.

In order to analyse the overall discrimination capacity of various terms we transform the corpus in a way that represents news as a vector with entries given by the tf-idf measure. The F-score of the training set is calculated. Moreover, the analysis is limited to the 50000 most frequent terms as measured by their occurrence over the corpus. A 1−0 measure is applied, i. e., if a word occurs in a given document it is counted only once.

Thus given the feature vector x_j of the j-th training example and n_k the total number of training-instances in class k the F-score of the i-th feature reads [21]

$$F_i = \frac{\sum_{k=1}^{l} \left(\hat{x}_i^k - \hat{x}_i\right)^2}{\sum_{k=1}^{l} \frac{1}{n_k - 1} \sum_{j=1}^{n_k} \left(x_{j,i}^k - \hat{x}_i^+\right)^2} \tag{4.11}$$

with l the number of classes, the average of averages $\hat{x}_i$ and $\hat{x}_i^k$ of the respective dataset. Thus, the nominator is the inter-class variance whereas the denominator sums the variances within each class. However, it is well-known that the F-score is not capable to measure mutual information among various features.

Figure 27 depicts the logarithmic, non-zero F-score of individual features where targets are calculated with respect to normalized future 1, 5 and 10 minute log returns y_t. The class map is given by

$$g(y_t) = \begin{cases} y_t < -1 \\ |y_t| \leqslant 1 \\ y_t > 1 \end{cases} . \tag{4.12}$$

The plot is truncated for features that show a vanishing F-score.

The F-score drops super-exponentially for the approximately 5000 features with the highest F-score. This is followed by a region that is linear with respect to $\log F_i$. Only for features beyond the first 30000 their F-score drops again super-exponentially. Table 16 lists the first 72 words that correspond to the features with the highest F-score. It is read in a line by line fashion, i. e., *commentari* has the highest F-score, followed by *announc* and *would*.

However, while it is clear that the the importance of features in terms of their F-score drops faster than exponentially for the first 5000 features giving way to an exponential behaviour for the following 25000, it is not obvious how to gauge the cut for the selection of features. Thus, various formulations of Support Vector Machines are trained using a different number of features. Finally, the version with the lowest out-of-sample error is chosen.

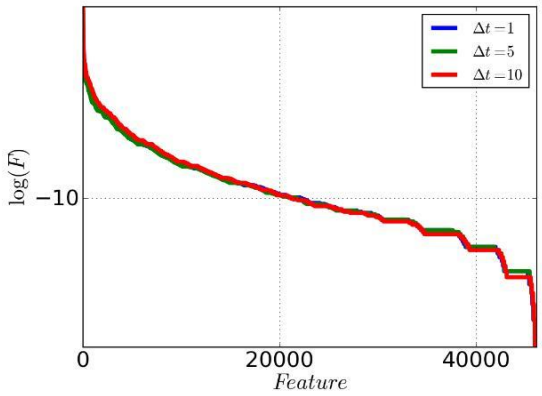


Figure 27: The logarithmic and non-zero F-score of individual features ordered by magnitude drops faster then exponentially for the first 5000 features which is followed by exponential behaviour. The F-score drops again super-exponentially after the first 30000 features.

commentari	announc	would	command	estim	analyst
conquer	trend	can	minor	invest	billion
befor	today	analysi	futur	fix	news
import	busi	follow	most	use	mild
repair	will	bank	reclam	mode	non-stat
expect	more	new	equalis	pricing	million
work	target	perform	inform	addict	valu
buy	smart	howev	concern	fun	down
control	focus	ratio	technolog	trade	rate

Table 16: Terms with highest F-score. Notice that the ordering is given in a line by line fashion.

4.5 SUPPORT VECTOR MACHINES AND FINANCIAL TEXT MINING

Having established the pre-processing steps that lead to a feature representation which can be used by standard machine learning algorithms and the relevant timescales for which news in the dataset have impact on the time series of the return various formulations of Support Vector Machines are applied. Targets are log returns of midprices across several

minute time horizons that are standardized by a daily estimator of the corresponding stock's volatility, i. e.,

$$y_t = \log\left(\frac{s_t}{s_{t-1}}\right)\frac{\sqrt{T}}{\hat{\sigma}} \tag{4.13}$$

with T being the intraday time interval and the estimator of the daily volatility calculated as a 14 day exponential average $\hat{\sigma} = \mathrm{EMA}_{14}(\omega)$ where

$$\omega = \sup\left\{|\alpha-\beta| : \ \alpha,\beta \in \{\mathrm{open}^{\pm},\mathrm{close}^{\pm},\mathrm{high}^{\pm},\mathrm{low}^{\pm}\}\right\}, \tag{4.14}$$

and $\pm$ indicates succeeding trading days.

We use the 1, 5 and 10 minute returns following the release of the news, i. e., for news published in the interval $t_0 \in [\tau-1,\tau)$ returns are calculated for the interval $[\tau,\tau+\Delta t]$. Moreover, a 3-class approach is pursued, i. e., class labels are assigned according to equation (4.12). Thus, news having no impact larger than one standard deviation are discriminated from those leading to a sharp increase or decrease of the future log return.

Three Support Vector Machines are trained, i. e., a one-against-one ansatz is used [52]. The final class label is assigned using a majority vote. Inputs are rescaled such that they lie in $[-1,1]$ which keeps features with a larger numerical range from being overemphasized. Moreover, this prevents numerical instabilities in the calculation of kernels [19].

Model selection, i. e., the adjustment of free parameters, such as C, is done using a 5-fold cross validation. The model with the lowest cross-validation error is chosen. To this end a simple grid-search is applied. Only after this the corresponding machine learning algorithm is trained using 2/3 of the chronologically ordered news dataset. The performance of the Support Vector Machine is measured using the remaining news. With respect to classification the out-of-sample accuracy is given by the number of right classifications over the number of testing examples.

4.5.1 *Linear and Nonlinear Classification*

We start training a linear L_2-regularized L_2-Support Vector Classifiers using LIBLINEAR [34] that solves the following primal problem

$$\min_w \frac{1}{2}w^T w + C\sum_{i=1}^{n}\left(\max(0,1-y_i w^T x_i)\right)^2. \tag{4.15}$$

The choice of linear kernels is based on the notion that the feature dimension is very large. Thus, a nonlinear kernel does not improve the out-of-sample error significantly. Figure 28 illustrates the out-of-sample testing accuracy for various dimensions of the feature space and various

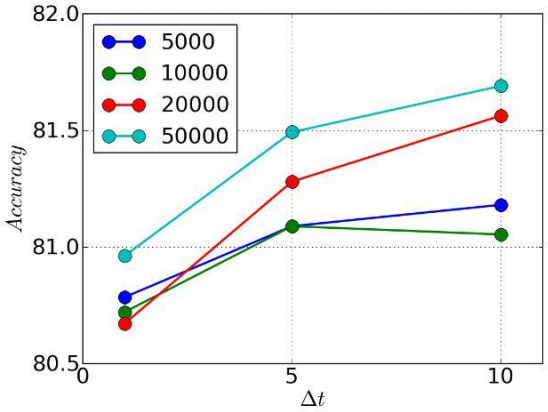


Figure 28: Out-of-sample accuracy of L_2-regularized L_2-Support Vector Classifiers using a different number features and forecasting horizons

forecast horizons. Best results corresponding to a cross-validation accuracy of 85.14% are obtained for 10 minute time horizons and 50000 features.

The out-of-sample classification accuracy improves for larger input dimensions, i. e., we achieve best results for the 50000 most frequent words in the news corpus. Moreover, the accuracy is better for the 10 minute interval than for the 1 minute interval.

While it is certainly true that without using any information whatsoever the probability to assign the right label would be 33%, it is mandatory to to consider the inclusive probabilities as given by the global target distribution in order to decide whether the learning algorithm is capable of individualizing its forecasts with respect to the information structure encoded in the news.

In particular, concerning the 10 minute interval the inclusive probability for $|y_t| <$ is 84.63%, for $y_t < -1$ it is 7.46% and for $y_t > 1$ it is 7.91%. Performing a Monte Carlo simulation with $n = 10000$ repetitions the classification accuracy itemized by its cost matrix is stated in table 29.

Truth/Prediction	$\|y_t\| < 1$	$y_t < -1$	$y_t > 1$
$\|y_t\| < 1$	0.7090 ± 0.0016	0.0663 ± 0.0013	0.0668 ± 0.0012
$y_t < -1$	0.0662 ± 0.0005	0.0062 ± 0.0004	0.0062 ± 0.0004
$y_t > 1$	0.0668 ± 0.0005	0.0062 ± 0.0004	0.0063 ± 0.0004

Table 17: Classification accuracy if class labels are assigned according to the probabilities induced by the global target distribution. Results are obtained using a Monte Carlo simulation with $n = 10000$ repetitions.

Thus, the total classification accuracy is 72% given by the trace of the cost matrix. Moreover, the probability to classifiy news with large impacts correctly, i. e., $|y_t| > 1$ is 50%.

This is compared with the cost matrix of the Support Vector Classification using 10 minute log returns. The result is stated in table 18.

Truth/Prediction	$\|y_t\| < 1$	$y_t < -1$	$y_t > 1$
$\|y_t\| < 1$	0.8120	0.0116	0.0116
$y_t < -1$	0.0780	0.0026	0.0010
$y_t > 1$	0.0795	0.0015	0.0023

Table 18: Classification accuracy for the linear L_2-regularized L_2-SVC on 10 minute forecast horizons. The total accuracy is 81.69%. The accuracy of forecasts $|y_t| > 1$ if the true returns also have large impact is 66.67%.

The total accuracy increases to 81.69%. Moreover, it is striking that the accuracy to correctly predict $|y_t| > 1$ if the news actually belongs to one of the two classes is 66.67%.

In order to verify the introductory statement that the accuracy of linear kernel Support Vector Classification approaches the accuracy of nonlinear kernels as the dimension of the feature space increases, nonlinear SVCs are trained. In particular, the open-source software libSVM [19] is used, i. e., we apply the formulation discussed in section 1.6.3 using radial kernel functions

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|_2) \tag{4.16}$$

on 1 and 5 minute forecast horizons.

Figure 29 depicts the grid-search in the cross-validation phase in order to obtain optimal parameters C and γ with respect to the out-of-sample error.

Cross-validation accuracies of 84.37% and 84.55% are obtained, respectively. Compared to 84.22% and 84.29% for the linear L2-regularized L_2-SVC on 1 and 5 minute time horizons no substantial improvement is reported which supports our introductory statement.

As a result, a simple strategy is proposed. Given a trading signal $|y_t| > 1$ the strategy goes long or short at the best bid and ask in the minute following the publication of a certain news. Moreover, this new position is evened up at market price at the end of the forecasting horizon. Constant volumes and no trading costs (both implicit and explicit) are assumed. For the 14124 news in the test sample 466 trading signals for the 1 minute time horizon, 477 for the 5 minute horizon and 432 for the 10 minute horizon are generated.

Finally, a profit of 41.9bps, 17.5bps and 46.3bps per roundtrip can be reported for the respective forecasting horizons.

4.6 CONCLUSION

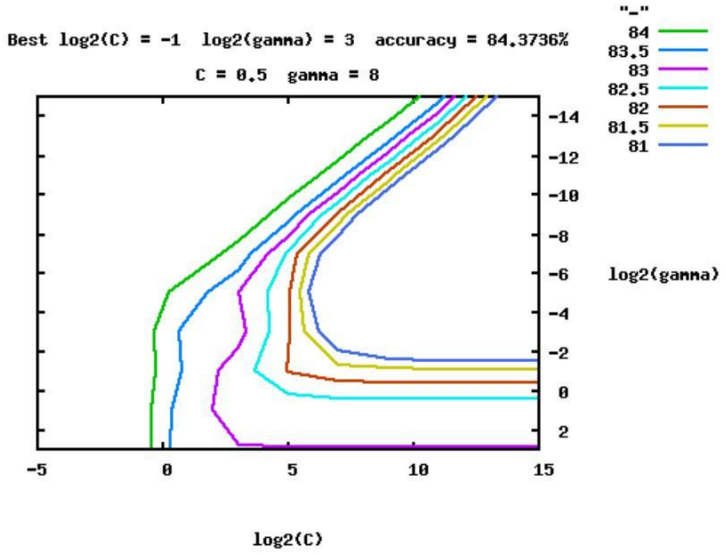
This chapter introduced a coherent approach to financial text mining. In order to quantify relevant response patterns due to the release of news that are non-ambiguously linked to certain equities a singular-value decomposition of return and return dispersion matrices is performed. It is found that for news published during the continuous trading in Europe time series of return dispersion are impacted for the minutes around the release with the strongest reaction immediately at the publication time. However, for certain news an increased return dispersion as measured by the second and third right singular vectors is observed both for the minutes prior as well as after the release. Higher order profiles for log return time series that can be mapped to equities for which news were published during the respective non-trading period exhibit a similar pattern. However, response patterns can no longer be mapped to the release of single news but are rather a measure of the accumulated information over the non-trading period.

Having established that response patterns are relevant only for the minutes around the release of company-specific news an approach is introduced to map textual information into a representation that can be used by standard machine learning algorithms. To this end the tf-idf measure is motivated on information theoretic grounds following Aizawa [1]. Moreover, technical aspects such as the the tokenization of words and word-stemming are relevant for an optimal feature representation. The former problem is efficiently tackled applying a two-step procedure using the unsupervised multilingual sentence boundary detection due to Kiss and Strunk [50] which is followed by the Treebank word tokenization. Stemming is performed using the well known Porter algorithm [74].

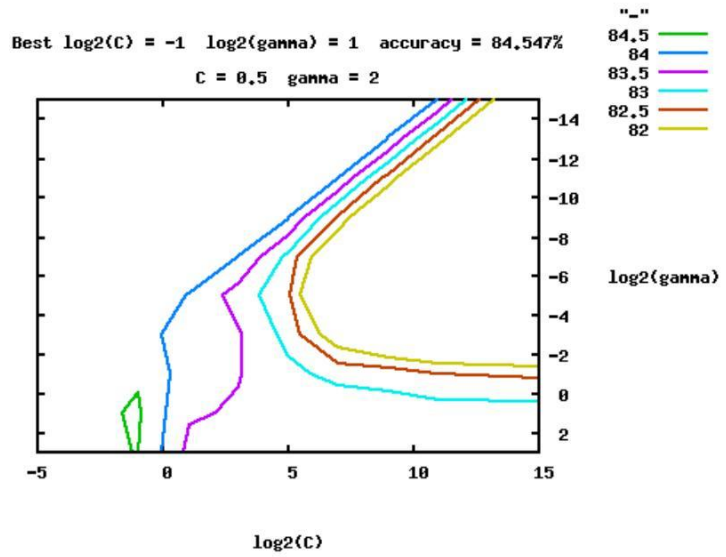
Given the resulting feature representation the *bag-of-words* model is justified by an analysis of the probabilities to observe single terms. Moreover, the capactity to discriminate with respect to categorical targets is analysed introducing the F-score of single features. A list of words with the highest F-score is stated explicitly.

Various formulations of Support Vector Machines are trained. It is found that the out-of-sample accuracy of variants using linear kernels approaches the accuracy of non-linear kernels due to the high-dimensional feature space. Additionally, the accuracy increases for 10 minute time horizons is compared with 1 minute time horizons.

Finally, a simple strategy is proposed that enters positions at constant volumes omitting any transaction costs. A performance of 41.9bps, 17.5bps and 46.3bps per roundtrip given the respective time horizons can be reported.



(a) $\Delta t = 1$



(b) $\Delta t = 5$

Figure 29: Grid search of the parameters C and γ in training a Support Vector Classifier with radial kernels on 5 and 10 minute time horizons of future log returns. The accuracy is given by the 5-fold cross-validation error.

Part IV

APPENDIX

A

PROBABILITY DISTRIBUTIONS

A.1 THE NORMAL DISTRIBUTION

The multivariate Gaussian or normal distribution $\mathcal{N}(\mu, \Sigma)$ of a random variable X is given by

$$f_N(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{A.1}$$

with $\mu \in \mathbb{R}^n$ and a positive semidefinite matrix $\Sigma \in \mathbb{R}^{n\times n}$. We call μ the location parameter and Σ the covariance matrix. The univariate density function with mean μ and variance $\Sigma = \sigma$ reduces to

$$\mathcal{N}(\mu, \sigma) \equiv f_X(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{A.2}$$

Various location parameters such as median, mean and modus coincide. It is the only distribution for which vanishing correlation of two random variables implies independence. The normal distribution belongs to the class of exponential distributions

$$f_X(x|\Theta) = h(x)e^{\eta^T(\Theta)T(x)-A(\Theta)} \tag{A.3}$$

with $h(x) = 1/\sqrt{2\pi}$, the natural parameter $\eta = (\mu/\sigma^2, -1/(2\sigma^2))^T$, the sufficient statistic $T_\sigma(x) = (x, x^2)^T$ and the partition function $A(\eta) = -\eta_1^2/4\eta_2 + 1/2\ln|1/2\eta_2|$.

Its characteristic function is given by

$$\phi_X(t|\mu\sigma) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2} \tag{A.4}$$

which implies that all centralized moments exist. Moreover, Gaussian random variables are closed under linear transformations, i. e., for independent and Gaussian $\{X_i\}$ it holds

$$\alpha_0 + \sum_i \alpha_i X_i \stackrel{d}{=} \mathcal{N}(\tilde{\mu}, \tilde{\sigma}), \tag{A.5}$$

with $\tilde{\mu} = \alpha_0 + \sum_i \alpha_i\mu_i$ and $\tilde{\sigma} = \sum_i \alpha_i^2\sigma_i^2$. The reverse is also true. More generally, the normal distribution is infinite divisible and stable with $\alpha = 2$.

A.2 STUDENT-t DISTRIBUTION

The Student's t-distribution exhibits asymptotically Pareto-like tails. Its density distribution is given by

$$f_X(x|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{2}\right)^{-\frac{\nu+1}{2}} \tag{A.6}$$

with the Gamma function $\Gamma(\cdot)$ and the number of degrees of freedom ν. As $\nu \to \infty$ the Student t-distribution approaches the normal distribution. Moreover, $\nu = 1$ is the standard Cauchy distribution. The Student's t-distribution falls into the class of generalised hyperbolic distributions.

Its characteristic function is given by

$$\phi_X(t|\nu) = \frac{K_{\frac{1}{2}\nu}(\sqrt{\nu}|t|)(\sqrt{\nu}|t|)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)2^{\frac{1}{2}\nu-1}}, \tag{A.7}$$

with the modified Bessel function of second kind $K_\lambda(\cdot)$. Thus, moments of the distribution are defined for all $k \leqslant \nu$. All odd orders vanish, in particular, its expectation value and skewness. All even orders are given by

$$\mathbb{E}[X^k] = \prod_{i=1}^{k/2}\frac{2i-1}{\nu-2i}\nu^{k/2}. \tag{A.8}$$

The excess kurtosis is $\gamma_1 = 6(\nu-4)^{-1}$ assuming it exists.

Three parameter versions with a location and scale exist due to the compounding with a normal distribution.

A.3 CLASSICAL TEMPERED STABLE DISTRIBUTION

Classical tempered stable (CTS) distributions are also known as truncated Lévy flights [54] and the CGMY model [18]. Roskiński [81] generalized them to the class of tempered stable distributions. The idea is to temper the Lévy measure of a stable distribution using an exponentially decaying function such that all moments exist.

Thus, the CTS distribution is given by the Lévy triplet $(0, \nu, \gamma)$ with

$$\nu(dx) = (C_+e^{-\lambda_+ x}\mathbb{1}_{x>0} + C_-e^{-\lambda_+|x|}\mathbb{1}_{x<0})\frac{dx}{|x|^{\alpha+1}} \tag{A.9}$$

$$\gamma = \mu - \int_{|x|>1} x\nu(dx) \tag{A.10}$$

and $\gamma \in \mathbb{R}$, $C_\pm, \lambda_\pm > 0$ and $\alpha \in (0,2)$, $\alpha \neq 1$. From the existence of a Lévy triplet it follows that the distribution is infinite divisible.

The characteristic function of a CTS distributed random variable $X \sim \mathrm{CTS}(\alpha, C_\pm, \lambda_\pm, m)$ is given by

$$\begin{aligned}\phi_{\mathrm{CTS}}(t) = \exp(itm + C_+\Gamma(-\alpha)((\lambda_+ - it)^\alpha - \lambda_+^\alpha) \\ + C_-\Gamma(-\alpha)((\lambda_- + it)^\alpha - \lambda_-^\alpha)),\end{aligned} \tag{A.11}$$

with $m \in \mathbb{R}$. Equation (A.11) can analytically be continued $\{t \to z \in \mathbb{C} : -\lambda_- \leqslant \mathrm{Im}(z) \leqslant \lambda_+\}$ such that exponential moments exist. In partciular, the CGMY distribution is obtained for $C_+ = C_-$ whereas truncated Lévy flights are recovered for $\lambda_+ = \lambda_-$ [47].

Moreover, the CTS distribution with zero mean and unit variance called standard CTS distribution is given by

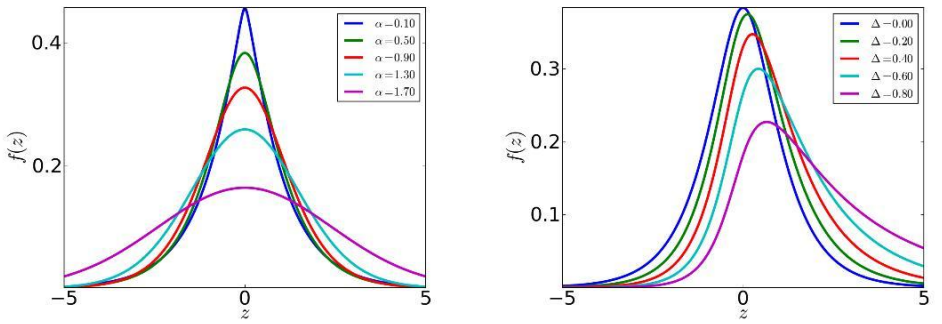
$$m = -\Gamma(1-\alpha)(C_+\lambda_+^{\alpha-1} - C_-\lambda_-^{\alpha-1}) \tag{A.12}$$

$$C = C_\pm = \frac{1}{\Gamma(2-\alpha)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})}. \tag{A.13}$$

Concerning the tail behaviaour Kim et al. [85] state upper and lower bounds by

$$k\frac{e^{-2\bar{\lambda}x}}{\bar{\lambda}x^{\alpha+1}} \leqslant P(|X-m| \leqslant x) \leqslant \frac{K}{x^2} \tag{A.14}$$

as $x \to \infty$, with k and K being not a function of x and $\bar{\lambda} = \min(\lambda_+, \lambda_-)$. Figure 30 depicts the behaviour of the density function of a CTS distributed random variable. In particular, figure 30a illustrates the influence of a varying α on the overall shape of the distribution whereas figure 30b depicts the influence of different $\lambda_\pm$ on the skewness.



(a) Varying α with $\lambda_\pm = 1$ (b) Varying $\lambda_\pm = 1 \pm \Delta$ and $\alpha = 0.5$

Figure 30: CTS density distributions for $m = 0$ and $C_\pm = 1$. Figure 30a illustrates the effect of a varying shape parameter α whereas figure 30b depicts the induced skewness due relative variations in $\lambda_\pm$.

A.4 VARIANCE GAMMA DISTRIBUTION

A Variance Gamma (VG) distribution is a normal distribution where the variance itself is a random variable $\sigma^2 V$ where V is gamma distrbuted and $\sigma^2 \in \mathbb{R}_+$. Thus, the density distribution of V is given by

$$f_V(v|m,\nu) = \left(\frac{m}{\nu}\right)^{\frac{m^2}{\nu}} \frac{v^{\frac{m^2}{\nu}-1}\exp(-\frac{m}{\nu}v)}{\Gamma\left(\frac{m^2}{\nu}\right)},\ v>0, \tag{A.15}$$

with the expected value m. Moreover, we have $m = 1$ as the scale of the corresponding VG distribution has already been factored out as σ^2.

The unconditional density distribution of random variable X that is VG distributed is given by

$$\begin{aligned} f_{VG}(x|\sigma,\Theta,\nu) &= \mathbb{E}_V[\mathcal{N}(\Theta v, \sigma^2 v)|V] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_0^\infty \frac{1}{\sqrt{v}}\exp\left(-\frac{(x-\Theta v)^2}{2\sigma^2 v}\right)\frac{v^{\nu^{-1}-1}\exp(-\frac{v}{\nu})}{\nu^{\nu^{-1}}\Gamma(\nu^{-1})}dv. \end{aligned} \tag{A.16}$$

The skewness of the distribution enters via the parameter Θ whereas ν controls for the kurtosis. In particular, negative Θ induce negative skewness and large ν increase the excess kurtosis. The density can be written in closed-form to read

$$f_{VG}(x|\sigma^2,\Theta,\nu) = \frac{2\exp\left(\frac{\Theta x}{\sigma^2}\right)}{\sqrt{2\pi}\sigma\nu^{1/\nu}\Gamma(1/\nu)}\left(\frac{x^2}{2\sigma^2\nu^{-1}+\Theta^2}\right)^{\frac{1}{2\nu}-\frac{1}{4}} K_{\frac{1}{\nu}-\frac{1}{2}}(C) \tag{A.17}$$

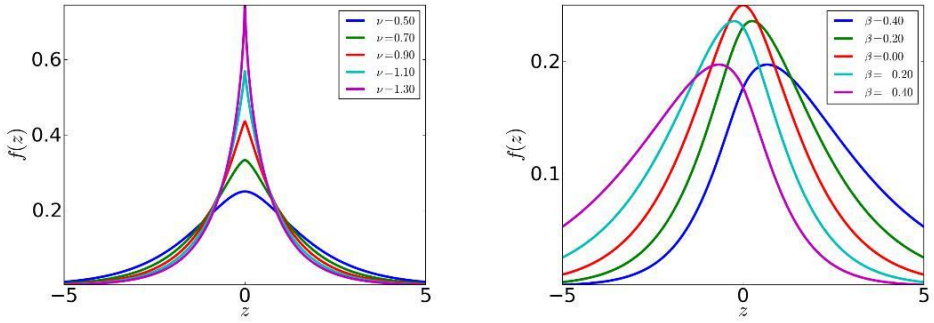
$$C = \frac{1}{\sigma^2}\sqrt{x^2(2\sigma^2/\nu+\Theta^2))},$$

with the modified Bessel functions of seconds kind $K_n(\cdot)$. A symmetric version of equation (A.17) was used by Madan and Seneta [61] to model stock market returns.

Moreover, the measure of the Lévy process with VG distributed increments has three representations in terms of a Brownian motion with random time scales, the difference of two gamma processes and a measure change to take into account relative risk aversion [62]. We state only the first one as it corresponds to our motivation of the distribution. Hence. given the Lévy triplet (σ,ν,Θ) the measure might read

$$k_X(x)dx = \frac{\exp(\Theta x/\sigma^2)}{\nu|x|}\exp\left(-\sqrt{\frac{2}{\nu}+\frac{\Theta^2}{\sigma^2}}\frac{|x|}{\sigma}\right)dx. \tag{A.18}$$

Madan et al. [62] applied the VG process in the context of option pricing such that the Black-Scholes formula is a special case of a more general representation. In particular, parameters to control for kurtosis and skewness allow to take into account empirical facts such as volatility smiles and skewness premia. We note that the VG process is a pure jump process with no martingale component.



(a) Varying ν, $\beta = 0$ (b) Varying β, $\nu = \frac{1}{2}$

Figure 31: VG density distributions for $\mu = 0$ and $\sigma = 1$. Figure 31a illustrates the effect of different parameters ν, i. e., different excess kurtosis, whereas figure 31b illustrates the influence the parameter β on the skewness of the distribution.

The characteristic function is given by

$$\phi_{VG}(t) = \left(1 - i\Theta\nu t + \sigma^2 t^2 \nu/2\right)^{-1/\nu}. \tag{A.19}$$

Thus, the first moments are $\mathbb{E}[X] = \Theta$, $\mathrm{Var}[X] = \Theta\nu + \sigma^2$ and

$$\mathbb{E}[(X - \mathbb{E}[X])^3] = 2\Theta^3\nu^2 + 3\sigma^2\Theta\nu \tag{A.20}$$

$$\mathbb{E}[(X - \mathbb{E}[X])^4] = 3\sigma^4(1+\nu) + 6\sigma^2\Theta^2(2\nu^2 + \nu) + 3\Theta^4\nu^2(2\nu + 1). \tag{A.21}$$

Vanishing Θ implies both vanishing skewness and expected value. Moreover, for a symmetric VG distribution the kurtosis becomes $\kappa = 3(1+\nu)$ such that ν controls the relative excess kurtosis.

We note that equation (A.17) is not capable to control for location and skewness independent of each other. In particular, non-zero location implies skewness. Therefore, we use a slight modification leading to the 4 parameter density distribution

$$f_{VG}(x|\mu,\sigma,\beta,\nu) = \frac{(\sigma^2 - \beta^2)^{1/\nu} e^{\beta(x-\mu)}}{\sqrt{\pi}\Gamma(1/\nu)(2\sigma)^{\frac{1}{\nu}-\frac{1}{2}}} |x-\mu|^{\frac{1}{\nu}-\frac{1}{2}} K_{\frac{1}{\nu}-\frac{1}{2}}(\sigma|x-\mu|), \tag{A.22}$$

where μ is the location, σ the scale, β the skewness and ν amounts to the shape. Expected value and variance are given by

$$\mathbb{E}[X] = \mu + \frac{2\beta}{\nu(\sigma^2 - \beta^2)} \tag{A.23}$$

$$\mathrm{Var}[X] = \frac{2}{\nu(\sigma^2 - \beta^2)} \left(1 + \frac{2\beta^2}{\sigma^2 - \beta^2}\right). \tag{A.24}$$

Figure 31 depicts various VG density distributions. For example, figure 31a exhibits the influence of the shape parameter ν whereas figure 31b illustrates the induced skewness due to β.

Finally, a multivariate generalization can be found under the name generalized, asymmetric Laplace distribution [55].

B

STATISTICAL TESTING

Statistical testing addresses the question if a set of data is compatible with a certain hypothesis. Thus, the null-hypothesis H_0 is tested for statistical significance, i. e., one assigns a probability to the occurrence of the data given H_0 is true. Statistical testing is possible both for the Frequentist and Bayesian approach to statistics, where the latter tests for statistical significance based on the posterior probability.

The typical approach is the following. A null-hypothesis H_0 and one or more alternative-hypotheses H_i, $i > 0$ are defined. With respect to H_0 a test statistic ζ is calculated. Given the confidence interval α and the density distribution $f(\zeta)$ of the random variable ζ, H_0 is rejected at the confidence level α if it is larger than a critical value that separates the region that contains the fraction $1-\alpha$ of the probability mass of $F(\zeta)$.

Eventually, there are two kind of errors. Rejecting H_0 in favor of an alternative hypothesis is called error of first kind with a probability α. Errors of second kind accept H_0 if H_i, $i > 0$ is true. Hereby, probability for errors of seconds kind β is given by the area below $f_{H_i}(\zeta)$ that falls into the acceptance region of H_0.

The remainder of this chapter describe the t, the Kolmogorov-Smirnov, the Anderson-Darling and Kupiec's proportion of failures test as applied within the thesis.

B.1 STUDENT t-TEST

The one-sample t-test compares the sample mean $\bar{x}$ of a normally distributed measurement $\{x_i\}$ to a certain theoretical location parameter μ. The test statistic

$$t = \frac{\bar{x} - \mu}{s\sqrt{n}} \tag{B.1}$$

with sample size n and sample standard deviation s follows a Student-t distribution with $n-1$ degrees of freedom. The corresponding p-value is given by

$$P\left(t \leqslant 1 - \frac{\alpha}{2}\right) = \int_{-\infty}^{1-\alpha/2} f_{n-1}(t')dt' \tag{B.2}$$

and the hypothesis $H_0 : \bar{x} = \mu$ is rejected at the confidence level α if

$$|t| > P\left(t \leqslant 1 - \frac{\alpha}{2}\right). \tag{B.3}$$

The above test describes the two-sided t-test. However. one-sided and two sample as well as multivariate variants exist.

B.2 KOLMOGOROV-SMIRNOV-TEST

The one-sample Kolmogorov-Smirnov (KS) test checks whether an empirical distribution is compatible with a certain theoretical distribution. Kolmogorov proposed to calculate the test statistic as the maximal distance between corresponding quantiles of the empirical and theoretical distribution, i. e.,

$$d_n = \sup_x |F_n(x) - F(x)|, \tag{B.4}$$

with the empirical distribution function $F_n = 1/n \sum_{i=1}^{n} \mathbb{1}_{x_i \leqslant x}$ for n observations x_i. Hereby, $\mathbb{1}$ denotes the indicator function. The hypothesis H_0 that the data follows a distribution is rejected with respect to p-values that are calculated using Kolmogorov's CDF. Values for finite samples are calculated numerically according to [87]. An asymptotic form is given by

$$\lim_{n\to\infty} F(\sqrt{n} d_n \leqslant x) = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} \exp\left(-\frac{(2i-1)^2\pi^2}{8x^2}\right). \tag{B.5}$$

Concerning the ordered set $\{x_i | x_i \leqslant x_j,\ i < j\}$ Smirnov's variant of the KS test calculates the test statistic d_n separately for the lower and upper part of observations.

B.3 ANDERSON-DARLING-DISTANCE

The Anderson-Darling (AD) is another goodness-of-fit test that can be used to compare empirical and theoretical distribution. The test statistic is calculated as the appropriately weighted squared area between the empirical and theoretical distribution, i. e.,

$$A_n = n \int_{-\infty}^{\infty} w(x)(F_n(x) - F(x))^2 dx, \tag{B.6}$$

with the empirical distribution $F_n(x) = 1/n \sum_{i=1}^{n} \mathbb{1}_{x_i \leqslant x}$ and weights that penalise deviations in the tail of the distribution, i. e.,

$$w(x) = \frac{1}{F(x)(1 - F(x))}. \tag{B.7}$$

Inducing the ordered set $\{x_i | x_i \leqslant x_j,\ i < j\}$ the test statistic can be re-expressed exploiting the staircase character of the empirical distribution $F_n(x)$ to read

$$A_n = -n - \sum_{i=1} n \frac{2i-1}{n} \left(\ln\left(F(x_i)\right) + \ln\left(1 - F(x_{n+1-i})\right)\right). \tag{B.8}$$

The hypothesis H_0 that the measurement follows a theoretical distribution is rejected with respect to p-values that are calculated numerically using the method proposed in [66]. Asymptotic values are given by

$$\lim_{n\to\infty} F\left(\sqrt{n}d_n \leqslant x\right) = \frac{\sqrt{2\pi}}{x}\sum_{i=0}^{\infty}\binom{\frac{1}{2}}{i}(4i+1)\exp\left(-\frac{(4i+1)^2\pi^2}{8x}\right)\times \int_0^{\infty} e^{\frac{x}{8(1+\omega^2)}-\frac{\omega^2(4i+1)^2\pi^2}{8x}}\,d\omega. \tag{B.9}$$

B.4 KUPIEC'S PROPORTION OF FAILURES

One of the easiest test to verify if a number of VaR exceptions is compatible with a certain confidence interval α is Kupiec's proportion of failures (POF) test [56]. Thus, given a number of losses k that exceed VaR at a given significance level η and the total number of observations n the null hypothesis

$$H_0 : \hat{p} \equiv \frac{k}{n} = \alpha \tag{B.10}$$

has to be tested. For example, at the 95% confidence interval one expects 5% exceptions with feasible deviations due to the finite sample size. Moreover, the number of exceptions k is a binomial distributed random variable. The test statistic is calculated as a likelihood-ratio

$$\Lambda = -2\ln\left(\frac{(1-\alpha)^{n-k}\alpha^k}{(1-\hat{p})^{n-k}\hat{p}^k}\right). \tag{B.11}$$

Given H_0 is correct Λ is χ^2_1-distributed with one degree of freedom. Thus, one rejects the model if Λ is larger than $f_{\chi^2_1}(\alpha)$.

However, one of the major drawbacks of Kupiec's POF test is the unconditional treatment of exceptions. In particular, changes in volatility and/or correlation might lead to clustered VaR exceptions which indicates that the model cannot capture the new environment and has to be rejected. A Test that takes into account conditional expectations is, for example, Christoffersen's interval forecast test [22].

BIBLIOGRAPHY

[1] A. Aizawa. An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, 39(1):45–65, 2003. doi: 10.1016/S0306-4573(02)00021-3. (Cited on pages 104 and 115.)

[2] Alvaro Almeida, Charles Goodhart, and Richard Payne. The effects of macroeconomic news on high frequency exchange rate behavior. *Journal of Financial and Quantitative Analysis*, 33:383–408, 1998. doi: 10.2307/2331101. (Cited on page 51.)

[3] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Clara Vega. Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, 93:38–62, 2001. (Cited on page 42.)

[4] D. Applebaum. *Lévy processes and stochastic calculus*. Cambridge studies in advanced mathematics. Cambridge University Press, 2004. (Cited on pages 27 and 28.)

[5] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3), 1999. (Cited on page 78.)

[6] Louis Bachelier. Théorie de la Spéculation. *Annales Scientifiques de L'Ecole Normale Supérieure*, 17:21–88, 1900. (Cited on page 40.)

[7] Richard T Baillie and Tim Bollerslev. Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies*, 58(3): 565–85, 1991. (Cited on page 44.)

[8] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, 1999. (Cited on pages 22, 24, and 29.)

[9] Fischer Black. Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, 1976. (Cited on pages 43 and 46.)

[10] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973. doi: 10.2307/1831029. (Cited on pages 38 and 40.)

[11] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *CoRR*, abs/1010.3003, 2010. URL `http://arxiv.org/abs/1010.3003`. (Cited on page 97.)

[12] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986. doi: 10.1016/0304-4076(86)90063-1. (Cited on pages 10, 45, 46, and 48.)

[13] Tim Bollerslev, Jun Cai, and Frank M. Song. Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US treasury bond market. *Journal of Empirical Finance*, 7:37–55, 2000. (Cited on page 51.)

[14] Chris Brooks, Simon P. Burke, and Gita Persand. Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting*, 17(1):45–56, 2001. doi: 10.1016/S0169-2070(00)00070-4. (Cited on page 48.)

[15] John Y. Campbell, Andrew W. Lo, A. Craig MacKinlay, and Andrew Y. Lo. *The Econometrics of Financial Markets*. Princeton University Press, 1996. (Cited on page 38.)

[16] Geronimo Cardano. *Liber de Ludo Alae*. 1526. (Cited on page 21.)

[17] Bradley P. Carlin and Thomas A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, 2nd edition, 2000. (Cited on page 23.)

[18] Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. The Fine Structure of Asset Returns: An Empirical Investigation. *The Journal of Business*, 75(2):305–332, 2002. (Cited on pages 40 and 122.)

[19] C-C Chang and C-J Lin. LIBSVM: a library for support vector machines. *Science*, 2(3):1–39, 2001. (Cited on pages 112 and 114.)

[20] Carl R. Chen, Nancy J. Mohan, and Thomas L. Steiner. Discount rate changes, stock market returns, volatility, and trading volume: Evidence from intraday data and implications for market efficiency. *Journal of Banking & Finance*, 23(6):897–924, 1999. (Cited on page 51.)

[21] Y. W. Chen and C. J. Lin. *Combining SVMs with various feature selection strategies*. Springer, 2006. (Cited on page 110.)

[22] Peter F. Christoffersen, Anil Bera, Jeremy Berkowitz, Tim Bollerslev, Frank Diebold, Lorenzo Giorgianni, Jin Hahn, Jose Lopez, and Roberto Mariano. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1997. (Cited on page 129.)

[23] Timothy Cook and Thomas Hahn. The information content of discount rate announcements and their effect on market interest rates.

Journal of Money, Credit and Banking, 20(2):167–180, 1988. (Cited on page 51.)

[24] Antonio R. Damasio. *Descartes' error : emotion, reason, and the human brain*. G.P. Putnam, 1994. (Cited on page 97.)

[25] P. Dirac. The physical interpretation of quantum mechanics. *Proceedings of the Royal Society London*, pages 1–39, 1942. (Cited on page 21.)

[26] K. Dowd. *Beyond value at risk: the new science of risk management*. Wiley series in financial engineering. Wiley, 1998. (Cited on page 77.)

[27] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/BF02288367. (Cited on page 54.)

[28] Louis H. Ederington and Wei Guan. Forecasting volatility. *Journal of Futures Markets*, 25(5):465–490, 2005. doi: 10.1002/fut.20146. (Cited on page 53.)

[29] Louis H. Ederington and Jae Ha Lee. How markets process information: News releases and volatility. *The Journal of Finance*, 48(4): 1161–1191, 1993. (Cited on pages 9 and 51.)

[30] R. F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50, 1986. (Cited on page 46.)

[31] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. doi: 10.2307/1912773. (Cited on pages 10, 43, and 44.)

[32] Eugene F. Fama. The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34–105, 1965. doi: 10.2307/2350752. (Cited on page 40.)

[33] Eugene F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25, 1970. doi: 10.2307/2325486. (Cited on page 38.)

[34] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008. (Cited on page 112.)

[35] Michael J. Fleming and Eli M. Remolona. Price formation and liquidity in the U.S. treasury market: The response to public information. *The Journal of Finance*, 54(5):1901–1915, 1999. doi: 10.1111/0022-1082.00172. (Cited on pages 9 and 51.)

[36] B.E. Fristedt and L.F. Gray. *A modern approach to probability theory*. Birkhäuser Boston, 1996. (Cited on page 29.)

[37] Galileo Galilei. *Considerazione sopra il Giuco dei Dadi*. 1718. (Cited on page 21.)

[38] G. Gidófalvi and C. Elkan. Using news articles to predict stock price movements. Technical report, 2003. (Cited on page 98.)

[39] Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993. doi: 10.2307/2329067. (Cited on page 47.)

[40] Paul Harrison. Similarities in the Distribution of Stock Market Price Changes between the Eighteenth and Twentieth Centuries. *Journal of Business*, 71(1):55–79, 1998. (Cited on page 40.)

[41] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2nd edition, 2009. (Cited on pages 33 and 34.)

[42] Martin Hilbert and Priscila López. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332 (6025):60–65, 2011. doi: 10.1126/science.1200970. (Cited on page 11.)

[43] Christian Huygens. *de Ratiociniis in Alea Ludo*. 1657. (Cited on page 21.)

[44] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. (Cited on page 24.)

[45] Leonidas S. Junior and Italo De Paula Franca. Correlation of financial markets in times of crisis. 2011. URL `http://arxiv.org/abs/1102.1339v1`. (Cited on page 63.)

[46] John Maynard Keynes. *Treatise on Probability*. Macmillan And Co., 1921. (Cited on page 22.)

[47] Young Shin Kim, Svetlozar T. Rachev, Michele Leonardo Bianchi, and Frank J. Fabozzi. Financial market models with Lévy processes and time-varying volatility. *Journal of Banking & Finance*, 32(7):1363–1378, 2008. (Cited on page 123.)

[48] Young Shin Kim, Svetlozar T. Rachev, Michele Leonardo Bianchi, Ivan Mitov, and Frank J. Fabozzi. Time series analysis for financial market meltdowns. Technical report, 2010. (Cited on page 40.)

[49] Y.S. Kim, D.M. Chung, S.T. Rachev, and M.L. Bianchi. The modified tempered stable distribution, garch models and option pricing. *Probability and Mathematical Statistics*, 29(1):91–117, 2009. (Cited on page 40.)

[50] Tibor Kiss and Jan Strunk. Unsupervised Multilingual Sentence Boundary Detection. *Comput. Linguist.*, 32(4):485–525, 2006. doi: 10.1162/coli.2006.32.4.485. (Cited on pages 106 and 115.)

[51] Lev B. Klebanov, Tomasz J. Kozubowski, and Svetlozar T. Rachev. *Ill-Posed Problems in Probability and Stability of Random Sums*. Nova Publishers, 2006. (Cited on page 29.)

[52] Stefan Knerr, Léon Personnaz, and Gérard Dreyfus. Single-layer learning revisited: A stepwise procedure for building and training a neural network. In F. Fogelman Soulié and J. Hérault, editors, *Neurocomputing: Algorithms, Architectures and Applications*, volume F68 of *NATO ASI Series*, pages 41–50. Springer-Verlag, 1990. (Cited on pages 36 and 112.)

[53] Andrei Kolmogorow. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933. (Cited on page 21.)

[54] Ismo Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E*, 52:1197–1199, 1995. doi: 10.1103/PhysRevE.52.1197. (Cited on pages 40 and 122.)

[55] Samuel Kotz, Tomasz Kozubowski, and Krzystof Podgórski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser Boston. (Cited on pages 29 and 126.)

[56] Paul H Kupiec. Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 3:73–84, 1995. doi: 10.3905/jod.1995.407942. (Cited on page 129.)

[57] Victor Lavrenko, Matthew D. Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Mining of concurrent text and time series. In *Knowledge Discovery and Data Mining*, 2000. (Cited on page 97.)

[58] Victor Lavrenko, Matthew D. Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Language models for financial news recommendation. In *International Conference on Information and Knowledge Management*, pages 389–396, 2000. doi: 10.1145/354756.354845. (Cited on page 97.)

[59] H. P. Luhn. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4):309–317, 1957. (Cited on page 104.)

[60] David J. C. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 1st edition, 2003. (Cited on page 30.)

[61] Dilip B. Madan and Eugene Seneta. The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business*, 63(4):511–524, 1990. (Cited on pages 10 and 124.)

[62] Dilip B. Madan, Peter Carr, and Eric C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2:79–105, 1998. (Cited on pages 124 and 125.)

[63] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963. doi: 10.2307/2350970. (Cited on pages 10, 38, 40, 43, and 67.)

[64] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 1 edition, 2008. (Cited on pages 105 and 107.)

[65] Harry Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1): 77–91, 1952. doi: 10.2307/2975974. (Cited on pages 10, 38, and 40.)

[66] George Marsaglia and John Marsaglia. Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software*, 9(2):1–5, 2004. (Cited on page 129.)

[67] Robert C. Merton. Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 4(1), 1973. doi: 10.2307/3003143. (Cited on page 38.)

[68] M. A. Mittermayer and G. Knolmayer. Text mining systems for market response to news: A survey. *Proceedings of the IADIS European Conference Data Mining 2007*, 2007. (Cited on pages 11 and 98.)

[69] Marc-André Mittermayer and Gerhard F. Knolmayer. Newscats: A news categorization and trading system. In *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, pages 1002–1007. IEEE Computer Society, 2006. doi: 10.1109/ICDM.2006.115. (Cited on page 97.)

[70] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991. doi: 10.2307/2938260. (Cited on pages 46 and 47.)

[71] Jussi Nikkinen and Petri Sahlström. Scheduled domestic and US macroeconomic news and stock valuation in Europe. *Journal of Multinational Financial Management*, 14(3):201–215, 2004. doi: 10.1016/j.mulfin.2003.01.001. (Cited on pages 9 and 52.)

[72] John R. Nofsinger and Richard W. Sias. Herding and feedback trading by institutional and individual investors. *Journal of Finance*, 54:2263–2295, 1999. doi: 10.1111/0022-1082.00188. (Cited on page 97.)

[73] Desh Peramunetilleke and Raymond K. Wong. Currency exchange rate forecasting from news headlines. In *Thirteenth Australasian Database Conference (ADC2002)*. ACS, 2002. (Cited on page 98.)

[74] Martin F. Porter. An algorithm for suffix stripping. *Program: Electronic Library & Information Systems*, 40(3):211–218, 1980. doi: 10.1108/00330330610681286. (Cited on pages 107 and 115.)

[75] Martin F. Porter. Snowball: A language for stemming algorithms, 2001. URL `http://snowball.tartarus.org/texts/introduction.html`. (Cited on page 107.)

[76] G. Pui Cheong Fung, J. Xu Yu, and Wai Lam. Stock prediction: Integrating text mining approach using real-time news. In *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pages 395–402, 2003. doi: 10.1109/CIFER.2003.1196287. (Cited on page 98.)

[77] Svetlozar T. Rachev and Stefan Mittnik. *Stable Paretian Models in Finance*. Wiley New York, 2000. (Cited on pages 10, 29, and 40.)

[78] Svetlozar T. Rachev, Stefan Mittnik, Frank J. Fabozzi, Sergio M. Focardi, and Teo Jašić. *Financial Econometrics*. Wiley Finance. (Cited on pages 38, 42, 45, and 46.)

[79] C. Radhakrishna Rao. Minimum variance and the estimation of several parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43(02):280–283, 1947. doi: 10.1017/S0305004100023471. (Cited on page 31.)

[80] S. E. Robertson and K. S. Jones. Relevance weighting of search terms. *J. Am. Soc. Inf. Sci.*, 27(3):129–146, 1976. doi: 10.1002/asi.4630270302. (Cited on page 104.)

[81] Jan Rosiński. Tempering Stable Processes. *Stochastic Processes and Their Applications*, 117:677–707, 2007. doi: 10.1016/j.spa.2006.10.003. (Cited on pages 10, 40, and 122.)

[82] Stephen A. Ross. Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy. *Journal of Finance*, 44(1):1–17, 1989. (Cited on pages 9, 42, and 53.)

[83] G. Salton and M. J. Mcgill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1986. (Cited on page 104.)

[84] William F. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3):425–442, 1964. doi: 10.2307/2977928. (Cited on page 38.)

[85] Young Shin Kim, Svetlozar T. Rachev, Michele Leonardo Bianchi, and Frank J. Fabozzi. Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking & Finance*, 34(9):2096–2109, 2010. (Cited on page 123.)

[86] James Thomas and Katia Sycara. Integrating genetic algorithms and text learning for financial prediction. In *Proceedings of the Genetic and Evolutionary Computing 2000 Conference Workshop on Data Mining with Evolutionary Algorithms*, pages 72–75, 2000. (Cited on page 98.)

[87] Jingbo Wang, Wai Wan Tsang, and George Marsaglia. Evaluating Kolmogorov's Distribution. *Journal of Statistical Software*, 8, 2003. (Cited on page 128.)

[88] Walter Wasserfallen. Macroeconomics news and the stock market: Evidence from Europe. *Journal of Banking & Finance*, 13(4–5):613–626, 1989. doi: 10.1016/0378-4266(89)90033-2. (Cited on page 51.)

[89] Jeffrey M. Wooldridge. Estimation and inference for dependent processes. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 45, pages 2639–2738. Elsevier, 1986. (Cited on page 48.)

[90] Beat Wüthrich, D. Permunetilleke, S. Leung, Vincent Cho, Jian Zhang, and W. Lam. Daily Prediction of Major Stock Indices from Textual WWW Data. In *Knowledge Discovery and Data Mining*, pages 364–368, 1998. (Cited on pages 11 and 97.)

[91] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965. doi: 10.1016/S0019-9958(65)90241-X. (Cited on page 21.)

ACKNOWLEDGMENTS

My utmost gratitude goes to Prof. Svetlozar T. Rachev and Prof. Michael Feindt who gave me the opportunity to do this thesis. I would like to thank Prof. Rachev for his constant support regarding my research. I have benefited – though doubtless insufficiently – from his deep understanding of finance.

I thank Prof. Feindt for his patient guidance and encouragement over the years. It is his contagious sense for curiosity that has led to a working atmosphere were creativity actually is nurtured.

This thesis was conducted at Lupus Alpha NeuroBayes® Short Term Trading Fund at Phi-T® GmbH. I would like to thank Jochen Bossert and Dr. Andreas Heiss financial support and an efficient infrastructure.

I was very lucky to work in an inspiring environment. I express my sincere gratitude to Dr. Jan Fränkle, Dr. Christian Scherrer, Dr. Christian Haag, Dr. Bruno Daniel, Dr. Pascal Böhi, Dr. Markus Kreer, Dr. Martina Reber, Alexander Beck, Daniel Martschei and Martin Hahn for enlightening discussions and their good sense of humor.

Moreover, I would like to thank Dr. Young Shin (Aaron) Kim for constructive criticism, valuable discussions and encyclopedic supply of background material.

During the last weeks of this thesis, proof-reading was very much appreciated. I thank Dr. Pascal Böhi, Dr. Bruno Daniel, Dr. Jan Fränkle, Dr. Markus Kreer and Martin Hahn for critical suggestions.

Finally, my thanks go to my family for their unfailing encouragement throughout my life. Especially, I would like to thank Vanessa Kim Ahlefeld for her love and support.