University of Economics, Prague

Faculty of Finance and Accounting

Department of Banking and Insurance
Field of study: Financial Engineering

MASTER THESIS

Portfolio Value at Risk and Expected Shortfall using High-frequency data

Author of the Master thesis: Bc. Marek Zváč
Supervisor of the Master thesis: Ing. Milan Fičura
Year of Defence: 2017

Declaration of Authorship

The author hereby declares that he compiled the thesis *"Portfolio Value at Risk and Expected Shortfall using High-frequency data"* independently, using only the resources and literature properly marked and included in the bibliography. The thesis has not been used to obtain a different or the same degree.

Prague, March 8, 2017

Signature

Acknowledgments

The author is grateful to Ing. Milan Fičura for his comments and valuable feedback and my family for their priceless support.

Abstract

The main objective of this thesis is to investigate whether multivariate models using High-frequency data provide significantly more accurate forecasts of Value at Risk and Expected Shortfall than multivariate models using only daily data. Our objective is very topical since the Basel Committee announced in 2013 that is going to change the risk measure used for calculation of capital requirement from Value at Risk to Expected Shortfall. The further improvement of accuracy of both risk measures can be also achieved by incorporation of high-frequency data that are rapidly more available due to significant technological progress. Therefore, we employed parsimonious Heterogeneous Autoregression and its asymmetric version that uses high-frequency data for the modeling of realized covariance matrix. The benchmark models are chosen well established DCC-GARCH and EWMA. The computation of Value at Risk (VaR) and Expected Shortfall (ES) is done through parametric, semi-parametric and Monte Carlo simulations. The loss distributions are represented by multivariate Gaussian, Student's t, multivariate distributions simulated by Copula functions and multivariate filtered historical simulations. There are used univariate loss distributions: Generalized Pareto Distribution from EVT, empirical and standard parametric distributions. The main finding is that Heterogeneous Autoregression model using high-frequency data delivered superior or at least the same accuracy of forecasts of VaR to benchmark models based on daily data. Finally, the backtesting of ES remains still very challenging and applied Test I. and II. did not provide credible validation of the forecasts.

JEL Classification	C52, C53, C58, G1, G17, G21
Keywords	Portfolio,Value at Risk, Expected Shortfall, Realized covariance, HAR, Copula, EVT, High-frequency data

Author's e-mail	marek.zvac@countermail.com
Supervisor's e-mail	milan.ficura@isis.vse.cz

Abstrakt

Hlavním cílem této práce je zjistit, zda vícerozměrné modely s použitím vysokofrekvenčních dat poskytují výrazně přesnější předpovědi Value at Risk a Expected Shortfall než vícerozměrné modely pouze s pomocí denních data. Náš cíl je velmi aktuální, protože v roce 2013 Basilejský výbor oznámil, že se chystá změnit rizikovou míru používanou pro výpočet kapitálových požadavků z Value at Risk na Expected Shortfall. Další zlepšení přesnosti obou rizikových měr může být také dosaženo začleněním vysokofrekvenční údajů, které jsou mnohem více k dispozici vzhledem k významnému technologickému pokroku. Jako reprezentativní model, který využívá vysokofrekvenční data pro modelování realizované kovarianční matice, jsme vybrali heterogenní autoregresi a její asymetrickou verzi. Jako benchmark jsou vybrány dobře zavedené modely DCC-GARCH a EWMA. Výpočet Value at Risk a Expected Shortfall se provádí pomocí parametrické, semi-parametrické metody a Monte Carlo simulace. Vícerozměrné rozdělení ztrát jsou reprezentovány Gaussovým, Studentovým rozdělením, simulovaným rozdělením z copula funkcí a filtrovaných historických simulací. Jako jednorozměrné rozdělení byly použity generalizované Paretovo rozdělení z EVT, empirické a standartní parametrické rozdělení. Hlavním zjištěním je, že heterogenní autoregrese s použitím vysoko frekvenčních dat dodala lepší nebo alespoň stejnou přesnost prognóz Value at Risk jako benchmark modely s použitím denních dat. Nakonec backtesting Expected Shortfall zůstává stále velmi náročný a aplikace testů I. a II. neposkytla věrohodnou validaci předpovědí.

Klasifikace JEL C52, C53, C58, G1, G17, G21

Klíčová slova Portfólio,Value at Risk, Expected Shortfall, Realizovaná kovariance, HAR, Copula, EVT, Vysoko frekvenční data

E-mail autora marek.zvac@countermail.com

E-mail vedoucího práce milan.ficura@isis.vse.cz

Contents

List of Tables

List of Figures

Acronyms

ADCC	Asymmetric Dynamic Conditional Correlation
AR	Autoregressive model
ARCH	Autoregressive Conditional Heterosckedasticity
ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARMA	Autoregressive Moving Average
APARCH	Asymmetric Power ARCH
CARE	Conditional Autoregressive Expectile
CAW	Conditional Autoregressive Wishart
CCC	Constant Conditional Correlation
cDCC	corrected Dynamic Conditional Correlation
cRDCC	corrected realized Dynamic Conditional Correlation
DCC	Dynamic Conditional Correlation
EGARCH	Exponential GARCH
ES	Expected Shortfall
EVT	Extreme Value Theory
EWMA	Exponential weighted moving average
FHS	Filtered Historical Simulations
GARCH	Generalized Autoregressive Conditional Heterosckedasticity
GAS	Generalized Autoregressive Score
GEV	Generalized Extreme Value distribution
GO-GARCH	Generalized Orthogonal GARCH
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
GPD	Generalized Pareto Distribution
HAR	Heterogeneous Autoregression
HEAVY	High-frEquency-bAsed VolatilitY
HFD	High-frequency data
iid	independent and identically distributed

LHAR Leveraged Heterogeneous Autoregression

MLE Maximum Likelihood Estimation

OLS Ordinary Least Squares

POT Peak-Over-Threshold

PSD Positive Semi-Definiteness

QMLE Quasi Maximum Likelihood Estimation

RCOV Realized Covariance

RCOR Realized Correlation

RV Realized Variance

RVOL Realized Volatility

SV Stochastic Volatility

VaR Value at Risk

VAR Vector Autoregression

VARMA Vector Autoregressive Moving Average

VARFIMA Vector Autoregressive Fractionally Integrated Moving Average

WAR Wishart Autoregression

Chapter 1

Introduction

Since the first implementation of VaR in the early '90 by J.P. Morgan, the VaR became the key modern risk measure from the internal and regulatory point of view for financial institutions. The further development of financial econometrics in the field of risk measure was particularly focused on the investigation of VaR. The key parameters of VaR have become volatility estimated by standard deviation from the sample of daily closing prices and covariance as the representative of linear dependence also estimated from the same type of data as volatility. However, the recent financial crisis in 2007-2009 showed that neither VaR that is in effect only the lower or upper quantile of the loss distribution nor just using daily closing prices is sufficient risk approach in current conditions.

The academic research was aware of these main issues and hence, since 1997 there has been proposed the alternative to VaR called ES that studies the average of loss distribution given that VaR was exceeded. Thus, ES is more informative about the possible risk according to certain probability. Simultaneously to ES, since 1998, there has been started a deep research about the utilization of prices sampled with higher frequency than one day until the finest frequency that is transaction by transaction. The importance of ES has been recently even magnified as Basel Committee in 2013 announced that ES is going to replace VaR measure for calculation of capital requirements.

In case of high-frequency data, meaning prices, the rapid technological progress allowed to boost significantly the computation power resulting in the substantial volume of trading. Many markets turned to such liquidity that intraday information become statistically relevant also for the measurement of volatility and covariance that is currently known as realized measures.

These events give the main motivation for this master thesis to investigate ES besides of VaR as we are currently in the transition period from VaR to ES from regulator point of view and recent advent of utilization of high-frequency data also for risk management purposes. Additionally, our investigation is from the portfolio perspective because in practice we are usually interested in various assets at least for the diversification purposes as basic technique to minimize the risk.

The main objectives of this thesis is to investigate whether multivariate models using High-frequency data provide significantly more accurate forecasts of VaR and ES than multivariate models using only daily data. The investigation will be carried out through answering following questions: What model and approach provides the most accurate forecasts of VaR and ES? Does the best model and approach of VaR perform similarly also in the forecasting of ES? What is the difference between the two approaches for various market volatility periods (stable versus turbulent period)?

We choose standard practice of financial econometrics for the methodology. Our benchmark models for modeling of covariance matrix use daily prices represented be the Exponential weighted moving average (EWMA) model with estimated parameter λ by RiskMetrics and well-established Dynamic Conditional Correlation (DCC)-Generalized Autoregressive Conditional Heterosckedasticity (GARCH) with its asymmetric versions. The representation of model using high-frequency data is the one of class called Heterogeneous Autoregression (HAR) due to its parsimony and proven good performance from the other researchers. We implement all standard methods of calculation meaning parametric, semi-parametric and non-parametric. The latter one will be using advanced econometric approaches such as Extreme Value Theory (EVT) and copula functions. Finally, we check the validity and accuracy of models by using the most common backtests and model selection represented by loss functions in case of VaR.

However, the conducted research is limited in the certain areas. The first limitation lies in the type of products used in the portfolio. The portfolio consists of only linear products such as futures and spot prices. The reason is that applied models do not capture correctly nonlinear dependency between the price of product and the underlying variables. The second limitation is that the agent using VaR and ES measures is a price taker and he is able to close out its entire position for the market price from the used data set. Therefore the liquidity adjustment of VaR and ES is omitted. The third limitation is that we investigate only passive risk management application and we do not study the active one such as incremental, marginal and component VaR and ES. The another limitation is assumption that circuit breakers applied on futures products in our portfolio will remain in the same structure also for the future implying that we do not expect the structural change from the regulator. The latest example how such an assumption can be strong is the unexpected exit of peg on Swiss franc by Swiss National Bank. It caused the unseen volatility in the entire history of currency trading since the floating regime was established. The last limitation is in the size of portfolio and variety of assets since the used portfolio contains only four and low correlated assets, specifically S&P500 futures, Crude oil futures, Spot gold, EURUSD currency pair. In the case of larger portfolio or high correlated assets, the different approaches or models would be more suitable. Nevertheless, the recommended length of master thesis by itself is limitation because there exist much higher number of models, distributions and backtests that would cover our objectives.

Our contribution according to review of literature is that there has not been published or found by author yet which investigate the application of high-frequency data in terms of realized measure in multivariate space in order to estimate ES measure. Furthermore, the we provide a comprehensive comparison of the difference between high-frequency and daily data according to all standard methods of calculation VaR and ES. Such a scope has not been conducted yet to the best author's knowledge.

The thesis is structured as follows: Chapter 2 gives besides of enlightening literature review a necessary theoretical background of realized and risk measure. Chapter 3 covers the introduction of variance-covariance models, loss distribution and backtesting and model selection methods applied in empirical analysis Chapter 4 provides tha investigation of our data sets, application of chosen models and presentation of the empirical results Chapter 5 summarizes our findings.

Chapter 2

Theoretical background

2.1 Literature review

The history of VaR measure is well-documented in the working paper Holton (2002) where he considers the origins of VaR in the first risk metric - standard deviation of simple return proposed in the portfolio theory of Markowitz (1952) and capital requirements imposed by the New York Stock Exchange around 1922.

The standard deviation of asset returns was the first estimator of dispersion so-called volatility. It was simple to estimate but there was a major drawback since it could assign the same value to different probability distributions of returns and differently risky assets were considered as the same one. In the late of 1980's and the beginning of 1990's the financial institutions started using current known VaR measure defined as *a high quantile of the financial returns distribution of a portfolio over a certain time horizon* Cont (2001) or in other words the maximum loss of the portfolio at a given confidence level and time horizon. The high popularity and wide adaption of VaR came in 1994 when J. P. Morgan released for free the technical document describing in details their internal computation of VaR named RiskMetrics™ J.P.Morgan (1996). Since then, the VaR has been perceived as main benchmark risk measure in the financial industry.
Nevertheless, the VaR represents still only the quantile and therefore its main drawback is that VaR does not say what a loss can be made when the VaR is breached. Additionally, Artzner *et al.* (1997) and Artzner *et al.* (1999) has shown that VaR even does not satisfy one of the axioms of the coherence since it is not sub-additive meaning that in some cases a diversified portfolio of assets can obtain higher VaR than would be the sum of individual VaRs of the same assets. As a solution, Artzner *et al.* (1999) proposed new risk measure called ES as expected value of a loss given that VaR was exceeded or in other words, mean value of the worst losses exceeding given confidence level. It can be easily seen there is connection between the ES ES computation and VaR.

The VaR can be decomposed as a conditional mean plus a product of conditional volatility and the high quantile of the distribution function of the innovation process. Thus, the conditional volatility and the distribution function of the innovation process are the key parameters in the VaR measure since the conditional mean of the financial returns distribution is close to zero for the short time horizon (e.g. one day what is a typical prediction horizon for VaR).

The systematic modeling of conditional volatility was founded by the introduction of parametric Autoregressive Conditional Heterosckedasticity (ARCH) model of Engle (1982) and its generalized version GARCH of Bollerslev (1986). Particularly, univariate family of GARCH models became widely used in the modeling of conditional volatility for its ability to accommodate the most of the stylized facts of volatility such as volatility clustering, asymmetric impact of the asset returns on the conditional volatility well-known as leverage effect, time-varying higher moments, fat tails and long memory (persistence) of the innovations Cont (2001), Jondeau *et al.* (2007, pp. 10-26). The further improvement of the VaR estimation is achieved by the choice of the second key parameter, the distribution of the innovations. The significance gains in the accuracy of VaR can be gained with Skew-Student's t distribution that allows for skewness and fat tails according to Kuester *et al.* (2006) as non-Gaussian distribution of asset returns was already discovered by Mandelbrot (1963). However, for each of asset class those stylized facts applies in different magnitude. The comprehensive overview of univariate VaR approaches can be found in Engle & Manganelli (2001), Dowd (2005), Kuester *et al.* (2006) and Alexander (2009).

Regarding to ES in an univariate dimension, an excellent survey of estimation methods of ES for various probability distributions is provided in Nadarajah *et al.* (2014) including the list of available software and relating packages. Another assessment of the all established models for VaR in terms of accuracy of ES estimation is in a study of Righi & Ceretta (2015) resulting in the preference of conditional models, particularly GARCH with Filtered Historical Simulations (FHS) distribution and an interesting finding that accuracy of ES estimation depends on the accuracy of VaR estimation. Obviously further references for ES estimation can be found in the mentioned papers or literature overview mentioned for univariate estimation of VaR.

In the case of Portfolio VaR and ES, we are interested in the modeling of the conditional portfolio volatility (if we consider only elliptical multivariate distribution of innovations) what is the most popular approach in the academic research. The conditional portfolio volatility is obtained either directly from portfolio returns or from the covariance matrix of individual asset returns. The most popular methods of modeling covariance matrix for VaR purpose are multivariate RiskMetrics^{TM} and multivariate GARCH. RiskMetrics^{TM} applies EWMA methodology with estimated parameter $\lambda = 0.94$ J.P.Morgan (1996). The first well-known representatives of multivariate GARCH became diagonal VECH of Bollerslev *et al.* (1988) and BEKK of Engle & Kroner (1995) but those models suffers with the curse of the dimensionality. This problem was resolved in the model Constant Conditional Correlation (CCC) of Bollerslev (1990) and time-variant conditional covariance called DCC of Engle (2000). As is the case with univariate modeling, the conditional covariance matrix experiences not only asymmetric behaviour of variances but also covariances, especially during stress events . This characteristics is more analyzed and incorporated in the Asymmetric Dynamic Conditional Correlation (ADCC) GARCH model of Cappiello *et al.* (2006). Other viewpoint on DCC model has Aielli (2013) in his paper where he claims that DCC estimator of the correlation can be inconsistent and he suggests corrected Dynamic Conditional Correlation (cDCC) and proves that cDCC is uniformly unbiased. The reference guide to the majority of univariate and multivariate GARCH models and their alternatives is provided in the glossary of Bollerslev (2008), more detailed overview of multivariate GARCH models can be found in the review article of Laurent *et al.* (2006) and multivariate concepts in calculation of VaR in McNeil *et al.* (2015), Christoffersen (2011) and Alexander (2008).

Besides of multivariate EWMA and Conditional Correlation GARCH models for estimation of covariance matrix, there are other approaches such as multivariate versions of Stochastic Volatility (SV) and R (RVOL). A shortcoming of multivariate stochastic model is the complexity of estimation Laurent *et al.* (2006) .The latter option brings a novel concept of understanding and estimating volatility and covariance. All aforementioned models treat daily volatility and covariance as non-observable variables (latent). However, daily realized volatility and covariance are estimated, in their naive version, as the sum of the squared intraday returns and the sum of the products of intraday returns, respectively. Based on this methodology, realized volatility and covariance are, in principle, observable and estimators become non-parametric (model-free). The new paradigm of volatility and covariance estimators perceived as realized ones was introduced in the seminal work of Andersen & Bollerslev (1998) and Andersen *et al.* (1999). The theoretical framework of realized volatility and covariance can be found in Andersen *et al.* (2003) , Barndorff-Nielsen & Shephard (2004) and comprehensive summary in McAleer & Medeiros (2008) and Bauwens *et al.* (2012).

An essential property of forecasted covariance matrix for computation of VaR is the Positive Semi-Definiteness (PSD). This property is not guaranteed from the naive version of realized covariance due to market microstructure noise and hence, one of the proposed augmented realized covariance estimators satisfying mentioned property is multivariate realized kernel of Barndorff-Nielsen *et al.* (2011). The other methods insuring PSD property include the Matrix Logarithm transformation of realized covariance of Bauer & Vorkink (2007) and adapting long-memory univariate model Heterogeneous Autoregression of realized volatility of Corsi (2009) on the individual elements of transformed realized covariance matrix. The similar technique is used by Chiriac & Voev (2011) which applies Cholesky decomposition of realized covariance and forecasting Cholesky factors by a multivariate long-memory Vector Autoregressive Fractionally Integrated Moving Average (VARFIMA) model and univariate HAR model. The drawback of previous two approaches was a lost of interactions between variances and covariances. In a different approach Gourieroux *et al.* (2009) modeled entire realized covariance matrix by the Wishart Autoregression (WAR) and further extensions by block WAR and HAR-WAR can be found in Bonato *et al.* (2009) or asymmetric version of WAR in Jin & Maheu (2010). The dynamic generalization of the models of Gourieroux *et al.* (2009) and Jin & Maheu (2010) was proposed by Golosnoy *et al.* (2012) as a Conditional Autoregressive Wishart (CAW).

A new class of multivariate models modeling realized covariance matrix can be considered also a High-frEquency-bAsed VolatilitY (HEAVY) model of Noureldin *et al.* (2012). There is a wide body of additional literature focusing on estimation of realized covariance matrix and hence above paragraph represents just only the selection of the most established approaches nowadays.

Completely different multivariate approaches that are not based on covariance matrix, as the main parameter describing dependence and one of the parameters describing entire multivariate distribution, are represented by Monte Carlo simulations. One approach uses returns of each asset are modeled by stochastic process and another common approach is through Copulas which estimates marginal distributions and dependence structure separately. The last unique method which is not Monte Carlo simulations is the multivariate EVT that estimates directly only the tail of the multivariate distribution. A disadvantage of multivariate extreme value theory is the curse of dimensionality Rocco (2011) and Goix *et al.* (2015). The most promising concept from Monte Carlo simulations is the Copula which allows to construct multivariate distributions that do not have even analytical form. According to knowledge of author, there are only three papers of Fengler & Okhrin (2012),Fengler & Okhrin (2016) and Brechmann *et al.* (0) proposing utilization of high-frequency data in Copula estimation. First one establishes the term Realized Copula as *"the copula structure materialized in realized covariance estimated from within-day high-frequency data"*. Second one continues in the fashion of first paper and just extend number of competitor models by dynamic Copula models such as DCC, Patton (2004), the Generalized Autoregressive Score (GAS), realized GAS and realized covariance models of Bauer & Vorkink (2007) and Chiriac & Voev (2011). The last one uses vine copulas which provides more flexible modeling of dependencies among the asset returns.

The applications of high-frequency data in risk management set off, as usual in an univariate dimension. The pioneering paper was of Giot & Laurent (2004) where they did not find superior perfomance of VaR forecasts of ARFIMAX-RV model in comparison with Asymmetric Power ARCH (APARCH) model using Skew-Student's t distribution. However, one of the outcomes was that neither daily log returns nor intraday log returns have Normal distribution. The most complex research was conducted by Kruse (2006) who analyzed 107 models including ARCH, Realized volatility, Stochastic volatility models based on Normal, Skew-Student's t distributions and FHS or EVT approach. His finding was that benchmark models RiskMetrics$^{\mathrm{TM}}$ and GARCH with Normal distribution were not significantly outperformed by any other model. Nonetheless, he chose FHS approach with Realized volatility model as the best performing model in VaR forecasting. Another rewarding paper is Louzis *et al.* (2014) providing extensive literature review of 11 papers inspecting whether realized volatility models give higher accuracy of VaR estimates than ARCH-type models. They highlighted that around 60% of papers prefer VaR models employing high-frequency data than ARCH models based on daily data. Furthermore, they pointed out that nearly all of researchers used only full parametric approach in computation of VaR what means that quantiles of innovations were computed always from the same parametric distributions such as Normal, Student's t or Skew-Student's t. Therefore different estimates of VaR were mainly driven only by the estimation of dispersion computed either as standard deviation from daily data or squared realized variance. Contrary to this practice, Louzis *et al.* (2014) applied similarly to Kruse (2006) also FHS, EVT approach and in addition to standard statistical backtesting seen in other papers, they examined the models even in terms of Basel II accuracy, regulatory accuracy and capital efficiency. Their empirical analysis showed that any volatility models using Skew-Student's t distribution or EVT method have the best results in complying the statistical and regulatory accuracy. Regarding to the efficiency of VaR, the results are in favor of realized volatility model which is the Asymmetric HAR model together with EVT method. Finally, they provide conclusions also from the regulatory and financial institutions point of view.

Unlike the VaR, academic research dedicate much less attention to ES risk measure in terms of high-frequency data in univariate dimension. There were found only four papers analyzing the impact of intraday returns on ES measure.

The paper of Watanabe (2011) concludes that the Realized GARCH with Skew-Student's t distribution performs in estimation of ES superior to Student's t or Normal distribution and also toExponential GARCH (EGARCH) model using daily data. The type of realized estimator meaning naive realized variance or realized kernel did not show any significant difference in the forecasts of ES. That result of the Realized GARCH was confirmed by Contino & Gerlach (2014) where they state that the Realized GARCH provided more accurate estimates of ES in all eight stock indices than its daily counterpart. On the other hand, the preferred distributions depended on significance level when on 5% Skew-Student's t distribution was solely preferred but on 1 % the choice would be also either Student's t or Normal distribution.

Besides of standard deviation and realized volatility, there is another estimator of volatility called range-based estimator which takes into account either only high and low price or including open and close price, for further reference Chou *et al.* (2010). The incorporation of realized range estimator was done by Chao & Richard (2014) who favor the realized range estimor using Realized GARCH and Markov Chain Monte Carlo estimation and forecasting approach with Student's t distribution of innovations, as the most accurate at forecasting 1% ES in comparison with realized volatility estimator, daily data based GARCH, Historical simulations and the Conditional Autoregressive Expectile (CARE) indirect GARCH model. The last study comes from Bee *et al.* (2016) who examine EVT approach with filtration of returns by asymmetric Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) using daily data and high-frequency data based HAR and its extensions such as inclusion of jumps and asymmetry. From the ES perspective, there is not significant distinction between HAR models and GJR-GARCH. Different case is VaR where HAR models prove higher accuracy than GJR-GARCH increasing with the longer time horizon.

Finally, we assess the literature of papers investigating potential benefits of high-frequency data in estimation of portfolio VaR and ES in a multivariate dimension. This area of research is even smaller than univariate one. We can divide papers again in two groups, those do compare the performance against daily data based models and those that do not.

Let's start with the first group initiated by McMillan *et al.* (2008) who found the preferred model is univariate GARCH estimated on the raw intraday portfolio returns to multivariate Vector Autoregression (VAR) model using the same type of data or other models using daily data. Following innovative report of Fengler & Okhrin (2012) showed that proposed realized copula managed to adopt quickly to volatile events thanks to utilization of high-frequency data and enabled sufficient capturing of non-trivial tail-dependence structures in comparison with Gaussian copulas and hence, realized copulas were superior to other models. Another very comprehensive comparison is due to Candila (2013) where he evaluated rolling realized covariance, CAW models versus BEKK, DCC-GARCH and Generalized Orthogonal GARCH (GO-GARCH) models without finding the significant difference of forecasting portfolio VaR. The only one result considering the most appropriate models such as DCC-GARCH and RiskMetrics^{TM} based on daily is in the master thesis of Čech (2013). Even though he included not only basic multivariate HAR model based on Cholesky decomposition but also more advanced WAR models of Bonato *et al.* (2009). The biggest sample of data consisting of 52 stocks of the largest U.S. financial institutions is in Boudt *et al.* (2014) with the most accurate VaR forecasts recorded by model utilizing high-frequency data by corrected realized Dynamic Conditional Correlation (cRDCC) on Cholesky decomposed realized covariance (Liquidity sorting type) using Autoregressive Fractionally Integrated Moving Average (ARFIMA) model. Generally cRDCC performed good and Scalar-BEKK or HEAVY significantly worse in comparison cRDCC or cDCC. The last one comparing article is Fengler & Okhrin (2016) with the same conclusion as in Fengler & Okhrin (2012) namely with Cholesky decomposition irrespective to marginal distribution or type of realized copula but in the context of additional high-frequency and daily data models.

The second group of papers that did not assessed performance of portfolio VaR between high-frequency and daily data models can serve as the inspiration for further assessment, specifically Bonato *et al.* (2009), Bauwens *et al.* (2014) and Brechmann *et al.* (0).

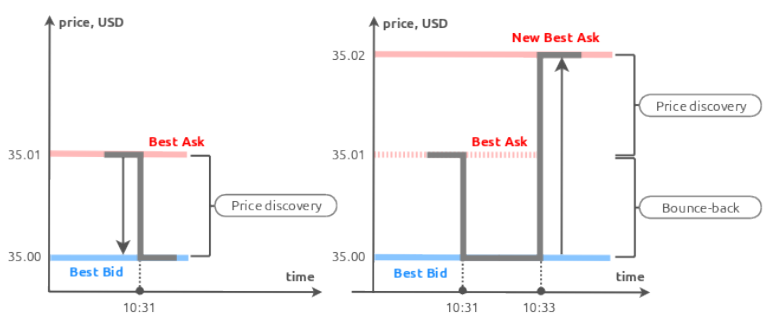
The common aspects of all papers evaluating models using multivariate realized measure are that the most popular multivariate distributions are the elliptical Normal and Student-t apart from simulations applied in Copula models, as the data are used almost solely only sets of currency pairs and stocks. The exception is the master thesis of Čech (2013) who included commodities such as futures Light Crude NYMEX and futures Gold COMEX. Overall, there were four papers in favor of high-frequency data based models, one paper preferring daily data based models, one paper that was irresolute to any kind of data and three papers estimating VaR portfolio exclusively on high-frequency data based models. Moreover, the most popular backtesting methods were (Un)conditional coverage tests of Kupiec (1995) and Christoffersen (1998) and Dynamic Quantile test of Engle & Manganelli (2004). To the best knowledge of author, there is only one paper of estimating VaR and ES portfolio Ubukata & Watanabe (2015) but entirely for purposes of hedging performance and not risk management one.

2.2 High-frequency data

High-frequency data (HFD) are generally considered as data (prices of trades, quotes or close prices of financial assets) sampled with higher frequency than one day. The most finest frequency is trade-by-trade or tick what is a definition of the smallest price increment in case of organized market. Such data are know as an ultra high-frequency data or alternatively as tick data. There is a wide range of applications of HFD with several examples mentioned in Tsay (2012) for instance price discovery process, bid-ask dynamics, algorithmic trading or designing trading strategies. Nevertheless, the interesting benefits can be found also in the risk management field. The availability or the intraday prices introduces new properties not encountered in homogenously spaced daily or lower frequencies.

- First property is an univariate one. HFD, especially tick data are contaminated by **market microstructure noise** which consist of bid-ask bounce (transaction prices tend to bounce between bid and ask quotes what cause false impression of increased volatility and subsequently such a recorded volatility is upward biased Bauwens *et al.* (2012)), price discretization, irregular trading, etc. In order to reduce this noise for the purpose of estimation Realized Variance (RV) or Realized Covariance (RCOV), there were suggested various methods. The most simple one is to sample data from lower frequency with equidistant intervals i.e. 5 or 20 minutes, see Hansen & Lunde (2006) or Liu *et al.* (2015). This is called *calendar time sampling* or *sparse sampling* and its disadvantage is a significant loss of HFD and a necessity to decide about the optimal sampling frequency. Another approach is to use all available data even on the tick level and apply kernel based techniques as mentioned in Hansen & Lunde (2006).

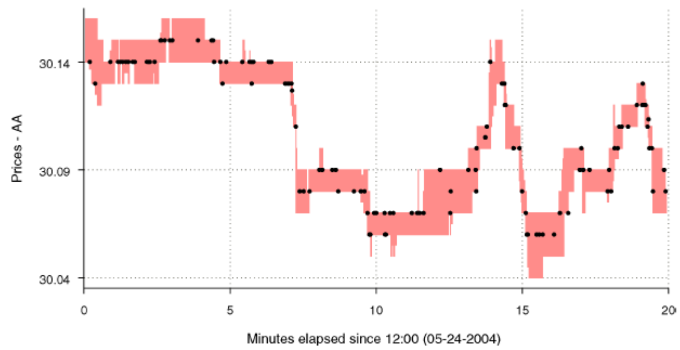
Figure 2.1: Bid Ask Bounce



Source: Snowfall Systems (2014)

- Second property appears in the multivariate dimension where **non-synchronous trading** occurs among several financial assets on high-frequency basis. One method how to deal with this issue is to sample HFD at fixed time when trades of all assets occurred i.e. to use close prices on minute basis what would cause again the loss of data. Barndorff-Nielsen *et al.* (2011) suggests a different approach, so called *refresh-time*, that synchronizes data based on the least traded asset. The synchornization could be considered both an advantage and

Figure 2.2: Bid and Ask Quotes (defined by the shaded area) and Actual Transaction Prices Over 20-Minute Subperiods on April 24, 2004 for AA



Source: Hansen & Lunde (2006)

its weak point. Both methods require liquid traded assets. However, asynchronicity is not solely an issue for HFD but it can become also for daily data from different exchanges and platforms which are recorded at different time due to different time-zones, trading hours or other reasons.

- Third property is derived from the second one. HFD are asynchronously traded, then covariance among assets with increasing sampling frequency has downward bias. This is called ***Epps effect*** by Epps (1979) who observed it in the stock prices.

2.3 Realized measures

"We designate the class of estimators of quadratic (co) variation based on the high frequency data as ''realized measures''." Bauwens *et al.* (2012). As we will show, the theory of quadratic (co) variation is the corner stone of derived RV and RCOV estimators. The foundations of realized measures were laid in the papers of Andersen *et al.* (2003) and Barndorff-Nielsen & Shephard (2004) with comprehensive summary in Andersen *et al.* (2011) and Bauwens *et al.* (2012) where we drew mostly the inspiration for following subsections. Very good job with explanation of realized measures for purpose of master thesis was already written in the master thesis of Čech (2013).

2.3.1 Realized Volatility

Let logarithmic asset price increment $\mathrm{d}S_t$ be determined by the following a continuous time mean diffusion semi-martingale process of the form:

$$\mathrm{d}S_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, 0 \geq t \geq T, \tag{2.1}$$

where μ_t is an instantaneous drift (predictable locally bounded variation process), σ_t is an instantaneous volatility (strictly positive and stochastic volatility process) and W_t is a standard Brownian motion. If we assume the discrete interval of length of 1 what would represent one day and set an instantaneous drift $\mu_t = 0$ (non-zero drifts are relevant for longer time horizons than one day and they are easy to incorporate into the model) then:

$$r_t \equiv S_t - S_{t-1} = \int_{t-1}^{t} \sigma(u) \mathrm{d}W(u), \tag{2.2}$$

and

$$r_t \sim N(0, IV_t), \tag{2.3}$$

where IV_t denotes the *integrated variance* (quadratic variation if we assume no jumps, then $QV_t = IV_t$) what is a key factor for risk management.

$$IV_t \equiv \int_{t-1}^{t} \sigma^2(u) \mathrm{d}u. \tag{2.4}$$

However, IV_t is not observable and therefore, Andersen & Bollerslev (1998) and Andersen *et al.* (1999) popularized an idea of an precise estimator of IV_t based on sums of intraday squared returns and called it *realized variance*[1]. Realized variance in its naive version on day t and at intraday frequency $\triangle$ is defined as

$$RV_t(\triangle) \equiv \sum_{j=1}^{N(\triangle)} (S_{t-1+j\triangle} - S_{t-1+(j-1)\triangle})^2 = \sum_{j=1}^{M} r_{t,j}^2, \tag{2.5}$$

subsequently *realized volatility*,

$$RVOL_t(\triangle) = \sqrt{RV_t(\triangle)}, \tag{2.6}$$

where $S_{t-1+j\triangle} \equiv S(t-1+(j-1)\triangle)$ denotes the intraday log-price at the end of the jth interval on day t, and $N(\triangle) \equiv 1/\triangle = M$ what is a number of intraday prices or returns. For instale, $M = 1380$ for 1-minute returns in a 23-hour market, corresponding to $\triangle = 1/(23 \cdot 60) \approx 0.000724$. When we let $\triangle$ to go to zero, RV estimator will approach the integrated variance (convergence in probability) defined in equation 2.4 . In order to construct RV as unbiased and consistent estimator, the intraday returns cannot be serially correlated and no market microstructure noise Bauwens *et al.* (2012) described in subsection 2.2 can be present. These conditions are not met in empirical HFD. There are several adequate solutions for this problem. First one is to find out optimal sampling frequency which reduce market microstructure and ensure acceptable bias-variance trade-off for RV estimator. Second one is about to choose more advanced estimator than in equation 2.5. The last one is to include additional characteristics of price generating

[1] We can find in the original papers or in some other cited literature the term *realized volatility* or *integrated volatility* as interchangeable terms for *realized variance*, resp. *integrated variance*. In order to keep consistency and distinguish volatility and variance, we will stick in this master thesis with notation of volatility as a square root of variance.

process such as jumps in the estimator which were omitted in the semimartingale assumption in equation 2.1 when jumps are detected in empirical HFD.

First solution can be proceed through the *signature volatility plots* which set on y-axis computed average RVOL over the different sampling frequencies and x-axis the decreasing sequence of sampling frequency (i.e. from 1 min to 120 min) and choose sampling frequency where the curve shows stabilized average RVOL. Liquid assets should have downward sloping curve and illiquid assets upward one see Andersen *et al.* (1999). Moreover, RVOL does not need to yield precise estimates for illiquid financial assets and feasible alternative to daily data based estimator is to use range-based estimator of volatility proposed in Christoffersen (2011). Suggestions for second and third solution can be found in the paper of McAleer & Medeiros (2008) and in Chapters 17. and 18. of book Bauwens *et al.* (2012) with comparison of different estimators using methods such as subsampling or kernel based one or to be robust against jumps.

The theory of construction of RV implies using only intraday returns from market open to market close what omits overnight return information that can be significant for risk management purposes. The relevant adjustments of RV are proposed Christoffersen (2011, pp.108-109).

Stylized facts

The list of stylized facts of RV was provided in Christoffersen (2011).

- *"RV is a more precise indicator of daily variance than is the daily squared return"* if conditions are met.
- *"RV has large positive autocorrelations for many lags." This feature is also called as long-memory or high persistence.*
- *"The log of RV is approximately normally distributed."*
- *"The daily return divided by the square root of RV is close to independent and identically distributed (iid) standard Gaussian."*

2.3.2 Realized Covariance

In case of multivariate dimension, we assume that prices follow a multivariate semimartingale proces:

$$\mathrm{d}\mathbf{S}_t = \boldsymbol{\mu}_t \mathrm{d}t + \boldsymbol{\Omega}_t \mathrm{d}\mathbf{W}_t, 0 \geq t \geq T, \tag{2.7}$$

where $\mathbf{S}_t$ and $\boldsymbol{\mu}_t$ is the N dimensional vector of the log prices, respectively instantaneous drifts, while $\boldsymbol{\Omega}_t$ is the $N \times N$ càdlàg process such that $\boldsymbol{\Sigma}_t = \boldsymbol{\Omega}_t \boldsymbol{\Omega}_t'$ is the instantaneous covariance matrix and $\mathbf{W}_t$ is a N dimensional vector of independent standard Brownian motion processes. Likewise the instantaneous volatility is difficult to estimate, it applies also for instantaneous covariance and hence, in practice we are interested in the estimation of *integrated covariance* (quadratic covariation without presence of jumps, $ICOV_t = QCOV_t$) for a day t,

$$ICOV_t \equiv \int_{t-1}^{t} \mathbf{\Sigma}(u)\mathrm{d}u. \tag{2.8}$$

The estimator of $ICOV_t$ is *realized covariance* computed for a day t as the sum of the products of intraday returns $\boldsymbol{r}_{j,t} = \boldsymbol{S}_{t-1+j\triangle} - \boldsymbol{S}_{t-1+(j-1)\triangle}, j = 1, \ldots, d(\triangle)$ when we assume that $\boldsymbol{\mu}(t) = \mathbf{0}$. RCOV in its naive version is defined as

$$RCOV_t(\triangle) \equiv \sum_{j=1}^{N(\triangle)} (\boldsymbol{S}_{t-1+j\triangle} - \boldsymbol{S}_{t-1+(j-1)\triangle})(\boldsymbol{S}_{t-1+j\triangle} - \boldsymbol{S}_{t-1+(j-1)\triangle})' = \sum_{j=1}^{M} \boldsymbol{r}_{t,j}\boldsymbol{r}'_{t,j}, \tag{2.9}$$

and *realized correlation* between asset i and j for day t,

$$RCOR_t = \frac{RCOV_{i,j,t}}{RVOL_{i,t}RVOL_{j,t}} \tag{2.10}$$

Similarly to univariate case, if we let $\triangle$ to go to zero, RCOV estimator will converge to the integrated covariance matrix of the continuous stochastic volatility process on day t under conditions that there are no market microstructure noise, Epps effect and asset returns are linearly independent, so RCOV will be an unbiased and consistent estimator. Obviously, these conditions are hardly fulfilled in reality. Additional requirement is that number of sampled intraday returns for one day is not lower than number of assets N, otherwise $RCOV_t$ would become singular. Singularity or mentioned market microstructure noise cause that $RCOV_t$ is not guaranteed to be PSD while PSD being a desired property for application in risk management.

The simplest way how to mitigate market microstructure noise, Epps effect, non-PSD is to apply calendar time sampling with fairly low (optimal) frequency, synchronization at fixed clock time and use number of assets that do not exceed the number of observations for one day. More advanced method was proposed in Barndorff-Nielsen *et al.* (2011) who constructed multivariate realised kernel that is robust to noise, guarantee PSD characteristic and uses refresh time technique to synchronize multivariate HFD. Other estimators which are even robust to jumps can be found in chapters 1., 13. and 17. of bookBauwens *et al.* (2012). The theory of construction of RCOV implies using only intraday returns from market open to market close what omits overnight return information that can have substantial impact for risk management purposes.

2.4 Risk measures

The term "Risk" does not have unambiguous definition, for instance the *Concise Oxford English Dictionary* defines it as *"hazard, a chance of bad consequences, loss or exposure to mischance"* McNeil *et al.* (2015). In this master thesis we will focus on financial risk, particularly market

risk [2] that has very apt definition as *"the risk of a change in the value of a financial position or portfolio due to changes in the value of the underlying components on which that portfolio depends, such as stock and bond prices, exchange rates, commodity prices, etc."* McNeil *et al.* (2015). Mentioned "underlying components" are perceived also as *risk factors.* Market risk represents possible downside what means a possible loss but it is not associated exclusively with decline of price because the profit of short position is made thanks to decline of price.
Now we can set up a framework for modeling the value of a financial position or portfolio and its change. Let introduce the formal definitions based on Danielsson (2011).

Definition 2.1 (**Risk measure**). *A risk measure is a mathematical method for computing risk.*

Definition 2.2 (**Risk measurement**). *A number that captures risk. It is obtained by applying data to a risk measure.*

Exceptional explaining guidance through risk measures was written in chapter 2. McNeil *et al.* (2015) where we pick up the most important parts for the purpose of this master thesis. Considering the change of portfolio value as $\triangle V_t = V_t - V_{t-1}$, then the *loss* is $L_t := -\triangle V_t$ where we can neglect time value of money if time horizon $\triangle$ is short, i.e. one day. The L_t is random value from point of t and its distribution is called *loss distribution.* Modeling of V_t is done as a function of time and a N-dimensional random vector $\boldsymbol{Z_t} = (Z_{t,1}, \ldots, Z_{t,d})$ of risk factors,

$$V_t = f(t, \boldsymbol{Z_t}) \tag{2.11}$$

for some measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. Observable risk factors of random vector $\boldsymbol{Z_t}$ takes realized value $\boldsymbol{z_t}$ at time t while the portfolio V_t has realized value $f(t, \boldsymbol{z_t})$. The decision about the choice of risk factors and of f is crucial in risk modeling and it depends on the characteristics of portfolio, available data and on the desired level of sophistication to achieve.
Let assume an increment in risk factors as $\triangle \boldsymbol{Z}_t = \boldsymbol{Z}_t - \boldsymbol{Z}_{t-1}$. Then portfolio loss has following form

$$L_t = -(f(t, \boldsymbol{z}_t + \triangle \boldsymbol{Z}_t) + f(t-1, \boldsymbol{z}_{t-1})), \tag{2.12}$$

we see that the distribution of of loss is given by the distribution of the risk factor increment $\triangle \boldsymbol{Z}_t$. In order to find a change of portfolio loss, we can assume that f is differentiable, then we can use a first-order approximation $L_t^{\triangle}$ of the portfolio loss of the form

$$L_t^{\triangle} = -\left(f_{t-1}(t-1, \boldsymbol{z}_{t-1}) + \sum_{i=1}^{d} f_{zi}(t-1, \boldsymbol{z}_{t-1}) \triangle \boldsymbol{Z}_{t,i} \right), \tag{2.13}$$

where $L_t^{\triangle}$ denotes to linearized portfolio loss, the subscripts on f denote partial derivatives. Thanks to first-order approximation, we can represent the portfolio loss as a linear function of

[2] Other categories of risk are: credit and operational (main interest in the banking industry), then liquidity (can be closely connected with market risk when agent is capable to influence significantly the risk factors) and model risk, see definitions in McNeil *et al.* (2015)

the risk factor increments. However such an approximation gives accurate result if the risk factor increments are small (this is achieved if we use only short time horizon, i.e. one day) and if the portfolio value is almost linear in the risk factors, so f would have small second derivative (this is achieved when we use as risk factors only financial assets which have payoff diagram linear).

Practitioners in risk management usually use logarithmic prices as risk factors and their difference gives geometric return, i.e. $\triangle Z_{t,i} = \ln S_{t,i} - \ln S_{t-1,i}$, where $S_{t,i}$ is the price process of asset i. Portfolio value can be computed as $V_t = \sum_{i=1}^{d} w_i S_{t,i}$ where w_i is the weight of asset i and $\sum_{i=1}^{d} w_i = 1$. Linearization of the portfolio loss allow us to compute first two moments of the distribution of $\triangle L_t$ as

$$E(L_t^{\triangle}) = -V_t \boldsymbol{w}' \boldsymbol{\mu} \tag{2.14}$$

$$Var(L_t^{\triangle}) = V_t^2 \boldsymbol{w}' \boldsymbol{\Sigma} \boldsymbol{w} \tag{2.15}$$

Currently the most popular risk measures in financial institutions and especially their regulators are based on loss distribution. These risk measures allow to compute the conditional or unconditional loss distribution of the portfolio over given horizon $\triangle t$, incorporate netting and diversification effects. The representatives are VaR and ES. Nevertheless, the level of accuracy depends on how properly are these loss distributions estimated and hence, we find more approaches. The simplest one is an *analytical method* using well-known defined statistical distributions such as Normal or Student's t, *historical simulations* using empirical distribution and *Monte Carlo simulations* using an explicit parametric model for risk factors or Copula function to construct unique multivariate distributions from various different marginal distributions and their dependency structure separately. The issue of loss distributions is more discussed in the subsection 3.2.

The latest inventions or suggestions as risk measures are *Expectiles* with attractive features proposed by Emmer *et al.* (2015), *Shortfall Deviation* and with combination of ES called *Shortfall Deviation Risk* proposed by Righi & Ceretta (2015).

2.4.1 Value at Risk

VaR is mostly adapted risk measure in financial industry and its regulation since Basel regulatory framework chose VaR as main measure to calculate capital requirements. We continue with previous concept of loss distribution of a portfolio of financial assets and a fixed time horizon $\triangle t$ defined as: $F_L(l) = P(L \leq l)$. As pointed out in McNeil *et al.* (2015) maximum possible loss of such portfolio neglects any probability information, so its information value is very limited. Therefore VaR combines these two terms as *"maximum loss that is not exceeded with a given high probability"* McNeil *et al.* (2015) over time horizon $\triangle t$. Formal definition is following.

Definition 2.3 (Value at Risk). *Given some condifence level $\alpha \in (0, 1)$, the VaR of a portfolio* *with loss L at the confidence level α is given by the smallet number l, such that the probability that the loss L exceeds l is no larger than $1-\alpha$. Formally,*

$$\mathrm{VaR}_\alpha = \mathrm{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : P(L > 1)1-\alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \leq \alpha\}. \tag{2.16}$$

Statistically speaking VaR is a *quantile* of the loss distribution. The time horizon[3] considered for a bank's trading desk is a one day, ten business days for calculation of capital requirements and one year for credit and operational risk management. The confidence level is usually within range 95

The choice of parameters depends how risk averse is the agent or its regulator, than the availability of data and computational burden as higher confidence level and longer time horizon will be more demanding for data and computation capacity.

Regarding to multivariate dimension, there are basically two approaches Andersen *et al.* (2011). First one is called *portfolio-level* where portfolio return is calculated as $r_{p,t} = \sum_{i=1}^{d} w_{t,i} r_{t,i}$ of $i = 1, \dots, d$ and thereinafter considered as univariate time series. The drawback is, we cannot track and compute the dependency between assets.

***Example* 2.1.** This is an example how VaR is derived analytically under condition $F =$ Gaussian distribution. Let consider $L \overset{iid}{\sim} F(\mu, \sigma^2)$ with confidence level α and time horizon one day[4], then

$$P(L \leq l) = P\left(\frac{L-\mu}{\sigma} \leq \frac{l-\mu}{\sigma}\right) = P\left(L^{std} \leq \frac{l-\mu}{\sigma}\right) = \alpha, \tag{2.17}$$

where $L^{std} \sim F(0,1)$, F is the distribution of the standardized loss L^{std} or *innovations*. Then $P(L^{std} \leq F_\alpha^{-1}) = \alpha$, so

$$\frac{l-\mu}{\sigma} = F_\alpha^{-1} \tag{2.18}$$

then VaR is defined

$$\mathrm{VaR}_\alpha(L) = \mu + \sigma F_\alpha^{-1} \qquad \text{or} \qquad \mathrm{VaR}_\alpha(L) = \mu + \sigma q_\alpha(F) \tag{2.19}$$

[3] while the portfolio is held unchanged

[4] In practice, the most common time horizon is one day. However, for capital requirements there is a requirement to calculate VaR for ten business days. If we apply the same method of calculation as we describe below and in Chapter 3, then in order to have non-overlapping data, our data will shrink by ten times what would dramatically influence the quality of forecasted VaR. The simple solution is scaling by *square-root-of-time* rule, where we scale up day ahead forecast of volatility to h business days by square root of h, i.e. ten business days $\mathrm{VaR}_\alpha^{(h)}(L) = \mu + \sqrt{10}\sigma F_\alpha^{-1}$ what implies $\mathrm{VaR}_\alpha^{(h)}(L) = \sqrt{10}\mathrm{VaR}_\alpha^{(1)}(L)$. This approach holds only under the conditions that risk-factor change distribution is iid and Gaussian. Otherwise square-root-of-time is only an approximation. The limitation is overcome by Monte Carlo simulation when risk-factor change is simulated M times for time horizon h and resulting h day loss distribution is used for calculation of VaR and also ES since the same technique described above applies also for ES measure. More detailed explanation can be found in McNeil *et al.* (2015, pp. 349-351) where the inspiration for this foot note was taken from.

Second approach is called *asset-level* which takes into account an N-dimensional loss distribution and it allows us to compute covariance matrix that tracks the linear dependency and hence, active risk management. We will use in this master thesis asset-level.

We can follow an 2.1 but in multivariate dimension where we assume $L^{std} \sim F(\mathbf{0}, \boldsymbol{I_d})$ and the VaR will follow of the form:

$$\mathrm{VaR}_\alpha(L) = \boldsymbol{w'\mu} + \sqrt{\boldsymbol{w'\Sigma w}}q_\alpha(F) \tag{2.20}$$

Such a linear transformation is possible only for elliptical distribution of loss, for instance: Normal and Student's t according to Jondeau *et al.* (2007).

Methods of VaR computation

1. **Variance-covariance or parametric**. This methods is based on the estimation of covariance matrix and mean vector of asset losses proceeded with further linear transformation to univariate dimension of portfolio mean and variance. It requires an assumption of elliptical loss distribution. Since mean vector is usually assumed to be equal to zero for short time horizon, then this method depends solely on the technique of estimation of covariance matrix and choice about the loss distribution. Since loss distribution is assumed to be described completely by its parameters that is the reason this method is called also as *parametric*. Portfolio-level VaR requires only model for estimation of variance and asset-level, obviously, estimation of covariance matrix.

2. **Semi-parametric**. We can consider this method as combination of parametric model and empirical loss distribution. Typical representative on asset-level approach would be Multivariate FHS of Christoffersen (2011, pp. 194-195). For the portfolio-level VaR it would be univariate FHS, EVT[5] and Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR) of Engle & Manganelli (2004).

3. **Non-parametric**. There is only one method for portfolio-level VaR called Historical simulations.

4. **Monte Carlo simulations**. This method complies with description mentioned already in section 2.4. Additionall remark, there is a possibility to combine various models, for instance for standardization of losses, it can be used volatility model (i.e. GARCH) and EVT together for more accurate further estimation with Copula function.

Coherency

Since introduction of VaR in the early '90 and its wide range of applications, there was still a lack of theoretical framework what a good risk measure should constitute. The turning point was done by seminal works of Artzner *et al.* (1997) and Artzner *et al.* (1999) who defined axioms (a

[5] Depending on the method of estimation, EVT can belong either to parametric or semi-parametric class

list of properties) of risk measures which fulfillment would provide us a coherent risk measure.

Let consider risk measures to be a real-valued functions defined on a linear space of random variables $\mathcal{M}$, assumed to include constants where elements L_1 and L_2 of $\mathcal{M}$ represent losses of any two portfolios and risk measure denotes to $\varrho(.)$ McNeil *et al.* (2015, pg. 73).

Axiom 2.1 (**Monotonicity**)**.** For $L_1, L_2 \in \mathcal{M}$ such that $L_1 \leq L_2 \implies \varrho(L_1) \leq \varrho(L_2)$.

Portfolio with higher losses implies higher riskiness, i.e. higher capital requirements.

Axiom 2.2 (**Translation invariance**)**.** For all $L \in \mathcal{M}$ and every $lin\mathbb{R}$ we have $\varrho(L + l) = \varrho(L) + l$.

If we add or subtract a deterministic loss l to a portfolio loss L, the riskiness of such a portfolio will be changed exactly by that quantity l.

Axiom 2.3 (**Subaddivitity**)**.** For all $L_1, L_2 \in \mathcal{M}$ we have $\varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2)$.

This axiom is the most discussed from all of them because it states the riskiness of diversified portfolio should be the exact or lower than non-diversified portfolio. If a risk measure does not hold this axioms, it means diversification of agent's portfolio lead to higher capital requirements what is exactly in contradiction of agent's motivation to diversify trading activities.

Axiom 2.4 (**Positive homogeneity**)**.** For all $L \in \mathcal{M}$ and every $\lambda > 0$ we have $\varrho(\lambda L) = \lambda\varrho(L)$.

If axiom 2.3 holds, then this axiom is justified but only till to certain extend of λ. The reason is for very high values of multiplier λ the agent concentrates a significant position in such a portfolio and if he decides to liquidate its position in short time horizon, it will have huge impact on price (this holds also for opposite direction if an agent wants to create a significant position for short time horizon). Hence, the risk measure should add to market risk also a liquidity risk, so we get $\varrho(\lambda L) > \lambda\varrho(L)$. Problem is that it contradictory to subadditivity and it was the reason to introduce convex risk measure Embrechts *et al.* (2016).

We say the risk measure $\varrho(.)$ is *coherent* when it satisfies all four above axioms.

Axiom 2.5 (**Convexity**)**.** For all $L_1, L_2 \in \mathcal{M}$ and all $\lambda \in [0, 1]$ we have $\varrho(\lambda L_1 + (1 - \lambda)L_2) = \lambda\varrho(L_1) + (1 - \lambda)\varrho(L_2)$.

The idea is again the diversification should reduce the riskiness.

We say the risk measure $\varrho(.)$ is *convex* when it satisfies axioms 2.1, 2.2 and 2.5. Every risk measure which is coherent is also convex but vice versa it does not hold.

Advantages

The list of advantages of VaR is well elaborated in Dowd (2005, pp. 11-13). The summary of advantages can be characterized as VaR is the first risk measure which can be computed across different position and risk factors and hence, it allows to aggregate the risk of subpositions into one portfolio risk using probability, so an agent can join computed risk with level of probability and finally, the output is a single number in the unit of ”lost money” that is easily understood through the all departments in the agent’s structure and its stakeholders such as regulator or clients. Moreover VaR is easier to estimate for heavy-tail loss distributions and backtest than ES McNeil *et al.* (2015, pg. 77).

Disadvantages

Each coin has two sides and VaR is not an exception. The following list of disadvantages was provided in Danielsson (2011, pp. 80-85).

1. **VaR is only a quantile on the loss distribution.** This is a conceptual flaw of VaR because it gives us no information what a loss an agent can face to if the VaR is exceeded on given confidence level. The most extreme situation can happen in the credit risk when an occurrence of defaults starts beyond VaR confidence level, then VaR is not able to detect any risk and returns zero. The other implications is that VaR would be the same for two financial assets but due to different tails in loss distributions the breach of VaR will cause different losses, so these assets would be equally risky only till VaR’s confidence-level.

2. **VaR is is not a coherent risk measure** due to violence of subadditivity axiom when loss distributions of financial assets does not come from multivariate Gaussian distribution. The empirical loss distribution of financial assets was found already by Mandelbrot (1963) is non-Gaussian and since then it became an stylized fact[6]. Therefore an agent will face to increased capital requirements due to diversification.

3. **VaR is easy to manipulate.** VaR as defined as it is above due to assumption of linear payoff of underlying risk factors. The involvement of non-linear risk factors in the portfolio such as options will cause artificial decrease of VaR but the real riskiness will be increased Danielsson (2011, pp. 84-85).

2.4.2 Expected Shortfall

The first and second diasadvantange of VaR let Artzner *et al.* (1999) to propose a new risk measure named *Expected Shortfall* which does not suffer those disadvantages. The ES answers the question what is an expected loss given that VaR was exceeded or in other words, mean value of the worst losses exceeding given confidence level.

[6]It was confirmed by many researchers but with technological development, the financial markets become more and more efficient and currently, the most liquid markets such as foreign exchange market have the shape of loss distribution the closest to Gaussian one

Definition 2.4 (Expected Shortfall). For a loss L with $E(|L|) < \infty$ and distribution function F_L, the ES at confidence level $\alpha \in (0, 1)$ is defined as

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)\mathrm{d}u, \tag{2.21}$$

where $q_u(F_L)\mathrm{d}u = F_L^{\leftarrow}(u)$ is the quantile function of F_L. Then connection between ES and VaR is

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\mathrm{d}u, \tag{2.22}$$

or simple notation

$$\mathrm{ES}_\alpha(L) = \mathrm{E}[L|L > \mathrm{VaR}_\alpha] \tag{2.23}$$

When we follow an example 2.1, then ES would be defines for univariate and multivariate dimension as

$$\mathrm{ES}_\alpha(L) = \mu + \sigma\mathrm{ES}_\alpha(F) \qquad \text{resp.} \qquad \mathrm{ES}_\alpha(L) = \boldsymbol{w}'\boldsymbol{\mu} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}}\mathrm{ES}_\alpha(F) \tag{2.24}$$

Based on above formulas, we conclude that we can use the same methods of computation as in case of VaR.

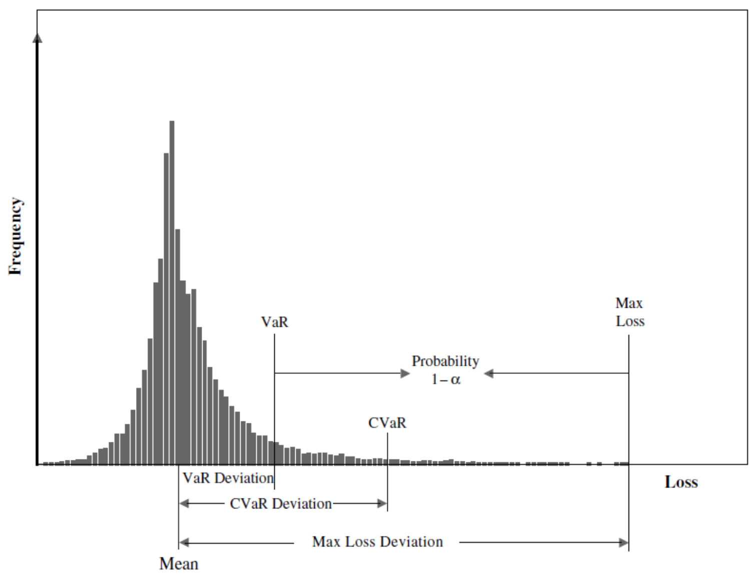
Advantages

Unlike VaR, ES is a coherent risk and convex measure and reflects the tail of loss distribution beyond VaR McNeil *et al.* (2015, pg. 77).

Disadvantages

ES posses also drawbacks that did not occure by VaR for instance, it is more difficult to estimate for heavy-tail loss distributions and to backest McNeil *et al.* (2015, pg. 77). We will discuss more the issue of ES backtesting in subsection 3.4.

Figure 2.3: VaR and ES = Conditional VaR (CVaR)



Source: Sarykalin *et al.* (2014)

Chapter 3

Methodology

The purpose of this chapter is to present individual methodological approaches chosen for this master thesis. We will apply each method of computation of VaR and ES mentioned in the subsection 2.4.1 separately for data sampled on daily and high-frequency basis. Lastly, we describe backtesting methods for each risk measure in order to test the validity of each model and compare to each other.

3.1 Model Choice

In this subsection we characterize models which are used in all three mentioned methods of computation ofVaR and ES. We decided to use Multivariate version of EWMA and DCC-GARCH as representatives using daily data. They were chosen based on their high popularity for estimation of covariance matrix, especially in among academics, and often are set as benchmark models using daily data. These models are able to estimate volatility or covariance matrix which are the key parameters for parametric method or also used for standardization of financial asset returns for semi-parametric and Monte Carlo simulation methods. The opponent to them will be Multivariate HAR model using high-frequency data that is also very popular and parsimonious model to estimate.

The continuation of the theoretical framework from section 2.4 and from now on notation will be following.

According to McNeil *et al.* (2015, pg. 338), we introduce the *loss operator* at time t, written $l_{[t]} : \mathbb{R}^d \to \mathbb{R}$ which maps risk-factor changes into losses[1]. Portfolio loss on will be given as $L_t = l_{[t-1]}(\boldsymbol{r}_t)$ where $\boldsymbol{r}_t = \triangle \boldsymbol{Z_t} = \ln \boldsymbol{S}_t - \ln \boldsymbol{S}_{t-1}$ where $\boldsymbol{S}_t$ vector of prices of financial assets in the portfolio at time t.

[1] For further references see Chapter 9. in McNeil *et al.* (2015)

Additional property of estimation of risk measures is that they use the recent available data set ($\mathscr{F}_t$ denotes sigma algebra) and therefore they become conditional risk measures which uses *conditional loss distribution* what is the *"the distribution of the loss operator $l_{[t-1]}(.)$ under $F_{\boldsymbol{r}|\mathscr{F}_{t-1}}$, that is, the distribution with distribution function $F_{L_t|\mathscr{F}_{t-1}}(l) = P(l_{[t-1]}(\boldsymbol{r}_t) \leq l|\mathscr{F}_{t-1})$.* McNeil *et al.* (2015, pg. 339).The disadvantage of conditional approach is that if used conditional distributional function does not contain high volatility period, then resulting conditional risk measure can be underestimated. The opposite approach is an unconditional one which is based on assumption that $\boldsymbol{r}_t$ forms a stationary[2] multivariate time series. So we need to *"estimate a stationary distribution function $F_{\boldsymbol{r}}$ of the time series and the evaluate the unconditional loss distribution of L_t"* McNeil *et al.* (2015, pg. 339). If $\boldsymbol{r}_t$ are iid then we get $F_{\boldsymbol{r}_t|\mathscr{F}_{t-1}} = F_{\boldsymbol{r}}$. However, as we mentioned in the section 2.1, compounded returns of financial assets usually show volatility clustering what violates iid condition, so previous equation does not hold McNeil *et al.* (2015, pg. 339).

3.1.1 EWMA

First model is EWMA which was suggested by RiskMetrics$^{\text{TM}}$ J.P.Morgan (1996).

Portfolio-level We consider setup proposed by RiskMetrics$^{\text{TM}}$. We assume that, given $\mathscr{F}_{t-1}$, $r_{p,t} = \mu_t + \sigma_t \epsilon_t$, $r_{p,t} \sim N(\mu_t, \sigma_t^2)$, and $\epsilon_t \sim N(0,1)$, where N denotes to univariate Gaussian distribution, ϵ_t is innovations, $\mu_t = 0$, decay factor $lambda = 0.94$ for 1-day time horizon and .

The conditional variance is estimated recursively by

$$\sigma_{p,t}^2 = \lambda \sigma_{p,t-1}^2 + (1-\lambda) r_{p,t-1}^2, 0 < \lambda < 1 \tag{3.1}$$

The decay factors represent the weight of the previous observations and decay exponentially. The initial value of σ_p^2 is usually set as the unconditional variance of the data and its influence is negligible after about 30 days (known as *burn time*) Danielsson (2011, pg. 60). If we use λ estimated by RiskMetrics$^{\text{TM}}$, then we have model-free approach. On the other hand, if we choose to estimate λ, then we can consider EWMA model as a special version of IGARCH(1,1) model and use it for estimation of λ for further reference with example see Tsay (2012, pp.252-255).

Asset-level We augment the assumptions from univariate to multivariate dimension that $\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{A}\boldsymbol{\epsilon}_t$, $\boldsymbol{r}_t \sim N_d(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{\epsilon}_t \sim N_d(\boldsymbol{0}, \boldsymbol{I}_d)$, where N_d denotes to multivariate Gaussian distribution and $\boldsymbol{\mu}_t = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_t = \boldsymbol{A}\boldsymbol{A}'$.

The conditional covariance matrix is estimated recursively by

[2] The multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *strictly* stationary if $(\boldsymbol{X}_{t'_1}, \ldots, X_{t'_n}) \stackrel{\text{d}}{=} (\boldsymbol{X}'_{t_1+k}, \ldots, X'_{t_n+k})$ for all $t_1 \ldots, t_n, k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$. The multivariate time series $(\boldsymbol{X}_t)_t \in \mathbb{Z}$ is *covariance* stationary (or *weakly* if the first two moments exist and satisfy $\boldsymbol{\mu}(t) = \boldsymbol{\mu}, t \in \mathbb{Z}$ and $\Gamma(t,s) = \Gamma(t+k, s+k), t, s, k \in \mathbb{Z}$. A strictly stationary multivariate time series with finite covariance matrix is covariance stationary but vice versa it does not hold. Moreover, it is possible to define infinite-variance processes (including certain multivariate ARCH and GARCH processes) that are strictly stationary but not covariance stationary McNeil *et al.* (2015, pg. 540).

$$\mathbf{\Sigma}_t = \lambda \mathbf{\Sigma}_{t-1} + (1-\lambda)\boldsymbol{r}'_{t-1}\boldsymbol{r}_{t-1} \tag{3.2}$$

The decay factor 0.94 was estimated by RiskMetrics$^{\text{TM}}$ but decay factor $lambda = 0.94$ for 1-day time horizon.

Advantages It is simple and fast estimation of large covariance matrices and required positive semi-definiteness is always guaranteed (conditional portfolio volatility is non-negative).

Disadvantages It does not allow to incorporate leverage effect, counterfactual longer-horizon forecasts Christoffersen (2011), heavy dependence on the accuracy of estimation of λ which has the same degree for all elements of the covariance matrix and lack of mean-reversion Andersen *et al.* (2011).

3.1.2 DCC-GARCH

DCC-GARCH is the model which estimates the conditional covariance matrix as its decomposition into diagonal matrix of conditional volatility and correlation matrix and it allows to incorporate the most of stylized facts about financial asset returns.

GARCH

Initially, we describe univariate estimation of conditional variance for portfolio-level approach which is applied later in asset-level approach.

Symmetric - GARCH(1,1) The GARCH was introduced by Bollerslev (1986) which become the most common used model in practice from the rich family of GARCH models drawn up by Bollerslev (2008).

We assume that, given $\mathscr{F}_{t-1}$, $r_{p,t} = \mu_t + \sigma_t \epsilon_t$, $r_{p,t} \sim N(\mu_t, \sigma_t^2)$, and $\epsilon_t \sim N(0,1)$, where N denotes to univariate Gaussian distribution, ϵ_t is innovations, $\mu_t = 0$, then GARCH(s[3],q) is defined as

$$\sigma_{p,t}^2 = \omega + \sum_{i=1}^{s} \alpha_i r_{p,t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{p,t-j}^2 \tag{3.3}$$

and the simplest GARCH(1,1) which employes one lag is defined as

$$\sigma_{p,t}^2 = \omega + \alpha r_{p,t-1}^2 + \beta_j \sigma_{p,t-1}^2 \tag{3.4}$$

where $\omega > 0, \alpha \leq 0, \beta \leq 0$ (positive volatility is ensured) and $\alpha + \beta < 1$ what implies that the unconditional variance of r_t is finite (covariance stationarity is ensured) and it is defined as $\sigma_p^2 = \frac{\omega}{1-\alpha-\beta}$ (if $\alpha + \beta = 1$, then σ^2 is infinite and otherwise undefined) while its conditional variance σ_t^2 is time variant Tsay (2012). The distribution of ϵ is not limited only to Gaussian one

[3] Usual label is p but we use p as portfolio in this master thesis

but in the context of financial asset returns and risk measure, there are often used Student's*t*, Generelized Error Distribution, etc.

Asymmetric - GJR-GARCH(1,1) The asymmetric version of GARCH(1,1) with acronym GJR is able to accommodate the leverage effect and is commonly used in the practice besides of EGARCH. It was introduced by Glosten *et al.* (1993).

The model GJR-GARCH(1,1) follows of the form:

$$\sigma_{p,t}^2 = \omega + \alpha r_{p,t-1}^2 + \gamma r_{p,t-1}^2 I(r_{p,t-1} < 0) + \beta_j \sigma_{p,t-1}^2 \tag{3.5}$$

where $I(.)$ denotes the indicator function.

Advantages The GARCH models offer wide flexibility to adapt to a specific characteristics of financial asset returns such as volatility clustering and leverage effect and at the same time they stay parsimonious (other models can accommodate even long-memory or other stylized facts). Moreover, it adds that its forecasts revert back to the long-run variance. The estimation is based on Quasi Maximum Likelihood Estimation (QMLE) that allows to violate assumption of conditional Gaussian distribution if mean and variance are correctly specified with the cost that estimates of QMLE are less efficient than Maximum Likelihood Estimation (MLE) estimates Christoffersen (2011).

Disadvantages Difficulty to forecast the entire conditional distribution Christoffersen (2011) or possible issues in the optimization if likelihood function has multiple local minima or numerical instability, especially for higher magnitute of lags Danielsson (2011). The papers specialized in the issues of GARCH can be find in the Further resources of chapter 4. in Christoffersen (2011).

DCC

We now turn to the asset-level approach using DCC model introduced by Engle (2000) to model covariance matrix. Unlike EWMA, this model consists of two components that are estimated separately. First component is a diagonal matrix of conditional volatilities denotes to $\boldsymbol{D}_t$ and second component is time variant conditional correlation matrix denotes to Γ_t.

Symmetric - DCC Basic version of DCC model is its symmetric version. We use the assumptions $\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{A}\boldsymbol{\epsilon}_t$, $\boldsymbol{r}_t \sim N_d(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{\epsilon}_t \sim N_d(\boldsymbol{0}, \boldsymbol{I}_d)$, where N_d denotes to multivariate Gaussian distribution and $\boldsymbol{\mu}_t = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_t = \boldsymbol{A}\boldsymbol{A}'$. The covariance matrix is given by:

$$\Sigma_t = \boldsymbol{D}_t \Gamma_t \boldsymbol{D}_t, \tag{3.6}$$

where

$$\boldsymbol{D}_t = \begin{pmatrix} \sigma_{t,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{t,d} \end{pmatrix} \tag{3.7}$$

and

$$\Gamma_t = diag\{\boldsymbol{Q}_t\}^{-1/2}\boldsymbol{Q}_t diag\{\boldsymbol{Q}_t\}^{-1/2} \tag{3.8}$$

with $\boldsymbol{Q}_t$ given by:

$$\boldsymbol{Q}_t = (1-\alpha-\beta)\bar{\boldsymbol{Q}} + \alpha(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \beta\boldsymbol{Q}_{t-1} \tag{3.9}$$

where $\bar{\boldsymbol{Q}} = E[\boldsymbol{\epsilon_t}\boldsymbol{\epsilon}'_t$ is the unconditional correlation matrix of epsilons that are given $\boldsymbol{\epsilon}_t = \boldsymbol{D}_t^{-1}\boldsymbol{r}_t$, α and β are non-negative real numbers satisfying $0 < \alpha + \beta < 1$ to ensure positive definiteness , stationarity and that models becomes mean reverting.

Asymmetric - ADCC As volatility exhibits asymmetric response to positive and negative shocks, the similar case is also for covariance where the negative shock can induce disproportional joint reaction of financial assets if the same magnitude of the shock would be positive. The asymmetric version of DCC (scalar) was introduced by Cappiello *et al.* (2006) as adjusted version of equation 3.9 and as a special case of Asymmetric generalized DCC .

$$\boldsymbol{Q}_t = (\bar{\boldsymbol{Q}} - \alpha^2\bar{\boldsymbol{Q}} - \beta^2\bar{\boldsymbol{Q}} - \gamma^2\bar{\boldsymbol{N}}) + \alpha^2(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \gamma^2(\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1}) + \beta\boldsymbol{Q}_{t-1} \tag{3.10}$$

where $\boldsymbol{\eta}_t = I(\epsilon_t < 0) \circ \epsilon_t$, $I(.)$ is a $d \times 1$ indicator function whole $\circ$ denotes the Hadamard product, $\bar{\boldsymbol{N}} = E(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t)$ and if condition $\alpha^2 + \beta^2 + \delta\gamma^2 < 1$ holds, where $\delta =$ maximum eigenvalue $[\bar{\boldsymbol{Q}}^{-1/2}\bar{\boldsymbol{N}}\bar{\boldsymbol{Q}}^{-1/2}]$, then $\boldsymbol{Q}_t$ is guaranteed to be positive definite.

Advantages We can estimate each variance of d financial assets separately using different GARCH models and then time depending conditional correlation matrix. This models handle to model also a large covariance matrix.

Disadvantages The parameters α and β are constant values meaning that *"the conditional correlations of all assets are driven by the same underlying dynamics - often an unrealistic assumption"* Danielsson (2011).

3.1.3 HAR

In the section 2.3 we defined estimators of RV and RCOV. We now illustrate how these estimators can be forecasted with HAR model of Corsi (2009) for portfolio-level and its multivariate version of Chiriac & Voev (2011) for asset-level. The choice of this model is based on its ability to capture the one of key stylized facts that is high persistence and its simplicity in comparison with initially proposed ARFIMA model in Andersen *et al.* (1999).

The meaning of long-memory feature is that high autocorrelation function decays slowly at a hyperbolic rate and characterized by fractional integration whereas short-memory model's (Autoregressive model (AR) or Autoregressive Moving Average (ARMA)) autocorrelation function would decay faster at an exponential rate.

The economic interpretation of long-memory property of volatility was proposed in the work of Müller *et al.* (1997) who introduced heterogeneous market hypothesis. The main idea is that *"different market agent types or components perceive, react to, and cause different types of volatility."*. The meaning of "different" is in terms of time resolution that volatility is measured, i.e. short-term trades (intraday traders, dealers or market makers) watch, trade, analyze the situation on high-frequency basis wheres long-term traders (financial institutions) monitor and trade on daily or lower frequency. This heterogeneous behaviour of market participants cause aggregation of these different volatilities and so high persistence according to Bauwens *et al.* (2012, pg.364). This idea has been followed in Corsi (2009), who introduced symmetric HAR model as *"an additive cascade model of realized volatility aggregated at different time horizons"* [pg. 364]Bauwens2012.

Symmetric

Portfolio-level Univariate version of HAR model is not an exact long-memory model such as ARFIMA but only approximate because it is type of AR-model estimated with Ordinary Least Squares (OLS) method.

We assume that, given $\mathscr{F}_{t-1}$, $r_{p,t} = \mu_t + \sigma_t \epsilon_t$, $r_{p,t} \sim N(\mu_t, \sigma_t^2)$, and $\epsilon_t \sim N(0,1)$, where N denotes to univariate Gaussian distribution, ϵ_t is innovations, $\mu_t = 0$ and $\sigma_t^2 = RV_t^{(d)}$, then HAR is defined as

$$RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \nu_t, \nu_t \overset{iid}{\sim} N(0, \sigma_\nu^2) \tag{3.11}$$

where individual components of HAR are calculated as daily, weekly and monthly realized volatility in terms of business days and vector of parameters $\boldsymbol{\beta} = (c, \beta^{(d)}, \beta^{(w)}, \beta^{(m)})$ is estimated by the OLS regression.

$$\begin{aligned} RV_t^{(d)} &\equiv RV_t \\ RV_t^{(w)} &\equiv RV_{t-4,t} = \frac{[RV_{t-4,t} + RV_{t-3,t} + RV_{t-2,t} + RV_{t-1,t} + RV_{t-4,t}]}{5} \\ RV_t^{(m)} &\equiv RV_{t-20,t} = \frac{[RV_{t-20,t} + RV_{t-19,t} + \cdots + RV_t]}{21} \end{aligned} \tag{3.12}$$

Given the log normal property of RV

$$\ln(RV_{t+1}^{(d)}) = c + \beta^{(d)} \ln(RV_t^{(d)}) + \beta^{(w)} \ln(RV_t^{(w)}) + \beta^{(m)} \ln(RV_t^{(m)}) + \nu_t \tag{3.13}$$

Asymmetric

We can find different implementations of leverage effect into HAR model, i.e. Corsi & Reno (2009b) or in Chapter 15. Bauwens *et al.* (2012) including incorporation of jumps. However, we chose the version of asymmetric HAR introduced in the article of Allen *et al.* (2014) that even implement the asymmetry of realized volatility of risk-factors as well as asymmetry of volatility of realized volatility and hence, they call it *a dually asymmetric realized volatility model.* For the scope of this master thesis, it will be satisfying to incorporate only the former degree of asymmetry that has the same construction as the asymmetry in GJR-GARCH(1,1). Simplified or "singular" asymmetric HAR of DARV-HAR suggest by Allen *et al.* (2014) is defined as

$$\begin{aligned} RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} \\ + \gamma^{(d)} r_{p,t}^{(d)} I(r_{p,t}^{(d)} < 0) + \gamma^{(w)} r_{p,t}^{(w)} I(r_{p,t}^{(w)} < 0) \\ + \gamma^{(m)} r_{p,t}^{(m)} I(r_{p,t}^{(m)} < 0) + \nu_t, \quad \nu_t \overset{iid}{\sim} N(0, \sigma_\nu^2) \end{aligned} \tag{3.14}$$

where $I(.)$ is an indicator function, $r_{p,t}^{(d)}$ is a daily financial asset return, $r_{p,t}^{(w)}$ and $r_{p,t}^{(m)}$ are cumulated returns for weekly, respectively monthly period and vector of parameters $\boldsymbol{\beta} = (c, \beta^{(d)}, \beta^{(w)}, \beta^{(m)}, \gamma^{(d)}, \gamma^{(w)}, \gamma^{(m)})$ is estimated by the OLS regression.

Asset-level One of basic approaches how to model realized covariance was proposed by Chiriac & Voev (2011). The model constitutes of the Cholesky decomposition of RCOV into Cholesky factors $\boldsymbol{A}_t$ (lower triangular matrix) for which $\boldsymbol{A}_t'\boldsymbol{A}_t = \Sigma_t$. Chiriac & Voev (2011) suggest to model the dynamics of the individual components of lower triangular matrix by using m univariate HAR models described in portfolio-level approach. Thanks to Cholesky decomposition, the resulting RCOV is guaranteed to be PSD.

We assume $\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{A}\boldsymbol{\epsilon}_t$, $\boldsymbol{r}_t \sim N_d(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{\epsilon}_t \sim N_d(\boldsymbol{0}, \boldsymbol{I}_d)$, where N_d denotes to multivariate Gaussian distribution, $\boldsymbol{\mu}_t = \boldsymbol{0}, \boldsymbol{\Sigma}_t = \boldsymbol{A}\boldsymbol{A}'$ and $d = 2$. The covariance matrix is given by:

$$\boldsymbol{A}_t = \begin{pmatrix} a_{RCOV,t,11} & 0 \\ a_{RCOV,t,21} & a_{RCOV,t,22} \end{pmatrix} \tag{3.15}$$

$$\begin{aligned} a_{RCOV,t+1,11}^{(d)} = c + \beta^{(d)} a_{RCOV,t,11}^{(d)} + \beta^{(w)} a_{RCOV,t,11}^{(w)} + \beta^{(m)} a_{RCOV,t,11}^{(m)} + \nu_{t,11} \\ a_{RCOV,t+1,21}^{(d)} = c + \beta^{(d)} a_{RCOV,t,21}^{(d)} + \beta^{(w)} a_{RCOV,t,21}^{(w)} + \beta^{(m)} a_{RCOV,t,21}^{(m)} + \nu_{t,21} \\ a_{RCOV,t+1,22}^{(d)} = c + \beta^{(d)} a_{RCOV,t,22}^{(d)} + \beta^{(w)} a_{RCOV,t,22}^{(w)} + \beta^{(m)} a_{RCOV,t,22}^{(m)} + \nu_{t,22} \end{aligned} \tag{3.16}$$

where $a_{RCOV,t,11}^{(d)}, a_{RCOV,t,11}^{(w)}, a_{RCOV,t,11}^{(m)}$ is calculated in the same way as in equations 3.12. Logarithmic form of RCOV does not exist since covariance component can obtain negative value.

In case of asymmetric multivariate HAR the procedure is similar, we replace symmetric HAR for its asymmetric version defined in Equation 3.14 for forecasting RV and keep symmetric HAR for forecasting RCOV. Therefore such a asymmetric multivariate HAR is a counterpart to DCC-GJR-GARCH.

Advantages Parsimonious model with low demanding estimation that has a good performance in reproduction of persistence of realized (co)variance. Multivariate version is immune to curse of dimensionality in the number of parameters of the model in comparison with other models for RCOV Bauwens *et al.* (2012).

Disadvantages It can happen that volatility clustering of realized volatility appears and then extension of HAR model by GARCH to model innovations ν_t of HAR can be suitable solution, known has HAR-GARCH model Bauwens *et al.* (2012, pg. 369). In case of asset-level approach, *"a drawback of this approach is that the dynamic linkages among the variance and covariance series (e.g., volatility spillovers) is neglected"* Chiriac & Voev (2011).

3.2 Loss Distribution

As we mentioned in Section 2.4 that our focus is on the risk measure based on conditional loss distribution, we will review now applied loss distributions in this master thesis. First two, specifically, *Gaussian* and *Student's t*, belong to class of elliptical distributions that can be estimated through its parameters either in univariate or multivariate dimension. Following two, *FHS* and *EVT*, are non-elliptical and are estimated semi-parametrical. FHS will be presented for univariate as wells as for multivariate dimension. Strictly speaking, EVT is the method to analyze the tail of loss distribution and the most used distributions of tail are Generalized Extreme Value distribution (GEV) or Generalized Pareto Distribution (GPD). In this master thesis we will only work with EVT in univariate dimension jointly with Monte Carlo simulations via Copula in order to construct multivariate loss distribution.

3.2.1 Elliptical Distributions

Let's define an elliptical distributions as *"distributions with densities which are constant on ellipsoids"*Kyselá (2016, pg. 26) and their properties according to McNeil *et al.* (2015, pp. 200-203) and Jondeau *et al.* (2007, pp. 223-230).

Definition 3.1. *An d-dimensional vector $\boldsymbol{X}$, from $\boldsymbol{X} = \boldsymbol{\mu}+\boldsymbol{\Sigma}^{1/2}\boldsymbol{Y}$, is considered to be elliptically distributed with location vector $\boldsymbol{\mu}$ of size $d \times 1$ and dispersion matrix $\boldsymbol{\Sigma}$ of size $d \times d$, if the density is*

$$g(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} f^{(d)}((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{3.17}$$

where the spherical vector $\boldsymbol{Y}$ has d-dimensional density generating function $f^{(d)}$.

We denote $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f^{(d)})$ with following properties.

Linear combinations *"If we take linear combinations of elliptical random vectors, then these remain elliptical with the same characteristic generator $f^{(d)}$. Let's take any $\boldsymbol{B} \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$"* McNeil *et al.* (2015, pg. 202). We can show that

$$\boldsymbol{BX} + \boldsymbol{b} \sim E_k(\boldsymbol{B\mu} + \boldsymbol{b}, \boldsymbol{B\Sigma B}', f^{(d)}). \tag{3.18}$$

If we apply vector of portfolio weights $\boldsymbol{w} \in \mathbb{R}^d$, then

$$\boldsymbol{w}'\boldsymbol{X} \sim E_1(\boldsymbol{w}'\boldsymbol{\mu}, \boldsymbol{w}'\boldsymbol{\Sigma w}, f^{(d)}). \tag{3.19}$$

Marginal distributions *"Marginal distributions of $\boldsymbol{X}$ must be elliptical distributions with the same characteristic generator. Using the $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$ notation and again extending this notation naturally to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$"* McNeil *et al.* (2015, pg. 202).,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad , \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{3.20}$$

we have that $\boldsymbol{X}_1 \sim E_k(\mu_1, \Sigma_{11}, f^{(d)})$ and $\boldsymbol{X}_2 \sim E_k(\mu_2, \Sigma_{22}, f^{(d)})$.

Further properties can be found in Chapter 6. in McNeil *et al.* (2015).

Risk measurement for elliptical risk factors Suppose that $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, f^{(d)})$ and let $\mathcal{M}$ be the space of linear portfolios $\mathcal{M} = \{L : L = m + \boldsymbol{w}'\boldsymbol{X}, m \in \mathbb{R}, \boldsymbol{w} \in \mathbb{R}^d\}$. For any risk measure ϱ that satisfy axioms 2.2,2.4 and is law-invariant[4], then on $\mathcal{M}$ the following property hold McNeil *et al.* (2015, pg. 295).

- For any $L = m + \boldsymbol{w}'\boldsymbol{X} \in \mathcal{M}$ we have

$$\varrho(L) = m + \boldsymbol{w}'\boldsymbol{\mu} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma w}}, \tag{3.21}$$

  where $\boldsymbol{Y} \sim S_1(f^{(d)})$, i.e. a univariate symmetric around 0 spherical distribution with generator $f^{(d)}$.

Further properties can be find in Chapter 8. McNeil *et al.* (2015). This was an excursion how and why we can calculate VaR and ES as linear combination of risk factors that have Gaussian and Student's t multivariate distribution that are elliptical distributions[5]

Gaussian

One of the stylized fact about financial asset returns was that they commonly did not come from Gaussian (Normal) distribution neither at daily frequency nor higher-frequency than daily.

[4]Such a risk measure $\varrho(L)$ if it depends on L only via its distribution function F_L, i.e. VaR and ES McNeil *et al.* (2015, pg. 295)

[5]For enthusiastic reader willing to learn about skewed elliptical distributions, we can refer to Chapter 16. in Jondeau *et al.* (2007).

However Gaussian distribution is still used in quantitative finance for its appealing mathematical features and simplicity to estimate because it is fully described by its first two moments.

Let's set $f^{(d)}(u) = \frac{e^{-u/2}}{(2\pi)^{d/2}}$ and we get joint density of multivariate Gaussian distribution given by

$$g(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}}|\boldsymbol{\Sigma}|^{-1/2}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{3.22}$$

Multivariate Gaussian distribution is deemed attractive due to its properties. These properties are the same properties mentioned in the section about elliptical distributions. Moreover, if the risk factors are iid, then we can conclude sample mean and covariance matrix are good of estimators of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, respectively. However, financial asset returns are not described very well by multivariate Gaussian distribution. It is mainly due to its dependency introduced by the covariance matrix equals to zero in the tails of the distributions. Another disadvantage arises from the symmetrical construction omitting any leverage effect. Particularly, the first shortcoming is the crucial for portfolio VaR and ES as one of stylized fact about multivariate distribution of financial asset returns suggests that *"Extreme returns in one series often coincide with extreme returns in several other series"* McNeil *et al.* (2015) especially if such returns are member of the same category of financial asset i.e. stocks (even more if these stocks are from the same industry). The latter limitation can be fixed by using *skewed elliptical distribution.*

Portfolio-level: parametric VaR and ES Suppose that the loss distribution F_L is univariate Gaussian with forecasted conditional mean μ_{T+1} and variance σ^2_{T+1} and confidence level α.

$$\mathrm{VaR}_{p,\alpha,T+1}(L) = \mu_{p,T+1} + \sigma_{p,T+1}\Phi_\alpha^{-1} \tag{3.23}$$

$$\mathrm{ES}_{p,\alpha,T+1}(L) = \mu_{p,T+1} + \sigma_{p,T+1}\frac{\phi(\Phi_\alpha^{-1})}{1-\alpha} \tag{3.24}$$

where Φ_α^{-1} denotes the quantile of standard univariate Gaussian distribution and ϕ is the density of standard univariate Gaussian distribution.

Asset-level: parametric VaR and ES Suppose that the multivariate loss distribution F_L is multivariate Gaussian with forecasted conditional mean vector $\boldsymbol{\mu}_{T+1}$ and covariance matrix $\boldsymbol{\Sigma}_{T+1}$, vector of portfolio weights $\boldsymbol{w}$ and confidence level α.

$$\mathrm{VaR}_{\alpha,T+1}(L) = \boldsymbol{w}'\boldsymbol{\mu}_{T+1} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{T+1}\boldsymbol{w}}\,\Phi_\alpha^{-1} \tag{3.25}$$

$$\mathrm{ES}_{\alpha,T+1}(L) = \boldsymbol{w}'\boldsymbol{\mu}_{T+1} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{T+1}\boldsymbol{w}}\,\frac{\phi(\Phi_\alpha^{-1})}{1-\alpha} \tag{3.26}$$

Student's *t*

The solution for serious limitation of multivariate Gaussian distribution can be provided by employing multivariate Student's t distribution which allows dependency in the tails besides of heavier tail than in univariate Gaussian distribution.

Let's set $f^{(d)}(u) = \left(1 + \frac{u}{\nu}\right)^{-(\nu+d)/2}$, $\nu > 0$ and we get joint density of multivariate Student'st distribution given by

$$g(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{(\pi\nu)^{d/2}\,\Gamma\left(\frac{\nu}{2}\right)} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}} \tag{3.27}$$

We say that $\boldsymbol{X}$ has a multivariate Student's t distribution with ν degrees of freedom where ν stands for the parameter determining the kurtosis or in other words, fat-tailedness of the Student's t distribution. We denote $\boldsymbol{X} \sim t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $\mathrm{cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}\frac{\nu}{\nu-2}$ where covariance matrix is defined if $\nu > 2$ and mean vector if $\nu > 1$. As ν approaches large value, Student's t distribution converge to the normal distribution.

According to Jondeau *et al.* (2007, pg. 226) we should rule out the possibility of ν being constant because that would mean each distribution of the financial asset returns would have the same fat tails and the dependence in tails can be biased. Due to this fact this scenario is overly simplified and should be avoided. If we assume that the degrees of freedom ν_i is different for each financial asset, then multivariate Student's t will not be an elliptical distribution anymore. Therefore we will need to employ i.e. Monte Carlo simulations through Copula to sample such a multivariate Student's t distribution.

The multivariate Student's t distribution is a special case of the Gaussian variance mixture distribution with the mixing variable W from the inverse gamma distribution $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ that is equivalent to $\frac{\nu}{W} \sim \chi^2_\nu$. Further theory about the construction of multivariate Student's t distribution can fined in Chapter 6. in McNeil *et al.* (2015) or from a bit different angle in Jondeau *et al.* (2007, pp.225-229) or Dowd (2005, pp.159-160).

Portfolio-level: parametric VaR and ES Suppose that the loss distribution F_L is univariate Gaussian with forecasted conditional mean μ_{T+1} and variance σ^2_{T+1} and confidence level α and degrees of freedom ν.

$$\mathrm{VaR}_{p,\alpha,T+1}(L) = \mu_{p,t+1} + \sigma_{p,T+1}\sqrt{\frac{\nu-2}{\nu}}\,t^{-1}_{\nu,\alpha} \tag{3.28}$$

$$\mathrm{ES}_{p,\alpha,T+1}(L) = \mu_{p,T+1} + \sigma_{p,T+1}\sqrt{\frac{\nu-2}{\nu}}\left(\frac{g_\nu(t^{-1}_{\nu,\alpha})}{1-\alpha}\right)\left(\frac{\nu + (t^{-1}_{\nu,\alpha})^2}{\nu-1}\right) \tag{3.29}$$

where t_α denotes the quantile of standard univariate Student's t distribution and g_ν is the density of standard univariate Student's t distribution .

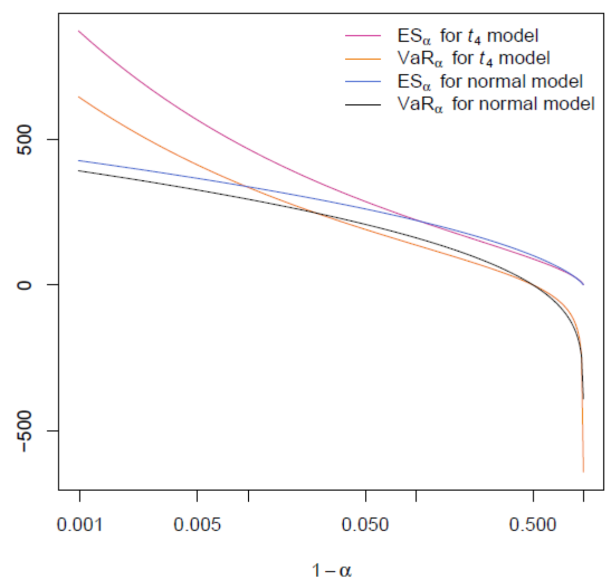
Asset-level: parametric VaR and ES Suppose that the multivariate loss distribution F_L is multivariate Student's t with forecasted conditional mean vector $\boldsymbol{\mu}_{T+1}$ and covariance matrix $\boldsymbol{\Sigma}_{T+1}\frac{\nu}{\nu-2}$, vector of portfolio weights $\boldsymbol{w}$, confidence level α and degrees of freedom ν .

$$\mathrm{VaR}_{\alpha,T+1}(L) = \boldsymbol{w}'\boldsymbol{\mu}_{T+1} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{T+1}\boldsymbol{w}}\sqrt{\frac{\nu-2}{\nu}}t^{-1}_{\nu,\alpha} \tag{3.30}$$

$$\mathrm{ES}_{\alpha,T+1}(L) = \boldsymbol{w}'\boldsymbol{\mu}_{T+1} + \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{T+1}\boldsymbol{w}}\sqrt{\frac{\nu-2}{\nu}}\left(\frac{g_\nu(t^{-1}_{\nu,\alpha})}{1-\alpha}\right)\left(\frac{\nu+(t^{-1}_{\nu,\alpha})^2}{\nu-1}\right) \tag{3.31}$$

McNeil *et al.* (2015) has shown in Figure 3.2 that at relatively low confidence levels, say $95-97.5\%$ portfolio-level VaR using Gaussian vs Student's t distribution for $\nu = 4$, the Student's t distribution has even lower VaR than Gaussian one despite of higher likelihood of large losses for Student's t. Difference becomes apparent for higher confidence level than 97.5. On the other hand, risk measure ES reflects already on those relatively lower confidence levels the significant difference in the riskiness between Gaussian and Student's t distribution.

Figure 3.1: Comparison between VaR_α and ES_α from Gaussian and Student's t with $\nu = 4$ for $\alpha \in (0.001, 0.5)$



Source: Embrechts *et al.* (2016)

The calculation of analytical VaR and ES can be characterized as easy and fast computation at the cost of high simplification and possible significant departure from the empirical evidence.

3.2.2 Filtered Historical Simulations

The *Filtered Historical Simulations* belongs to class of semi-parametric method of calculation of VaR and ES and it was suggested by Barone-Adesi *et al.* (1998), Barone-Adesi *et al.* (1999).

The approach is based on first phase that is standardization of (filtration) financial asset returns via forecasted conditional mean and (co)variance and instead of analytical specification of the standardized loss (innovations) distribution, the second phase uses historical (or empirical) standardized loss distribution which is model-free[6]. FHS is the one of the ways how to avoid of misspecification of loss distribution but still incorporate forecasts of conditional mean and (co)variance by sophisticated models and hence, it is more advanced method than naive *historical simulations.*

In order to compute VaR and ES, we only need to decide what model to use for forecasts of conditional mean and (co)variance and what size of window (length) of past observed financial asset returns to take in order to construct empirical distribution of innovations.

The problematic spot is in the choice of size of window to be adequately representing the distribution, so it should contain also the period of increased volatility in order not to underestimate the risk. On the other hand, if there happened some structural break in the market for in favor of volatility drop or reverse, then long size of window could cause overestimate of the risk or underestimate the risk. The same applies if the size of window is too short. Therefore it is a very tricky decision about the optimal size of window. One of the recommendation mentioned in Christoffersen (2011) but for naive historical simulation is between 250 and 1000 business days.

Detailed discussion of advantages and disadvantages can be found in Dowd (2005, pp. 99-101). Following computing methods of FHS are based on Part V. Value-at-Risk Models in Andersen *et al.* (2009) with various extensions.

Portfolio-level: semi-parametric VaR and ES We assume that, given $\mathscr{F}_{t-1}$, $r_{p,t} = \mu_t + \sigma_t \epsilon_t$, $r_{p,t} \sim G(\mu_t, \sigma_t)$, $\epsilon_t \sim G(0,1)$, where G denotes the univariate empirical distribution of innovations ϵ_t, $\{\epsilon_t\}_t^T$ is the set of empirical innovations and window size (t,T), then portfolio-level VaR and ES is given for long position by

$$\mathrm{VaR}_{p,\alpha,T+1} = \mu_{p,T+1} + \sigma_{p,T+1} Percentile\left\{\{\epsilon_t\}_t^T, 100(1-\alpha)\right\}, \tag{3.32}$$

where the *Percentile* is the function that returns an empirical quantile $\epsilon_{100(1-\alpha)}$.

$$\mathrm{ES}_{p,\alpha,T+1} = \mu_{p,T+1} + \sigma_{p,T+1} \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} \epsilon_t I\left(\epsilon_t < \frac{\mathrm{VaR}_{p,\alpha,T+1}}{\sigma_{p,T+1}}\right),$$

(3.33)

where $I(.)$ is the indicator function returning a 1 if the argument is true and zero otherwise.

[6] we do not make even iid assumption

Asset-level: semi-parametric VaR and ES We assume $\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{A}\boldsymbol{\epsilon}_t$, $\boldsymbol{r}_t \sim G_d(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{\epsilon}_t \sim G_d(\boldsymbol{0}, \boldsymbol{I}_d)$, where G_d denotes to multivariate empirical distribution of innovations $\boldsymbol{\epsilon}_t$, $\Sigma_t = D_t \Gamma_t D_t$ and we assume $\boldsymbol{\mu}_t = \boldsymbol{0}$.

First step is to remove volatilities from the vector of portfolio return by

$$\boldsymbol{z}_t = D_t^{-1} \boldsymbol{r}_t, \tag{3.34}$$

where D_t is an $d \times d$ diagonal matrix of volatility estimates. Then we create a data set of empirical dynamically uncorrelated innovations as

$$\boldsymbol{z}_t^D = \Gamma_t^{-1/2} \boldsymbol{z}_t, \tag{3.35}$$

where, $\Gamma_t^{-1/2}$ is the cholesky decomposition of conditional correlation matrix Γ_t. When we finish a data set of empirical uncorrelated shocks$\{\boldsymbol{z}_t^D\}_{t=1}^T$, we draw a random vector, called $\boldsymbol{z}_{i,T+1}^D$ from this data set and re-apply forecasted $\Gamma_{T+1}^{-1/2}$ and D_{T+1}. Afterwards, we compute simulated vector of portfolio return as

$$\boldsymbol{r}_{i,T+1} = D_{T+1} \Gamma_{T+1}^{-1/2} \boldsymbol{z}_{i,T+1}^D. \tag{3.36}$$

It will be repeated T times. Finally, we transform logarithmic returns to arithmetic returns since logarithmic returns are not portfolio additive $\boldsymbol{R}_{i,T+1} = \exp(\boldsymbol{r}_{i,T+1}) - 1$, consequently we convert arithmetic portfolio return back to logarithmic one and compute VaR and ES given by

$$\mathrm{VaR}_{\alpha,T+1} = Percentile \left\{ \{\ln(\boldsymbol{w}' \boldsymbol{R}_{i,T+1} + 1)\}_{i=1}^T, 100(1-\alpha) \right\}, \tag{3.37}$$

$$\mathrm{ES}_{\alpha,T+1} = \frac{1}{(1-\alpha)T} \sum_{i=1}^T \ln(\boldsymbol{w}' \boldsymbol{R}_{i,T+1} + 1) I \left(\ln(\boldsymbol{w}' \boldsymbol{R}_{i,T+1} + 1) < \mathrm{VaR}_{T+1}\right).$$

(3.38)

3.2.3 Extreme Value Theory

Given the interest in risk measures such as VaR and ES, we are exposed to study and analysis of extreme events which Dowd (2005, pg. 189) characterizes as low-probability and high-impact events resulting in large changes of risk-factors. Thus, practitioners in financial risk management need to deal with extreme values that are observed rarely[7]. Unlike modeling of entire loss distribution discussed in previous subsections, EVT is *"a branch of probability concerned with limiting laws for extreme values in large samples ... describing the behaviour of sample maxima and minima, upper-order statistics and sample values exceeding high thresholds" McNeil et al. (2015, pg. 135).* Our interest will be in the analysis of tail of loss distributions that are well-studied by

[7] As Dowd (2005, pg. 189) points out, this issue is present not only in financial risk management but also in other fields, i.e. hydrology *"where engineers have long struggled with the question of how high dikes, sea walls and similar barriers should be to contain the probabilities of floods within reasonable limits."*

models of *block maxima* and *peaks-over-threshold* denoted Peak-Over-Threshold (POT).

The *block maxima* method initially divides the sample of risk changes into blocks and observed values within inside the block are assumed to be iid. Then, the largest[8] observed value is selected and subsequently the sample of these largest values across all blocks will construct the tail distribution. The most used tail distributions in risk management are *Gumbel*, *Fréchet* and *Weibull*. If we generalize these distributions, we get GEV distribution which is able to describe each of these distributions according to selected *tail*[9] parameter. The disadvantage of this method is the significant waste of data leading to requiring rather very large data sets - a requirement that is rarely met in financial markets. These reasons motivate to consider the second method POT described briefly in following subsection. The more detailed explanation of EVT and its models exceeds the scope of this master thesis and hence, we refer to comprehensive book devoted to only EVT Embrechts *et al.* (1997) or a more handy summary in Chapter 5. of McNeil *et al.* (2015).

Peak Over Threshold

but use all of those exceeding Contrary to *Block maxima* method, POT does not divide full sample of data into blocks but use all of those exceeding determined by high threshold and those ones construct the tail of a distribution. There are two approaches to model the tail of a distribution. First one is the full parametric model so-called *Generalized Pareto distribution* and its alternative is semi-parametric the *Hill* approach using the Hill estimator. As McNeil *et al.* (2015, pp. 161-162) has shown in their Monte Carlo experiment to estimate the 99% VaR that GPD method is more robust than Hill method and that was the reason why we implemented only GPD method in this master thesis.

Generealized Pareto Distribution Let's consider an iid random variable X with distribution function$F(x)$ and u is a threshold of X (i.e. risk-factor changes), then the excess distribution over threshold u (for upper tail) is given by:

$$F_u(x) = Pr(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{3.39}$$

for $x > 0$. It gives us the probability that X exceeds the threshold u by no more than x, given that threshold u was violated. The distribution of X itself (i.e. Gaussian, Student's t, etc.) does not play crucial role since we are only interested in the excesses. The important conclusion driven by Gnedenko-Pickands-Balkema-deHaan (GPBdH) theorem is that as $u \to \infty$, $F_u(x)$ converges to the GPD $G_{\xi,\beta}(x)$:

Definition 3.2 (Generalized Pareto distribution). The distribution function of GPD is

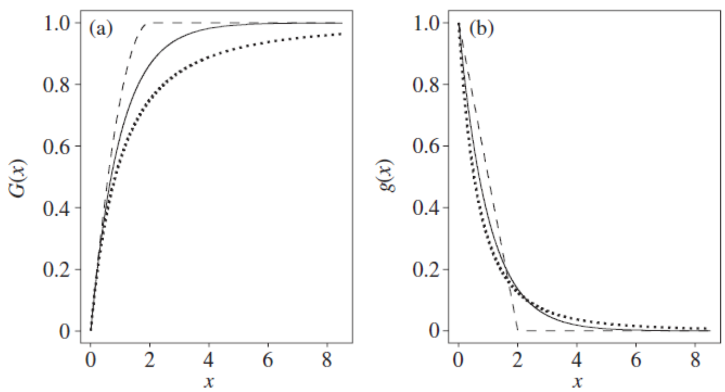
[8] Depending on the position if it is long, then largest negative value otherwise positive value
[9] Inverse tail index is called *shape* parameter

defined as

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0, \end{cases} \tag{3.40}$$

where $\beta > 0$, and $x \leq 0$ when $\xi \leq 0$ and $0 \geq x \geq -\beta/\xi$ when $\xi < 0$. The parameter ξ is the *shape* or *tail index* parameter that can be positive (common for heavy-tail data), zero or negative and β is a positive *scale* parameter. The latter one is the same one as mentioned in GEV distribution Dowd (2005).

Figure 3.2: (a) Distribution function of the GPD in three cases: the solid line corresponds to $\xi = 0$ (exponential); the dotted line to $\xi = 0.5$ (a Pareto Distribution); and the dashed line to $\xi = -0.5$ (Pareto type II). The scale parameter β is equal to 1 in all cases. (b) Corresponding densities.



Source: McNeil *et al.* (2015)

Another important result of GPBdH theorem is that *"the distribution of excess losses always has the same form (in the limit, as the threshold gets high), pretty much regardless of the distribution of the losses themselves."* Dowd (2005, pg. 202). Therefore considering that threshold is optimally set, we can say that GPD is a natural model for the excess losses.

Selection of the optimal threshold and estimation The optimal threshold u should be a sufficiently high that the GPD is an appropriate fit and simultaneously keep enough exceeding observations in order to provide reliable estimates of the GPD parameters. In order to decide about optimal trade-off between those two conditions, we can construct *mean excess function* (MEF) defined as $e(u) = E(X - u|X > u)$ and visually select the threshold u where the MEF starts to be linear McNeil *et al.* (2015). The estimation of GPD paramaters can be done by either MLE or probability-weighted moments McNeil *et al.* (2015).

Apart from the issue of selection of threshold, there is often a departure from iid assumption in context of time series of raw financial asset returns. The solution is either to apply GEV distribution to block maxima method but with new issue that what length the block should have or we treat the tail of the distribution as a conditional one and filter raw financial asset returns with the conditional volatility model (i.e. GARCH or HAR) and eliminate the time dependency. Therefore the EVT-POT with GPD is applied rather on innovations[10] than raw financial asset returns Dowd (2005).

Portfolio-level: parametric VaR and ES We assume that, given $\mathscr{F}_{t-1}$, $r_{p,t} = \mu_t + \sigma_t \epsilon_t$, $r_{p,t} \sim G(\mu_t, \sigma_t)$, $\epsilon_t \sim G(0,1)$, where G denotes the univariate distribution of innovations ϵ_t with number of innovations exceeding the threshold u is N_u and total number of innovations n[11], then GPD is estimated on innovations with portfolio-level VaR and ES estimates are given by

$$\mathrm{VaR}_{p,\alpha,T+1} = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{N_u/n}\right)^{-\xi} - 1\right), \tag{3.41}$$

$$\mathrm{ES}_{p,\alpha,T+1} = \frac{\mathrm{VaR}_{p,\alpha,T+1}}{1-\xi} + \frac{\beta - \xi u}{1-\xi}. \tag{3.42}$$

Asset-level: semi-parametric VaR and ES As it was mentioned in Section 2.1, the multivariate EVT is cumbersome, much less spread in risk management field than all other discussed methods and lastly, the theoretical and practical background would certainly exceed the scope of this master thesis. However, the easiest way how to implement EVT-POT in multivariate dimension is to use univariate EVT-POT for additional filtration with the entire distribution (upper and lower tail estimated parametrically as GPD and the interior with non-parametric kernel or FHS) and apply on residuals the Monte Carlo simulations with Copula functions to model multivariate loss distribution.

3.2.4 Monte Carlo - Copula

The main pitfalls of covariance (normalized covariance is Pearson's correlation coefficient computed as $\rho(X,Y) = \frac{Cov(X,Y}{\sqrt{Var(X)Var(Y)}}$ which takes values in $[-1,1]$ and henceforth we use only term *correlation*) pointed out in Dowd (2005, pg. 146) as concept of dependency measure is that it describes only linear dependence what is a good measure for elliptical distributions such as Gaussian and Student's t since they are fully specified by their mean vector, covariance matrix and characteristic generator function. Therefore if there is a non-linear dependency among risk-factors, then correlation is not able to detect it and is equal to zero. So we cannot say that zero correlation means generally no dependency. The only exception is multivariate Gaussian distribution. Its dependency is fully described linearly and correlation coefficient implies also general dependency. Additionally, correlation is not invariant under non-linear transformations of the risk-factors McNeil *et al.* (2015, pg. 239).

The more critical pitfall is the correlation usage itself when we move beyond elliptical distributions where correlation does not need to be even defined since correlation requires finite variances

[10] This term is used interchangeably with standardized loss
[11] In Subsection 3.2.2 we used notation T

what is not necessary condition for non-elliptical distributions (particularly heavy-tailed distributions with infinite variance such as Student's t with $\nu < 2$). Anyway, if the correlation is defined, then it is not necessary that correlation takes values in $[-1, 1]$.

Thus, we need firstly dependency measure that is less restrictive than correlation and the statistical concept how to model the multivariate distribution when we are able to estimate separately univariate distributions (known as *marginal distributions* or *margins*) relatively well and join them through the function that is using less restrictive alternative to correlation. The construction of multivariate distribution is finished by the Monte Carlo simulations of this concept. The concept is known as *"bottom-up approach to multivariate model building"* McNeil *et al.* (2015, pg. 221) and is called *copula*. Copula gives us a possibility to link the combinations of all sorts of marginal distributions into joint (multivariate) distribution even though these joint distributions have not been analytically defined yet.

Definition 3.3 (**copula** McNeil *et al.* (2015, pg. 221))**.** A d-dimensional copula is a distribution function $[0, 1]^d$ with standard uniform marginal distributions.

We reserve the notation $C(\mathbf{u}) = C(u_1, \ldots, u_d)$ for the multivariate distribution functions that are copulas. Hence C is a mapping of the form $C : [0, 1]^d \to [0, 1]$, i.e. a mapping of the unit hypercube into the unit interval. The following three properties must hold.

(1) $C(u_1, \ldots, u_d)$ is increasing in each component u_i.

(2) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}, u_i \in [0, 1]$.

(3) For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0, 1]^d$ with $a_i \leq b_i$ we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i1}, \ldots, u_{did}) \leq 0, \tag{3.43}$$

where $u_{j1} = a_j$ and $u_{j1} = b_j$ for all $j \in \{1, \ldots, d\}$. As we can see from the second property, the copula requires the marginal distributions to be uniform but our risk-factors have standard Gaussian distribution, Student's t or any other and hence, we need to recall the quantile and probability transformation that we will use later in empirical part of our master thesis in simulations.

Proposition 3.1. *Let F be a distribution function and let $F^{\leftarrow}$ denote its generalized inverse, i.e. the function $F^{\leftarrow}(u) = inf\{x : F(x) \geq u\}$.*

(1) ***Quantile transformation****. If $U \sim U(0, 1)$ has a standard uniform distribution, then $P(F^{\leftarrow}(U) \leq x) = F(x)$.*

(2) ***Probability transform****. If X has distribution function F, where F is a continuous univariate distribution function, then $F(X) \sim U(0, 1)$.*

The connection between the joint distribution and a copula was shown in Sklar's theorem.

***Theorem* 3.1** (**Sklar's Theorem** Sklar (1959))**.** *Let F be a joint distribution function with margins $F_1, \dots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \dots, x_d \in \bar{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \tag{3.44}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $\mathrm{Ran}F_1 \times \mathrm{Ran}F_2 \times \dots \mathrm{Ran}F_d$*, where* $\mathrm{Ran}F_i = F_i(\bar{\mathbb{R}})$ *denotes the range of F_i. Conversely, if C is a copula and $F_1, \dots, F_d$ are univariate distribution functions, then the function F defined in Equation 3.44 is a joint distribution function with margins $F_1, \dots, F_d$.*

The importance of Sklar's theorem lies in two aspects. Firstly, it shows that each multivariate distribution function has a copula. Secondly, the Equation 3.44 shows that coupling copula C and univariate distribution functions allows us to construct the multivariate distribution function F.

If we evaluate Equation 3.44 at the arguments $x_i = F_i^{\leftarrow}(u_i), 0 \leq u_i \leq 1, i = 1, \dots, d$, and use Proposition A.3 (viii) in McNeil *et al.* (2015, pg. 642), we obtain

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \tag{3.45}$$

The Equation 3.45 shows how copulas are extracted from multivariate distribution functions with continuous margins. Furthermore, Equation 3.45 tells us that copula convey dependence on quantile scale since copula represents the joint probability of X_1 being below its u_1 quantile and X_2 below its u_2 quantile (contrary to previous methods of dependence which were on risk-factor changes scale) McNeil *et al.* (2015, pg. 224). Another advantage of copula of a distribution is its invariance property under *strictly increasing* transformation of the marginals McNeil *et al.* (2015, pg. 224) since we transform all underlying margins to uniform margins before estimating copula.

As we mentioned in the beginning of this subsection, Pearson's correlation coefficient would not be the suitable correlation estimate, particularly for non-elliptic multivariate distributions, which we could construct via copula function. Therefore the alternative was proposed to be rank correlation that are *scalar measures of dependence that depend only on the copula of a bivariate distribution and not on the marginal distributions, unlike linear correlation, which depends on both* and they are able to calibrate copulas to empirical data McNeil *et al.* (2015, pp. 243-244).

The representatives of rank's correlation are *Kendall's tau* and *Spearman's rho*. According to McNeil *et al.* (2015, pg. 244)] *"the both can be understood as a measure of concordance for bivariate random vectors. Two points in $\mathbb{R}^2$, denoted by (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$ are said to be concordant if $(x_1 - \tilde{x}_1)((x_2 - \tilde{x}_2) > 0$ and to be disconcordant if $(x_1 - \tilde{x}_1)((x_2 - \tilde{x}_2) < 0$.*

Definition 3.4 (**Kendall's tau** McNeil *et al.* (2015, pg. 244))**.** Consider a random vector (X_1, X_2) and an independent copy $(\tilde{X}_1, \tilde{X}_2)$, then the Kendall's tau is defined as

$$\rho_\tau(X_1, X_2) = P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0] - P[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0]. \tag{3.46}$$

Simply said, the Kendall's tau is the probability of concordance minus the probability of discordance for used pairs.

Definition 3.5 (**Spearman's rho** McNeil *et al.* (2015, pg. 245))**.** Consider a random vector (X_1, X_2) with continuous marginal distributions F_1 and F_2, then the Spearman's rho is defined as

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)). \tag{3.47}$$

what is *"the linear correlation of the probability transformed random variables which for continuous random variables is the linear correlation of their unique copula"* Kyselá (2016, pg. 27).

Simply said, the Kendall's tau is the probability of concordance minus the probability of discordance for used pairs.

Rank correlations have symmetric dependence property that is taking values in $[-1, 1]$. They give zero value for independent random values but what does not necessary mean generally independence. The main advantage is that rank correlation can allocate any value on its range $[-1, 1]$ from a specified bivariate distribution constructed by any combination of continuous marginal distributions in contrast to Pearson's rho.

The family of copula functions is very wide and they are divided in three categories *fundamental* (copulas with special dependence structure), *implicit* (copulas derived from their multivariate distributions i.e. elliptic ones) and *explicit* (copulas with simple closed-form what is a subset of *Archimedean* copulas) but some copulas can be assigned to both implicit and explicit category. The advantages of implicit copulas are quite easily extended into higher dimension than bivariate but on the other hand, they are symmetric and do not have closed form expressions in comparison to Archimedean Bauwens *et al.* (2012, pg. 248).

Further division can be between static and dynamic copulas that have time-varying dependence parameter and some examples of dynamic copulas are provided in Bauwens *et al.* (2012, pp. 304-308). The more advanced copulas are *vine copulas.* These are multivariate copulas that *can be decomposed into a cascade of iteratively conditioned bivariate copulas* see Bauwens *et al.* (2012, pp. 313-315) or *extreme value* copula see McNeil *et al.* (2015, pg. 591-598).

Gaussian copula

Definition 3.6 (**Gaussian copula** McNeil *et al.* (2015, pg. 226))**.** If $\boldsymbol{Y} \sim N_d(\boldsymbol{\mu}, \sigma)$ is a multivariate normal random vector, then its copula is a so-called *Gaussian copula.* Since the operation of standardizing the margins amounts to applying a series of strictly increasing transformations. The copula of $\boldsymbol{Y}$ is exactly the same as the copula $\boldsymbol{X} \sim N_d(\boldsymbol{0}, \mathrm{P})$, where P is the correlation matrix of $\boldsymbol{Y}$ (Pearson's rho). This copula is given by

$$C_{\mathrm{P}}^{Ga}(\boldsymbol{u}) = \boldsymbol{\phi}_{\mathrm{P}}(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)), \tag{3.48}$$

where ϕ dones the standard univariate normal distribution function and $\boldsymbol{\phi}_{\mathrm{P}}$ denotes the joint distribution function of $\boldsymbol{X}$, so Gaussian copula is equivalent to a multivariate Gaussian distribution. The Gaussian copula is parametrized by the correlation matrix P (we can use analytically either Pearson's rho, Kendaull's tau or Spearman's rho, see Jondeau *et al.* (2007, pg. 246)) and for bivariate dimension it would be $\rho = \rho(X_1, X_2)$. The fundamental copulas are special cases of Gaussian copula see McNeil *et al.* (2015, pg. 227). The main limitation of Gaussian copula is that it does not allow tail dependence[12] and symmetry.

t-copula

The example of copula with tail dependence is t-copula that belongs also to implicit and symmetry class of copulas.

Definition 3.7 (***t*-copula** McNeil *et al.* (2015, pg. 228))**.** The copula is constructed by the process as it was described in Gaussian copula. The copula is given by

$$C^t_{\nu,\mathrm{P}}(\boldsymbol{u}) = \boldsymbol{t}_{\nu,\mathrm{P}}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)), \tag{3.49}$$

where t_ν is the distribution function of a standard univariate t distribution with ν degrees of freedom. $\boldsymbol{t}_{\nu,\mathrm{P}}$ is the joint distribution function of the vector $\boldsymbol{X} \sim t_d(\nu, 0, \mathrm{P})$ and P) is a correlation matrix (we can use analytically either Pearson's rho or Kendaull's tau, see Jondeau *et al.* (2007, pg. 248)).

Explicit copulas

Explicit copulas are characterized by their simple closed forms but have difficulties to be extended to higher dimensions. We will mention just the most commonly used. *Gumbel* copula is asymmetric with significant upper tail dependence without lower tail dependence. The opposite tail dependencies than Gumbel has the *Clayton* copula. If we want to generate some dependence in the independent tails, then we can use their rotated versions Jondeau *et al.* (2007, pp. 251-252). These copulas are useful for the financial assets which behave similarly to only direction of shocks i.e. stocks from the same sector. Other explicit copulas are *Frank*, *Plackett*, *Marshall-Olkin*, see Jondeau *et al.* (2007, pp. 246-254).

Estimation

One-step estimator Exact MLE is a method that estimates the joint likelihood of parameters associated to marginal distributions as well as to copula functions. In other words, we estimate

[12] *"Quantile dependence focuses on the tails of the distribution. If X and Y are random variables with distribution functions F_X and F_Y , there is quantile dependence in the lower tail at threshold α, whenever $P[Y \leq F_Y^{-1}(\alpha)|X \leq F_X^{-1}(\alpha)]$ is different from zero. Finally, tail dependence is obtained as the limit of this probability, as we go arbitrarily far out into the tails."* Bauwens *et al.* (2012, pg. 298). The perfect example how to misuse the Gaussian copula was application for pricing and risk management of credit derivatives that contains high tail dependence of defaults. This practice contributed to the latest financial crises in 2007-2009 McNeil *et al.* (2015, pg. 14).

all parameters simultaneously. Obviously, this method faces to computational burden when number of parameters is large.

Two-step estimator The computationally friendly alternative is to split vector of parameters between those for margins and copula functions and estimate them separately. There are three common methods to estimate parameters of margins in order to transform the original data vectors to uniformly distributed vectors.

1. **Parametric.** We select an appropriate parametric model(s) describing underlying margins and fit them. This is called also as *inference functions for margins* (IFM).

2. **Semi-parametric.** Instead of relying on full description of margins by parametric models, we estimate the tail of margins by EVT (i.e. via GPD and the interior of margin can be estimated parametrically or non-parametrically.

3. **Non-parametric.** No assumptions of margins will be made and instead, they will be estimated through empirical distribution function.

Asset-level: non-parametric VaR and ES The formulas of VaR and ES are the same as in Asset-level in Subsection 3.2.2 only T will represent a number of simulations.

3.3 Backtesting of VaR

Since our objective is to forecast VaR day-ahead, we need to introduce backtesting methods. We understand[13] under the term backtesting as quantitative check of the significance of the forecasts from the out-of-sample[14] against the realization of the losses. Nonetheless, backtesting methods do not pick up the best model from the set of candidate models, given data. Therefore we introduce also the methods for model selection.

The basic and primitive test for backtesting and model selection is so-called *violation ratio*. It is defined for long position where loss and VaR obtain negative values as $VR = \frac{\sum_{t=1}^{T} I_t(L_t < VaR_t)}{(1-\alpha)T}$ where $n = [t, T]$ is length of out-of-sample forecasts, $n_1 = \sum_{t=1}^{T} I_t$ is number of failures violations, number of non-failures is $n_0 = n - n_1$, numerator is total observed violations, denominator is total expected violations, $I_t(.)$ is an indicator function of $failures$. The book of Danielsson (2011, pg. 147) advices as a rule of thumb for VaR model to be precise is that $VR \in [0.8, 1.2]$.

The accurate VaR model is achieved when *proportion of failures* $PoF = [\sum_{t=1}^{T} I_t(L_t < VaR_t)]/T$ is equal to coverage level $(1-\alpha)$ and I_t follows an iid Bernoulli process.

[13]The backtesting does not have formal definition and we can find various definition, see Roccioletti (2016, pg. 44)

[14]Sometimes is referred as testing window what is the data sample with m forecasts of period $[T+1, \ldots, T*]$ using in-sample data of period $[t, \ldots, T]$.

3.3.1 Unconditional coverage test

Unconditional coverage test was introduced by Kupiec (1995) with null hypothesis whether PoF is statistically equal to coverage level $(1-\alpha)$, ignoring the history of the indicator function and the alternative hypothesis is non-equivalence. Under the null hypothesis $E(I_t) = (1-\alpha)$ and assuming the independence, Kupiec (1995) proposes the likelihood ratio (LR)

$$LR_{uc} = 2[\ln((1-PoF)^{n_0} PoF^{n_1}) - \ln((1-(1-\alpha))^{n_0}(1-\alpha)^{n_1})] \tag{3.50}$$

which is asymptotically distributed $\chi^2(1)$.

3.3.2 Conditional coverage test

It is clear that previous test lacks of detecting failure clustering if the VaR model cannot react adequately i.e. to volatility clustering. We would expect that failures of correct VaR model are spread over time and so to be independent. Danielsson (2011, pg. 155). Thus, Christoffersen (1998) proposes a joint test of previous unconditional coverage test plus test of first order independence of the failure process called *conditional coverage* test. Under the null hypothesis is that failures are independently distributed through time and at the same time PoF equals to coverage level $(1-\alpha)$ and the alternative hypothesis is at least one of these equalities does not hold, the LR statistic is given by

$$LR_{cc} = 2[\ln((1-\hat{p}_{01})^{n_{00}} \hat{p}_{01}^{n_{01}} (1-\hat{p}_{11})^{n_{10}} \hat{p}_{11}^{n_{11}}) - \ln((1-(1-\alpha))^{n_0}(1-\alpha)^{n_1}) \tag{3.51}$$

which is asymptotically distributed $\chi^2(1)$. where according to Louzis *et al.* (2014) *"p_{ij} is the transition probability between two consecutive observations from state i to state j assuming a first-order Markov chain probability transition matrix between the two possible states (a successful VaR estimation, or a failure), $n_i j$ is the number of all occurrences of transitions from state i to state j, with $i, j = 0, 1$ and $\hat{p}_{ij} = n_{ij} / \sum_{j=0}^{1} n_{ij}$ are the maximum likelihood estimates for p_{ij}."*

3.3.3 Dynamic Quantile test

Engle & Manganelli (2004) point out that Christoffersen test can detect the serial correlation of failures (they show that it is relatively easy to generate such failures with iid property). They proposed more powerful test called *dynamic quantile* test. They defined a new variable $Hit_t = I_t - (1-\alpha)$ to use in OLS regression in order to test if $E(Hit_t) = 0$ (what is the unconditional coverage test) and at the same if the Hit_t is serially uncorrelated with the past information. Therefore we run the regression $Hit_t(1-\alpha) = \delta + \sum_{k=1}^{K} \beta_k Hit_{t-k}(1-\alpha) + \epsilon_t$ and test the joint hypothesis $H_0 : \delta = \beta_1 = \cdots = \beta_k = 0$ for all lags k. The Hit_t are serially uncorrelated over time if the β are 0 and the PoF is correct if $\delta = 0$ and the alternative hypothesis is at least one of these equalities does not hold. The test statistic is given by $\frac{\hat{\boldsymbol{\lambda}} \boldsymbol{X}' \boldsymbol{X} \hat{\boldsymbol{\lambda}}}{(1-\alpha)\alpha} \sim \chi^2_{K+1}$ where $\hat{\boldsymbol{\lambda}} = (\delta, \beta_1, \ldots, \beta_K)'$ is the vector of estimated parameters of the OLS model and $\boldsymbol{X}$ is the matrix of explanatory variables. We are going to choose lag $k = 5$ as it was chosen in Louzis *et al.* (2014).

3.3.4 Loss Function

All previous three tests are backtesting methods and they do not have ability to compare different VaR models but only validate. We now present loss functions that calculate a loss (magnitude of failure) per observation for each model and the model with the smallest average loss is considered as the best one. The loss function is interchangeable with the term "scoring function" that will be mentioned in backtesting methods of ES.

Regulatory Loss function

The magnitude of failures is also a concern for regulators and Lopez (1998) proposed *quadratic or regulatory loss function* that is mainly used in the assessment of bank internal models of VaR. It is defined for long position as

$$RLF = \begin{cases} 1 + (L_t - \mathrm{VaR}_{t,\alpha})^2 & \text{if } L_t < \mathrm{VaR}_{t,\alpha} \\ 0 & \text{if } L_t \geq \mathrm{VaR}_{t,\alpha} \end{cases}$$

We can see that larger failures are penalized more heavily due to squared distance between failure and VaR. However, RLF measure prefer the VaR model that favours too conservative VaR models due to fact that RLF does not penalize non-failures.

Firm Loss function

If we include penalization of the non-failures meaning penalization of agent's opportunity cost of its reserved capital, then we talk about the *economic or firm loss function* FLF. There were proposed many FLF as well as RLF and we can find exhausting overview of RLF and FLF loss functions for VaR model selection in article of López Martín *et al.* (2015). We consider the most apt FLF newly suggested by López Martín *et al.* (2015) called $FABL$ but it is a questionable how much it is "new" since from the methodological point of view it is a subset of FLF mentioned in Jondeau *et al.* (2007, pg. 343), we call it as JPR. Both are defined as

$$FLF_{FABL} = \begin{cases} (\mathrm{VaR}_{t,\alpha} - L_t)^2 & \text{if } L_t < \mathrm{VaR}_{t,\alpha} \\ (L_t - \mathrm{VaR}_{t,\alpha})\beta & \text{if } L_t \geq \mathrm{VaR}_{t,\alpha} \end{cases}$$

$$FLF_{JPR} = \begin{cases} |L_t - \mathrm{VaR}_{t,\alpha}|^\gamma & \text{if } L_t < \mathrm{VaR}_{t,\alpha} \\ |L_t - \mathrm{VaR}_{t,\alpha}| \times i & \text{if } L_t \geq \mathrm{VaR}_{t,\alpha} \end{cases}$$

where in both cases $i = \beta$ are considered to be interest rate and López Martín *et al.* (2015) uses in empirical analysis the key interest rate of European Central Bank and the Deutsche Bundesbank's interest rate for the previous period . In our case, we are going to apply RLF and FLF_{JPR} with $\gamma = 2$ and instead of constant interest rate i, we choose classical *risk premium* from Capital Asset Pricing Model (CAPM) which will be time-varying such as $rp_t = (r_{m,t} - r_{f,t})$ where $r_{m,t}$ is the market return and $r_{f,t}$ is the risk free interest. The reason is that we assume the agent's capital reserved for VaR yields at least the risk free return. Moreover, we assume zero transaction cost.

Asymmetric Loss function

The RLF favors only the most conservative VaR models. Despite of the correction by penalization of opportunity cost of reserved capital, FLF might favor the model that is weakly conservative due to minimization of often exhibited opportunity cost. The opportunity cost events have expected large sample itself due to high confidence level of VaR. In case, there are even less than expected violations, the opportunity cost intensifies. Additionally, the impact can be increased if agent has high opportunity cost. As a result, we might get from FLF function the inverse model selection to RLF.

The required balance between high disproportion between violations and non-violations and their impacts, can be found in *asymmetric loss function* (AFL) suggested by González-Rivera *et al.* (2004). The asymmetric loss function is defined for confidence level α as

$$ALF = \begin{cases} ((1-\alpha) - 1)(L_t - \mathrm{VaR}_{t,\alpha}) & \text{if } L_t < \mathrm{VaR}_{t,\alpha} \\ (1-\alpha)(L_t - \mathrm{VaR}_{t,\alpha}) & \text{if } L_t \geq \mathrm{VaR}_{t,\alpha} \end{cases}$$

Suppose the $\mathrm{VaR}_{t+1,97.5\%}$ is -5% and realization is -7%. The loss will be $((1 - 0.975) - 1)(-2) = 1.95$ and if $\mathrm{VaR}_{t+2,97.5\%}$ remains -5% and realization will be -2%. The loss will be only $(1 - 0.975)(-3) = 0.075$. Due to fact that non-violations occur more often, they are less penalized than violations that happen rarely.

ALF is mainly used in the Model Confidence Set (MCS) that serves for advanced model selection of VaR models but the MCS is out of the scope of this master thesis. Further reference about the MCS is presented in Hansen *et al.* (2011).

3.4 Backtesting of ES

Since VaR was established as the main risk measure for the calculation of capital requirements, there was a little research about the backtesting methods of ES. One of pioneering ES backtests was for instance a test of McNeil & Frey (2000, pg. 294) ased on exceedance residuals that can be considered as t-test. Following research brought on light a fundamental question whether ES is backtestable since Gneiting (2011) showed that ES lacks a mathematical property called *elicitability* while VaR does have it. The meaning of eliticability got absolutely new dimension when in October 2013 Basel Committee on Banking Supervision issued the revision of market risk framework which contained also a replacement of 99% VaR for 97.5% ES Committee (2013, pg. 18). This sparked a new global discussion among scholars, research and practitioners about ES backtesting because Basel Committee did not suggest that time any backtesting method (neither if it exists) for ES but to keep backtesting 99% and 97.5% VaRs. These circumstances motivated research to investigate how and if ES can be backtested.

Definition 3.8 (**Elicitability**)**.** A statistical function $psi(Y)$ of a random variable Y is defined as elicitable if it minimizes the expected value of a scoring function$S$ that is strictly consistent.

$$\psi(Y) = \underset{x}{\operatorname{argmin}}\, E[S(x,Y)] \tag{3.56}$$

where the representative of scoring function S can be for instance a squared error[15] $S(x,Y) = (x-Y)^2$, x denote the point forecasts and Y the realization. If the ψ is elicitable, then we can backtest the performance of the predictions $y_1,\ldots,y_T$ and their realizations $x_1,\ldots,x_T$ through model

$$\bar{S} = \frac{1}{T}\sum_{t=1}^{T} S(x_t, y_t). \tag{3.57}$$

The elicitability of $\mathrm{VaR}_\alpha(Y)$ is given through scoring function $S(x,y) = (I(x \geq y)-\alpha)(x-y)$ and so it can be shown that $\mathrm{VaR}_\alpha(Y) = \underset{x}{\operatorname{argmin}}\, E[(I(x)-\alpha)(x-y)]$ equals to $\alpha = \int_{-\infty}^{x} f_Y(y)\mathrm{d}y$ and $x = F_Y^{-1}(\alpha)$. The full proof can be found in Wimmerstedt (2015, pg. 13). Contrary to VaR, Gneiting (2011) showed it is not possible to find minimizing scoring function for ES and hence, ES is not elicitable.

As Acerbi & Szekely (2014a) points out that the most of people understood Gneiting (2011) that ES is not backtestable at all and they explain it was further strengthen by statement of Embrecht *"ES cannot be back-tested because it fails to satisfy elicitability ... If you held a gun to my head and said: 'We have to decide by the end of the day if Basel 3.5 should move to ES, or do we stick with VaR', I would say: 'Stick with VaR' "* said in 2013 at Imperial College. The opposition to these statements was formalized in the article of Acerbi & Szekely (2014b) where the authors firstly argue that the property of elicitability has to do only with model selection in order to choose the best model among competitors and additional argument is that currentlyVaR is backtested without exploiting its elicitability property. Therefore they suggested three ES non-parametric tests using Monte-Carlo simulations even with missing elicitability property because it is not needed for backtesting of ES. Another insightful article of Emmer *et al.* (2015) showed that ES is conditionally elicitable and proposed another non-parametric without need of Monte-Carlo simulations.

The great overview of ES backtesting methods is written in the master thesis of Wimmerst-

[15] Other famous scoring functions are absolute error $S(x,Y) = |x-Y|$, absolute percentage $S(x,Y) = |(x-Y)/Y|$ or relative error $S(x,Y) = |(x-Y)/x|$.

edt (2015) including implementation of four of them. The other were not chosen due to their parametric assumptions and requirement of large out-of-sample samples. The conclusion of that master thesis[16] is that backtesting of ES is possible but the complexity is significantly higher compared to the backtesting of VaR and further research is needed. Nonetheless, the author prefers the test of Emmer *et al.* (2015) where ES is backtested through approximation of several VaR levels.

We are going to implement first two tests proposed by Acerbi & Szekely (2014b) due to their non-parametric, simulation properties and possibility to backtest on just one confidence level contrary to test of Emmer *et al.* (2015) which is designed for four or even more confidence levels what increase computational burden.

Both tests of Acerbi & Szekely (2014b) assume that out-of-sample period consits of days $t = 1, \ldots, T$ and X_t represents a bank's profit and loss that has a *real* (unknowable) distribution F_t and it is forecasted by a model *predictive* distribution P_t which is also used for computation of the VaR and ES. The random variables $\vec{X} = \{X\}$ are assumed to be independent, but not identically distributed. There is no restriction on variability of F_t and P_t over time. We denote the value of the risk measures as $\mathrm{VaR}_{\alpha,t}^F$ and $\mathrm{ES}_{\alpha,t}^F$ when $X \sim F$. Under null hypothesis it is generally assumed that the predicted ES is correct, while the alternative hypotheses is that the predicted ES is underestimated and so following tests are only one-tailed tests what is the their disadvantage from the agent's point of view.

3.4.1 Test 1: testing ES after VaR

Let ES defined as

$$\mathrm{ES}_{\alpha,t} = -E[X_t | X_t + \mathrm{VaR}_{\alpha,t} < 0] \tag{3.58}$$

Previous equation can be rearranged to

$$E\left[\frac{X_t}{\mathrm{ES}_{\alpha,t}} + 1 | X_t + \mathrm{VaR}_{\alpha,t} < 0\right] = 0 \tag{3.59}$$

We assume that $VaR_{\alpha,t}$ has been already tested (necessary requirement for Test 1) and hence, we can test the magnitude of the realized violations against the model predictions. Let's define an indicator function of a $\mathrm{VaR}_{\alpha,t}$ violations $I_t = (X_t + \mathrm{VaR}_{\alpha,t} < 0)$ and if $N_t = \sum_{t=1}^{T} I_t > 0$ as the number of violations. Then we construct the test statistics given by

$$Z_1(\vec{X}) = \frac{\sum_{t=1}^{T} \frac{L_t I_t}{\mathrm{ES}_{\alpha,t}}}{N_T} + 1 \tag{3.60}$$

The null hypothesis is

$$H_0 : P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t \tag{3.61}$$

[16]Further thesis dedicated to elicitability and implementation of backtesting of ES are Roccioletti (2016) and Jäger (2015). Both include the MATLAB codes in their appendices.

where $P_t^{[\alpha]}(x) = \min(1, P_t(x)/\alpha)$ is the distribution tail for $x < -\mathrm{VaR}_{\alpha,t}$. The alternatives are

$$\begin{aligned} H_1: \quad & \mathrm{ES}^F_{\alpha,t} \geq \mathrm{ES}_{\alpha,t}, \quad \text{for all } t \text{ and} > \text{for some } t \\ & \mathrm{VaR}^F_{\alpha,t} = \mathrm{VaR}_{\alpha,t}, \quad \text{for all } t \end{aligned} \tag{3.62}$$

We can see from H_1 that predicted $\mathrm{VaR}_{\alpha,t}$ is assumed to be correct and that is the reason why Test 1 should be proceed only after non-rejected backtest of $\mathrm{VaR}_{\alpha,t}$. As (Acerbi & Szekely 2014b, pg. 4) mentioned *"this test is in fact completely insensitive to an excessive number of exceptions as it's an average taken over exceptions themselves."*. Under these conditions $E_{H_0}[Z_1|N_T > 0] = 0$ and $E_{H_1}[Z_1|N_T > 0] < 0$ meaning that expected value of $Z_1(\vec{X})$ is zero and if it negative, then $\mathrm{ES}_{\alpha,t}$ is underestimated and if positive, then $\mathrm{ES}_{\alpha,t}$ is overestimated. The proofs can be found in Proposition A.2 Acerbi & Szekely (2014b).

3.4.2 Test 2: testing ES directly

Let's write ES as an unconditional expectation

$$\mathrm{ES}_{\alpha,t} = -E\left[\frac{X_t I_t}{\alpha}\right] \tag{3.63}$$

the above equation can be converted to test statistic

$$Z_2(\vec{X}) = \frac{\sum_{t=1}^{T} \frac{L_t I_t}{\mathrm{ES}_{\alpha,t}}}{T\alpha \mathrm{ES}_{\alpha,t}} + 1 \tag{3.64}$$

and following hypotheses were suggested

$$\begin{aligned} H_0: \quad & P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t \\ H_1: \quad & \mathrm{ES}^F_{\alpha,t} \geq \mathrm{ES}_{\alpha,t}, \quad \text{for all } t \text{ and} > \text{for some } t \\ & \mathrm{VaR}^F_{\alpha,t} \geq \mathrm{VaR}_{\alpha,t}, \quad \text{for all } t \end{aligned} \tag{3.65}$$

We can see that Test 2 jointly test $\mathrm{VaR}_{\alpha,t}$ as well as $\mathrm{ES}_{\alpha,t}$. Unfortunately, it remains one-tailed test. Under conditions $E_{H_0}[Z_2] = 0$ and $E_{H_1}[Z_2] < 0$ meaning that expected value of $Z_2(\vec{X})$ is zero and if it is negative, then $\mathrm{ES}_{\alpha,t}$ is underestimated and if positive, then $\mathrm{ES}_{\alpha,t}$ is overestimated. The proofs can be found in Proposition A.3 Acerbi & Szekely (2014b). Test 2 jointly evaluates frequency and magnitude of α-tail events as shown by the relationship

$$Z_2 = 1 - (1 - Z_1)\frac{N_T}{T\alpha} \tag{3.66}$$

3.4.3 Finding the significance

Testinf the significance was proposed by Acerbi & Szekely (2014b, pg. 6) to be done through Monte Carlo simulations where it is simulated the distribution P_Z under H_0 to compute p-value $p = P_Z(Z(\vec{x}))$ of a realization $Z(\vec{x})$:

- simulate independent $X_t^i \sim P_t, \quad \forall t, \forall i = 1, \ldots, M$
- compute $Z^i = Z(\vec{X}^i)$ where Z represents either Z_1 or Z_2
- estimate $p = \sum_{i=1}^{M}(Z^i < Z(\vec{x}))/M$

where M is sufficient number of simulations. Given a significance level of test ϕ, the test is non-rejected if $p > \phi$ or rejected if otherwise.

Quite exceptional finding is shown in Acerbi & Szekely (2014b, pg. 8) that critical levels of Z_2 were discovered to be remarkably stable across different distribution types. It means that usage of critical level in tab:3.Z2 would eliminate the need of Monte Carlo simulations and storage of all predictive distributions. The authors remark that for extra heavy tails, meaning $\nu < 5$, the critical levels significantly diverge from previous stable level and Test 2 would be more penalizing.

Table 3.1: 5% and 0.01% significance thresholds for Z_2 across Student's t distributions with different ν and location

	Significance					
	5% location			0.01% location		
ν	-1	0	1	-1	0	1
3	-0.78	-0.82	-0.88	-3.9	-4.4	-5.5
5	-0.72	-0.74	-0.78	-1.9	-2.0	-2.3
10	-0.70	-0.71	-0.74	-1.8	-1.9	-1.9
100	-0.70	-0.70	-0.72	-1.8	-1.8	-1.9
Gaussian	-0.70	-0.70	-0.72	-1.8	-1.8	-1.9

Source: Acerbi & Szekely (2014b, pg. 10)

Chapter 4

Empirical analysis

Having defined necessary theoretical background and introduced selected methodological approaches for this master thesis, we can move to their application on empirical data in this chapter. All computational work of this master thesis was done in `RStudio version 1.0.136` using language `R version 3.3.2`.

4.1 Data analysis

Our portfolio consists of the most liquid representatives of major financial asset classes denominated in the U.S. dollars. Specifically, the data employed in this thesis are E-mini futures S&P 500, Light Crude Oil futures, Spot gold and spot EURUSD for the time period from January 1, 2008 to June 15, 2015.

E-mini futures with underlying stock index S&P 500[1] (denoted with the ticker symbol ES) are traded on the Chicago Mercantile Exchange. The notional value of one contract is 50 times the value of the SP 500 stock index quoted in U.S. dollars. The source of data is Tickdatamarket[2].

Light Crude Oil futures (denoted with the ticker symbol CL) are traded on the New York Mercantile Exchange and represent a blend of several U.S. domestic streams of light sweet crude oil with the delivery point in Cushing, Oklahoma. The notional value of one contract is 1,000 barrels of the Light Crude Oil futures price quoted in U.S. dollars per barrel. The source of data is Tickdatamarket.

Spot gold (denoted with the symbol XAU) is mainly traded through London, in Over-the-Counter (OTC) transactions. *"The governance of this market is maintained through the London Bullion Market Association's (LBMA) publication of the Good Delivery List. This is the list of accredited refiners, whose standards of production and assaying meet the requirements set out in the LBMA's Rules. Only bullion conforming to these standards is acceptable in settlement against transactions conducted between participants in the bullion market"* LBMA (2017). Furthermore,

[1] Standard & Poor's 500, it is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ

[2] see `http://www.tickdatamarket.com/`

LBMA provides clearing system, vaulting servis, good delivery, pricing and statistics. The price discovery takes place twice a day through auctions or so-called "fixings" by ICE Benchmark Administration (IBA) at 10:30 and 15:00 with the price set in U.S. dollars per fine troy ounce[3]. Our source of data is Dukascopy[4] and denoted with the symbol XAUUSD.

Spot EURUSD is the euro and U.S. dollar pair for the currencies of the European Union and the United States. We will use indirect quotation, for instance a quote of 0.70 EURUSD would mean that it takes 0.70 euros to purchase 1 U.S. dollar. Our source of data is forexhistorydatabase.com[5].

All HFD are sampled at 1 minute frequency with candle structure. It means we obtained open, high, low and close price at 1 minute frequency. In case of futures, we talk about continuous front month prices and otherwise spot prices.

4.1.1 Data processing

In the beginning, we check whether there are no zero prices or some abnormal high prices in terms of multiples. We did not find any such an error in our samples. The first decision about data processing is that we are going to work only with close prices. Since we have financial assets traded in different time zones and some of them are not available each minute, we need to apply price synchronization. We decide to synchronize them according to Central European Time zone (CET[6]) taking into account also the differences in dates of daylight saving time between USA and Central Europe.

Secondly, we keep only prices with the same timestamp according to minute, hour and day or in other words, in other words price synchronization at fixed clock time. Due to structural changes in the markets, we kept only prices between time 00:00:00 and 23:00:00 (CET) each trading day what are current standard trading hours on all four markets. Furthermore, we consider as eligible days for trading only from Monday till Friday.

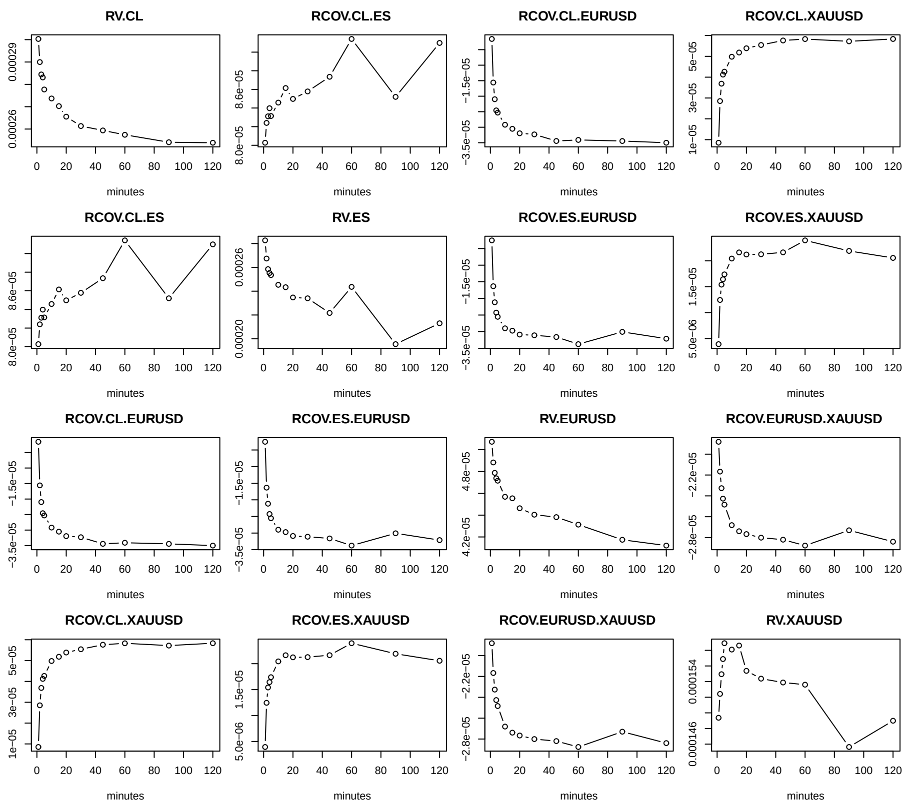
The last step is to compute logarithmic returns as $r_{i,t} = \ln(S_{i,t}) - \ln(S_{i,t-1})$ for $i = 1, \ldots, d$ assets (in our case $d = 4$) and create the time series at different time equidistant frequencies in order to decide about optimal frequency of sampling for the purpose of market microstructure noise and Epps effect reduction. Our decision tool about optimal sampling frequency is the signature covariance plot in Figure 4.1 estimated on full sample of data. We prefer some empirical evidence for the choice of optimal sampling frequency to blindly believe the choice written in articles or studies.[7]

Regarding to prices on daily basis, we are using daily closing prices recorded at 22:00 CET

[3] 1 troy ounce = 31.1034768 g ($\approx$ 1.0971 oz.)
[4] see `https://www.dukascopy.com/swiss/english/marketwatch/historical/`
[5] Unfortunately not available anymore.
[6] Central European Time is 1 hour ahead of Coordinated Universal Time (UTC)
[7] We have seen in some papers or thesis that researches just blindly picked up i.e. 5min frequency and automatically assumed to get an unbiased estimator while ignoring possible unique properties of their underlying financial assets or portfolio which would be against the choice of 5min frequency.

(pseudo-closing prices) for the purpose of correct[8] synchronization. We decided for the time 22:00 CET due to fact that it is closing time of major US stock exchanges and there is only 1 hour left for the closing of US trading session on futures markets. Daily returns are computed in the same manner as high-frequency returns.

Figure 4.1: Signature Covariance Plot



Source: Author's computation

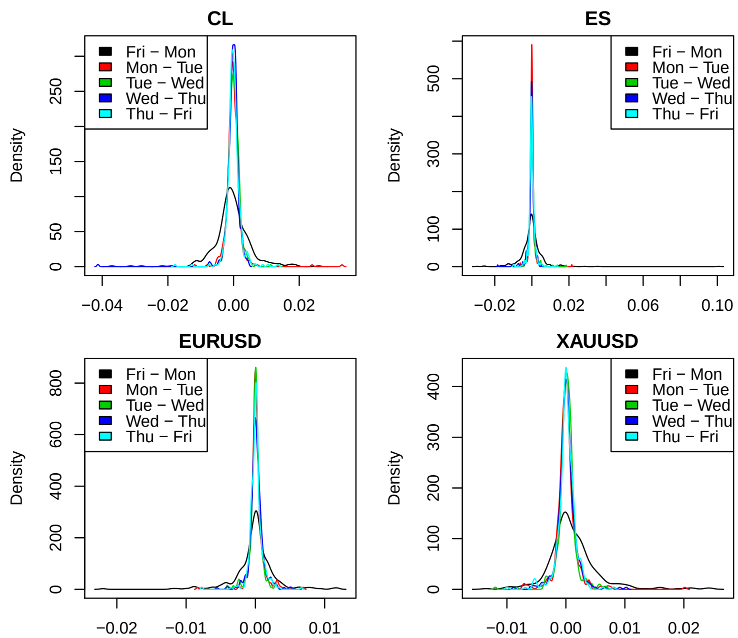
It is shown in Figure 4.1 that volatility and relatively also covariance is stabilized around 20 minute frequency. Thus, we assume that market microstructure noice is sufficiently eliminated at this frequency and so we choose t20 minute frequency for computation of realized covariance matrix defined in Equation 2.9. Additionally, Figure 4.1 provides an empirical evidence for Epps effect. We can observe that negative realized covariances (those including EURUSD) are biased upwards to zero with increasing frequency and positive covariances experience downward bias to zero with increasing frequency.

Another aspect of realized covariance is that implies only using intraday returns from market open to market close what omits overnight return information. In our case when when there

[8]We can find articles or thesis that are using daily closing so-called settlement prices from different exchanges or platforms and at the same time omitting the fact that these settlement prices are recorded at different time. Moreover, the exchanges publish several correction of settlement prices and therefore it can be nearly impossible to construct correct synchronization

is only 1 hour break in the trading during the week, we do not expect the high variability of overnight return wheres the weekend[9] break aggregates a lot of new information and higher variability of over the weekend return is expected. The Figure 4.2 matches with previous expectation where we can see negligible variability in overnight returns within week and just little significance overnight returns from Friday 23:00:00 to Monday 00:00:00 on CL and XAUUSD asset.

Figure 4.2: Densities of overnight returns



Source: Author's computation

Based on Figure 4.2 we omit overnight returns and leave the correction of this deficiency for further research. We are aware that deficiency can cause worse forecasting accuracy of VaR and ES from models using RCOV matrix.

4.1.2 Descriptive statistics

Summary of descriptive statistics of daily and HF returns for our portfolio tells is provided in Table 4.1.

Daily returns It is clear that the least volatile asset is the EURUSD and the closest parameters of skewness and kurtosis to Gaussian distribution but A-D test probably mainly due to fact that it is the most efficient market. The equivalence of empirical distribution of EURUSD to Gaussian

[9]Despite of the fact that weekend days are non-working days in the USA and other developed countries, the prices of our financial assets are reflecting and reacting also to business, political, environmental or military events happening during weekend. Additionally, it is familiar that working hours of companies in financial industry are often prolonged even to weekends. Both phenomena can result in new decisions (which were not known when market was closing on Friday) in trading or investing strategies transferred to markets immediately at the opening markets. The significant changes in prices are known as *gaps* due to large positive or negative in difference between Friday's close price and Monday's open price.

Table 4.1: Descriptive statistics for the full data set (1/1/2008–6/15/2015).

Data	**Asset**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**A-D^a pv**
Daily									
	CL	-0.00	0.02	0.00	-0.09	0.07	-0.23	1.57	0.00
	ES	0.00	0.02	0.00	-0.12	0.15	0.13	12.42	0.00
	EURUSD	0.00	0.01	0.00	-0.04	0.03	0.06	1.76	0.00
	XAUUSD	0.00	0.01	0.00	-0.09	0.11	-0.32	6.76	0.00
20min									
	CL	-0.00	0.00	0.00	-0.03	0.03	-0.03	9.63	0.00
	ES	0.00	0.00	0.00	-0.05	0.05	0.22	38.14	0.00
	EURUSD	-0.00	0.00	0.00	-0.02	0.01	-0.01	11.06	0.00
	XAUUSD	-0.00	0.00	0.00	-0.03	0.03	-0.12	18.02	0.00

Note: [a]Anderson-Darling test of normality with p-value.

Source: Author's computation

distribution is rejected as well as for all other assets. Such zero values of p-values of A-D tests indicate that multivariate distribution of risk-factors of our portfolio will not be Gaussian one.

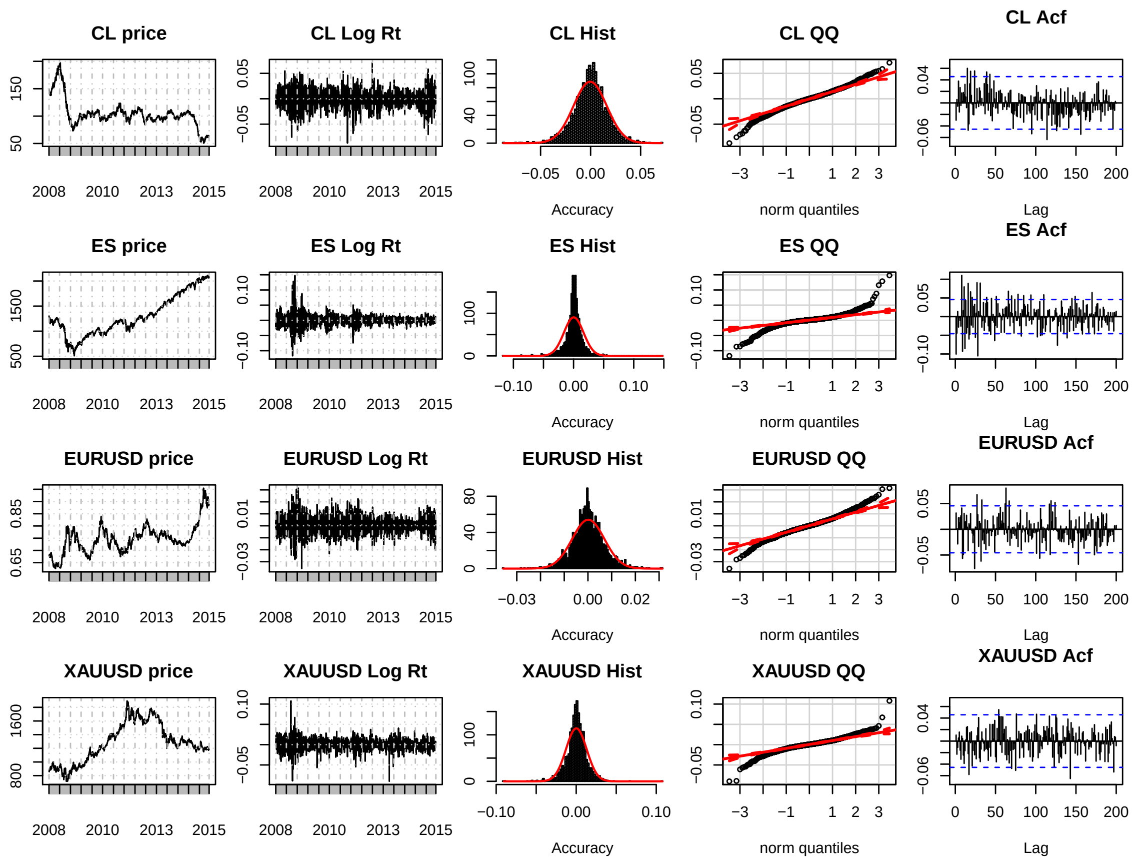
Interesting outcome is that ES has bigger extreme values meaning minimum and maximum and CL. This can be explained that stock markets face to larger daily shocks than CL. On the other hand, higher kurtosis of ES than CL tells that ES has fewer extreme returns from the tails and the returns are more concentrated around mean wheres CL returns are more dispersed around mean. XAUUSD is between CL as ES. It has considerable extreme returns but fewer than CL and more than ES. The explanation can be due to XAUUSD role as "safe haven" on financial and capital markets.

We can find support of previous statements also in Figure 4.3. The histogram with fitted Gaussian distribution and Q-Q plot visually confirm our opinion about the statistical properties of assets in our portfolio.

20min returns The obvious main difference between 20min and daily returns is seen in kurtosis. All assets have significantly higher kurtosis than at daily frequency. Other properties more or less copy the behavior of the daily frequency. The additional evidence for very strong departure from the assumption of Gaussian distribution is provided in Figure 4.4 through histogram and QQ-plot. Another interesting statistics is serial correlation which looks to be insignificant in both frequencies.

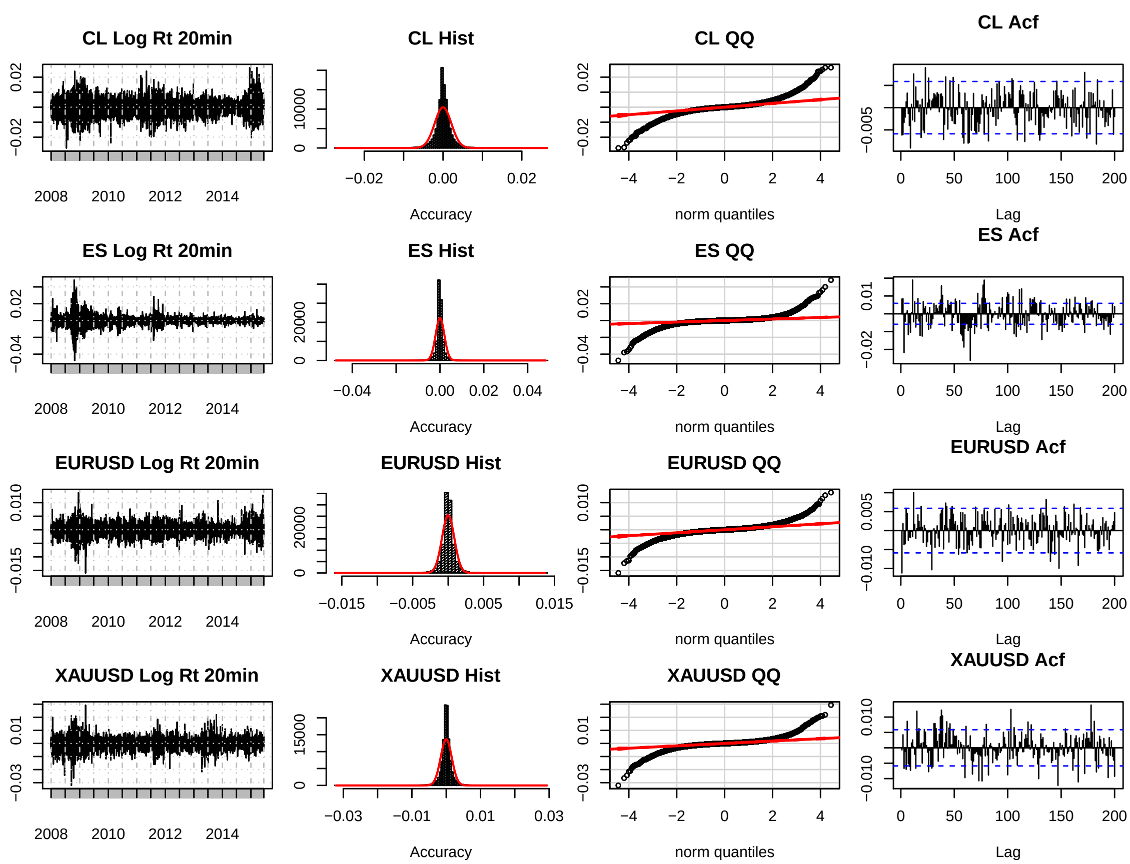
Volatility Our main interest is in volatility and linear dependence represented by RV, respectively RCOV. The visual inspection of Figure 4.5 confirm another stylized fact that squares of daily returns exhibit significant serial correlation or known as long-memory. Regarding to RV, we witness even stronger serial correlation until 150^th lag. Thus, we can evaluate that modeling of RV through HAR model will be an appropriate approach. Moreover, we can see from the

Figure 4.3: Summary statistics of daily returns



Source: Author's computation

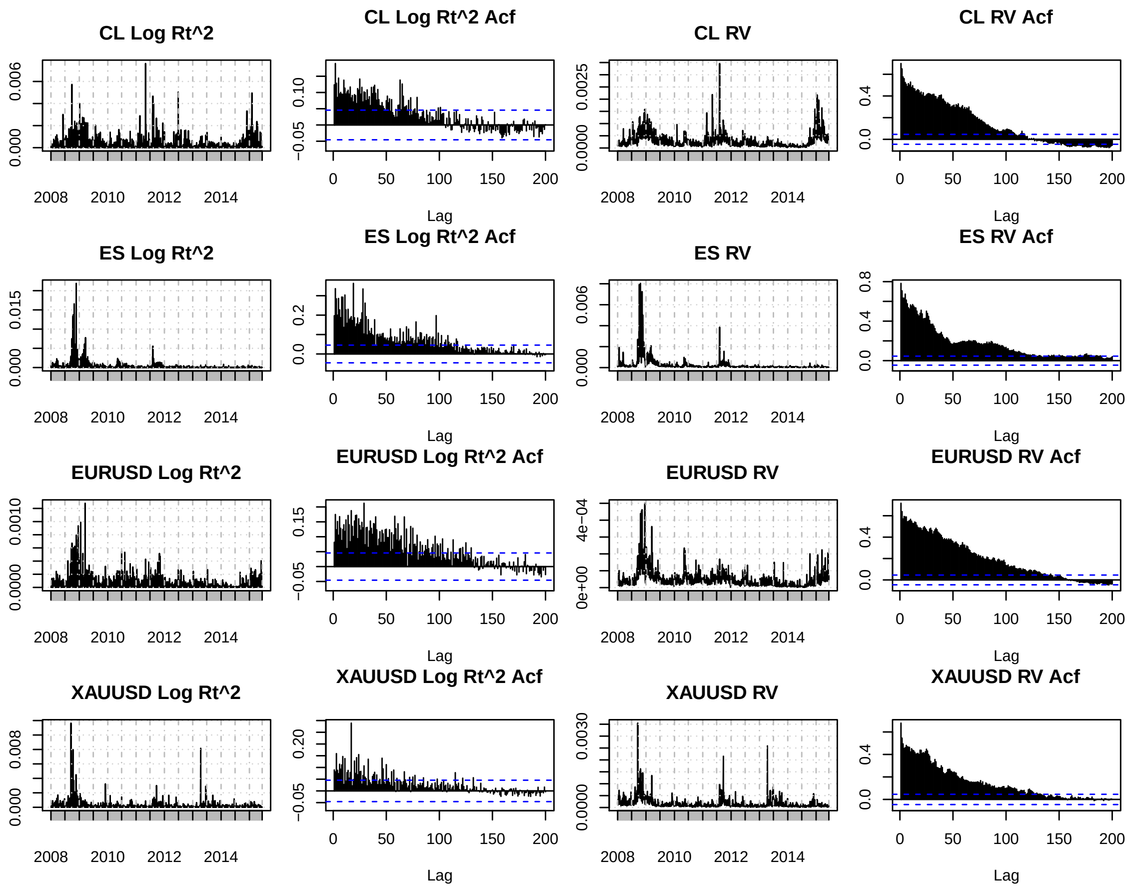
Figure 4.4: Summary statistics of 20min returns



Source: Author's computation

Figure 4.5 stress events in our assets. The most volatile year was definitely 2009 for all assets excluding CL. We will focus on the stress periods more in decision about the in-sample and out-of-sample in Subsection 4.2.1.

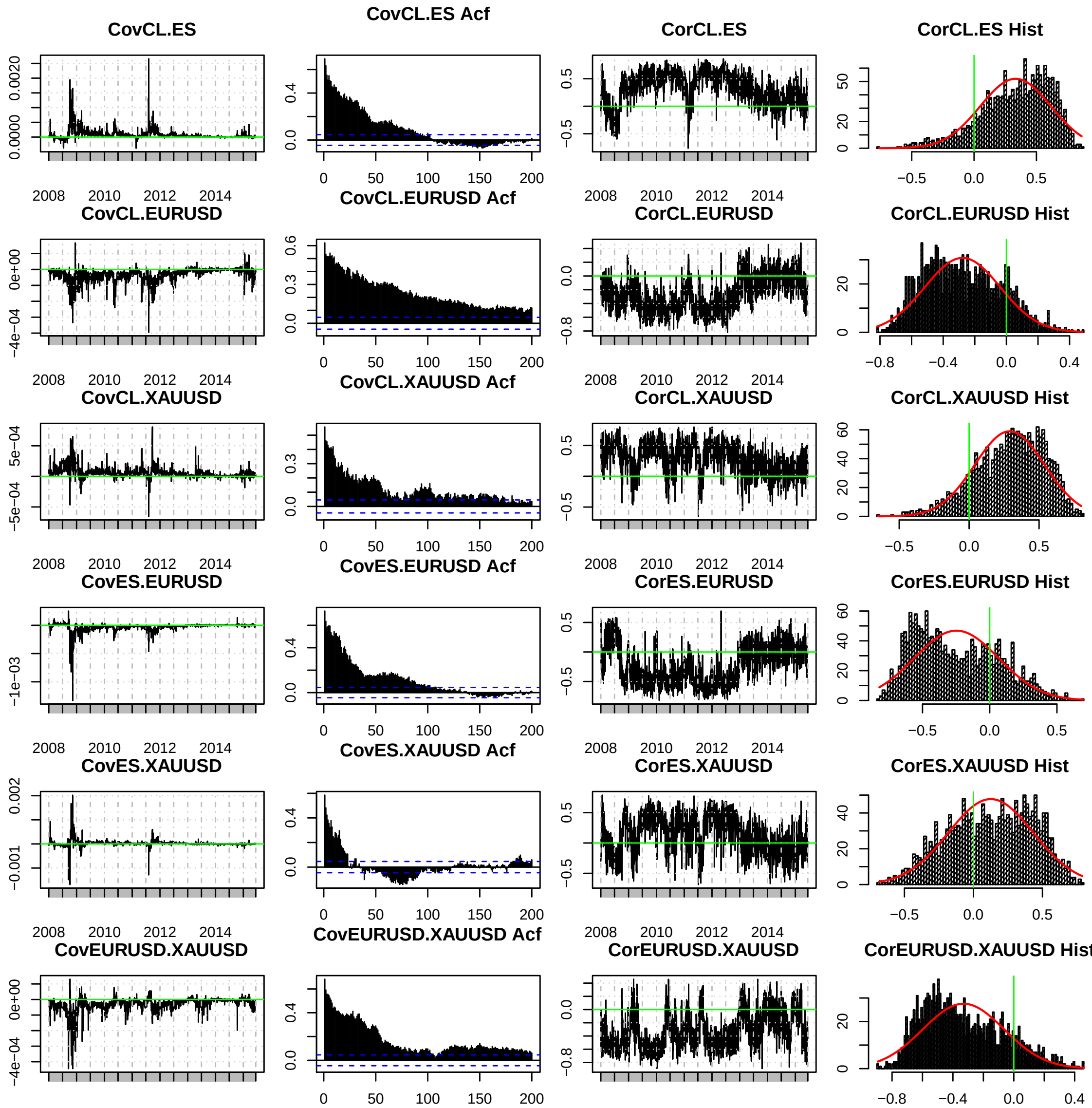
Figure 4.5: Daily squares vs High-frequency RV



Source: Author's computation

Covariance The last one is Figure 4.6 illustrating us the time development of interactions between individual couple of assets in our portfolio. The important property to study of RCOV is also its serial correlation. It is shown again that RCOV exhibit strong long-memory to RV. It give us another incentive for the modeling of RCOV through HAR model. We included also time varying Realized Correlation (RCOR) because it is easier to imagine for the linear dependence than RCOV. It is a nice evidence how correlation can behave steadily within certain range and after a few years it sets up a new range, for instance RCOR between CL and EURUSD or CL and ES. Thus, intended dynamic modeling of RCOV/RCOR in this master thesis will be more than appropriate. The most of histograms of RCOR are economically justifiable. The exception is RCOR between ES and EURUSD giving the bimodal distribution what can be quite interesting issue for further investigation and understanding.

Figure 4.6: RCOV and RCOR



Source: Author's computation

4.2 Application

After initial analysis of our data, we describe briefly the implementation of cross-validation technique as a basis for backtesting and model selection methods. Then, we outline the individual steps in the implementation of methodology introduced in Chapter 3.

4.2.1 In-sample and out-of-sample

The idea of cross-validation for the purpose of forecasting is about to split of data set between training sample so-called *in-sample* and testing sample so-called *out-of-sample*. The main goal is to estimate statistical model on in-sample data and consequently perform the forecasting which is compared against the out-of-sample. Such technique helps to avoid overfitting (in-sample and out-of-sample is actually identical sample). Moreover, it allows to measure the accuracy of the forecasts under realistic conditions because we usually do not know the future when we make forecast.

In terms of time series, unfortunately we cannot make several iterations of those splits of data because we have for each day only one realization. At least we divide our full sample into two subsamples that represent two different stress periods and subsequently we split each subsample between in-sample and out-of-sample. We choose the proportion for each in-sample as 67% of data and out-of-sample remaining 33%. We characterize three scenarios as *Full* sample meaning that we used all data we had. The second one is a subsample of time period with *High* volatility with some stress events and the third one is a subsample of time period with *Low* volatility.

1. **Full** The initial in-sample covers the period from January 3, 2008 until December 31, 2012 consisting of 1,240 business days. The period is distinguished by the latest financial crises in 2007-2009, then temporarily calm period in 2010 which was replaced by again high volatility due to European debt crises and again followed by low volatility in 2012. The out-of-sample covers the period from January 2, 2013 until June 12, 2015 consisting of 604 business days. There were no huge stress events apart from increased volatility on CL at the end of the out-of-sample.

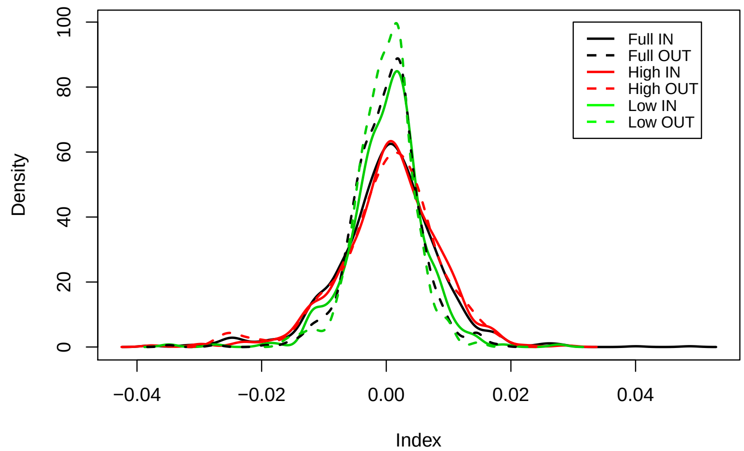
2. **High** The initial in-sample covers the period from January 2, 2009 until December 31, 2010 consisting of 501 business days. The out-of-sample covers the period from January 3, 2011 until December 30, 2011 consisting of 249 business days. As it was already said, the both samples exhibit the highest price movements.

3. **Low** The initial in-sample covers the period from January 3, 2012 until December 31, 2013 consisting of 495 business days. The out-of-sample covers the period from January 2, 2014 until December 31, 2014 consisting of 247 business days. Both samples are almost without no stress in the markets.

An useful illustration of all these scenarios is done through their densities in Figure 4.7. Both scenarios Low and High have almost identical densities for in-sample and out-of-sample and only

Full scenario shows that in-sample contained fat tails that were not repeated in out-of-sample anymore.

Figure 4.7: Densities of daily returns from in-samples and out-of-samples during different volatility scenarios



Source: Author's computation

However, the Figure 4.7 was only illustration because since we make only day ahead forecasts, we establish moving window in order to include always the latest information and leave out the oldest information. Our moving window has the length of initial in-sample.
For example, considering Full scenario, we make the first forecast for January 2, 2013 using data from January 3, 2008 until December 31, 2012. The next forecast for January 3, 2013 is done by using data from January 4, 2008 until January 2, 2013. Each forecast includes the re-estimation of the model. This is computationally demanding, especially if it includes Monte Carlo simulations. Thus, there is also practice to re-estimate the model by each i.e. 5^{th} or 10^{th} business day. It has additional advantage that estimated parameters are more stable. However, we stick with daily re-estimation in this master thesis.

4.2.2 Implementation of models

We briefly describe how we switched theoretically defined models in Chapter 3 into practice. The necessary component of the below models is the portfolio logarithmic return defined defined as weighted sum of the individual financial asset returns. In this master thesis we work with equally weighted portfolio $\boldsymbol{w}_i = 0.25$ for $i = 1, \dots, d$. Moreover, we assume that our portfolio is perfectly[10] re-balanced each day while we omit transaction cost[11].

1. **Parametric** models are pretty straightforward for the estimation and the fastest methods. HFD models consisting of HAR and Leveraged Heterogeneous Autoregression (LHAR)

[10]It is an unrealistic condition because futures are standardized products and we cannot trade arbitrary sizes but for the sake of simplicity the academy prefers this condition
[11]Another unrealistic condition omitting brokerage, exchange, clearing and other fees or costs connected with re-balancing

(asymmetric version) were estimated by regular OLS method which is implemented in each statistical application for regressions. EWMA model does not even require estimation if we use estimated parameter by RiskMetrics, so it is clear winner in terms of speed.

Regarding to DCC-GARCH and its asymmetric version, they are more sophisticated but there are available very comprehensive and well-documented `R` packages for univariate estimation `rugarch` and multivariate estimation `rmgarch`. Thanks to linear transformation of forecasted covariance matrix, there is left only the computation of quantile of distribution representing the innovations.

The only hurdle is to estimate properly the degrees of freedom for standardized Student's t distribution of innovations. We solved it that we estimated the degrees of freedom from the innovations that were filtered for the purpose of multivariate FHS since we wanted to estimate it in the multivariate dimension. The much faster approach would be in univariate dimension. First step is to calculate the portfolio return and fitted volatility (from fitted covariance matrix) and subsequently to divide such return by volatility and resulting outcome is univariate innovation.

2. **Semi-parametric** method is very well described in Chapter 3 and consists of only vector and matrix operations. Nevertheless, it is more time consuming than Parametric method because we work need to filtrate entire in-sample with fitted covariance matrix in order to get the empirical distribution of innovations.

3. **Monte Carlo - Copula** is unambiguously the most sophisticated approach applied in this master thesis. The level of sophistication can be easily increased thanks to simplicity of incorporation of other models. Our highest sophistication was achieved by joining of volatility models with EVT and consequently Copula. We proceeded following steps:

    (a) Individual creating innovations for each asset by dividing the historical daily returns by volatility fitted from either HFD model or model using daily data.

    (b) In case of incorporation of EVT, we fit semi-parametric distribution consisting of tails from GPD and Gaussian kernel for the interior part (using the package `spd`). The key parameter for the optimal fit of GPD is the threshold level. The most common way is the visual inspection of mean excess function described in Section 3.2.3. This method is not practical if we use moving window with daily re-estimation and require automation of the decision about threshold level. We follow findings from the Monte Carlo experiment in McNeil *et al.* (2015, pp. 161-162) where they concluded that optimal choice of the threshold level would be from the sample of 100-150 exceedances. Therefore, we decide to take 10% observations from each tail in case of Full scenario (124 observations) and 20% from each tail in case of High and Low scenario (100, resp. 99 observations).

    (c) Transformation of each asset's innovations to uniformly distributed variable by:

        i. Through empirical distribution function using function `pobs` from package copula

ii. Through standard distribution function such as Gaussian or Student's t (in this case, we estimated degrees of freedom from the innovations).

iii. Through semi-parametric distribution meaning GPD+Kernel fitted in previous step.

(d) Copula fit using the package `copula` with copula parameter estimator Kendall's tau.

(e) 10,000 simulations from the fitted copula

(f) Transformation of simulated uniformly distributed variables through quantile function of Gaussian, Student's t or previously estimated semi-parametric distribution.

(g) Backward filtration by forecasted day ahead volatility from the same model as in the beginning.

(h) Transformation of logarithmic returns to arithmetic ones because they are portfolio additive.

(i) Calculation of portfolio arithmetic return by weights. So we got final univariate time series.

(j) Transformation of portfolio returns back to logarithmic ones because they are time additive. This property is exploited in following backtesting methods.

(k) Final computation of the required quantile from 10,000 simulated portfolio logarithmic returns as day ahead forecast of VaR and average value of exceedances of VaR as day ahead forecast of ES

The summary of applied models is provided in Table 4.2. In case of distribution column next to copula models, the first distribution means transformation to uniform distribution (step (c)) and the second distribution means transformation from uniform distribution (step (f)).

The proceeded forecasting of VaR and ES from 18 models of HAR using HFD and 24 models using daily data is shown for each scenario in Figure 4.8, respectively Figure 4.9.

4.3 Results of backtesting

4.3.1 Value at Risk

Our evaluation strategy of VaR consits of two steps. In the first step, we conduct statistical tests, specifically unconditional and conditional coverage test and dynamic quantile test. In the second step, we assess significantly valid models according to their value of regulator, firm and asymmetric loss function. Afterwards, we sort the models according to their value of asymmetric loss function in ascending order. Finally, we briefly discuss our findings.

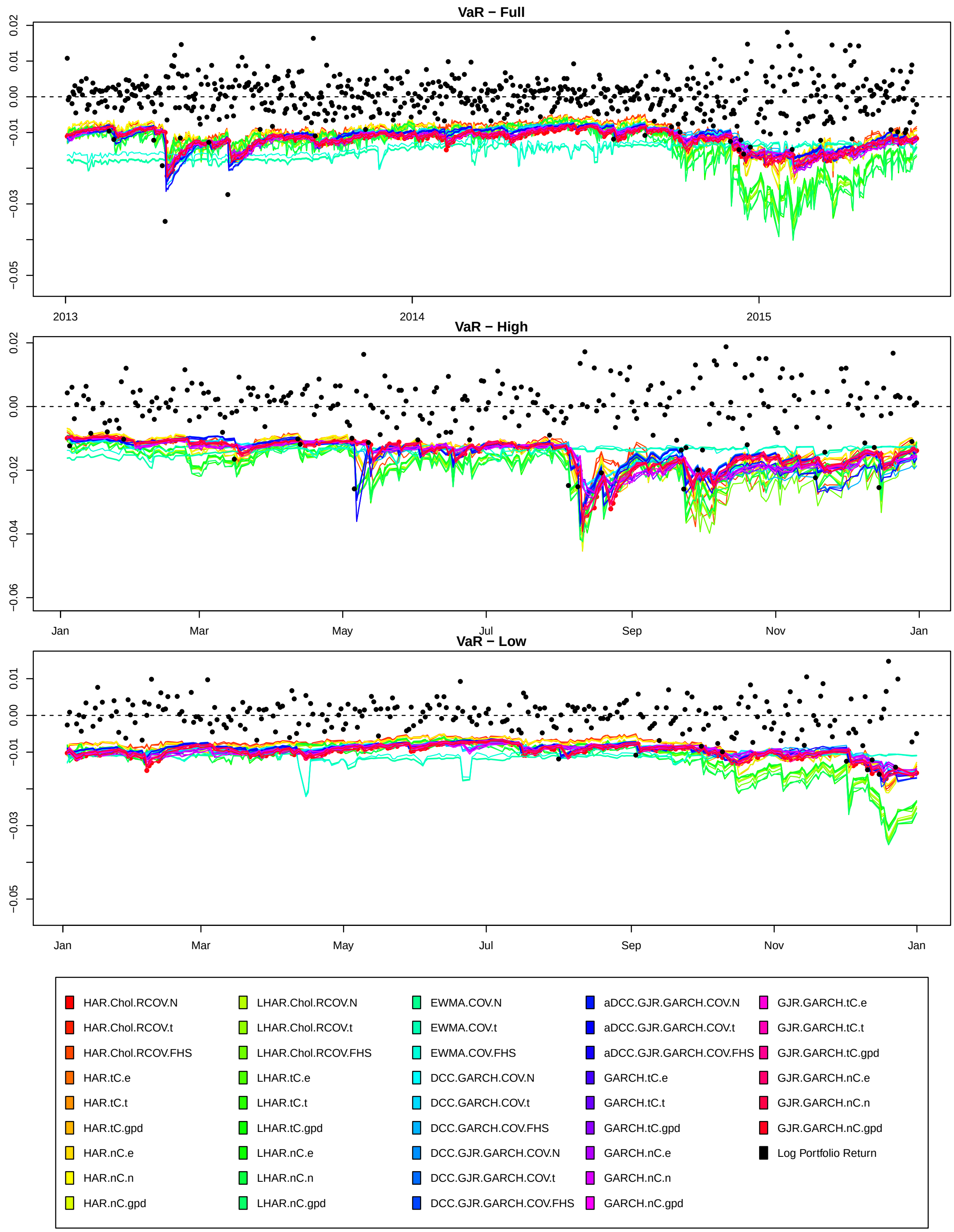
Before we start discussion about the results of our assessment of VaR forecasts, we explain a bit the implementation of loss functions. The important factor is the performance criterion of loss function. We calculate an average value of loss function as a performance criterion for all applied loss functions in this master thesis.

Table 4.2: The list of applied methods, distribution and resulting models.

VaR/ES Method	**Data**	**Variance-Covariance model**	**Distribution**	**Model name**
Parametric				
	20min			
		HAR.Chol	MV-Normal	HAR.Chol.RCOV.N
		HAR.Chol	MV-t	HAR.Chol.RCOV.t
		LHAR.Chol	MV-Normal	LHAR.Chol.RCOV.N
		LHAR.Chol	MV-t	LHAR.Chol.RCOV.t
	Daily			
		EWMA	MV-Normal	EWMA.COV.N
		EWMA	MV-t	EWMA.COV.t
		DCC-GARCH(1,1)	MV-Normal	DCC.GARCH.COV.N
		DCC-GARCH(1,1)	MV-t	DCC.GARCH.COV.t
		DCC-GJR-GARCH(1,1)	MV-Normal	DCC.GJR.GARCH.COV.N
		DCC-GJR-GARCH(1,1)	MV-t	DCC.GJR.GARCH.COV.t
		ADCC-GJR-GARCH(1,1)	MV-Normal	aDCC.GJR.GARCH.COV.N
		ADCC-GJR-GARCH(1,1)	MV-t	aDCC.GJR.GARCH.COV.t
Semi-parametric				
	20min			
		HAR.Chol	FHS	HAR.Chol.FHS
		LHAR.Chol	FHS	LHAR.Chol.FHS
	Daily			
		EWMA	FHS	EWMA.COV.FHS
		DCC-GARCH(1,1)	FHS	DCC.GARCH.COV.FHS
		DCC-GJR-GARCH(1,1)	FHS	DCC.GJR.GARCH.COV.FHS
		ADCC-GJR-GARCH(1,1)	FHS	aDCC.GJR.GARCH.COV.FHS
Monte Carlo				
	20min			
		HAR-GaussCopula	Normal-Normal	HAR.nC.n
		HAR-GaussCopula	Empirical-Normal	HAR.nC.e
		HAR-tCopula	t-t	HAR.tC.t
		HAR-tCopula	Empirical-t	HAR.tC.e
		HAR-EVT-GaussCopula	GPD+Kernel	HAR.nC.gpd
		HAR-EVT-tCopula	GPD+Kernel	HAR.tC.gpd
		LHAR-GaussCopula	Normal-Normal	LHAR.nC.n
		LHAR-GaussCopula	Empirical-Normal	LHAR.nC.e
		LHAR-tCopula	t-t	LHAR.tC.t
		LHAR-tCopula	Empirical-t	LHAR.tC.e
		LHAR-EVT-GaussCopula	GPD+Kernel	LHAR.nC.gpd
		LHAR-EVT-tCopula	GPD+Kernel	LHAR.tC.gpd
	Daily			
		GARCH(1,1)-GaussCopula	Normal-Normal	GARCH.nC.n
		GARCH(1,1)-GaussCopula	Empirical-Normal	GARCH.nC.e
		GARCH(1,1)-tCopula	t-t	GARCH.tC.t
		GARCH(1,1)-tCopula	Empirical-t	GARCH.tC.e
		GARCH(1,1)-EVT-GaussCopula	GPD+Kernel	GARCH.nC.gpd
		GARCH(1,1)-EVT-tCopula	GPD+Kernel	GARCH.tC.gpd
		GJR-GARCH(1,1)-GaussCopula	Normal-Normal	GJR.GARCH.nC.n
		GJR-GARCH(1,1)-GaussCopula	Empirical-Normal	GJR.GARCH.nC.e
		GJR-GARCH(1,1)-tCopula	t-t	GJR.GARCH.tC.t
		GJR-GARCH(1,1)-tCopula	Empirical-t	GJR.GARCH.tC.e
		GJR-GARCH(1,1)-EVT-GaussCopula	GPD+Kernel	GJR.GARCH.nC.gpd
		GJR-GARCH(1,1)-EVT-tCopula	GPD+Kernel	GJR.GARCH.tC.gpd

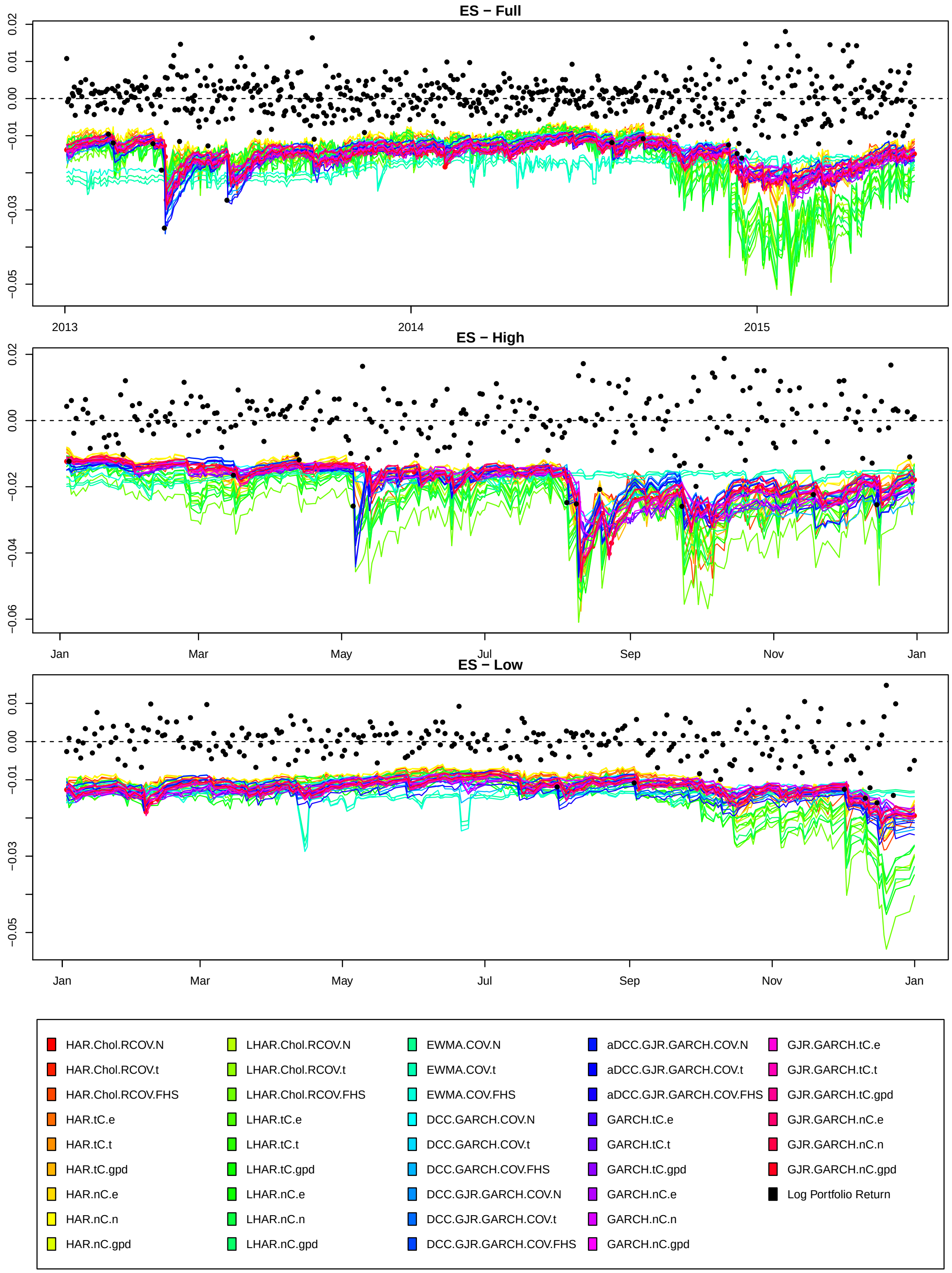
Source: Author's computation

Figure 4.8: Forecasts of day-ahead $\text{VaR}_{97.5\%}$ for all scenarios



Source: Author's computation

Figure 4.9: Forecasts of day-ahead $\mathrm{ES}_{97.5\%}$ for all scenarios



Source: Author's computation

The Regulator Loss Function (RLF) is calculated according to its definition in Subsection 3.3.4. The Firm Loss Function (FLF) is also calculated according to its previous definition with our adjustment of interest rate. Instead of interest rate, we calculate the difference between market return and risk free return as opportunity cost of reserved capital for VaR measure. We choose market return to be represented by daily log return of settlement prices of futures product ES1[12] provided by Quandl and risk free return is represented by daily log return of settlement prices of US Treasury notes futures TU1[13] with maturity 2 years provided by Quandl. Obviously, the agent would choose the yield of its i.e. trading desk or other opportunity that has better yield than reserved capital in some very liquid assets with very low yield.

Additionally to this, we calculated also a percentile of RLF and FLF in order to have better visual understanding how many competitors the each model beats. Percentiles of RLF and FLF are denoted with symbol %.

Regarding to statistical tests, we use significance level to be 5% and excluded all VaR models that did not pass at least one of these statistical tests (it is rather a strict condition). Finally, we recalculated again our loss functions percentiles in order to see how many left models are worse.

The final results are summarized in below tables and they reveal many interesting information. We start with individual scenarios and then with the global view across all scenarios.

Full scenario has the first places dominantly occupied by HAR models. The best two models according to FLF and ALF are parametric models of RCOV following multivariate Student's t and Gaussian distribution. It is a bit surprising that the first model has the Gaussian distribution since the evidence from our descriptive statistics suggested the non-Gaussian distribution. Anyway, from the third to sixth position, we see t and Gaussian copula models utilizing again HAR models. Overall, the HAR models on top position have also very favourable violation ratio close to one and also high p-values. The viewpoint from another angle are the values of RLF. They suggest exact opposite and rank the most conservative models as the best ones. Thus, our assumption was met in this particular case that FLF and RLF tend to suggest models in inverse way.

High scenario has assigned symmetric and asymmetric version of HAR again on top positions according to ALF as well as RLF. Although we need to be careful about the first model as it has p-value of Unconditional test only 0.08 what would be even rejected under significance level of 10%. Unlike first symmetric HAR, the very good performance can be seen by acLHAR in connection with Gaussian or Student's t copula and GPD distribution of tails. Both models have perfect violation ratio, high p-values and RLF gives them the lowest regulator loss. Nevertheless, all nine variants of LHAR are among top eleven models. Moreover, copula functions with GPD distribution among top four. Thus, we have an evidence that the choice of copula, GPD and

[12] Continuous front month product

[13] Continuous front month product

Table 4.3: VaR and ES Test I. results for Full sample and significant models.

Model	**Viol. ratio**	**UC pv**	**CC pv**	**DQ pv**	**RLF**	**RLF %**	**FLF**	**FLF %**	**ALF**	**ALF %**	**Z1**	**Z1 pv**
HAR.Chol.RCOV.N	1.3	0.22	0.44	0.62	0.019	0	0.243	100	0.039	73	-0.27	0.58
HAR.Chol.RCOV.t	1.2	0.46	0.64	0.66	0.018	15	0.244	94	0.039	73	-0.18	0.48
HAR.tC.e	1.1	0.82	0.72	0.61	0.019	0	0.248	61	0.039	73	-0.22	0.27
HAR.tC.t	1.1	0.82	0.72	0.61	0.019	0	0.248	61	0.039	73	-0.24	0.24
HAR.nC.e	1.1	0.63	0.70	0.65	0.019	0	0.247	76	0.039	73	-0.33	0.35
HAR.nC.n	1.1	0.63	0.70	0.66	0.019	0	0.246	88	0.039	73	-0.34	0.31
DCC.GJR.GARCH.COV.N	0.8	0.40	0.34	0.30	0.015	61	0.248	61	0.039	73	-0.30	0.28
DCC.GJR.GARCH.COV.t	0.7	0.26	0.22	0.29	0.015	61	0.250	45	0.039	73	-0.20	0.23
aDCC.GJR.GARCH.COV.N	0.8	0.40	0.34	0.30	0.015	61	0.249	55	0.039	73	-0.29	0.30
aDCC.GJR.GARCH.COV.t	0.7	0.16	0.13	0.17	0.014	94	0.250	45	0.039	73	-0.23	0.16
HAR.Chol.RCOV.FHS	1.3	0.33	0.55	0.56	0.018	15	0.244	94	0.040	42	-0.11	0.41
HAR.nC.gpd	0.7	0.16	0.13	0.16	0.015	61	0.257	9	0.040	42	-0.33	0.10
DCC.GARCH.COV.N	0.9	0.57	0.47	0.32	0.016	45	0.246	88	0.040	42	-0.30	0.27
DCC.GARCH.COV.t	0.8	0.40	0.34	0.26	0.015	61	0.247	76	0.040	42	-0.19	0.25
DCC.GJR.GARCH.COV.FHS	0.7	0.26	0.22	0.14	0.015	61	0.254	24	0.040	42	-0.17	0.16
aDCC.GJR.GARCH.COV.FHS	0.8	0.40	0.34	0.24	0.015	61	0.254	24	0.040	42	-0.14	0.18
GJR.GARCH.tC.e	0.7	0.26	0.22	0.16	0.017	24	0.252	33	0.040	42	-0.22	0.17
GJR.GARCH.tC.t	0.8	0.40	0.34	0.26	0.017	24	0.251	36	0.040	42	-0.20	0.20
GJR.GARCH.nC.e	0.8	0.40	0.34	0.24	0.017	24	0.248	61	0.040	42	-0.33	0.19
GJR.GARCH.nC.n	0.8	0.40	0.34	0.26	0.017	24	0.250	45	0.040	42	-0.34	0.19
HAR.tC.gpd	0.7	0.16	0.13	0.16	0.015	61	0.259	6	0.041	15	-0.29	0.08
GARCH.tC.e	0.8	0.40	0.34	0.20	0.017	24	0.248	61	0.041	15	-0.21	0.17
GARCH.tC.t	0.8	0.40	0.34	0.19	0.017	24	0.249	55	0.041	15	-0.22	0.16
GARCH.tC.gpd	0.7	0.26	0.22	0.13	0.016	45	0.251	36	0.041	15	-0.19	0.19
GARCH.nC.e	0.8	0.40	0.34	0.22	0.017	24	0.247	76	0.041	15	-0.36	0.16
GARCH.nC.n	0.8	0.40	0.34	0.22	0.018	15	0.247	76	0.041	15	-0.36	0.14
GARCH.nC.gpd	0.7	0.26	0.22	0.14	0.016	45	0.251	36	0.041	15	-0.24	0.19
GJR.GARCH.tC.gpd	0.6	0.09	0.07	0.06	0.016	45	0.254	24	0.041	15	-0.28	0.08
GJR.GARCH.nC.gpd	0.6	0.09	0.07	0.07	0.016	45	0.255	21	0.041	15	-0.32	0.08
LHAR.tC.e	0.6	0.09	0.07	0.14	0.015	61	0.257	9	0.042	6	-0.25	0.14
LHAR.nC.e	0.7	0.16	0.13	0.21	0.015	61	0.257	9	0.042	6	-0.36	0.20
LHAR.nC.n	0.6	0.09	0.07	0.14	0.015	61	0.256	18	0.042	6	-0.42	0.12
LHAR.Chol.RCOV.N	0.6	0.09	0.07	0.15	0.012	97	0.262	3	0.043	0	-0.29	0.34
LHAR.Chol.RCOV.FHS	0.6	0.09	0.07	0.15	0.012	97	0.264	0	0.043	0	-0.09	0.27

Source: Author's computation

Table 4.4: VaR and ES Test I. results for High sample and significant models.

Model	Viol. ratio	UC pv	CC pv	DQ pv	RLF	RLF %	FLF	FLF %	ALF	ALF %	Z1	Z1 pv
HAR.tC.gpd	1.8	0.08	0.17	0.42	0.014	67	0.835	37	0.059	93	-0.01	0.95
LHAR.tC.gpd	1.0	0.93	0.86	0.99	0.006	100	0.861	4	0.059	93	0.05	
LHAR.nC.gpd	0.8	0.61	0.79	0.99	0.007	93	0.856	7	0.059	93	-0.03	
HAR.nC.gpd	1.8	0.08	0.17	0.42	0.015	63	0.837	30	0.060	81	-0.07	0.93
LHAR.Chol.RCOV.N	1.3	0.49	0.60	0.95	0.009	85	0.853	19	0.060	81	-0.06	
LHAR.Chol.RCOV.t	1.1	0.76	0.78	0.98	0.009	85	0.856	7	0.060	81	0.01	0.91
LHAR.Chol.RCOV.FHS	1.3	0.49	0.60	0.64	0.007	93	0.880	0	0.061	63	0.23	
LHAR.tC.e	1.4	0.29	0.41	0.47	0.012	74	0.836	33	0.061	63	0.02	0.90
LHAR.tC.t	1.4	0.29	0.41	0.48	0.013	70	0.835	37	0.061	63	0.01	0.90
LHAR.nC.e	1.4	0.29	0.41	0.45	0.012	74	0.831	48	0.061	63	-0.10	
LHAR.nC.n	1.4	0.29	0.41	0.48	0.012	74	0.839	26	0.061	63	-0.10	0.90
GJR.GARCH.tC.gpd	1.4	0.29	0.41	0.65	0.022	48	0.807	56	0.062	52	-0.15	0.65
GJR.GARCH.nC.n	1.6	0.16	0.24	0.56	0.025	22	0.795	78	0.062	52	-0.24	0.70
GJR.GARCH.nC.gpd	1.4	0.29	0.41	0.65	0.022	48	0.807	56	0.062	52	-0.21	0.56
GJR.GARCH.tC.e	1.4	0.29	0.41	0.65	0.024	41	0.797	74	0.063	41	-0.17	0.58
GJR.GARCH.tC.t	1.4	0.29	0.41	0.65	0.025	22	0.803	63	0.063	41	-0.20	0.48
GJR.GARCH.nC.e	1.6	0.16	0.24	0.56	0.025	22	0.792	81	0.063	41	-0.26	0.67
DCC.GARCH.COV.FHS	1.4	0.29	0.41	0.62	0.024	41	0.819	52	0.064	26	-0.18	0.29
DCC.GJR.GARCH.COV.FHS	1.8	0.08	0.17	0.33	0.021	56	0.855	15	0.064	26	-0.11	0.73
aDCC.GJR.GARCH.COV.FHS	1.8	0.08	0.17	0.33	0.021	56	0.851	22	0.064	26	-0.11	0.70
GARCH.nC.gpd	1.6	0.16	0.24	0.45	0.025	22	0.801	67	0.064	26	-0.17	0.62
aDCC.GJR.GARCH.COV.t	1.8	0.08	0.13	0.40	0.026	15	0.834	44	0.065	7	-0.14	0.70
GARCH.tC.e	1.6	0.16	0.24	0.44	0.026	15	0.788	96	0.065	7	-0.15	0.53
GARCH.tC.gpd	1.6	0.16	0.24	0.42	0.025	22	0.790	93	0.065	7	-0.13	0.64
GARCH.nC.e	1.6	0.16	0.24	0.45	0.027	7	0.791	89	0.065	7	-0.27	0.51
GARCH.nC.n	1.6	0.16	0.24	0.43	0.028	0	0.787	100	0.065	7	-0.28	0.47
DCC.GARCH.COV.t	1.8	0.08	0.17	0.36	0.028	0	0.801	67	0.066	0	-0.15	0.59
GARCH.tC.t	1.6	0.16	0.24	0.40	0.027	7	0.792	81	0.066	0	-0.15	0.54

Source: Author's computation

asymmetric version of HAR are appropriate and accurate methods for modeling the extreme events for the purpose of forecasting VaR. We witness again the inverse phenomena when FLF assigns low average loss to symmetric model GARCH whereas this model does not beat any other competitor from the point of view ALF and FLF.

Table 4.5: VaR and ES Test I. results for Low sample and significant models.

Model	**Viol. ratio**	**UC pv**	**CC pv**	**DQ pv**	**RLF**	**RLF %**	**FLF**	**FLF %**	**ALF**	**ALF %**	**Z1**	**Z1 pv**
HAR.Chol.RCOV.N	1.1	0.74	0.77	0.53	0.004	0	0.201	24	0.028	83	-0.21	0.70
HAR.Chol.RCOV.t	1.1	0.74	0.77	0.53	0.004	0	0.203	17	0.028	83	-0.09	0.74
DCC.GJR.GARCH.COV.N	0.8	0.62	0.80	0.29	0.002	31	0.193	69	0.028	83	-0.10	0.84
DCC.GJR.GARCH.COV.t	0.8	0.62	0.80	0.29	0.002	31	0.194	55	0.028	83	0.00	0.86
aDCC.GJR.GARCH.COV.N	0.8	0.62	0.80	0.29	0.002	31	0.193	69	0.028	83	-0.09	0.86
aDCC.GJR.GARCH.COV.t	0.8	0.62	0.80	0.29	0.002	31	0.194	55	0.028	83	0.02	
HAR.Chol.RCOV.FHS	1.1	0.74	0.77	0.56	0.004	0	0.201	24	0.029	38	0.04	0.74
HAR.tC.e	1.1	0.74	0.77	0.61	0.003	10	0.204	7	0.029	38	-0.08	0.63
HAR.tC.t	1.1	0.74	0.77	0.60	0.003	10	0.204	7	0.029	38	-0.08	0.66
HAR.nC.e	1.1	0.74	0.77	0.55	0.003	10	0.204	7	0.029	38	-0.16	0.80
HAR.nC.n	1.1	0.74	0.77	0.58	0.003	10	0.202	21	0.029	38	-0.17	0.76
DCC.GJR.GARCH.COV.FHS	0.8	0.62	0.80	0.39	0.002	31	0.195	45	0.029	38	0.07	0.84
aDCC.GJR.GARCH.COV.FHS	0.8	0.62	0.80	0.36	0.002	31	0.194	55	0.029	38	0.14	
GJR.GARCH.tC.e	0.8	0.62	0.80	0.37	0.002	31	0.198	38	0.029	38	-0.01	
GJR.GARCH.tC.t	0.8	0.62	0.80	0.36	0.002	31	0.196	41	0.029	38	-0.00	
GJR.GARCH.tC.gpd	0.8	0.62	0.80	0.34	0.001	97	0.201	24	0.029	38	0.05	
GJR.GARCH.nC.e	0.8	0.62	0.80	0.37	0.002	31	0.195	45	0.029	38	-0.11	0.80
GJR.GARCH.nC.n	0.8	0.62	0.80	0.36	0.002	31	0.195	45	0.029	38	-0.09	
GJR.GARCH.nC.gpd	0.8	0.62	0.80	0.36	0.001	97	0.201	24	0.029	38	0.00	
HAR.tC.gpd	0.8	0.62	0.80	0.41	0.002	31	0.214	0	0.030	3	-0.01	
HAR.nC.gpd	0.8	0.62	0.80	0.40	0.002	31	0.213	3	0.030	3	-0.05	
DCC.GARCH.COV.N	1.0	0.94	0.86	0.56	0.003	10	0.187	93	0.030	3	-0.12	
DCC.GARCH.COV.t	0.8	0.62	0.80	0.38	0.002	31	0.187	93	0.030	3	-0.05	0.55
GARCH.tC.e	1.1	0.74	0.77	0.07	0.002	31	0.190	79	0.030	3	0.04	0.89
GARCH.tC.t	1.1	0.74	0.77	0.65	0.002	31	0.190	79	0.030	3	0.03	
GARCH.tC.gpd	0.8	0.62	0.80	0.40	0.002	31	0.194	55	0.030	3	-0.00	
GARCH.nC.e	1.1	0.74	0.77	0.07	0.002	31	0.190	79	0.030	3	-0.07	
GARCH.nC.n	1.1	0.74	0.77	0.06	0.002	31	0.189	90	0.030	3	-0.07	
GARCH.nC.gpd	0.8	0.62	0.80	0.39	0.002	31	0.193	69	0.030	3	-0.03	
DCC.GARCH.COV.FHS	0.8	0.62	0.80	0.41	0.003	10	0.187	93	0.031	0	0.10	

Source: Author's computation

Low scenario is the most balanced scenario where all models have almost indentical violation ratio closely around one. Additionally, all p-values of statistical tests are very similar. The loss functions have ambiguous vales and are not able to distinguish the models in the way we have seen in previous scenarios. This can be a nice example that if we validate and try to select the models on the time period without volatility, then it is very little robust decision about the best model. In this case, it looks that any of the validated models have sufficient accuracy in the forecasts of VaR.

Overall, the HAR models using high-frequency data showed very good performance in Full scenario and especially its asymmetric version in High scenario. The benchmark model DCC-GARCH performed relatively poorly and it required at least to incorporate its asymmetric version GJR in order to improve the performance. The superiority of GPD and copula in highly volatile times made an excellent job and performed the best as expected. The drawback of LHAR is its probably too conservative forecasting because we can see that it was rejected in Low scenario. Another proof is the visual inspection of Figure 4.8 where we see forecasts of LHAR are often much lower than forecasts of other models. Lastly, pioneering model RiskMetrics was rejected in all scenarios and nowadays we can say it is already obsolete approach.

The full results can be found in appendix, concretely Full scenario Table A.1, High scenario Table A.2 and Low scenario Table A.3.

4.3.2 Expected shortfall

The backtesting of ES was proceeded according to its mentioned definition and thep-value was found through 5,000 simulations by bootstrapping for both tests. Since the Test I. assumes valid forecasts of VaR, we included test statistics Z1 and its p-value in tables with VaR results.

The results of Test I. are quite surprising. If we use significance level 5%, then we will not reject any model from all three scenarios. Even if we increase significance level to 10%, then we will reject four models only from Full scenario. These models are GJR-GARCH and HAR using copula function with GPD distribution. So there can be some influence of GPD distribution.

Another interesting finding is that Full scenario has Z1 test statistic mostly negative and Low scenario has very close oscillation of Z1 around zero with significantly higher p-values than in Full scenario. Therefore, the forecasts of ES might indicate to be more valid in Low scenario. In case of High scenario, the p-values are very high for the best asymmetric HAR models and Z1 test statistic also close to zero.

The imperfection of Test I. is that it might happen that p-value is not even calculated (as we could see from our results) due to possible zero violation of VaR during bootstrapping process. This situation causes undefined mathematical operation (division by zero) returning an error. To sum up, the Test I. has still considerable limitations in order to be routinely used in practice.

The second backtest of ES was Test II that is testing the forecasts of VaR and ES mutually. The results are presented in Table 4.6. We could say that this test provided even more controversial results than Test I. Based on our simulated p-value from boostrapping, we did not reject any model from any scenario.

Nonetheless, we can have a look on the Table 3.1 of significance thresholds provided by the authors of test and compare with our Z2 test statistics.

Table 4.6: ES Test II. results for all scenarios

	Full		High		Low	
Model	**Z2**	**pv**	**Z2**	**pv**	**Z2**	**pv**
HAR.Chol.RCOV.N	-0.68	0.95	-1.52	0.80	-0.37	0.98
HAR.Chol.RCOV.t	-0.41	0.96	-1.29	0.77	-0.23	0.98
HAR.Chol.RCOV.FHS	-0.39	0.94	-0.88	0.70	-0.09	0.96
HAR.tC.e	-0.29	0.94	-1.21	0.79	-0.23	0.94
HAR.tC.t	-0.31	0.94	-1.12	0.83	-0.23	0.94
HAR.tC.gpd	0.14	0.92	-0.79	0.77	0.18	0.87
HAR.nC.e	-0.49	0.94	-1.66	0.72	-0.31	0.95
HAR.nC.n	-0.51	0.94	-1.46	0.79	-0.33	0.96
HAR.nC.gpd	0.12	0.94	-0.89	0.77	0.15	0.90
LHAR.Chol.RCOV.N	0.23	0.96	-0.36	0.64	0.63	0.99
LHAR.Chol.RCOV.t	0.36	0.96	-0.11	0.71	0.66	0.99
LHAR.Chol.RCOV.FHS	0.35	0.93	0.01	0.54	0.69	0.99
LHAR.tC.e	0.25	0.97	-0.42	0.65	0.66	0.99
LHAR.tC.t	0.30	0.98	-0.43	0.65	0.66	0.99
LHAR.tC.gpd	0.50	0.90	0.08	0.76	0.72	0.96
LHAR.nC.e	0.10	0.96	-0.59	0.66	0.62	0.99
LHAR.nC.n	0.15	0.97	-0.60	0.65	0.62	0.99
LHAR.nC.gpd	0.41	0.87	0.17	0.86	0.70	0.97
EWMA.COV.N	0.56	0.45	-1.45	0.39	-0.03	0.42
EWMA.COV.t	0.58	0.43	-1.33	0.36	0.00	0.38
EWMA.COV.FHS	0.48	0.42	-1.61	0.44	-0.09	0.38
DCC.GARCH.COV.N	-0.12	0.90	-1.37	0.66	-0.09	0.85
DCC.GARCH.COV.t	0.05	0.90	-1.03	0.70	0.15	0.90
DCC.GARCH.COV.FHS	0.02	0.85	-0.71	0.78	0.27	0.88
DCC.GJR.GARCH.COV.N	-0.03	0.91	-1.52	0.67	0.11	0.94
DCC.GJR.GARCH.COV.t	0.13	0.91	-1.30	0.64	0.20	0.92
DCC.GJR.GARCH.COV.FHS	0.15	0.89	-0.96	0.73	0.25	0.89
aDCC.GJR.GARCH.COV.N	-0.02	0.91	-1.34	0.73	0.12	0.93
aDCC.GJR.GARCH.COV.t	0.19	0.94	-1.02	0.77	0.21	0.91
aDCC.GJR.GARCH.COV.FHS	0.09	0.84	-0.97	0.74	0.30	0.90
GARCH.tC.e	0.04	0.78	-0.85	0.77	-0.09	0.72
GARCH.tC.t	0.03	0.76	-0.85	0.78	-0.10	0.73
GARCH.tC.gpd	0.13	0.70	-0.82	0.72	0.19	0.82
GARCH.nC.e	-0.08	0.80	-1.04	0.79	-0.21	0.75
GARCH.nC.n	-0.08	0.79	-1.05	0.79	-0.21	0.74
GARCH.nC.gpd	0.10	0.72	-0.87	0.76	0.17	0.84
GJR.GARCH.tC.e	0.11	0.82	-0.69	0.90	0.18	0.88
GJR.GARCH.tC.t	0.04	0.73	-0.73	0.88	0.19	0.89
GJR.GARCH.tC.gpd	0.24	0.78	-0.66	0.86	0.23	0.84
GJR.GARCH.nC.e	-0.06	0.78	-1.02	0.85	0.10	0.90
GJR.GARCH.nC.n	-0.06	0.77	-1.00	0.86	0.12	0.91
GJR.GARCH.nC.gpd	0.21	0.81	-0.74	0.86	0.19	0.85

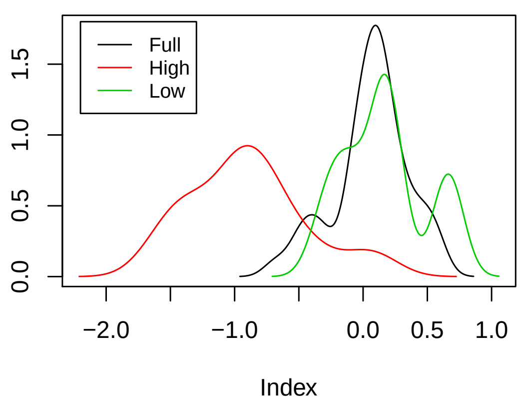
Source: Author's computation

We can observe on Full scenario the distribution of Z2 is centred around zero and quite dispersed with the highest value of 0.58 from EWMA.COV.t indicating overestimation of ES and -0.68 from HAR.Chol.RCOV.N indicating underestimation of ES.

The opposite situation is on High scenario where the majority of Z2 statistics are negative. It means that the majority of models underestimate the forecast of ES. The densities of Z2 from all our models can be seen in Figure 4.10. We can compare Z2 value of the best selected VaR model that was LHAR.tC.gpd with Z2 values of rejected models, i.e. family of EWMA models. LHAR had Z2 value 0.08 and EWMA family had in range of -1.33 and -1.61. We could say from this primitive comparison that EWMA models underestimate forecasts of ES. Anyway, it is very difficult to infer specific conclusion from Test II. On one hand, we have high positive p-values from all models and scenarios and on the other hand, we would reject 26 models of 42 for High scenario if we consider proposed critical values for Student's t with $\nu = 3$.

The Low scenario is the same situation as the Full scenario. We observe very high p-values and Z2 statistics surrounded around zero.

Figure 4.10: Z2 density for all scenarios



Source: Author's computation

Considering all findings from ES backtesting, we conclude that there is still plenty of room for further research about the backtesting methods of ES and it depends if there are any methods depending on the definition of backtesting and the issue of elicitability.

Moreover, the highest challange would be to find out the way of comparison various ES since there is currently no theoretical basis for this due to elicitability.

Thus, the promising alternative approach can be indirect backtesting through VaR such as one proposed in Emmer *et al.* (2015).

Chapter 5

Conclusion

In this master thesis we investigated the performance of forecasting accuracy of VaR and ES from the multivariate models based on high-frequency and daily data. The literature overview about current research focused on modeling of realized volatility and covariance with application in forecasting of VaR and ES can be found in Chapter 2 together with theoretical framework of realized measures, VaR and ES. Thanks to literature review, we found out that there has not been written any paper about the application of multivariate model using high-frequency data for forecasting of ES. Chapter 3 brings introduction of all applied variance-covariance models with their pros and cons. Moreover, we described also all used loss distributions because they are as important as estimation of volatility or covariance matrix but we could see in many papers that this fact is overlooked. The last subsection was dedicated description to standard back-testing methods and model selection methods represented by loss functions in case of VaR. The backtesting of ES was described by two recently proposed tests I. and II. Chapter 4 contains entire empirical analysis that we summarize in next paragraph.

Regarding to data analysis, we implemented the simplest estimator of realized covariance constructed by homogeneously spaced returns on 20 minutes frequency which were chosen from signature covariance plot, synchronization according to fixed time when all assets were traded and omitting the overnight returns. We found that overnight returns were only significant between Friday closing and Monday opening on CL and XAUUSD asset. Our synchronization technique resulted in high reduction of data as we had left only 69 observation per business day. Moreover, the long-memory effect was confirmed on all elements of realized covariance matrix that provided the support for our multivariate HAR model. Overall, the estimation of realized covariance matrix is still relatively in its infancy period and hence, the more advanced methods are very sophisticated with little documentation of their implementation in practice. From the practical point of view, one thinking about HFD sampled with very high frequencies must be also aware of substantial increased demand of computation power.

Subsequent modeling of realized covariance matrix was very efficient due to the parsimony and stability of multivariate HAR models. We tried to apply also ARFIMA model in the same fashion as multivariate HAR but it was very unstable estimation returning many errors. The estimation of multivariate distributions through copula functions is very well-documented and implemented also within R. The challenging point was the determination of threshold level for the GPD and we decided to set it as percentage of observations in order to get around 100 observations that should provide stable estimates of GPD.

Regarding to answers to our main objective and following questions, we are going to answer through our empirical results of our backtesting and model selection methods (the position of models were determined by the asymmetric loss function). All empirical results are derived for full sample called Full scenario and subsamples containing periods of high volatility called High scenario and low volatility called Low scenario.

The first question was *"What model and approach provides the most accurate forecasts of VaR and ES?"* The answer is that the most robust performance was achieved by utilization of HFD through univariate HAR using copula function either Gaussian or *t* in terms of forecasts of VaR.

The second question was *"Does the best model and approach of VaR perform similarly also in the forecasting of ES?"*. Unfortunately, we are not able to answer this question. The reason is that backtests of ES did not give credible results since both tests did not reject any model on significance level 5%. Moreover, the test I. did not even calculate p-value because the simulation via bootstrapping resulted in calculation of p-value that would include division by zero what is undefined mathematical operation. The both tests were rather disappointing and probably the backtesting approach by approximation of ES by VaR for different confidence levels can be better alternative as it was suggested by Emmer *et al.* (2015). The third question was *"What is the difference between the two approaches for various market volatility periods (stable versus turbulent period)?"*. The answer is there is significant difference. When we have a look on top models in High scenario, we can find the best performing model asymmetric version of univariate HAR called LHAR with Gaussian or *t* copula using GPD as marginal distribution. These models coped with the fat tails the best. Anyway, asymmetric version of univariate HAR was an excellent model in all its variations in High scenario. Another interesting result is from Low scenario where all models either using HFD or daily data performed relatively the same. It tells us that backtesting and selecting the models based on this scenario is very low robust.

The final answer for our main objective is that Heterogeneous Autoregression model using high-frequency data delivered superior or at least the same accuracy of forecasts of VaR to benchmark models (DCC-GARCH or EWMA) based on daily data. Nevertheless, EWMA model was the worst model from all because it was rejected in all scenarios and therefore it was not included in model selection.

The model selection based on loss functions revealed also interesting information. The regulatory loss function was giving more or less inverse preference of models than firm loss function.

This was the reason why we implemented asymmetric loss function as our decisive criterion to define the order of preference of models.

Another important finding about backtesting of ES is that depending on the definition of ”backtesting”, the backtesting might not exist or at least the model selection does not exist due to lack of elicitability what means there does not exist scoring function such as loss functions applied in the model selection of VaR.

The inspiration for further research can be found in the relaxing of limitations of this master thesis mentioned in Chapter 1 or implementation of more advanced methods mentioned in this master thesis that have not been implemented here.

Bibliography

Acerbi, C. & B. Szekely (2014a): "Testing expected shortfall." https://www.parmenides-foundation.org/fileadmin/redakteure/events/Workshop_Kondor_02._03.06.2014/Acerbi_ESBacktest_reduced_size.pdf. Accessed: 2017-03-01.

Acerbi, C. & B. Szekely (2014b): "Backtesting Expected Shortfall." *Technical report*, MSCI Inc.

Aielli, G. P. (2013): "Dynamic Conditional Correlation: On Properties and Estimation." *Journal of Business & Economic Statistics* **31(3)**: pp. 282–299.

Alexander, C. (2008): *Market Risk Analysis, Practical Financial Econometrics (Volume II).* Wiley, volume ii edition.

Alexander, C. (2009): *Market Risk Analysis, Value at Risk Models (Volume IV).* Wiley, volume iv edition.

Allen, D. E., M. McAleer, & M. Scharth (2014): "Asymmetric Realized Volatility Risk." *Journal of Risk and Financial Management* **7(2)**: pp. 1–80.

Andersen, T., T. Bollerslev, P. Christoffersen, & F. Diebold (2011): "Financial risk measurement for financial risk management." *Creates research papers*, Department of Economics and Business Economics, Aarhus University.

Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2003): "Modeling and forecasting realized volatility." *Econometrica* **71(2)**: pp. 579–625.

Andersen, T. G. & T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review* **39(4)**: pp. 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, & P. Labys (1999): "(Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation." *New York University, Leonard N. Stern School Finance Department Working Paper Seires 99-061*, New York University, Leonard N. Stern School of Business-.

Andersen, T. G., R. A. Davis, J.-P. Kreiss, & T. V. Mikosch (editors) (2009): *Handbook of Financial Time Series.* Springer, 2009 edition.

Artzner, P., F. Delbaen, J.-M. Eber, & D. Heath (1997): "Thinking coherently - generalised scenatios rather than var should be used when calculating regulatory capital." *RISK Magazine* **10(11)**: pp. 68–72.

Artzner, P., F. Delbaen, J.-M. Eber, & D. Heath (1999): "Coherent measures of risk." *Mathematical Finance* **9(3)**: pp. 203–228.

Barndorff-Nielsen, O. & N. Shephard (2004): "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics." *Econometrica* **72(3)**: pp. 885–925.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, & N. Shephard (2011): "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading." *Journal of Econometrics* **162(2)**: pp. 149–169.

Barone-Adesi, K. Giannopoulos, & L. Vosper (1999): "Var without correlations for portfolios of derivatives securities." *Journal of Futures Markets* **19**: pp. 583–602.

Barone-Adesi, G., F. Bourgoin, & K. Giannopoulos (1998): "Don't look back." *Risk* **11**: pp. 100–103.

Bauer, G. H. & K. Vorkink (2007): "Multivariate Realized Stock Market Volatility." *Technical report.*

Bauwens, L., M. Braione, & G. Storti (2014): "Forecasting comparison of long term component dynamic models for realized covariance matrices." *CORE Discussion Papers 2014053*, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Bauwens, L., C. M. Hafner, & S. Laurent (2012): *Handbook of Volatility Models and Their Applications*. Wiley, 1 edition.

Bee, M., D. J. Dupuis, & L. Trapin (2016): "Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective." *Journal of Empirical Finance* **36**: pp. 86 – 99.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* **31(3)**: pp. 307 – 327.

Bollerslev, T. (1990): "Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model." *The Review of Economics and Statistics* **72(3)**: pp. 498–505.

Bollerslev, T. (2008): "Glossary to ARCH (GARCH)." *CREATES Research Papers 2008-49*, Department of Economics and Business Economics, Aarhus University.

Bollerslev, T., R. F. Engle, & J. M. Wooldridge (1988): "A Capital Asset Pricing Model with Time-Varying Covariances." *Journal of Political Economy* **96(1)**: pp. 116–31.

Bonato, M., M. Caporin, & A. Ranaldo (2009): "Forecasting realized (co)variances with a block structure Wishart autoregressive model." *Technical report*.

Boudt, K., S. Laurent, A. Lunde, & R. Quaedvlieg (2014): "Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity." *CREATES Research Papers 2014-05*, Department of Economics and Business Economics, Aarhus University.

Brechmann, E. C., M. Heiden, & Y. Okhrin (0): "A multivariate volatility vine copula model." *Econometric Reviews* **0(0)**: pp. 1–28.

Candila, V. (2013): "A comparison of the forecasting performances of multivariate volatility models." *Working Papers* $3_2 28, Dipartimento di Scienze Economiche e Statistiche, Universit degli Studi di Salerno.$

Cappiello, L., R. F. Engle, & K. Sheppard (2006): "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns." *Journal of Financial Econometrics* **4(4)**: pp. 537–572.

Chao, W. & G. Richard (2014): "Forecasting risk via realized garch, incorporating the realized range." *Working Papers 2014-06*, University of Sydney Business School, Discipline of Business Analytics.

Chiriac, R. & V. Voev (2011): "Modelling and forecasting multivariate realized volatility." *Journal of Applied Econometrics* **26(6)**: pp. 922–947.

Chou, R. Y., H. Chou, & N. Liu (2010): "Range volatility models and their applications in finance." In "Handbook of Quantitative Finance and Risk Management," pp. 1273–1281. Springer.

Christoffersen, P. (1998): "Evaluating interval forecasts." *International Economic Review* **39(4)**: pp. 841–62.

Christoffersen, P. (2011): *Elements of Financial Risk Management, Second Edition*. Academic Press, 2 edition.

Committee, B. (2013): "Fundamental review of the trading book: A revised market risk framework." *Technical report*, Bank for International Settlements.

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

Contino, C. & R. Gerlach (2014): "Bayesian Tail Risk Forecasting using Realised GARCH." *Working Papers 2014-05*, University of Sydney Business School, Discipline of Business Analytics.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* .

Corsi, F. & R. Reno (2009b): "Volatility determinants: Heterogeneity, leverage, and jumps." *Technical report*, Working paper, available at.

Danielsson, J. (2011): *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*. Wiley, 1 edition.

Dowd, K. (2005): *Measuring Market Risk*. Wiley, 2 edition.

Čech, F. (2013): *Dynamic portfolio optimization during financial crisis using daily data and high frequency data*. Master's thesis, The Charles University in Prague, Czech Republic.

Embrechts, P., R. Frey, M. Hofert, & A. J. McNeil (2016): "Quantitative risk management." `https://github.com/qrmtutorial/qrm/blob/master/slides/qrm.pdf`. Accessed: 2016-09-23.

Embrechts, P., T. Mikosch, & C. Klüppelberg (1997): *Modelling Extremal Events: For Insurance and Finance*. London, UK, UK: Springer-Verlag.

Emmer, S., M. Kratz, & D. Tasche (2015): "What is the best risk measure in practice? A comparison of standard measures." *Journal of Risk* **18**: pp. 31–60.

Engle, R. & S. Manganelli (2004): "Caviar: Conditional autoregressive value at risk by regression quantiles." *Journal of Business Economic Statistics* **22**: pp. 367–381.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation." *Econometrica* **50(4)**: pp. 987–1007.

Engle, R. F. (2000): "Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models." *University of California at San Diego, Economics Working Paper Series qt56j4143f*, Department of Economics, UC San Diego.

Engle, R. F. & K. F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH." *Econometric Theory* **11(01)**: pp. 122–150.

Engle, R. F. & S. Manganelli (2001): "Value at risk models in finance." *Working Paper Series 0075*, European Central Bank.

Epps, T. W. (1979): "Comovements in stock prices in the very short run." *Journal of the American Statistical Association* **74**: pp. 291–298.

Fengler, M. & O. Okhrin (2012): "Realized Copula." *Economics Working Paper Series 1214*, University of St. Gallen, School of Economics and Political Science.

Fengler, M. R. & O. Okhrin (2016): "Managing risk with a realized copula parameter." *Computational Statistics & Data Analysis* **100(C)**: pp. 131–152.

Giot, P. & S. Laurent (2004): "Modelling daily Value-at-Risk using realized volatility and ARCH type models." *Journal of Empirical Finance* **11(3)**: pp. 379–398.

Glosten, L. R., R. Jagannathan, & D. E. Runkle (1993): “On the relation between the expected value and the volatility of the nominal excess return on stocks.” *Staff Report 157*, Federal Reserve Bank of Minneapolis.

Gneiting, T. (2011): “Making and evaluating point forecasts.” *Journal of the American Statistical Association* **106(494)**: pp. 746–762.

Goix, N., A. Sabourin, & S. Clémençon (2015): “Sparsity in Multivariate Extremes with Applications to Anomaly Detection.” *arXiv preprint arXiv:1507.05899* .

Golosnoy, V., B. Gribisch, & R. Liesenfeld (2012): “The conditional autoregressive Wishart model for multivariate stock market volatility.” *Journal of Econometrics* **167(1)**: pp. 211–223.

González-Rivera, G., T.-H. Lee, & S. Mishra (2004): “Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood.” *International Journal of Forecasting* **20(4)**: pp. 629 – 645.

Gourieroux, C., J. Jasiak, & R. Sufana (2009): “The wishart autoregressive process of multivariate stochastic volatility.” *Journal of Econometrics* **150(2)**: pp. 167 – 181. Recent Development in Financial Econometrics.

Hansen, P. & A. Lunde (2006): “Realized variance and market microstructure noise.” *Journal of Business Economic Statistics* **24**: pp. 127–161.

Hansen, P. R., A. Lunde, & J. M. Nason (2011): “The model confidence set.” *Econometrica* **79(2)**: pp. 453–497.

Holton, G. A. (2002): “History of Value-at-Risk: 1922-1998.” *Working papers*, Contingency Analysis.

Jäger, M. (2015): *Backtesting Expected Shortfall: A GARCH-EVT-Copula Approach*. Master’s thesis, University of Applied Sciences Berlin, Germany.

Jin, X. & J. M. Maheu (2010): “Modelling Realized Covariances and Returns.” *Working Papers tecipa-408*, University of Toronto, Department of Economics.

Jondeau, E., S.-H. Poon, & M. Rockinger (2007): *Financial modeling under non-gaussian distributions*. Springer Finance. Springer.

J.P.Morgan (1996): “RiskMetrics$^{\text{TM}}$ - Technical Document.” *Working papers*, J.P. Morgan.

Kruse, R. (2006): “Can realized volatility improve the accuracy of value-at-risk forecasts?” Working paper.

Kuester, K., S. Mittnik, & M. S. Paolella (2006): “Value-at-Risk Prediction: A Comparison of Alternative Strategies.” *Journal of Financial Econometrics* **4(1)**: pp. 53–89.

Kupiec, P. (1995): "Techniques for verifying the accuracy of risk measurement models." *Finance and Economics Discussion Series 95-24*, Board of Governors of the Federal Reserve System (U.S.).

Kyselá, E. (2016): *Modelling portfolios with heavy-tailed risk factors.* Master's thesis, The University of Economics in Prague, Czech Republic.

Laurent, S., L. Bauwens, & J. V. K. Rombouts (2006): "Multivariate GARCH models: a survey." *Journal of Applied Econometrics* **21(1)**: pp. 79–109.

LBMA (2017): "The london bullion market." `http://www.lbma.org.uk/market-overview`. Accessed: 2017-03-01.

Liu, L. Y., A. J. Patton, & K. Sheppard (2015): "Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes." *Journal of Econometrics* **187(1)**: pp. 293 – 311.

Lopez, J. (1998): "Methods for evaluating value-at-risk estimates." *Technical report*.

Louzis, D. P., S. Xanthopoulos-Sisinis, & A. P. Refenes (2014): "Realized volatility models and alternative value-at-risk prediction strategies." *Economic Modelling* **40**: pp. 101 – 116.

López Martín, C., S. Benito Muela, & P. Abad Moreno (2015): "The role of the loss function in value-at-risk comparisons." *Journal of Risk Model Validation* **9**: pp. 1–19.

Mandelbrot, B. (1963): "The variation of certain speculative prices." *The Journal of Business* **36**.

Markowitz, H. (1952): "Portfolio Selection." *Journal of Finance* **7(1)**: pp. 77–91.

McAleer, M. & M. Medeiros (2008): "Realized volatility: A review." *Econometric Reviews* **27(1-3)**: pp. 10–45.

McMillan, D. G., A. E. Speight, & K. P. Evans (2008): "How useful is intraday data for evaluating daily Value-at-Risk?" *Journal of Multinational Financial Management* **18(5)**: pp. 488–503.

McNeil, A. J. & R. Frey (2000): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach." *Journal of Empirical Finance* **7(3-4)**: pp. 271–300.

McNeil, A. J., R. Frey, & P. Embrechts (2015): *Quantitative Risk Management: Concepts, Techniques and Tools (Princeton Series in Finance).* Princeton University Press, revised edition.

Müller, U. A., M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, & J. E. von Weizsäcker (1997): "Volatilities of different time resolutions — analyzing the dynamics of market components." *Journal of Empirical Finance* **4(2–3)**: pp. 213 – 239. High Frequency Data in Finance, Part 1.

Nadarajah, S., B. Zhang, & S. Chan (2014): "Estimation methods for expected shortfall." *Quantitative Finance* **14(2)**: pp. 271–291.

Noureldin, D., N. Shephard, & K. Sheppard (2012): "Multivariate high-frequency-based volatility (heavy) models." *Journal of Applied Econometrics* **27(6)**: pp. 907–933.

Patton, A. J. (2004): "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation." *Journal of Financial Econometrics* **2(1)**: pp. 130–168.

Righi, M. B. & P. S. Ceretta (2015): "A comparison of expected shortfall estimation models." *Journal of Economics and Business* **78(C)**: pp. 14–47.

Roccioletti, S. (2016): *Backtesting Value at Risk and Expected Shortfall (BestMasters)*. Springer Gabler, 1st ed. 2016 edition.

Rocco, M. (2011): "Extreme value theory for finance: a survey." *Questioni di Economia e Finanza (Occasional Papers) 99*, Bank of Italy, Economic Research and International Relations Area.

Sarykalin, S., G. Serraino, & S. Uryasev (2014): "Fundamental review of the trading book: A revised market risk framework." *Technical report*, Institute for Operations Research and the Management Sciences (INFORMS).

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges." *De, Publications and l'Institut de Statistique de Paris* **8**: pp. 229–231.

Snowfall Systems, I. (2014): "Microstructure noise." `https://www.portfolioeffect.com/docs/glossary/models/microstructure/microstructure-noise`. Accessed: 2017-02-10.

Tsay, R. S. (2012): *An Introduction to Analysis of Financial Data with R.* Wiley, 1 edition.

Ubukata, M. & T. Watanabe (2015): "Evaluating the performance of futures hedging using multivariate realized volatility." *Journal of the Japanese and International Economies* **38**: pp. 148 – 171.

Watanabe, T. (2011): "Quantile Forecasts of Financial Returns Using Realized GARCH Models." *Global COE Hi-Stat Discussion Paper Series gd11-195*, Institute of Economic Research, Hitotsubashi University.

Wimmerstedt, L. (2015): *Backtesting Expected Shortfall: the design and implementation of different backtests.* Master's thesis, KTH Royal Institute of Technology in Stockholm, Sweden.

Appendix A

Tables

Table A.1: VaR and ES Test I. results for Full sample and all models.

Model	Viol. ratio	UC pv	CC pv	DQ pv	RLF	RLF %	FLF	FLF %	ALF	ALF %	Z1	Z1 pv
HAR.Chol.RCOV.N	1.3	0.22	0.44	0.62	0.019	0	0.243	100	0.039	78	-0.27	0.58
HAR.Chol.RCOV.t	1.2	0.46	0.64	0.66	0.018	12	0.244	95	0.039	78	-0.18	0.48
HAR.Chol.RCOV.FHS	1.3	0.33	0.55	0.56	0.018	12	0.244	95	0.040	54	-0.11	0.41
HAR.tC.e	1.1	0.82	0.72	0.61	0.019	0	0.248	68	0.039	78	-0.22	0.27
HAR.tC.t	1.1	0.82	0.72	0.61	0.019	0	0.248	68	0.039	78	-0.24	0.24
HAR.tC.gpd	0.7	0.16	0.13	0.16	0.015	49	0.259	20	0.041	29	-0.29	0.08
HAR.nC.e	1.1	0.63	0.70	0.65	0.019	0	0.247	80	0.039	78	-0.33	0.35
HAR.nC.n	1.1	0.63	0.70	0.66	0.019	0	0.246	90	0.039	78	-0.34	0.31
HAR.nC.gpd	0.7	0.16	0.13	0.16	0.015	49	0.257	24	0.040	54	-0.33	0.10
LHAR.Chol.RCOV.N	0.6	0.09	0.07	0.15	0.012	83	0.262	17	0.043	10	-0.29	0.34
LHAR.Chol.RCOV.t	0.5	0.04	0.03	0.08	0.012	83	0.264	12	0.043	10	-0.21	0.28
LHAR.Chol.RCOV.FHS	0.6	0.09	0.07	0.15	0.012	83	0.264	12	0.043	10	-0.09	0.27
LHAR.tC.e	0.6	0.09	0.07	0.14	0.015	49	0.257	24	0.042	22	-0.25	0.14
LHAR.tC.t	0.5	0.04	0.03	0.07	0.015	49	0.258	22	0.043	10	-0.32	0.07
LHAR.tC.gpd	0.4	0.01	0.00	0.07	0.011	93	0.270	7	0.046	0	-0.26	0.14
LHAR.nC.e	0.7	0.16	0.13	0.21	0.015	49	0.257	24	0.042	22	-0.36	0.20
LHAR.nC.n	0.6	0.09	0.07	0.14	0.015	49	0.256	32	0.042	22	-0.42	0.12
LHAR.nC.gpd	0.5	0.02	0.01	0.03	0.012	83	0.271	2	0.046	0	-0.26	0.22
EWMA.COV.N	0.4	0.01	0.00	0.00	0.007	98	0.271	2	0.044	5	-0.11	
EWMA.COV.t	0.4	0.01	0.00	0.00	0.006	100	0.273	0	0.044	5	-0.05	
EWMA.COV.FHS	0.5	0.02	0.01	0.00	0.008	95	0.266	10	0.043	10	-0.13	
DCC.GARCH.COV.N	0.9	0.57	0.47	0.32	0.016	37	0.246	90	0.040	54	-0.30	0.27
DCC.GARCH.COV.t	0.8	0.40	0.34	0.26	0.015	49	0.247	80	0.040	54	-0.19	0.25
DCC.GARCH.COV.FHS	0.9	0.57	0.07	0.00	0.015	49	0.251	46	0.041	29	-0.13	0.23
DCC.GJR.GARCH.COV.N	0.8	0.40	0.34	0.30	0.015	49	0.248	68	0.039	78	-0.30	0.28
DCC.GJR.GARCH.COV.t	0.7	0.26	0.22	0.29	0.015	49	0.250	56	0.039	78	-0.20	0.23
DCC.GJR.GARCH.COV.FHS	0.7	0.26	0.22	0.14	0.015	49	0.254	37	0.040	54	-0.17	0.16
aDCC.GJR.GARCH.COV.N	0.8	0.40	0.34	0.30	0.015	49	0.249	63	0.039	78	-0.29	0.30
aDCC.GJR.GARCH.COV.t	0.7	0.16	0.13	0.17	0.014	80	0.250	56	0.039	78	-0.23	0.16
aDCC.GJR.GARCH.COV.FHS	0.8	0.40	0.34	0.24	0.015	49	0.254	37	0.040	54	-0.14	0.18
GARCH.tC.e	0.8	0.40	0.34	0.20	0.017	20	0.248	68	0.041	29	-0.21	0.17
GARCH.tC.t	0.8	0.40	0.34	0.19	0.017	20	0.249	63	0.041	29	-0.22	0.16
GARCH.tC.gpd	0.7	0.26	0.22	0.13	0.016	37	0.251	46	0.041	29	-0.19	0.19
GARCH.nC.e	0.8	0.40	0.34	0.22	0.017	20	0.247	80	0.041	29	-0.36	0.16
GARCH.nC.n	0.8	0.40	0.34	0.22	0.018	12	0.247	80	0.041	29	-0.36	0.14
GARCH.nC.gpd	0.7	0.26	0.22	0.14	0.016	37	0.251	46	0.041	29	-0.24	0.19
GJR.GARCH.tC.e	0.7	0.26	0.22	0.16	0.017	20	0.252	44	0.040	54	-0.22	0.17
GJR.GARCH.tC.t	0.8	0.40	0.34	0.26	0.017	20	0.251	46	0.040	54	-0.20	0.20
GJR.GARCH.tC.gpd	0.6	0.09	0.07	0.06	0.016	37	0.254	37	0.041	29	-0.28	0.08
GJR.GARCH.nC.e	0.8	0.40	0.34	0.24	0.017	20	0.248	68	0.040	54	-0.33	0.19
GJR.GARCH.nC.n	0.8	0.40	0.34	0.26	0.017	20	0.250	56	0.040	54	-0.34	0.19
GJR.GARCH.nC.gpd	0.6	0.09	0.07	0.07	0.016	37	0.255	34	0.041	29	-0.32	0.08

Source: Author's computation

Table A.2: VaR and ES Test I. results for High sample and all models.

Model	Viol. ratio	UC pv	CC pv	DQ pv	RLF	RLF %	FLF	FLF %	ALF	ALF %	Z1	Z1 pv
HAR.Chol.RCOV.N	2.1	0.02	0.05	0.17	0.025	32	0.805	63	0.064	34	-0.20	0.83
HAR.Chol.RCOV.t	2.1	0.02	0.05	0.17	0.024	46	0.808	51	0.064	34	-0.10	0.87
HAR.Chol.RCOV.FHS	1.9	0.04	0.10	0.26	0.017	73	0.834	29	0.061	73	0.02	0.93
HAR.tC.e	2.1	0.02	0.05	0.17	0.023	54	0.808	51	0.063	51	-0.06	0.91
HAR.tC.t	1.9	0.04	0.10	0.23	0.023	54	0.807	56	0.063	51	-0.10	0.85
HAR.tC.gpd	1.8	0.08	0.17	0.42	0.014	78	0.835	24	0.059	95	-0.01	0.95
HAR.nC.e	2.2	0.01	0.02	0.07	0.023	54	0.809	49	0.064	34	-0.18	0.90
HAR.nC.n	2.1	0.02	0.05	0.17	0.022	61	0.812	46	0.062	63	-0.18	0.91
HAR.nC.gpd	1.8	0.08	0.17	0.42	0.015	76	0.837	20	0.060	88	-0.07	0.93
LHAR.Chol.RCOV.N	1.3	0.49	0.60	0.95	0.009	90	0.853	12	0.060	88	-0.06	
LHAR.Chol.RCOV.t	1.1	0.76	0.78	0.98	0.009	90	0.856	5	0.060	88	0.01	0.91
LHAR.Chol.RCOV.FHS	1.3	0.49	0.60	0.64	0.007	95	0.880	0	0.061	73	0.23	
LHAR.tC.e	1.4	0.29	0.41	0.47	0.012	83	0.836	22	0.061	73	0.02	0.90
LHAR.tC.t	1.4	0.29	0.41	0.48	0.013	80	0.835	24	0.061	73	0.01	0.90
LHAR.tC.gpd	1.0	0.93	0.86	0.99	0.006	100	0.861	2	0.059	95	0.05	
LHAR.nC.e	1.4	0.29	0.41	0.45	0.012	83	0.831	34	0.061	73	-0.10	
LHAR.nC.n	1.4	0.29	0.41	0.48	0.012	83	0.839	17	0.061	73	-0.10	0.90
LHAR.nC.gpd	0.8	0.61	0.79	0.99	0.007	95	0.856	5	0.059	95	-0.03	
EWMA.COV.N	1.9	0.04	0.10	0.00	0.035	2	0.731	100	0.069	2	-0.27	0.60
EWMA.COV.t	1.9	0.04	0.10	0.00	0.034	5	0.732	98	0.069	2	-0.21	0.69
EWMA.COV.FHS	2.1	0.02	0.02	0.00	0.037	0	0.733	95	0.070	0	-0.25	0.64
DCC.GARCH.COV.N	1.9	0.04	0.10	0.26	0.029	7	0.798	73	0.066	7	-0.23	0.64
DCC.GARCH.COV.t	1.8	0.08	0.17	0.36	0.028	12	0.801	68	0.066	7	-0.15	0.59
DCC.GARCH.COV.FHS	1.4	0.29	0.41	0.62	0.024	46	0.819	44	0.064	34	-0.18	0.29
DCC.GJR.GARCH.COV.N	2.1	0.02	0.02	0.01	0.029	7	0.827	41	0.066	7	-0.21	0.81
DCC.GJR.GARCH.COV.t	2.1	0.02	0.02	0.01	0.028	12	0.830	37	0.066	7	-0.10	0.84
DCC.GJR.GARCH.COV.FHS	1.8	0.08	0.17	0.33	0.021	68	0.855	10	0.064	34	-0.11	0.73
aDCC.GJR.GARCH.COV.N	1.9	0.04	0.10	0.25	0.028	12	0.830	37	0.066	7	-0.21	0.76
aDCC.GJR.GARCH.COV.t	1.8	0.08	0.13	0.40	0.026	27	0.834	29	0.065	22	-0.14	0.70
aDCC.GJR.GARCH.COV.FHS	1.8	0.08	0.17	0.33	0.021	68	0.851	15	0.064	34	-0.11	0.70
GARCH.tC.e	1.6	0.16	0.24	0.44	0.026	27	0.788	90	0.065	22	-0.15	0.53
GARCH.tC.t	1.6	0.16	0.24	0.40	0.027	22	0.792	80	0.066	7	-0.15	0.54
GARCH.tC.gpd	1.6	0.16	0.24	0.42	0.025	32	0.790	88	0.065	22	-0.13	0.64
GARCH.nC.e	1.6	0.16	0.24	0.45	0.027	22	0.791	85	0.065	22	-0.27	0.51
GARCH.nC.n	1.6	0.16	0.24	0.43	0.028	12	0.787	93	0.065	22	-0.28	0.47
GARCH.nC.gpd	1.6	0.16	0.24	0.45	0.025	32	0.801	68	0.064	34	-0.17	0.62
GJR.GARCH.tC.e	1.4	0.29	0.41	0.65	0.024	46	0.797	76	0.063	51	-0.17	0.58
GJR.GARCH.tC.t	1.4	0.29	0.41	0.65	0.025	32	0.803	66	0.063	51	-0.20	0.48
GJR.GARCH.tC.gpd	1.4	0.29	0.41	0.65	0.022	61	0.807	56	0.062	63	-0.15	0.65
GJR.GARCH.nC.e	1.6	0.16	0.24	0.56	0.025	32	0.792	80	0.063	51	-0.26	0.67
GJR.GARCH.nC.n	1.6	0.16	0.24	0.56	0.025	32	0.795	78	0.062	63	-0.24	0.70
GJR.GARCH.nC.gpd	1.4	0.29	0.41	0.65	0.022	61	0.807	56	0.062	63	-0.21	0.56

Source: Author's computation

Table A.3: VaR and ES Test I. results for Low sample and all models.

Model	Viol. ratio	UC pv	CC pv	DQ pv	RLF	RLF %	FLF	FLF %	ALF	ALF %	Z1	Z1 pv
HAR.Chol.RCOV.N	1.1	0.74	0.77	0.53	0.004	0	0.201	41	0.028	78	-0.21	0.70
HAR.Chol.RCOV.t	1.1	0.74	0.77	0.53	0.004	0	0.203	34	0.028	78	-0.09	0.74
HAR.Chol.RCOV.FHS	1.1	0.74	0.77	0.56	0.004	0	0.201	41	0.029	41	0.04	0.74
HAR.tC.e	1.1	0.74	0.77	0.61	0.003	7	0.204	27	0.029	41	-0.08	0.63
HAR.tC.t	1.1	0.74	0.77	0.60	0.003	7	0.204	27	0.029	41	-0.08	0.66
HAR.tC.gpd	0.8	0.62	0.80	0.41	0.002	27	0.214	22	0.030	15	-0.01	
HAR.nC.e	1.1	0.74	0.77	0.55	0.003	7	0.204	27	0.029	41	-0.16	0.80
HAR.nC.n	1.1	0.74	0.77	0.58	0.003	7	0.202	37	0.029	41	-0.17	0.76
HAR.nC.gpd	0.8	0.62	0.80	0.40	0.002	27	0.213	24	0.030	15	-0.05	
LHAR.Chol.RCOV.N	0.3	0.05	0.14	0.88	0.001	76	0.224	10	0.029	41	-0.15	
LHAR.Chol.RCOV.t	0.3	0.05	0.14	0.88	0.001	76	0.225	7	0.029	41	-0.04	
LHAR.Chol.RCOV.FHS	0.3	0.05	0.14	0.88	0.001	76	0.229	5	0.030	15	0.04	
LHAR.tC.e	0.3	0.05	0.14	0.88	0.001	76	0.220	12	0.028	78	-0.04	
LHAR.tC.t	0.3	0.05	0.14	0.88	0.001	76	0.220	12	0.028	78	-0.06	
LHAR.tC.gpd	0.3	0.05	0.14	0.88	0.000	98	0.232	2	0.032	7	0.12	
LHAR.nC.e	0.3	0.05	0.14	0.88	0.001	76	0.219	20	0.028	78	-0.16	
LHAR.nC.n	0.3	0.05	0.14	0.88	0.001	76	0.220	12	0.028	78	-0.17	
LHAR.nC.gpd	0.3	0.05	0.14	0.88	0.000	98	0.233	0	0.032	7	0.08	
EWMA.COV.N	1.0	0.94	0.30	0.01	0.003	7	0.201	41	0.035	0	-0.06	
EWMA.COV.t	1.0	0.94	0.30	0.01	0.002	27	0.202	37	0.035	0	-0.02	
EWMA.COV.FHS	1.1	0.74	0.38	0.03	0.003	7	0.201	41	0.034	5	0.04	
DCC.GARCH.COV.N	1.0	0.94	0.86	0.56	0.003	7	0.187	95	0.030	15	-0.12	
DCC.GARCH.COV.t	0.8	0.62	0.80	0.38	0.002	27	0.187	95	0.030	15	-0.05	0.55
DCC.GARCH.COV.FHS	0.8	0.62	0.80	0.41	0.003	7	0.187	95	0.031	12	0.10	
DCC.GJR.GARCH.COV.N	0.8	0.62	0.80	0.29	0.002	27	0.193	78	0.028	78	-0.10	0.84
DCC.GJR.GARCH.COV.t	0.8	0.62	0.80	0.29	0.002	27	0.194	68	0.028	78	0.00	0.86
DCC.GJR.GARCH.COV.FHS	0.8	0.62	0.80	0.39	0.002	27	0.195	61	0.029	41	0.07	0.84
aDCC.GJR.GARCH.COV.N	0.8	0.62	0.80	0.29	0.002	27	0.193	78	0.028	78	-0.09	0.86
aDCC.GJR.GARCH.COV.t	0.8	0.62	0.80	0.29	0.002	27	0.194	68	0.028	78	0.02	
aDCC.GJR.GARCH.COV.FHS	0.8	0.62	0.80	0.36	0.002	27	0.194	68	0.029	41	0.14	
GARCH.tC.e	1.1	0.74	0.77	0.07	0.002	27	0.190	85	0.030	15	0.04	0.89
GARCH.tC.t	1.1	0.74	0.77	0.65	0.002	27	0.190	85	0.030	15	0.03	
GARCH.tC.gpd	0.8	0.62	0.80	0.40	0.002	27	0.194	68	0.030	15	-0.00	
GARCH.nC.e	1.1	0.74	0.77	0.07	0.002	27	0.190	85	0.030	15	-0.07	
GARCH.nC.n	1.1	0.74	0.77	0.06	0.002	27	0.189	93	0.030	15	-0.07	
GARCH.nC.gpd	0.8	0.62	0.80	0.39	0.002	27	0.193	78	0.030	15	-0.03	
GJR.GARCH.tC.e	0.8	0.62	0.80	0.37	0.002	27	0.198	56	0.029	41	-0.01	
GJR.GARCH.tC.t	0.8	0.62	0.80	0.36	0.002	27	0.196	59	0.029	41	-0.00	
GJR.GARCH.tC.gpd	0.8	0.62	0.80	0.34	0.001	76	0.201	41	0.029	41	0.05	
GJR.GARCH.nC.e	0.8	0.62	0.80	0.37	0.002	27	0.195	61	0.029	41	-0.11	0.80
GJR.GARCH.nC.n	0.8	0.62	0.80	0.36	0.002	27	0.195	61	0.029	41	-0.09	
GJR.GARCH.nC.gpd	0.8	0.62	0.80	0.36	0.001	76	0.201	41	0.029	41	0.00	

Source: Author's computation