Market Risk Management Under the Fundamental Review of the Trading Book

Volatility Modeling and Historical Simulation of Value at Risk and Expected Shortfall

Julie A. S. Nilsen

Thesis submitted for the degree of
Master of Economic Theory and Econometrics
30 credits

Department of Economics
Faculty of Social Sciences

UNIVERSITY OF OSLO

May 2022

Market Risk Management Under the Fundamental Review of the Trading Book

Volatility Modeling and Historical Simulation of Value at Risk and Expected Shortfall

Julie A. S. Nilsen

Market Risk Management Under the FRTB – Volatility Modelling and Historical Simulation of Value at Risk and Expected Shortfall

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Acknowledgments

This thesis marks the end of my studies at the University of Oslo. I am grateful for the academic insight and experiences I have acquired and the quality of the education I have received. I have found these years to be both intriguing and rewarding.

I would like to extend my gratitude to my supervisor, Jin Cao, for his invaluable guidance and encouragement during the writing of this thesis; and for always being available to advise on any issues or findings. Furthermore, I wish to thank my colleagues at DNB for taking the time to provide me with insight and suggestions in the early stages of this thesis.

Any remaining errors or inaccuracies are my sole responsibility.

Oslo, May 2022

Julie A. S. Nilsen

Abstract

Under the forthcoming Fundamental Review of the Trading Book (FRTB), financial institutions using the Internal Models Approach must calculate their market risk capital requirements based on Expected Shortfall (ES), instead of Value at Risk (VaR). Backtesting will however still be VaR-based. This thesis compares volatility modeling of VaR and ES to the prevalent Historical Simulation (HS) while considering the VaR-based backtesting requirements of the new market risk framework and how it might create incentives within banks. The models are compared during the global financial crisis, the COVID-19 outbreak, and two periods of growth. The considered models are plain HS, RiskMetrics EWMA, mle EWMA, GARCH(1,1), volatility weighted HS (vwHS) using an EWMA filter and a GARCH filter. Both the Gaussian distribution and Student's t-distribution are evaluated for the parametric models. The VaR and ES forecasts of each model are generated from five trading portfolios that are representative of a Norwegian bank, subject to market risk from trading activities in the foreign exchange, interest rate, commodity, and equity markets. The VaR models are backtested using the unconditional coverage test by Kupiec (1995) and the conditional coverage test by Christoffersen et al. (2001). The ES models are backtested using the exceedance residuals test of McNeil & Frey (2000).

Though there does not seem to be a straightforward criterion for selecting proper models, the results indicate that the distributional risk properties of different portfolios affect the efficiency of models and that both VaR and ES could suffer from inaccurate estimates under certain conditions. While vwHS with an EWMA filter and GARCH(1,1) with t-innovations are found to be superior to plain HS under all market conditions, the findings suggest that the new market risk framework might provide wrong incentives within banks. The divergent performance of corresponding ES and VaR models implies that when implementing the FRTB, banks will face a trade-off between obtaining an accurate ES model at the cost of a less accurate VaR model; and obtaining a VaR model with good backtesting properties at the cost of a less accurate ES model. This implies that VaR-based backtesting could potentially discourage banks from implementing the best ES models, which could result in wrong capital numbers.

Keywords: Value at Risk, Expected Shortfall, Fundamental Review of the Trading Book, Market Risk, Historical Simulation, Volatility Modeling, Backtesting, Capital Requirements.

Content

List of Figures

List of Tables

Acronyms

ARCH Autoregressive Conditional Heteroscedasticity

BCBS Basel Committee of Banking Supervision

CC Conditional Coverage

ER Exceedance Residuals

ES Expected Shortfall

EWMA Exponentially Weighted Moving Average

FX Foreign Exchange

FRTB Fundamental Review of the Trading Book

GARCH Generalized Autoregressive Conditional Heteroskedasticity

HS Historical Simulation

IID Independent and Identically Distributed

IMA Internal Models Approach

MLE Maximum Likelihood Estimation

PDF Probability Density Function

SA Standardized Approach

SD Standard Deviation

sVaR Stressed Value at Risk

t-dist Student's t-distribution

UC Unconditional Coverage

VaR Value at Risk

VWHS Volatility Weighted Historical Simulation

1 Introduction

Within banking, quantifying risk represents the first step toward risk management and regulation. That is, banks must be able to quantify the risk they face to achieve an optimal balance between risk and earnings in a long-term perspective. Moreover, banks are required to hold capital reserves for the risk they take to ensure that they can cover unexpected losses and remain solvent during crises. Under the current market risk framework, banks using the Internal Models Approach (IMA) to calculate their market risk capital requirements are permitted to use their own Value at Risk (VaR) models, subject to certain standards. VaR is a risk measure that quantifies the worst expected financial loss over a certain horizon, within a fixed confidence level, making it a transparent measure of the potential loss (J. C. Hull 2012, 271).

Despite being a principal measure of market risk, VaR has several widely recognized shortcomings and has been on the receiving end of criticism from both academics and practitioners. Most importantly, VaR can lead to unwanted risk-taking as it fails to capture tail risk, i.e., high impact events with low probability do not affect VaR. This proved problematic in the wake of the 2007-2009 global financial crisis, whereby banks' VaR models did not accurately reflect the risk of loss, which resulted in insufficient capital reserves leading up to the crisis (Youngman 2009, 51). To ensure that banks are capturing tail risk events, the Basel Committee of Banking Supervision (BCBS) has proposed to replace VaR with Expected Shortfall (ES) in the «Fundamental Review of the Trading Book» (FRTB). When implemented in January 2023, the FRTB will replace the current market risk framework.

ES is defined as the expected loss, conditional on the loss being greater than VaR (J. C. Hull 2012, 274). ES accounts for tail risk by giving information about the probability of a large loss and the likely magnitude of these losses. However, as ES is more difficult to backtest than VaR, banks will need to *backtest* their ES models based on the corresponding VaR models under the FRTB (BCBS 2013, 100-103). Backtesting is a method used for model validation, where the former model-generated forecasts are compared to the realized profit and loss (BCBS 1996, 2-3). Banks will thus face regulatory incentives to obtain accurate VaR and ES estimates under the FRTB.

Regulatory incentives naturally affect a bank's decision-making process when deciding on an internal model specification. It can be argued that the great focus on regulatory capital since the financial crisis has impacted banks' risk management decisions. For instance, several studies show that VaR and ES estimates can be improved by using simulation approaches belonging to the Autoregressive Conditional Heteroscedasticity (ARCH) framework, which exploits the time-varying volatility of financial returns. By construction, such approaches are more responsive to changes in the market. Despite this, most banks use the less reactive Historical Simulation (HS) (European Banking Authority 2017, 32), which poorly deals with the volatility dynamics of financial returns. For instance, O'Brien and Szerszen (2017) find that a simple univariate Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model outperforms banks' HS models. Hansen and Lunde (2001) confirmed the relevance of the simple GARCH(1,1). When comparing 330 ARCH-type models on DM/USD exchange rate data, they find no evidence that GARCH(1,1) is outperformed by more sophisticated models. In a recent article by Deloitte (2020), the authors argue that the market volatility during the financial crisis and the COVID-19 outbreak could have been forecasted more accurately by applying an Exponentially Weighted Moving Average (EWMA) model compared to plain HS. Laurent and Firouzi (2017) find that volatility-weighted HS (vwHS) has better backtesting properties than plain HS, also during market stress. There are nevertheless good arguments against using volatility modeling approaches as the forecasted risk could be too volatile for banks trying to maintain stable capital levels. Laurent and Firouzi argue that the capital cost of increased model resilience is such that it provides wrong incentives within banks. This could explain why the plain HS is favored by the majority of banks that are using IMA.

The upcoming implementation of the FRTB gives rise to the question of how banks should model their risk to ensure that their VaR and ES models yield accurate estimates and whether new regulatory constraints might affect banks' internal model specifications.

1.1 Thesis Objective

The objective of this thesis is to contribute to the existing literature on the forecasting accuracy of VaR and ES, in light of the FRTB. Given that the FRTB is deeply rooted in addressing the issues arising from the financial crisis, this thesis investigates whether volatility modeling of VaR and ES can address a wider variety of market conditions and thereby improve banks' risk management compared to the widely used HS approach. Moreover, as banks must backtest their ES models based on VaR under the FRTB, the performance of corresponding VaR and ES models is compared to inform whether VaR-based backtesting might incentivize banks to choose certain model specifications.

1.2 Methodology

Following the standards outlined in the FRTB, 1-day VaR and ES forecasts at the 97.5% and 99% confidence levels will be generated and backtested. The length of the historical period used to generate the forecasts is set to the past 250 trading days[1], as this is what the FRTB favors and the most common practice among banks (European Banking Authority 2017, 32-33). The VaR models are backtested using the unconditional coverage test by Kupiec (1995) and the conditional coverage test by Christoffersen et al. (2001). The ES models are evaluated using the exceedance residuals test of McNeil & Frey (2000).

The considered simulation approaches are plain HS, RiskMetrics EWMA, maximum likelihood estimated (mle) EWMA, the GARCH(1,1), and vwHS using an EWMA filter and a GARCH filter. For the EWMA and GARCH(1,1) approach, which assume that the returns follow a given statistical distribution, both the Gaussian distribution and Student's t-distribution are evaluated. In total, nine model specifications are considered at two confidence levels.

The empirical analysis is carried out in R using time series of the daily closing prices of six indices and assets[2] during the period 01-01-2006 to 31-12-2021. The data are used to construct trading portfolios that are representative of a Norwegian bank, subject to market risk from trading activities in the foreign exchange, interest rate, commodity, and equity markets. The data are divided into four backtesting periods: (1) the global financial crisis, (2) the period 01-07-2009 to 19-02-2020, which is perceived by some as "the longest bull market in history", (3) the COVID-19 outbreak, and (4) the post COVID-19 outbreak. The chosen backtesting periods ensure that the models are evaluated during two different shocks and two periods of growth. This limits the possibility of a model performing well due to a good fit to certain market conditions.

1.3 Delimitations

This thesis does not seek to determine the most suitable model specification of VaR and ES, nor does it attempt to fit the best time-varying volatility model. Rather, the analysis is restricted to comparing the widely used plain HS approach to five given volatility modeling approaches while considering VaR-based backtesting. The choice of models is motivated by their parameter

[1] With about 250 trading days in a year, this corresponds to one year of data.
[2] The data were sourced from Yahoo! Finance, Investing.com, Bloomberg through Eikon, and the Federal Reserve Economic Data. Section 3.1 provides a detailed description of the data.

parsimony and by previous research where these models have performed well. Nine model specifications are considered in total, which should serve the purpose of this thesis.

The analysis is limited to 1-day forecasts that are estimated from a rolling window of the past 250 trading days. While a 10-day holding period is a common practice for capital purposes, banks will be obliged to backtest their ES models based on the 1-day 97.5% and 99% VaR under the FRTB. Following the standard outlined by the BCBS, 1-day forecasts at both confidence levels are generated and backtested.

The analysis is restricted to long positions, meaning that the risk of loss is associated with the asset decreasing in value. The tail of interest for estimation is hence the left tail of the profit and loss distribution. VaR and ES resulting from short positions, where the risk of loss is associated with the asset increasing in value and hence the right tail, will not be considered. It may be that the models that fit well to left tail risk do not fit equally well to right tail risk. However, as long positions are more common, as well as short positions falling outside of the scope of this thesis, only long positions and hence positive asset weights will be considered.

The chosen backtesting periods may seem somewhat arbitrary as banks are obliged by regulators to backtest their models daily using a backtesting period of the past 250 trading days. However, backtesting during a certain period is conditional on the market scenario during that backtesting period. While a backtesting window of the past 250 trading days would inform which model performs better on average, the four backtesting periods allow for a comparison of the volatility modeling approaches and plain HS under specific market conditions. The four backtesting periods were hence chosen to meet the objective of this thesis.

1.4 Contribution to Literature

In the previously mentioned study by Hansen and Lunde (2001), the authors concluded that different models do not fit different datasets equally well. If one can reasonably conclude that different models are suitable for different datasets, another reasonable assumption is that different models are suitable for different assets. While research focused on improving the predictive power of VaR and ES through more advanced approaches has been extensive, most of these studies have been carried out using data on only a limited number of assets, most commonly equities or foreign exchange. On the other hand, research that uses a variety of asset classes to construct trading portfolios is limited. For instance, Hansen and Lunde (2001) used data on DM/USD exchange rate and IMB stock returns to compare 330 ARCH-type models.

Using simulated data and a sample portfolio consisting of three U.S. stocks[3], Yamai and Yoshiba (2002b) described the advantages and disadvantages of ES over VaR. Yamai and Yoshiba (2002d) compared ES and VaR under market stress using simulated data and daily log returns of foreign exchange rates. When comparing GARCH(1,1) to plain HS, O'Brien and Szerszen (2017) used data on real total trading revenue from U.S. banks but did not identify the asset classes held by the banks. In the paper by Laurent and Firouzi (2017), the comparison of vwHS and plain HS was carried out using S&P 500 returns data only. In the previously mentioned studies, Deloitte (2020) is the only one to perform the analysis on securities belonging to several asset classes, with the assets being S&P 500, US Treasury Yields 5Y, EUR/USD exchange rate, and WisdomTree WTI Crude Oil. Their study is however limited to comparing EWMA to plain HS. Furthermore, while studies comparing different model specifications have been extensive, research that considers the regulatory requirements banks are constrained by when interpreting the results is scarce. While Laurent and Firouzi (2017) did consider how the stressed ES for capital purposes prescribed by the FRTB might provide wrong incentives within banks, their study did not consider the VaR-based backtesting requirements.

Rather than comparing VaR and ES model specifications in isolation, this thesis complements the existing literature by considering the VaR-based backtesting requirements of the FRTB and how it might create incentives within banks. Furthermore, the analysis is carried out using constructed trading portfolios of the asset classes foreign exchange, interest rate, commodity, and equities. This limits the possibility of a model performing well due to a good fit to a certain asset class. By conducting the analysis on a variety of asset classes while considering the backtesting constraint of the FRTB, this thesis allows for a reality dimension which in turn has wider implications for practical risk management.

Three major implications can be drawn from the analysis. Firstly, the predominant plain HS is outperformed by vwHS using an EWMA filter and GARCH(1,1) applying t-innovations with four degrees of freedom under all market conditions. However, even some of the approaches that are designed to account for time-varying volatility suffered from failed backtests in many cases. The findings further suggest that these backtesting exceptions can be partly explained by structural market changes rather than model deficiencies alone. This highlights the importance of the regulatory tools allowed by the FRTB. I.e., under extraordinary market conditions, regulatory flexibilities in assigning the backtesting multiplier[4] and in

[3] The considered stocks were General Electric, McDonald’s, and Intel.
[4] The capital measure is scaled up by a multiplier which is determined by the bank’s backtesting performance. See section 2.1 for details.

allowing banks to continue to use IMA even if they fail many backtests are important tools to safeguard against capital procyclicality.

Secondly, while there does not seem to be a straightforward criterion for selecting proper models, the results suggest that the distributional risk properties of different portfolios do affect the efficiency of models and that both VaR and ES could suffer from inaccurate risk estimates under certain conditions. In practice, this implies that one must be careful when interpreting outcomes from various models and that financial risk management should not depend on either risk measure alone. A combined approach is likely to be more sophisticated.

Thirdly, while the transition from VAR to ES ensures more conservative risk estimates, the results suggest that VaR-based backtesting might provide wrong incentives within banks. The divergent performance of corresponding ES and VaR models implies that when implementing the FRTB, banks will face a trade-off between obtaining an accurate ES model at the cost of a less accurate VaR model; and obtaining a VaR model with good backtesting properties at the cost of a less accurate ES model. This implies that VaR-based backtesting could potentially discourage banks from implementing the best ES models, which could result in wrong capital numbers.

1.5 Outline

The remainder of this thesis is structured as follows. Section 2 provides a theoretical background to market risk regulation within banking, the mathematical definitions, and properties of VaR and ES, as well as their strength and weaknesses. Moreover, this section reviews the commonly observed characteristics of financial returns and introduces the two main approaches used to calculate VaR and ES: the parametric and non-parametric approach. The considered simulation approaches are explained in this section. Section 3 provides an overview of the data and methodology used to estimate and evaluate the models and presents some descriptive statistics of the data. The results are presented in section 4, while the findings are summarized and discussed in section 5. Section 6 concludes the paper.

2 Theoretical Background

2.1 Market Risk Management Within Banking

Many banks have portfolios of traded instruments for short-term profits. These portfolios, referred to as trading books, are subject to *market risk*: the risk of loss due to adverse movements in market prices or exchange rates. This includes equity prices, interest rates, credit spreads, foreign exchange (FX), and commodity prices (BCBS 2019, 1). Under the current market risk framework, sophisticated banks with well-established risk management are allowed by regulators to use IMA. This includes calculating their capital charge for assets in the trading book based on their own VaR models, subject to certain standards (Youngman 2009, 51). The bank is free to choose what statistical approach their calculations should be based on, but VaR should be computed daily using a 99th percentile, one-tailed confidence interval, with the minimum holding period being 10 trading days[5]. The historical observation period used to calculate VaR is constrained to a minimum length of one year, corresponding to 250 trading days (BCBS 2009, 14). Capital calculation under the current framework is based on the sum of a current VaR and a stress calibrated VaR (sVaR). In practice, this has led to double counting of risk and has been accused of making market risk capital *procyclical*. Within the framework of regulatory capital, procyclicality arises when capital requirements increase during market stress, when banks' capital reserves may already be depleted. Regulations requiring additional capital may hence worsen market stress and further increase volatility in trading markets as banks liquidate their positions (Abboud, et al. 2021).

The significant trading book losses that banks incurred during the 2007-2009 global financial crisis highlighted the need for an improvement of the market risk framework. To this end, BCBS initiated the FRTB in 2012, which will replace the current market risk framework when implemented in January 2023[6] (BCBS 2019, 2-3). To ensure that banks are capturing tail risk events, the new IMA under FRTB replaces VaR with ES and changes the confidence level for capital purposes from 99% to 97.5%. As both VaR and sVaR will be replaced by a stressed 10-day[7] ES, capital requirements will be less likely to increase abruptly during market stress. Yet, there is still a potential for capital requirements to increase during market stress as the new

---

[5] Banks may use the 1-day VaR and scale this number up to the 10-day horizon by multiplying it with $\sqrt{10}$.

[6] While January 1st, 2023, is set by the BCBS as the global implementation deadline for the FRTB, local jurisdictions can decide the implementation timeline to better reflect local market requirements (BCBS 2020).

[7] 10 days is the liquidity horizon assigned by the FRTB to the most liquid risk factors. Less liquid market factors have liquidity horizons up to 120 days, though this results from 10-day computations that are scaled up (Laurent and Firouzi 2017, 4).

framework upholds the "backtesting multiplier" of the current framework. Backtesting is a method used for model validation, where former model-generated forecasts are compared to the realized loss (BCBS 2009). If the actual losses exceed the forecasts made by the model, this is considered a backtesting exception. Should a bank experience many exceptions, the backtesting multiplier starts to increase capital requirements. This is a way for regulators to ensure that poorly specified models are treated more conservatively (Abboud, et al. 2021).

Because ES is more difficult to backtest than VaR, banks will be obliged to backtest their ES models based on the corresponding 97.5% and 99% 1-day VaR (BCBS 2013, 101-103). Backtesting under the FRTB will have to be conducted for each trading desk[8] where banks must report the 1-day 97.5% ES and the number of VaR exceptions over the past 250 trading days at the 97.5% and 99% level (Laurent and Firouzi 2017, 4).

The backtesting methodology proposed by BCBS relies on testing the *coverage* of the risk model. I.e., comparing whether the number of backtesting exceptions is consistent with what one would expect at the given confidence level. For example, when forecasting a 99% VaR, the number of actual losses that exceed VaR should be around 1% of all cases, which corresponds to 2-3 days over 250 trading days. The model is then deemed reasonably accurate if the number of exceedances is around 2-3 days. Similarly, if the number of exceedances is significantly higher than 2-3 days, the validity of the model should be questioned. Should a bank experience many backtesting exceptions for a trading desk, the backtesting multiplier increases capital requirements and the bank could be forced by regulators to start using the more conservative Standardized Approach (SA) for this desk. To safeguard against capital procyclicality, the FRTB allows for regulatory flexibility in assigning the backtesting multiplier and allowing banks to continue using IMA even if they fail many backtests. Such regulatory measures were necessary during the outbreak of the COVID-19 pandemic. According to Abboud et al. (2021), U.S. banks experienced simultaneous backtesting exceptions during this period that would have increased capital requirements by $3.3 billion without intervention. U.S. regulators perceived that these exceptions reflected unusual market conditions rather than model misspecification. They thus responded by capping the multipliers to prior levels to avoid the exceptions triggering an increase in capital requirements at the height of the COVID-19 market volatility (Abboud, et al. 2021).

[8] A trading desk is a designated area of a financial institution where specific types of securities are sold and purchased. Many institutions have separate trading desks for the FX, commodities, fixed income, and equity markets. Institutions may further subdivide these markets (BCBS 2022).

2.2 Value at Risk

VaR is a risk measure that quantifies the largest potential loss of a portfolio over a specified time horizon, within a fixed confidence level (J. C. Hull 2012, 271). Mathematically, VaR is defined as (Artzner, et al. 1999, 216):

$$VaR^{1-\alpha}(L) = \min\{l : P(L > l) \leq 1 - \alpha\}, \qquad \alpha \in (0,1), \qquad (1)$$

Equation (1) states that the VaR of a portfolio at confidence level $1 - \alpha$, is given by the smallest number l, such that the probability of a loss, L, exceeding l is not larger than $1 - \alpha$ for some predetermined holding period, usually 1 or 10 days. In practical terms, if a bank's portfolio has a 10-day, 99% VaR equal to NOK 2.3 million, this means the probability of having a loss larger than 2.3 million does not exceed 1% over the next 10 days (Cao 2022, 372). VaR is hence a quantile of the loss distribution; the 99% VaR represents the 0.99 quantile of loss, as shown in Figure 1.

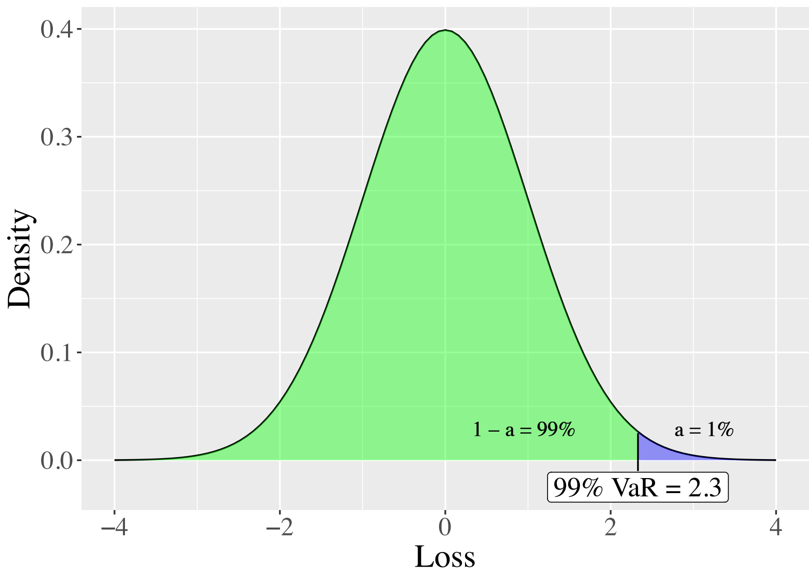

Figure 1: The 99% VaR of a standard normal loss distribution.

Despite its widespread use, VaR has several mathematical shortcomings. A well-cited paper by Artzner et al. (1999) describes what properties a quantitative risk measure should satisfy to be considered coherent. Among these properties is sub-additivity[9], which is an important property in financial risk management as it implies that diversification is beneficial. Artzners paper provided results that VaR fails to satisfy sub-additivity in general as the merger of two

[9] For a real-valued function $f: A \rightarrow B$ and elements $a, b \in A$, sub-additivity implies $f(a + b) \leq f(a) + f(b)$.

portfolios may have a larger VaR than the sum of the VaR of the individual portfolios (Artzner, et al. 1999). Another concern with VaR is its inability to capture the *tail risk* of the loss distribution, i.e., high impact events with low probability do not affect VaR. This is because VaR is simply a quantile of the loss distribution; it does not give any information about the size of the loss that exceeds the confidence level. Hence, it fails to capture the risk associated with the shape of the loss distribution beyond the confidence level. This is problematic if the return distribution of a portfolio has thick tails as this can incentivize banks to take on tail risk (BCBS 2013, 5).

2.3 Expected Shortfall

The BCBS acknowledged that VaR is not coherent as early as 2011 (BCBS 2011, 17-20). In the consultive papers on the FRTB, published in 2013, they proposed to replace VaR with ES[10] due to the many weaknesses identified with using VaR for determining regulatory capital requirements, including its inability to capture tail risk (BCBS 2013, 3). To calculate ES, it is necessary to first calculate VaR for the portfolio. Using the definition of VaR in equation (1), ES can be defined as:

$$ES^{1-\alpha}(L) = E[L|L \geq VaR^{1-\alpha}], \qquad (2)$$

Equation (2) states that the ES of a portfolio is equal to the average of all losses during a given time horizon, conditional on the loss being greater than VaR[11]. ES accounts for tail risk by giving information about the probability of a large loss and the likely magnitude of these losses. Furthermore, ES leads to more conservative risk estimates as the expected loss a day when VaR is exceeded is always larger than VaR. This is illustrated in Figure 2, which compares the 99% VaR with the 99% ES. As pointed out by Artzner et al. (1999), ES has better mathematical properties than VaR. If two portfolios are merged, ES usually decreases, reflecting the benefits of diversification. It certainly never increases as VaR occasionally does. Because ES assigns equal weights to losses beyond the confidence level, it is always sub-additive and hence a coherent risk measure (J. C. Hull 2012, 274-277).

[10] Sometimes referred to as "conditional VaR", "mean excess loss", or "tail VaR" (J. C. Hull 2012, 274).

[11] If the 10-day, 99% VaR is NOK 2.3 million, ES is the average amount lost over 10 days assuming that the loss is greater than NOK 2.3 million. For a standard normal loss distribution with VaR = 2.3, ES would be 2.67.

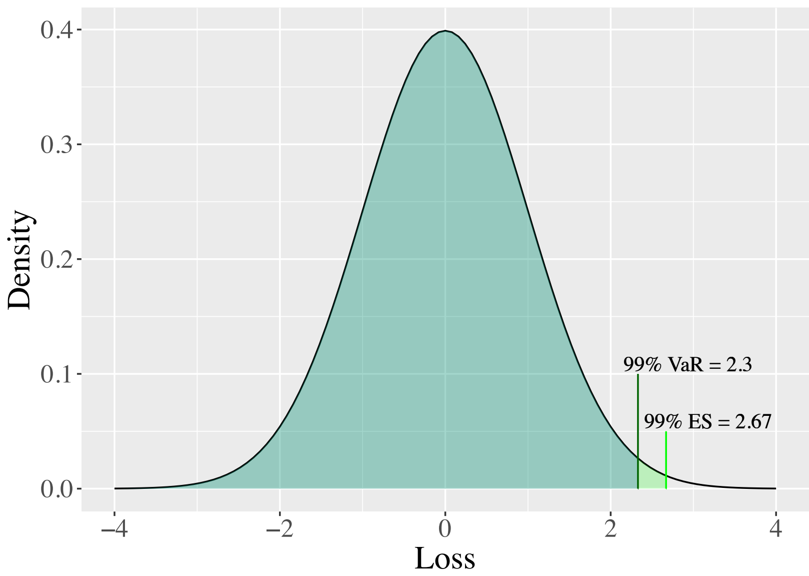

Figure 2: Comparison of the 99% VaR and 99% ES.

Despite being coherent, ES was rejected by the BCBS for a long time due to not being *elicitable*. If a risk measure is elicitable, it is possible to rank the risk model's performance, for example through backtesting. Backtesting ES is also more complicated than backtesting VaR as the distribution of the stochastic loss variable, i.e., the actual shortfall value, is needed to test whether ES estimations are derived from the same distribution. This has led VaR backtests to have a much stronger theoretical foundation than ES backtests. Thus, backtesting under the FRTB will still be based on VaR.

Another disadvantage is that larger data sets are generally required for ES to achieve the same level of accuracy as VaR. Specifically, Yamai and Yoshiba (2002b) found that around 1000 data points are required for the accuracies of the two risk measures to converge (Yamai and Yoshiba 2002b, 95). This makes ES less reliable than VaR for small data sets. Furthermore, they found that the estimation error of ES is larger than that of VaR when the losses have fat tails, especially at high confidence levels (Yamai and Yoshiba 2002b, 102). Their result implies that capital calculated from ES may be less stable than capital calculated from VaR. However, the FRTB changes the confidence level for the capital purposes from 99% to 97.5%, which should improve the accuracy of ES (Hull and White 2014, 12). Moreover, Yamai and Yoshiba (2002d) found that both VaR and ES have tail risk under extreme value distributions, though ES less than VaR. Their results show that even ES may underestimate the risk of securities with fat-tailed properties and high tail probability. They further argue that financial risk management should not depend entirely on VaR or ES and that it is essential to monitor diverse aspects of the profit and loss distribution (Yamai and Yoshiba 2002d, 233).

2.4 Stylized Facts of Financial Returns

Empirics show that returns from financial market variables, measured over daily time intervals, are characterized by the stylized facts *fat tails* and *volatility clustering*. Volatility clustering refers to the observation that volatility varies through time and its degree shows a tendency to persist, i.e., there are periods of low volatility and periods of high volatility (Bera and Higgins 1993, 309). Within econometrics, this phenomenon is referred to as Autoregressive Conditional Heteroscedasticity (ARCH). ARCH is an interesting property because it can be exploited to forecast the volatility of future periods – something that will be returned to later. The empirical distribution of financial returns is also more peaked and has fatter tails than the normal distribution would permit. Hence, the frequency of large positive or negative financial returns is higher than what would be expected if returns were normally distributed. Platen & Sidorowicz (2007), among others[12], identify the Student's t-distribution with about four degrees of freedom as the typical estimated log return distribution of diversified stock indices measured over a longer observation period. Owing to the observed high level of significance in their study, Platen & Sidorowicz argue that their result can be interpreted as a stylized fact (Platen and Sidorowicz 2007). These commonly observed characteristics imply both advantages and disadvantages of the different simulation approaches used to estimate VaR and ES and should be considered when deciding on a model specification. The simulation approaches used to estimate VaR and ES can be classified into the two main categories *non-parametric* approaches and *parametric* approaches.

2.5 The Non-parametric Approach

The non-parametric approach does not make any assumptions about the distribution of returns, rather it applies the empirical distribution of the data. Apart from its simplicity, the non-parametric approach has the advantage of overcoming the issue of the fat-tailed characteristics of financial asset returns.

2.5.1 Plain Historical Simulation

Among the methods belonging to the non-parametric approaches is the plain HS approach, which is the predominant method used to calculate VaR in the banking industry (Sharma 2012, 1). This approach uses data on actual historical returns of the portfolio and re-organizes the

[12] See (Markowitz and Usmen 1996a), (Markowitz and Usmen 1996b), and (Hurst and Platen 1997).

returns of the n previous trading days from worst to best. VaR is then the lowest return corresponding to the $1 - \alpha$ quantile. VaR can be estimated using HS according to equation (3), where r_τ is the q^{th} percentile lowest return in the sorted list of returns over the past n days. ES is estimated by HS according to equation (4), where r_i represent the i^{th} return in the sorted list of returns and τ is the number of days the loss exceeded the VaR.

$$VaR_t^q = r_\tau, \qquad (3)$$

$$ES_t^q = \frac{1}{\tau}\sum_{i=1}^{\tau} r_i\,, \qquad (4)$$

The main issue with HS is the trade-off associated with the length of the historical period used to generate the forecast. Large estimation errors are possible if the observation period covered by the data is too short. This calls for larger observation periods, which in turn could result in flatter risk measures that are less responsive to market movements (Hendricks 1996, 43-44). Furthermore, if the historical period does not contain a period of market stress, or if the current period of stress is very unlike past shocks, the estimates are likely to be biased. Another concern with HS is that it poorly deals with the time-varying volatility of financial returns. Considering volatility as a fixed constant through time might lead to underestimation of VaR and ES during stress (Laurent and Firouzi 2017, 10). This can be partially accommodated by considering shorter periods but at the price of large estimation errors. Shorter periods also increase the volatility of VaR estimates, which can result in procyclical capital numbers. The FRTB favors using a one-year period, which is the most common practice among banks. According to the March 2017 European Banking Authority benchmarking study, 58% of the respondents using IMA were using a one-year window to compute VaR while the other respondents were using lengthier periods. 66% of the banks in this study reported that they were using HS (European Banking Authority 2017, 32-33).

2.5.2 Volatility-Weighted Historical Simulation

Analogous to HS, vwHS relies on the empirical distribution of returns. However, vwHS accounts for the volatility dynamics of financial returns by adjusting each return according to equation (5). I.e., by scaling past returns by the most recent volatility forecast divided by the volatility estimate of the corresponding date. VaR and ES are then computed by HS from the empirical distribution of $\tilde{r}_t$. Hence, if the volatility in the current holding period is higher (lower) than average, the estimates for VaR and ES are adjusted upwards (downwards).

$$\tilde{r}_t = \left(\frac{\sigma_{t+1}}{\sigma_t}\right) r_t, \qquad (5)$$

In equation (5), the forecast of the next holding period's volatility, σ_{t+1}, is not directly observable and needs to be estimated. For this purpose, a GARCH or EWMA specification can be implemented according to equations (6) and (7), respectively. The GARCH and EWMA models are explained in more detail in the following sections.

$$\sigma_{T+1}^2 = \omega + \alpha r_t^2 + \beta \sigma_t^2, \qquad (6)$$

$$\sigma_{T+1}^2 = (1 - \lambda) r_t^2 + \lambda \sigma_t^2, \qquad (7)$$

Laurent and Firouzi (2017) find that vwHS with an EWMA filter outperforms plain HS, and better conforms with the expected number of VaR exceptions in the long run. Furthermore, they find that the vwHS-EWMA performs better than HS during periods of market stress, including the financial crisis and the European sovereign debt crisis (Laurent and Firouzi 2017). The results of Laurent and Firouzi (2017) indicate that the vwHS-EWMA model better deals with sudden increases in market volatility, which lowers the number of VaR exceptions and leads to more resilient risk models. However, they also find that this approach dramatically inflates the stressed ES under the FRTB and thus the capital charge compared to plain HS. As the results indicate that resilient models come at the price of increased capital charge, they argue that banks moving from plain HS to vwHS-EWMA would be unlikely once the FRTB is implemented (Laurent and Firouzi 2017, 24).

2.6 The Parametric Approach

The parametric estimation approach assumes that the returns follow a given statistical distribution, such as the Gaussian distribution or Student's t-distribution. To estimate VaR and ES, there are hence different formulas depending on the assumed distribution. Assuming that the returns, R_{t+1}, follow a normal distribution would imply $R_{t+1} \sim N(\mu, \sigma^2)$, where μ represents the mean of the distribution and σ^2 its variance. The next period's VaR is given by equation (8), where $q = 1 - \alpha$ and z_q represents the q$^{\text{th}}$ quantile for the normal distribution (McNeil, Frey and Embrechts 2005, 39-40). The corresponding ES is given by equation (9), where ϕ is the probability density function (pdf) of a standard normal variable (McNeil, Frey and Embrechts 2005, 45).

$$VaR_t^q(R) = \mu + \sigma z_q, \qquad (8)$$

$$ES_t^q(R) = \mu + \sigma \frac{\phi(z_q)}{1-q}, \qquad (9)$$

Using a normal distribution to model returns could lead to underestimation of risk as financial returns are characterized by fat tails. Fat tails may be modeled using the t-distribution. If one assumes that the returns follow a standard t-distribution with $\nu > 2$ degrees of freedom, VaR is given by equation (10) and ES by equation (11). t_ν^{-1} denotes the distribution function and g_ν the pdf of the standard t-distribution (McNeil, Frey and Embrechts 2005, 45-46).

$$VaR_t^q(L) = \mu + \sigma t_\nu^{-1}(q), \qquad (10)$$

$$ES_t^q(L) = \mu + \sigma \frac{g_\nu\left(t_\nu^{-1}(q)\right)}{1-q} \cdot \frac{\nu + \left(t_\nu^{-1}(q)\right)^2}{\nu - 1}, \qquad (11)$$

When using the parametric approach to forecast VaR and ES, the mean and volatility parameters need to be estimated. Given the short-term horizon involved in regulation, usually a 1-day or 10-day holding period, one can reasonably assume that the conditional mean has a marginal impact on the computation of risk measures and set its value equal to zero (Laurent and Firouzi 2017, 4). The most straightforward way of estimating the volatility, σ, is to assume that the past returns have constant volatility. However, this approach arguably fails to account for the volatility dynamics of financial returns. Like HS, the simple parametric method can be slow to incorporate new information during market changes. A possible solution is to model returns as having time-varying conditional volatility as opposed to constant volatility.

2.6.1 GARCH

Alternatives to the constant volatility assumption under the simple parametric approach are simulation approaches belonging to the ARCH framework, which exploits the volatility clustering inherent in financial returns. One such model is GARCH(p,q), proposed by Bollerslev (1986). Under the GARCH approach, returns are assumed not to be independent and identically distributed (i.i.d.). Rather, they are assumed to exhibit volatility clustering. The GARCH(p,q) model is defined by equation (12) – (14).

$$\sigma_{t+1}^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_t^2 + \sum_{i=1}^{q} \alpha_i r_t^2, \qquad \omega > 0, \alpha \geq 0, \beta \geq 0, \qquad (12)$$

$$r_t = \sigma_t z_t, \qquad (13)$$

$$z_t \sim i.i.d.\ D(0,1), \qquad (14)$$

In equation (12), the past return r_t, is assumed to be the squared conditional volatility times an i.i.d. variable, z_t, following a zero-mean and unit-variance distribution. z_t is usually assumed to be either standardized normal innovations or standardized t-innovations. σ_{t+1}^2 represents the estimated conditional volatility in period $t+1$. The coefficients, ω, α_i and β_j represent weights that reflect the presumed impact of each of the variables on the coming periods' volatility. The imposed constraints that these coefficients are larger than zero ensure positive conditional volatility. ω is the proportion of the volatility that is invariant across time. α captures the short-term volatility clustering effects while β captures the persistence of shocks on future volatility. Meaning that if α is large relative to β, then volatility will react quickly to market movements and appear spiky. If the reverse is true, volatility will appear to be persistent, remaining at the same level for longer (Dowd 2005, 132). The coefficients can be estimated by mle or set to fixed values. p and q represent how many previous periods of returns and conditional volatility should be included, i.e., the lags.

GARCH(p,q) models are not commonly used by banks to estimate VaR. Laurent and Firouzi (2017) argue that this is likely related to the large number of risk factors involved in the internal models of banks. Specifying, implementing, and monitoring a stable and meaningful GARCH model is therefore a huge challenge (Laurent and Firouzi 2017, 10). However, while multivariate volatility modeling might theoretically improve performance, O'Brien and Szerszen (2017) show that a simple univariate GARCH(1,1) outperforms banks' HS models. The results of Hansen and Lunde (2001) proved the relevance of the simple GARCH(1,1). When comparing 330 ARCH-type models on DM/USD exchange rate data, they find no evidence that a GARCH(1,1) is outperformed by more sophisticated models. Hansen and Lunde further argue that setting p and q to anything other than 1 will not yield a significant difference in the forecast (Hansen and Lunde 2001, 887). The GARCH(1,1) is shown by equation (6).

$$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta \sigma_t^2, \qquad (6)$$

When $\omega > 0$ and $\alpha + \beta < 1$, the volatility process is stationary, meaning that there will be some mean reversion towards a constant value (Laurent and Firouzi 2017, 6). When comparing the backtesting performance of GARCH(1,1) models to US banks' VaR models during the COVID-19 outbreak, Abboud et al. (2021) find that models with daily parameter updating performed best. Even though GARCH(1,1) outperformed the banks' VaR models in the study, they still experienced numerous exceptions during March 2020. The evidence of Abboud et al. (2021) points toward the market conditions being the main driver of these exceptions rather than material deficiencies in the considered models (Abboud, et al. 2021). The GARCH(1,1), which will be used in the analysis, will henceforth be referred to as GARCH.

2.6.2 EWMA

A simplified version of GARCH, the EWMA, was proposed by RiskMetrics (1996). In this model, the authors found that the decay factor, λ, which represents β in GARCH, should be optimally set to 0.94. Furthermore, they found that α in GARCH should be set to $1 - \lambda$; i.e., 0.06. The EWMA model sets ω constant to zero for simplification. The RiskMetrics EWMA model hence eliminates the need for coefficient estimation, although mle computation of λ is possible. Equation (7), which denotes the EWMA model, states that tomorrow's volatility is the weighted effect of today's volatility and today's squared return. Because the weight on the last period's conditional volatility, λ, is smaller than one, the effect of past returns decreases exponentially.

$$\sigma_{t+1}^2 = (1 - \lambda) r_t^2 + \lambda \sigma_t^2, \qquad \lambda < 1, \qquad (7)$$

A recent article by Deloitte (2020) shows that an EWMA estimation of VaR better accounts for volatility clustering and would have suffered far fewer backtesting exceptions during the financial crisis and the COVID-19 outbreak compared to plain HS, when looking at a variety of asset classes. Although some time series experienced failed backtests even with EWMA during the COVID-19 outbreak, EWMA still offered the best performance compared to plain HS (Deloitte 2020).

3 Data and Methodology

3.1 Data

The empirical analysis is carried out using a time series of the daily closing prices[13] on selected indices and assets, for every market day during the time-period 01-01-2006 to 31-12-2021. The indices and assets were selected to represent the trading book of a Norwegian bank, subject to market risk from trading activities in the FX, interest rate, commodity, and equity markets. To this end, the constructed trading desks consist of four portfolios: FX, equity, interest rate, and commodity. The fifth portfolio, which represents the full trading book of the bank, is a weighted portfolio of the four trading portfolios. The data were sourced from Yahoo! Finance, Investing.com, Bloomberg through Eikon, and the Federal Reserve Economic Data. The indices and assets, their weights, and source are presented in Table 1.

Portfolio	Asset	Weight	Acronym	Source
FX	USD/NOK	47.5%	USDNOK	Yahoo! Finance
	EUR/NOK	47.5%	EURNOK	Yahoo! Finance
	SEK/NOK	5%	SEKNOK	Yahoo! Finance
Equity	Oslo Stock Exchange Total Return Index	100%	OBX	Investing.com
Interest Rate	Norwegian Interbank Offered Rate 3 month		NIBOR	Bloomberg through Eikon
Commodity	Brent Crude Oil Prices		Brent	Federal Reserve Economic Data
Trading Book	FX	7%	TB	
	Equity	15%		
	Interest rate	75%		
	Commodity	3%		

Table 1: List of indices and assets, their weight, and source.

The FX portfolio consists of the three currency pairs USD/NOK, EUR/NOK, and SEK/NOK, weighted by 47.5%, 47.5%, and 5%, respectively. The two former pairs were chosen as USD and EUR each account for almost half of all FX funding in the Norwegian

[13] The closing price adjusted for dividends and stock splits were used whenever available. This concerns USD/NOK, EUR/NOK, SEK/NOK, and OBX.

banking sector. All other currencies account for less than 5% (Norges Bank 2019). SEK is assumed to be the most important among other currencies for the hypothetical bank. Equity-related risk in the trading book arises mainly due to market making in shares and equity derivatives on electronic marketplaces and to customer brokers (DNB 2020, 60). The OBX Index, which lists the 25 most liquid stocks on the main index of the Oslo Stock Exchange, was selected to represent the equity portfolio. Interest rate risk occurs when the financial instruments change value due to interest rate fluctuations. The interest rate portfolio consists of the three-month Norwegian Interbank Offered Rate (NIBOR). Commodities account for only a small fraction of the market risk arising from the trading activities of Norwegian banks. As oil is assumed to be the most important commodity (Danske Bank 2020, 45), the Brent Crude Oil price series was used for the commodity portfolio. A credit portfolio was not included in the analysis. This was motivated by a Pearson correlation test[14] on the three-month NIBOR and the Norwegian 10-year Government Bond Yield series, which yielded a correlation coefficient of 0.84. As the high correlation indicates that the credit spread risk is mostly captured by the interest rate portfolio, the credit portfolio was discarded.

The VaR and ES forecasts are estimated for each trading desk and the full trading book of the bank. Table 1 states the weight of each trading desk in the trading portfolio. These are 15% for the equity portfolio, 75% for the interest rate portfolio, 7% for the currency portfolio, and 3% for the commodity portfolio. The choice of weights was motivated by the reported market risk capital requirements by asset class in the 2020 annual risk reports of DNB (DNB 2020, 57-59) and Nordea (Nordea 2020, 145-146), and the reported VaR for trading-related activities in the 2020 annual risk report of Danske Bank (Danske Bank 2020, 46).

The time series of realized returns and forecasted VaR and ES estimates are divided into four time periods before running the backtests: (1) the global financial crisis, using data from 01-12-2007 to 30-06-2009, (2) the 01-07-2009 to 19-02-2020 time-period, henceforth referred to as the "bull market"[15] period, (3) the COVID-19 outbreak, covering the period 20-02-2020 to 30-09-2020, and (4) the post COVID-19 outbreak, going from 01-10-2020 to 31-12-2021. The backtesting periods ensure that the models are tested during two arguably different types of shocks, the financial crisis and the COVID-19 outbreak, as well as two periods of growth. The backtesting procedure is explained in greater detail in section 3.4.

[14] The results from the Pearson correlation test can be found in Appendix A.
[15] The time-period 01-07-2009 to 19-02-2020 is perceived by many as "the longest bull market in history" (Wigglesworth 2020).

3.2 Software

The analysis is carried out in R. The script can be found in Appendix B. The packages `rugarch` (Galanos 2022) and `quarks` (Letmathe 2022) were used to specify and estimate the forecasts of the parametric models, as well as the backtesting procedures; `PerformanceAnalytics` (Peterson and Carl 2020) for descriptive plots; `moments` (Komsta and Novomestky 2005) for descriptive statistics; `qqplot2` (Wickham 2009), `lemon` (Edwards, et al. 2020), and `cowplot` (Wilke 2020) for plotting; `tidyverse` (Wickham, Averick, et al. 2019), `xts` (Ryan, et al. 2020), `lubridate` (Spinu, Grolemund and Wickham 2021) and `writexl` (Ooms 2021) for data preparation, visualization, and exportation.

3.3 Estimation of VaR and ES

When analyzing financial time series, log returns are preferred due to being time-additive and lognormal. The returns are thus converted to log returns according to equation (15) before the forecasts are generated. r_t denotes the log return at time t, P_t the adjusted close price at time t, and $t - 1$ denotes the previous market day.

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \qquad (15)$$

Following the FRTB backtesting standard, 1-day forecasts at the 97.5% and 99% confidence levels are estimated for all models, using a rolling window of the past $n = 250$ trading days. For the non-parametric models, VaR and ES are estimated using HS by implementing a for-loop in R according to equations (3) and (4). vwHS-EWMA and vwHS-GARCH are specified and estimated using `rollcast`, belonging to the `quarks` package, according to equations (5) and (6), and equations (5) and (7), respectively. Following RiskMetrics (1996), a fixed value of $\lambda = 0.94$ is used to estimate vwHS-EWMA.

The parametric models are estimated using the Gaussian distribution and the t-distribution. Following Platen & Sidorowicz (2007), four degrees of freedom are applied for the t-distribution. Alternatively, the degrees of freedom could have been estimated using mle. However, this approach is computationally intense, and estimated parameters could suffer from estimation bias due to sampling uncertainty. Imposing four degrees of freedom for the t-distribution is motivated by the computational parsimony of this approach and the findings of Platen & Sidorowicz (2007). GARCH assuming the normal distribution, the nGARCH, is estimated according to equations (8), (9), and (12)-(14). GARCH assuming the t-distribution, the tGARCH, is estimated according to equations (10), (11), and (12)-(14). Two versions of EWMA are applied in addition to the normal distribution and t-distribution. The first approach follows RiskMetrics and uses a fixed value of $\lambda = 0.94$. The second approach obtains λ through mle. These models will henceforth be referred to as $EWMA_R$ and $EWMA_{ml}$, respectively. The GARCH and EWMA models are specified using `ugarchspec` and estimated using `ugarchroll`, both belonging to the `rugarch` package. Following Abboud et al. (2021), who found that models with daily parameter updating yielded the most accurate measures of risk, the GARCH and $EWMA_{ml}$ parameters are estimated and updated daily. Table 2 re-states the parameter choices. Due to the short holding period, the mean is assumed to be zero for the GARCH and $EWMA_{ml}$ models. Unfortunately, the `rugarch` package restricts from following a similar approach for the $EWMA_R$ models, as all parameters cannot be fixed. Thus, the mean will be assumed not to be zero for these models and will instead be estimated by mle.

Holding period, in days	1
Rolling window length, n	250
t-distribution degrees of freedom, ν	4
Decay factor, λ, fixed value	0.94
Bootstrap sample, used for ES backtesting	10 000
Seed used for bootstrapping	250

Table 2: Parameters used in the analysis.

3.4 Backtesting

Once the forecasts are obtained, the estimates are backtested to determine which models perform well during which periods and if the performance of corresponding VaR and ES models coincide. No matter the backtesting period used, backtesting suffers from sample dependence. I.e., if a backtest is performed over a specified 100-day period and yields exactly five exceptions, using another 100-day period might yield completely different results. This makes it difficult to assess the actual accuracy of the model. To address this, statistical tests are used to determine whether a model has passed or failed a backtest. The VaR models are backtested using the unconditional coverage test by Kupiec (1995) and the conditional coverage test by Christoffersen et al. (2000). The ES models are backtested using the exceedance residual test of McNeil & Frey (2000). These tests are explained in the following sub-sections.

3.4.1 Unconditional & Conditional Coverage Test for VaR

The unconditional coverage (UC) test by Kupiec (1995), tests whether the observed frequency of VaR exceedances is consistent with the expected exceedances at the given confidence level (Ghalanos 2022, 44). That is, when estimating $VaR^{1-\alpha}$, the days when the realized losses exceed the VaR should be around α% of all cases over a backtesting period of n days. If there are too many exceedances, the model is underestimating risk. Likewise, if there are too few exceedances, the model is overestimating risk. In both cases, the validity of the model should be questioned. Under the Null hypothesis of an accurate model, the number of exceedances, x, should be equal to $n \cdot \alpha$. The test can be conducted as a likelihood ratio test. Under the Null, the test statistic belongs to a χ^2 distribution with one degree of freedom. The test statistic, LR_{UC}, is given by equation (16), where N denotes the sample size and p the probability of an exceedance at the given confidence level (Ghalanos 2022, 44). A test statistic larger than the critical value[16] leads to a rejection of the Null hypothesis: "The model is correct".

$$LR_{UC} = -2 \ln\left(\frac{(1-p)^{N-x} p^x}{\left(1-\frac{x}{N}\right)^{N-x}\left(\frac{x}{N}\right)^x}\right), \qquad (16)$$

A shortcoming of the UC test is that it assumes that the exceedances are independent of each other. This assumption is likely not to hold during market stress as there might be several consecutive days of exceedances. As a result, the UC test might fail to reject a model that produces clustered VaR exceptions. While regulatory backtesting uses a UC measure[17] (Zhang and Nadarajah 2017, 9), clustering of exceptions is problematic as it might deplete a bank's capital reserves through the backtesting multiplier. The conditional coverage (CC) test of Christoffersen et al. (2001) will hence be included as it jointly tests for correct coverage and exception clustering. The test can be conducted as a likelihood ratio test where the test statistic, LR_{CC}, is asymptotically distributed as χ^2 with 2 degrees of freedom. Under the Null, the conditional and unconditional coverage are equal to the significance level, i.e., 1% for a 99% VaR. The UC and CC tests are implemented in R using `VaRTest`, belonging to the `rugarch` package.

[16] Alternatively, the Null is rejected when the probability is lower than a given significance level.
[17] In particular, the Traffic Light test is prescribed by the BCBS. The Traffic Light test is not considered in this thesis due to its low power and as it does not allow for a comparison of models (Zhang and Nadarajah 2017, 9).

3.4.2 Exceedance Residual Test for ES

McNeil & Frey (2000) were among the first to propose a backtesting procedure for ES. Their exceedance residual (ER) test defines exceedance residuals, r_{t+1}, as the difference between the next period's return, x_{t+1}, and the expected shortfall at time t, $ES_q^t(x_{t+1})$, conditional on x_{t+1} exceeding VaR at time t. Equation (17) denotes the exceedance residuals, also referred to as excess shortfall residuals (McNeil and Frey 2000, 294).

$$r_{t+1} = \frac{x_{t+1} - \widehat{ES}_q^t(x_{t+1})}{\hat{\sigma}_{t+1}}, \qquad (17)$$

If the model is correct, then under the Null these residuals should behave as an i.i.d. sample with mean zero and unit variance. The Null can be tested using a one-sided t-test against the alternative that the exceedance residuals have a mean greater than zero. A mean greater than zero would imply that ES is systematically underestimated, which McNeil and Frey remark as the most likely direction of failure (McNeil and Frey 2000, 294). Models can then be compared based on their p-values: high p-values indicates a valid model, whereas low p-values indicate that the validity of the model should be questioned. The p-values will be obtained using the bootstrap method proposed by Efron & Tibshirani (1993). This method should alleviate any bias concerning assumptions about the underlying distribution of the excess shortfall residuals (Ghalanos 2022, 44-45). The exceedance residuals test is implemented in R by using `ESTest`, belonging to the `rugarch` package using a bootstrap sample of $N = 10\,000$. As the VaR models are evaluated using both the UC and CC tests, two backtests of ES would ideally have been included in the analysis. As previously discussed, VaR backtests have a much stronger theoretical foundation than backtests of ES. While not all existing ES backtests have been implemented in R, some have been implemented but lack the same documentation as VaR backtests. The choice of the ER test of McNeil and Frey (2000) was motivated by the fact that this test was already implemented in R through the well-documented `rugarch` package.

3.5 Evaluation of models

Section 4 presents the backtesting results in turn for each of the four backtesting periods. The average performance of the models during the period will be presented first, followed by the portfolio level performance. The portfolio level VaR performance includes the UC and CC test statistics the models obtained for each of the individual portfolios. The portfolio level ES performance includes the obtained p-values from the ER test. Test statistics marked in gray indicate that the Null was rejected. The best performers, i.e., the models yielding the lowest test statistic for VaR models and the highest p-value for the ES models, will be marked by a box. All the best performers are marked in case of a tie. The test statistics that are neither gray nor boxed are test statistics where the Null was not rejected but the model did not yield the best result.

Following Karlsson and Zakrisson (2016), two average performance statistics will be obtained for the VaR models: the sum[18] of all UC test statistics during the period, $\sum LR_{uc}$, and the sum of all CC test statistics in the period, $\sum LR_{cc}$ (Karlsson og Zakrisson 2016, 17). Both $\sum LR_{uc}$ and $\sum LR_{cc}$ follow a χ_5^2-distribution[19] which yields a critical value of 11.1 at the 5% confidence level. To evaluate the average performance of the ES models, the average performance statistic for the ER test will be obtained using Fisher's method of combining p-values, which is calculated according to $X_{ES} = -2\sum_{i=1}^{k} \ln(p_i)$. The X_{ES} test statistic follows a χ_{10}^2-distribution[20], which has a critical value of 18.31 at the 5% level. To assess whether the Gaussian distribution or the t-distribution is more accurate for the parametric models, the statistic of all models belonging to a specific distribution will be summed for each period as $\sum\sum LR_{uc}$ and $\sum\sum X_{ES}$. These statistics were not tested for significance but used for internal comparison only. Finally, the results from a Jarque-Bera normality test and estimated nGARCH parameters are presented to help explain the performance of the models.

3.6 Descriptive Statistics

Table 3 states the descriptive statistics of the portfolio log returns during each of the backtesting periods. The reported statistics are the length of observations in days, n, the mean, standard deviation (sd), kurtosis, skewness, and the minimum and maximum log return. Considering the zero-mean assumption in the model specifications, it is reassuring to see that the mean of the log returns is zero or very close to zero for all portfolios during the four time periods. The sd lies within the range of 0.02 to 0.09 during the COVID-19 outbreak and 0.01 to 0.03 during the other periods. The kurtosis is higher than 3 for all portfolios. This indicates that the distributions of log returns have fatter tails compared to the normal distribution, which has a kurtosis of 3. It is worth noting that the FX portfolio possessed very fat tails during the Bull Market and COVID-19 outbreak, with a kurtosis of 382.18 and 57.62, respectively.

---

[18] I.e., the sum of the obtained test statistics for each of the 5 portfolios.

[19] This follows from the fact that if $LR_i \sim \chi_1^2$, then $\sum_{i=1}^{5} LR_i \sim \chi_5^2$.

[20] In general terms, the Fisher test statistic follows a χ_{2k}^2-distribution.

Period	Portfolio	n	mean	sd	kurtosis	skewness	min	max
Financial Crisis	Equity	391	0.00	0.03	4.69	-0.37	-0.11	0.11
	Interest Rate	412	0.00	0.02	8.63	-0.91	-0.14	0.08
	FX	361	0.00	0.01	17.23	-0.01	-0.08	0.08
	Commodity	396	0.00	0.03	7.10	0.15	-0.17	0.18
	Trading Book	350	0.00	0.03	4.96	-0.13	-0.10	0.13
Bull Market	Equity	2673	0.00	0.01	5.27	-0.16	-0.06	0.06
	Interest Rate	2776	0.00	0.01	25.64	0.80	-0.13	0.19
	FX	2766	0.00	0.01	382.18	0.15	-0.24	0.24
	Commodity	2693	0.00	0.02	5.47	0.17	-0.08	0.11
	Trading Book	2622	0.00	0.01	5.24	-0.21	-0.06	0.06
COVID19 Outbreak	Equity	154	0.00	0.02	6.52	-1.08	-0.09	0.05
	Interest Rate	160	-0.01	0.05	14.24	-2.29	-0.30	0.12
	FX	160	0.00	0.02	57.62	2.84	-0.17	0.22
	Commodity	155	0.00	0.09	21.92	-1.76	-0.64	0.41
	Trading Book	151	0.00	0.02	6.30	-1.04	-0.09	0.05
Post COVID19 Outbreak	Equity	315	0.00	0.01	5.18	0.02	-0.03	0.05
	Interest Rate	327	0.00	0.03	5.15	0.43	-0.11	0.12
	FX	327	0.00	0.01	21.23	0.16	-0.06	0.06
	Commodity	298	0.00	0.02	7.76	-0.95	-0.13	0.07
	Trading Book	292	0.00	0.01	5.56	-0.14	-0.04	0.05

Table 3: Descriptive statistics of the log returns.

The skewness is not equal to 0 for any of the portfolios, but during three of the four periods, most values fall between -0.5 and 0.5. This indicates that the underlying distributions are fairly symmetrical. During the COVID-19 outbreak, however, all values are less than -1 or greater than 1, indicating a highly skewed distribution. As expected, most of the portfolios are left skewed[21] during the two crisis periods, while positive skewness[22] is more common during the bull market and post COVID-19 outbreak. The difference between the minimum and maximum for each portfolio denotes the range of the log returns. The highest range belongs to the commodity portfolio during the financial crisis, the FX portfolio during the bull market period, the commodity portfolio during the COVID-19 outbreak, and the interest rate portfolio during the post COVID-19 outbreak.

Figure 3 shows the time series of log returns of all five portfolios. Time-varying volatility is evident from the time series. All portfolios except the FX portfolio seem to have at least some clusters of high volatility during the financial crisis and the COVID-19 outbreak.

[21] Negative values indicate that the tail is on the left side of the distribution, extending toward negative values.
[22] Positive values indicate that the tail is on the right side of the distribution, extending toward positive values.

Plots of the portfolio price series and histograms of log returns, which illustrate the descriptive statistics in Table 3, can be found in Appendix A.

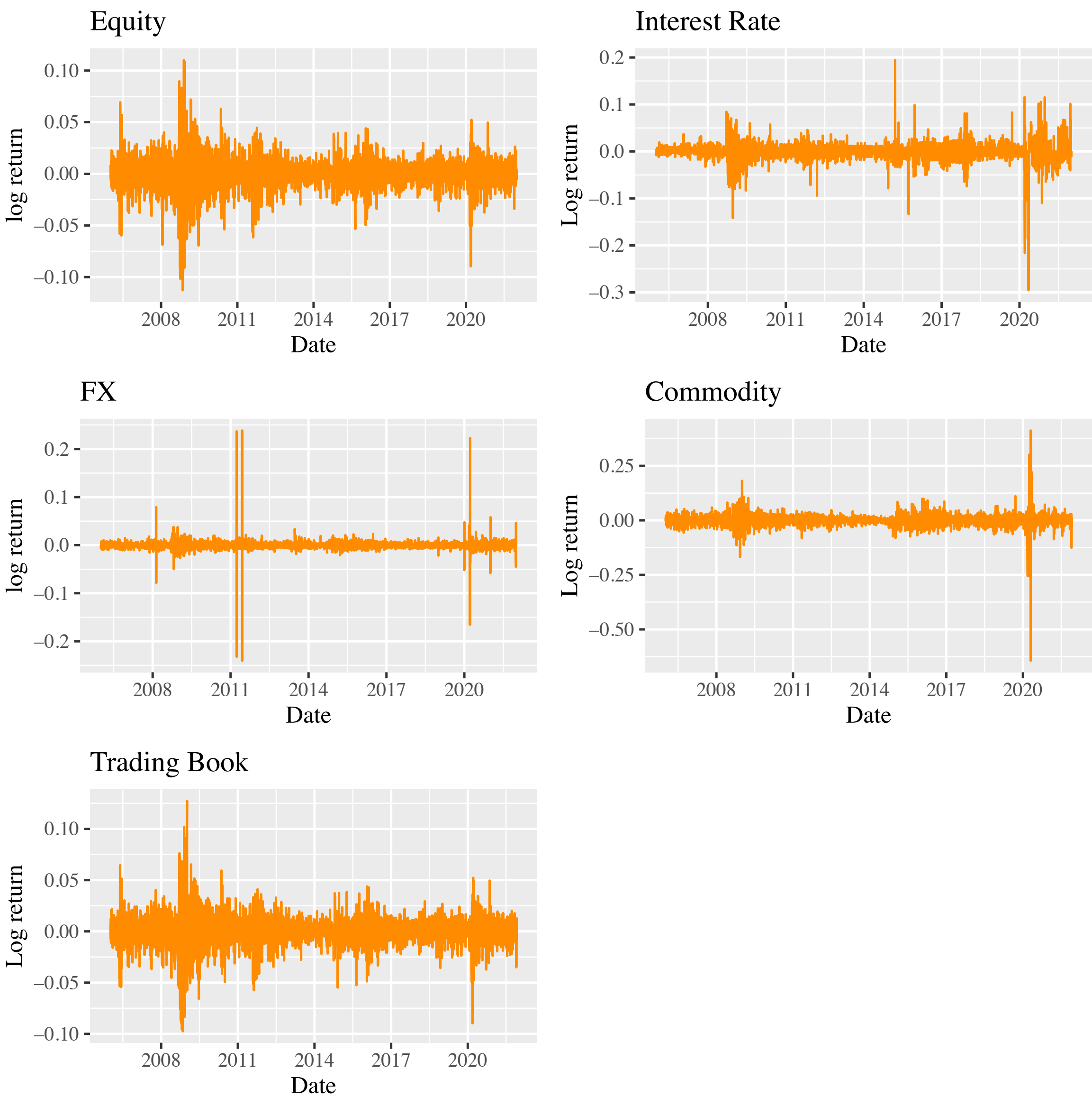

Figure 3: Time series of log returns.

4 Results

4.1 The Global Financial Crisis

Looking at tables 4 and 5, the VaR models clearly performed worse than their corresponding ES models during the financial crisis, as indicated by the high frequency of failed VaR tests. Furthermore, it can be concluded that while the plain HS had an inaccurate coverage, failed to account for clustering of exceptions, and systematically underestimated ES at the 99% level, even some of the volatility modeling approaches failed to keep up with the actual risk during the financial crisis. As will be shown in section 4.5, this can be partly explained by structural market changes rather than model deficiencies alone. Nevertheless, the tGARCH and vwHS-EWMA are both vast improvements over the plain HS and the only approaches that remained satisfactory from both a VaR and ES backtesting perspective at both levels. It should be noted that while the vwHS-EWMA VaR model had great results at both levels for the UC and CC tests, and even the best results at the 97.5% level, its corresponding ES model yielded a rather high test statistic. Especially at the 99% level, the model was very close to being rejected. The tGARCH is the only approach that obtained consistently low test statistics for its VaR and ES model at both levels. The $tEWMA_{ml}$ is the top-performing ES model at the 97.5% level while the $nEWMA_{ml}$ is the top-performing ES model at the 99% level. Surprisingly enough, as their corresponding VaR models failed both the UC and CC tests at both levels. At the 97.5% level, the $nEWMA_R$ is the only model that failed all three tests. At the 99% level, the models to fail all three tests were the nGARCH, $nEWMA_R$, and plain HS.

		UC		CC		ER	
q=0.975		$\sum LR_{UC}$	$\sum\sum LR_{UC}$	$\sum LR_{CC}$	$\sum LR_{CC}$	$-2\sum \ln(p_i)$	$\sum\sum X_{ER}$
Normal	GARCH	13.40		17.13		18.01	
	$EWMA_R$	107.65	$\boxed{156.54}$	109.12	$\boxed{173.30}$	33.97	22.82
	$EWMA_{ml}$	35.50		40.05		4.27	
t-dist.	GARCH	7.09		9.46		4.27	
	$EWMA_R$	130.89	170.34	143.38	190.16	14.76	$\boxed{19.08}$
	$EWMA_{ml}$	32.36		37.31		$\boxed{0.77}$	
HS	Plain	88.78		117.27		15.05	
	vwEWMA	$\boxed{2.92}$		$\boxed{6.63}$		14.57	
	vwGARCH	30.92		34.10		15.47	

Table 4: Financial Crisis average performance statistics, 97.5% level. Time period: 01-07-2007 to 30-06-2009. Test statistics marked in gray indicate a rejection of the H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

q=0.99		UC $\sum LR_{UC}$	UC $\sum\sum LR_{UC}$	CC $\sum LR_{CC}$	CC $\sum\sum LR_{CC}$	ER $-2\sum \ln(p_i)$	ER $\sum\sum X_{ER}$
Normal	GARCH	17.24	181.48	21.28	180.88	23.01	23.24
	$EWMA_R$	122.49		111.00		18.34	
	$EWMA_{ml}$	41.75		48.59		$\boxed{0.22}$	
t-dist.	GARCH	6.67	$\boxed{105.99}$	7.91	$\boxed{113.65}$	2.43	$\boxed{5.63}$
	$EWMA_R$	96.45		102.30		6.27	
	$EWMA_{ml}$	$\boxed{2.87}$		$\boxed{3.44}$		3.20	
HS	Plain	72.89		76.27		18.71	
	vwEWMA	4.62		9.20		17.45	
	vwGARCH	45.28		47.60		16.18	

Table 5: Financial Crisis average performance statistics, 99% level. Time period: 01-07-2007 to 30-06-2009. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

The $\sum\sum LR_{UC}$ and $\sum\sum LR_{UC}$ test statistics show that the Gaussian distribution yielded the overall lowest test statistic among the VaR models at the 97.5% level. A surprising result, as one might expect the t-distribution to fit better to a period of market stress. Looking at Table 4 it is evident that this result follows from the poor performance of the $tEWMA_R$ model. At the 99% level, the t-distribution obtains the lowest overall test statistic. For the ES models, the t-distribution is the best fit at both levels.

4.1.1 Portfolio Level VaR Backtesting Results

Tables 6 and 7 states the results from the UC and CC tests at the portfolio level during the financial crisis. The vwHS-EWMA is the only VaR model to pass all UC and CC tests at both confidence levels for all five portfolios. Recalling the results in Table 4, the $tEWMA_R$ was the overall worst performing VaR model at the 97.5% level. The model's poor performance further resulted in the t-distribution yielding higher $\sum\sum LR_{UC}$ and $\sum\sum LR_{UC}$ test statistics than the Gaussian distribution at the 97.5% level. Looking at tables 6 and 7, this owes to the model yielding particularly bad estimates of risk for the commodity and trading book portfolio. The average worst performers, the $nEWMA_R$, $tEWMA_R$, and plain HS are the worst performers also at the portfolio level. The $nEWMA_R$ failed the UC and CC tests for all five portfolios, while the $tEWMA_R$ failed the UC and CC tests for all portfolios, except the CC test of the FX portfolio at the 99% level. The plain HS had an accurate coverage and prevented clustered exceptions for the FX portfolio only. Looking at tables 6 and 7, the average top-performing tGARCH did fail the UC and CC tests for the interest rate portfolio at the 97.5% level.

	Normal			Student t			HS		
	GARCH	EWMA$_R$	EWMA$_{ml}$	GARCH	EWMA$_R$	EWMA$_{ml}$	Plain	vwEWMA	vwGARCH
q = 0.975									
FX	0.91	24.16	13.74	0.41	19.73	0.91	4.93	0.91	10.23
Equity	0.48	25.30	0.48	0.90	20.54	1.66	19.12	0.48	7.03
Interest rate	5.48	10.34	10.34	8.81	23.04	19.34	25.73	1.23	10.23
Commodity	1.55	20.75	8.19	0.40	32.17	9.71	15.98	0.12	2.33
TB	4.97	27.11	2.74	0.07	35.40	2.74	23.02	0.18	1.11
q = 0.99									
FX	4.01	23.30	12.25	0.86	7.71	0.86	1.33	0.48	14.78
Equity	0.97	23.79	2.00	3.11	15.80	0.00	15.80	0.97	4.89
Interest rate	6.05	14.73	20.00	0.18	28.18	1.68	14.73	0.18	14.78
Commodity	1.92	30.10	4.76	0.00	25.36	0.25	13.76	0.25	6.54
TB	4.29	30.58	2.74	2.51	21.40	0.07	27.27	2.74	4.29

Table 6: Financial Crisis UC test statistics. Test statistics marked in gray indicate rejection of H0: “the model is correct”. The best statistics are marked by a box. The critical values are 5.02 for the 97.5% models, and 6.63 for the 99% models.

	Normal			Student t			HS		
	GARCH	EWMA$_R$	EWMA$_{ml}$	GARCH	EWMA$_R$	EWMA$_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	1.61	22.88	14.06	1.11	23.46	1.61	6.41	1.61	10.89
Equity	1.25	27.04	1.25	1.15	22.83	2.70	27.14	1.29	7.04
Interest rate	15.64	12.83	12.64	9.53	25.63	19.38	37.55	2.22	10.89
Commodity	2.58	21.27	8.19	0.73	32.35	9.72	16.20	0.75	3.52
TB	5.06	25.09	3.91	0.44	37.11	3.91	29.97	0.77	1.77
q = 0.99									
FX	4.37	24.79	15.96	0.89	8.28	0.89	1.54	4.33	15.28
Equity	1.16	25.95	2.25	3.12	16.84	0.08	16.22	1.16	5.32
Interest rate	10.92	15.72	22.18	0.30	29.28	1.92	15.21	0.30	15.28
Commodity	2.17	22.02	5.18	0.08	25.50	0.38	14.68	0.38	7.06
TB	4.66	22.53	3.03	3.52	22.39	0.16	28.63	3.03	4.66

Table 7: Financial Crisis CC test statistics. Test statistics marked in gray indicate rejection of H0: “the model is correct”. The best statistics are marked by a box. The critical values are 7.38 for the 97.5% models, and 9.21 for the 99% models.

4.1.2 Portfolio Level ES Backtesting Results

Recalling the results from Table 4, the nEWMA$_R$ was the only ES model to fail the overall ER test at the 97.5% level. Looking at Table 8, this resulted from the nEWMA$_R$ obtaining inaccurate risk estimates for the equity, commodity, and trading book portfolio, while also obtaining quite low p-values for the other portfolios. While their overall performance statistics were not rejected at the 97.5% level, the tEWMA$_R$ failed the ER test for the interest rate portfolio, and the plain HS failed the ER test for the equity and trading book portfolio. Recalling the results from Table 5, the nGARCH, nEWMA$_R$, and the plain HS failed the ER test at the 99% level. For the nGARCH, Table 8 shows that this is a result of the model's inability to reflect the risk associated with the interest rate portfolio, and the low p-value it obtained for the commodity portfolio. The nEWMA$_R$ was rejected for the equity, commodity, and trading book portfolio. The plain HS was not rejected for any portfolio at this level but did obtain rather low p-values for some portfolios. Looking at Table 8, the ES vwHS-EWMA being close to rejection by the average ER statistic follows from the model obtaining low p-values for the FX and commodity portfolio at both levels.

	Normal			Student t			HS		
	GARCH	EWMA$_R$	EWMA$_{ml}$	GARCH	EWMA$_R$	EWMA$_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	0.07	0.10	0.97	0.29	0.11	0.86	0.39	0.13	0.20
Equity	0.27	0.01	0.98	1.00	0.92	1.00	0.02	0.37	0.60
Interest rate	0.06	0.07	0.81	0.85	0.02	0.96	0.21	0.37	0.13
Commodity	0.12	0.02	0.87	0.48	0.25	0.85	0.33	0.11	0.14
TB	0.69	0.02	0.98	1.00	0.82	0.98	0.00	0.35	0.20
q = 0.99									
FX	0.05	0.08	0.99	0.58	0.35	0.51	0.24	0.09	0.19
Equity	0.55	0.01	0.99	1.00	0.99	1.00	0.13	0.39	0.52
Interest rate	0.01	0.13	0.98	0.69	0.41	0.77	0.14	0.14	0.31
Commodity	0.06	0.00	0.94	0.74	0.33	0.52	0.11	0.06	0.04
TB	0.61	0.00	0.99	1.00	0.93	0.99	0.18	0.55	0.25

Table 8: Financial Crisis ER p-values. P-values marked in gray indicate rejection of H0: "the model is correct". The best p-values are marked by a box.

4.2 Bull Market

Tables 9 and 10 states the average performance statistics during the bull market period. Looking at the results, it can be concluded that most VaR models failed to accurately measure risk during this period. Only the nGARCH and tGARCH passed the UC and CC tests at both confidence levels, though their test statistics were very close to rejection. Nevertheless, the tGARCH is the only approach to pass all three tests at both confidence levels during this period. The vwHS-EWMA passed all three tests at the 97.5% level only. The plain HS had the best coverage among the VaR models but failed the CC test at both levels. Indicating that while the coverage of the model was satisfactory, it produced clustered VaR exceptions. Its corresponding ES model just passed the ER test. During the financial crisis, the $tEWMA_{ml}$ failed both VaR tests but was the best performing ES model at the 97.5% level. Looking at Table 9, this is the case also during the bull market period. The $nEWMA_{ml}$ yields good results for its ES model at both levels, and even the best results at the 99% level. Its corresponding VaR model on the other hand fails both tests at both levels. The 97.5% $tEWMA_R$ and 97.5% vwHS-GARCH were the only approaches to fail all three tests. The nGARCH did not do too good either, its ES test failed the test at both levels while its VaR model just passed the UC and CC tests. At both confidence levels, the t-distribution is the best fit among the parametric VaR and ES models.

		UC		CC		ER	
q=0.975		$\sum LR_{UC}$	$\sum\sum LR_{UC}$	$\sum LR_{CC}$	$\sum\sum LR_{CC}$	$-2\sum \ln(p_i)$	$\sum\sum X_{ER}$
Normal	GARCH	10.40	219.30	$\boxed{10.03}$	221.55	23.74	28.23
	$EWMA_R$	146.00		136.99		14.53	
	$EWMA_{ml}$	62.90		74.52		4.49	
t-dist.	GARCH	10.46	$\boxed{206.14}$	11.02	$\boxed{202.46}$	4.44	$\boxed{8.32}$
	$EWMA_R$	139.16		124.97		20.78	
	$EWMA_{ml}$	56.52		65.77		$\boxed{3.88}$	
HS	Plain	$\boxed{1.57}$		21.02		15.51	
	vwEWMA	10.04		11.01		10.23	
	vwGARCH	16.22		22.14		20.60	

Table 9: Bull Market average performance statistics, 97.5% level. Time period: 01-07-2009 to 19-02-2020. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

		UC		CC		ER	
q=0.99		$\sum LR_{UC}$	$\sum\sum LR_{UC}$	$\sum LR_{CC}$	$\sum\sum LR_{CC}$	$-2\sum \ln(p_i)$	$\sum\sum X_{ER}$
Normal	GARCH	10.20	212.16	$\boxed{9.45}$	221.48	32.24	33.77
	$EWMA_R$	117.71		117.63		16.00	
	$EWMA_{ml}$	84.57		94.56		$\boxed{1.54}$	
t-dist.	GARCH	10.25	$\boxed{146.05}$	10.71	$\boxed{157.88}$	4.05	$\boxed{10.18}$
	$EWMA_R$	118.17		124.06		10.35	
	$EWMA_{ml}$	17.65		23.11		6.13	
HS	Plain	$\boxed{5.76}$		20.93		15.47	
	vwEWMA	13.61		15.48		14.72	
	vwGARCH	20.58		30.12		18.30	

Table 10: Bull Market average performance statistics, 99% level. Time period: 01-07-2009 to 19-02-2020. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

4.2.1 Portfolio Level VaR Backtesting Results

The above results showed that the plain HS had the best coverage but was rejected by the CC test at both levels. Looking at the portfolio level VaR results in tables 11 and 12, this resulted from the model having the best coverage for all individual portfolios at the 97.5% level while producing clustered exceptions for the equity portfolio. At the 99% level, its coverage was good for all portfolios while it produced clustered exceptions for the interest rate portfolio. The vwHS-EWMA, which passed all three average tests at the 97.5% level but failed the average UC and CC tests at the 99% level, passed both tests for all individual portfolios. Looking at tables 11 and 12, the model failing the average tests at the 99% level follows from the rather high test statistics it obtained for the FX portfolio. The worst VaR models in tables 9 and 10, the nEWMA$_R$ and tEWMA$_R$, had inaccurate coverage and clustered exceptions for all individual portfolios.

	Normal			Student t			HS		
	GARCH	EWMA$_R$	EWMA$_{ml}$	GARCH	EWMA$_R$	EWMA$_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	4.99	38.1	4.19	5.30	29.33	4.87	0.01	1.98	2.87
Equity	8.53	26.6	23.12	1.25	26.70	24.02	0.01	1.82	2.58
Interest rate	0.45	22.6	3.36	0.62	31.06	0.62	0.45	1.58	2.87
Commodity	4.89	25.8	12.60	2.07	18.87	4.37	1.09	3.89	3.96
TB	6.46	32.9	19.64	1.22	33.2	22.64	0.00	1.63	3.95
q = 0.99									
FX	1.77	20.87	0.02	1.37	26.44	6.01	0.19	4.39	5.05
Equity	10.29	20.7	22.93	2.08	23.6	4.98	0.90	2.36	5.05
Interest rate	6.24	24.66	18.30	0.88	24.06	0.00	1.76	1.32	2.34
Commodity	9.94	21.7	18.53	2.26	26.86	0.34	2.71	3.41	4.82
TB	8.64	29.78	24.80	3.67	17.2	6.30	0.20	2.13	3.32

Table 11: Bull Market UC test statistics. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 5.02 for the 97.5% models, and 6.63 for the 99% models.

	Normal			Student t			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	5.37	21.86	5.96	5.42	24.88	5.37	$\boxed{0.76}$	4.36	2.96
Equity	10.54	31.10	27.20	1.64	19.88	25.29	10.60	2.68	2.96
Interest rate	$\boxed{0.45}$	28.06	3.42	0.84	27.92	2.12	3.78	2.68	5.68
Commodity	5.13	28.06	13.47	2.59	26.29	4.41	$\boxed{1.10}$	6.50	4.63
TB	8.49	27.90	24.47	2.23	26.01	28.58	4.78	$\boxed{1.90}$	5.91
q = 0.99									
FX	2.09	28.86	0.55	2.47	28.47	7.18	$\boxed{0.85}$	5.15	7.25
Equity	9.29	31.70	31.54	2.27	26.57	7.26	$\boxed{2.74}$	3.28	5.32
Interest rate	6.27	20.06	18.30	1.26	24.92	$\boxed{1.13}$	13.04	4.58	7.25
Commodity	9.94	21.06	18.53	2.47	25.29	$\boxed{1.02}$	3.66	4.44	5.97
TB	8.69	15.94	25.64	2.23	18.82	6.52	$\boxed{0.64}$	3.03	4.32

Table 12: Bull Market CC test statistics. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 7.38 for the 97.5% models, and 9.21 for the 99% models.

4.2.2 Portfolio Level ES Backtesting Results

Looking at the portfolio level ES backtesting results in Table 13, the nGARCH, which was rejected by the average ER statistic at both levels, only managed to accurately reflect the risk of loss associated with the FX portfolio. The obtained p-value for this portfolio was however close to rejection at both levels. The $tEWMA_R$ failed the average ER test at the 97.5% level but passed at the 99% level. Looking at Table 13, this follows from the model's inability to reflect the risk associated with the FX and interest rate portfolio. The vwHS-GARCH also failed the average ER test at the 97.5% level. From Table 13 it can be concluded that this is a result of the model's inability to estimate the risk associated with the equity portfolio. The plain HS, which was not rejected by the average ER test, was rejected by the ER test for the equity portfolio at the 99% level. Analogous to its corresponding VaR model, the ES plain HS yields the best test result for the FX portfolio at the 97.5% level.

	Normal			Student t			HS		
	GARCH	EWMA$_R$	EWMA$_{ml}$	GARCH	EWMA$_R$	EWMA$_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	0.07	0.10	0.28	0.29	0.01	0.39	0.39	0.13	0.20
Equity	0.01	0.01	0.99	1.00	0.99	0.99	0.13	0.45	0.02
Interest rate	0.00	0.70	0.39	0.39	0.01	0.38	0.23	0.61	0.20
Commodity	0.00	0.00	0.99	0.98	0.31	0.99	0.16	0.48	0.20
TB	0.01	0.00	0.99	1.00	1.00	0.99	0.23	0.35	0.14
q = 0.99									
FX	0.06	0.08	0.57	0.58	0.35	0.14	0.24	0.09	0.19
Equity	0.02	0.01	1.00	1.00	1.00	0.99	0.00	0.26	0.12
Interest rate	0.00	0.42	0.83	0.26	0.01	0.37	0.20	0.53	0.19
Commodity	0.01	0.00	0.99	0.94	0.54	0.92	0.13	0.32	0.22
TB	0.01	0.00	0.99	0.93	0.98	0.99	0.07	0.16	0.11

Table 13: Bull Market ER p-values. P-values marked in gray indicate rejection of H0: "the model is correct". The best p-values are marked by a box.

4.3 COVID-19 Outbreak

Looking at the average performance test statistics in tables 14 and 15, the VaR models clearly failed to keep up with the actual risk during the COVID-19 outbreak. At the 97.5% level, the vwHS-EWMA and tEWMA$_{ml}$ are the only VaR models to pass both the UC and CC tests. They are also the only models to pass all three tests at the 97.5% level. At the 99% level, no models pass all three tests and all VaR models except the tEWMA$_{ml}$ and wwHS-EWMA failed the UC test. The ES models did much better at reflecting the risk of loss during this period. Only the nGARCH is rejected by the average ER test at the 97.5% level, and only the vwHS-EWMA is rejected at the 99% level. The nEWMA$_{ml}$ is the best performing ES model at both confidence levels while its VaR model failed the UC and CC tests. The worst performing VaR models are the nEWMA$_R$, tEWMA$_R$, and plain HS. Contrary to the financial crisis, no approach stands out as an obvious improvement over the plain HS. While the vwHS-EWMA and tEWMA$_{ml}$ pass all three tests at the 97.5% level, no model passed all three tests at the 99% level. The t-distribution yields the lowest overall test statistic among the parametric VaR and ES models, at both confidence levels.

q=0.975		UC		CC		ER	
		$\sum LR_{UC}$	$\sum\sum LR_{UC}$	$\sum LR_{CC}$	$\sum\sum LR_{CC}$	$-2\sum \ln(p_i)$	$\sum\sum X_{ER}$
Normal	GARCH	17.51		23.12		25.04	
	$EWMA_R$	108.63	154.06	114.98	175.56	11.98	34.97
	$EWMA_{ml}$	27.93		37.46		9.93	
t-dist.	GARCH	21.94		26.94		12.08	
	$EWMA_R$	67.92	99.75	71.79	116.40	10.27	24.41
	$EWMA_{ml}$	9.84		10.67		12.80	
HS	Plain	59.27		102.86		16.08	
	vwEWMA	5.22		7.31		16.28	
	vwGARCH	21.81		27.67		16.35	

Table 14: COVID-19 Outbreak average performance statistics, 97.5% level. Time period: 20-02-2020 to 30-09-2020. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

q=0.99		UC		CC		ER	
		$\sum LR_{UC}$	$\sum\sum LR_{UC}$	$\sum LR_{CC}$	$\sum\sum LR_{CC}$	$-2\sum \ln(p_i)$	$\sum\sum X_{ER}$
Normal	GARCH	25.96		32.46		10.08	
	$EWMA_R$	129.04	207.04	135.82	219.43	16.55	15.89
	$EWMA_{ml}$	52.04		66.15		5.81	
t-dist.	GARCH	28.21		31.22		8.22	
	$EWMA_R$	47.91	86.38	52.66	99.42	11.98	23.71
	$EWMA_{ml}$	10.26		15.55		15.50	
HS	Plain	38.58		61.33		9.88	
	vwEWMA	10.72		12.75		23.19	
	vwGARCH	17.99		22.07		16.88	

Table 15: COVID-19 outbreak average performance statistics, 99% level. Time period: 20-02-2020 to 30-09-2020. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

4.3.1 Portfolio Level VaR Backtesting Results

The above results showed that the average worst performers were the $nEWMA_R$, $tEWMA_R$, and plain HS. Looking at the portfolio level VaR results in tables 16 and 17, the $nEWMA_R$ poorly reflected the risk associated with most of the portfolios. The coverage of the model was only satisfactory for the equity portfolio at the 97.5% level, while it avoided clustered exceptions for the interest rate portfolio only. At the 97.5% level, the $tEWMA_R$ had an accurate coverage and independent exceptions for the FX portfolio only, while at the 99% level it had an accurate coverage and independent exceptions for the FX and equity portfolio only. For all other portfolios, the results were unsatisfactory. The plain HS had an accurate coverage only for the FX portfolio at both levels, and the FX, equity, and trading book portfolio at the 99% level. Looking at Table 16, the plain HS produced clustered exceptions for all portfolios at the 97.5% level, while only avoiding clustered exceptions for the equity and trading book portfolio at the 99% level. The $tEWMA_{ml}$ and vwHS-EWMA are the only models to pass both tests for all five portfolios at both confidence levels. Tables 14 and 15 stated that the tGARCH failed the average UC and CC tests at both levels. Looking at tables 16 and 17, this resulted from the model's inaccurate coverage and clustered exceptions for the interest rate and commodity portfolio at the 97.5% level and the interest rate portfolio at the 99% level. For the other portfolios, the results were satisfactory. In fact, the 99% tGARCH had the best coverage and lowest CC statistic for the FX portfolio.

	Normal			t-distribution			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	0.00	10.78	0.89	0.24	4.57	0.24	3.19	0.24	8.76
Equity	5.16	35.41	3.52	2.14	14.35	2.14	7.05	0.32	1.06
Interest rate	4.56	4.76	10.78	10.78	16.31	2.78	27.44	2.19	8.76
Commodity	4.09	31.73	9.05	5.09	23.40	2.46	14.29	2.09	2.09
TB	3.69	25.94	3.69	3.69	9.29	2.27	7.30	0.37	1.14
q = 0.99									
FX	0.26	16.64	7.18	0.09	2.57	0.09	0.98	0.98	2.63
Equity	4.94	37.93	10.48	4.94	5.48	3.53	4.94	2.76	4.94
Interest rate	7.05	7.18	10.05	13.21	15.31	2.67	10.05	3.05	2.63
Commodity	8.63	37.76	13.63	4.89	15.86	0.89	17.53	1.08	2.72
TB	5.08	29.52	10.70	5.08	8.70	3.08	5.08	2.86	5.08

Table 16: COVID-19 Outbreak UC test statistics. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 5.02 for the 97.5% models, and 6.63 for the 99% models.

	Normal			t-distribution			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	$\boxed{0.21}$	12.58	1.36	0.56	7.21	0.56	12.12	0.56	10.42
Equity	5.55	37.03	4.19	3.20	14.55	3.20	12.83	$\boxed{0.66}$	2.64
Interest rate	6.88	5.72	18.46	10.78	16.31	2.46	47.86	$\boxed{2.19}$	8.76
Commodity	6.05	33.06	9.11	8.05	23.41	$\boxed{2.14}$	17.07	3.17	3.17
TB	4.43	26.60	4.34	4.34	10.31	2.30	12.98	$\boxed{0.72}$	2.69
q = 0.99									
FX	0.27	17.60	7.66	$\boxed{0.14}$	2.77	$\boxed{0.14}$	5.53	1.10	2.84
Equity	5.27	25.76	11.54	5.27	9.15	4.02	7.20	$\boxed{2.97}$	5.27
Interest rate	10.17	7.66	20.90	13.21	15.31	4.80	20.90	4.05	$\boxed{2.63}$
Commodity	11.32	38.20	14.32	7.16	16.05	3.16	20.40	$\boxed{2.56}$	5.91
TB	5.42	31.61	11.73	5.42	9.38	3.42	7.30	$\boxed{2.07}$	5.42

Table 17: COVID-19 Outbreak CC test statistics. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 7.38 for the 97.5% models, and 9.21 for the 99% models.

4.2.2 Portfolio Level ES Backtesting Results

In tables 14 and 15, the 97.5% nGARCH and the 99% vwHS-EWMA were the only ES models to be rejected by the average ER test. From Table 18 it can be concluded that the poor average performance of the 97.5% nGARCH is a result of its low p-values for the trading book, equity, and commodity portfolio. The nGARCH was also the only model to fail a backtest at the 97.5% level, with that being the trading book portfolio. Even though the model was not rejected by the average ER test at the 99% level, the model was rejected for the equity and trading book portfolio in Table 18. For the 99% vwHS-EWMA, the failed average ER test owes to the model's inability to reflect the risk associated with the interest rate portfolio and the low p-value it obtained for the commodity portfolio. The plain HS, which passed the average ER test, did not pass the ER test for the equity and trading book portfolio at the 99% level. The $tEWMA_R$, whose VaR model only managed to reflect the risk of loss for the FX and equity portfolio, did not fail a single ER test. Moreover, while the vwHS-EWMA was the top-performing VaR model, its corresponding ES model failed the ER test for the interest rate portfolio at the 99% level and was close to rejection for the interest rate portfolio at the 97.5% level and the commodity portfolio at the 99% level.

	Normal			Student t			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	0.39	0.41	0.44	0.41	0.47	0.39	0.46	0.52	0.51
Equity	0.06	0.71	0.66	0.40	0.55	0.07	0.14	0.13	0.09
Interest rate	0.13	0.05	0.48	0.17	0.12	0.35	0.36	0.08	0.51
Commodity	0.06	0.43	0.19	0.19	0.38	0.90	0.08	0.30	0.12
TB	0.02	0.40	0.20	0.45	0.50	0.09	0.14	0.18	0.10
q = 0.99									
FX	0.42	0.41	0.49	0.35	0.38	0.33	0.41	0.59	0.47
Equity	0.00	0.49	0.67	0.72	0.28	0.44	0.00	0.10	0.07
Interest rate	0.14	0.01	0.55	0.39	0.21	0.22	0.11	0.01	0.47
Commodity	0.11	0.31	0.44	0.22	0.40	0.15	0.13	0.06	0.14
TB	0.00	0.41	0.69	0.76	0.28	0.09	0.00	0.13	0.10

Table 18: COVID-19 Outbreak ER p-values. P-values marked in gray indicate rejection of H0: “the model is correct”. The best p-values are marked by a box.

4.4 Post COVID-19 Outbreak

Tables 19 and 20 state the average performance statistics from the post COVID-19 backtesting period at the 97.5% and 99% levels, respectively. The $nEWMA_R$ and $tEWMA_R$ failed the UC and CC tests at both levels, while the plain HS failed both tests at the 97.5% level. The $nEWMA_{ml}$ failed the UC test at both levels and only just passed the CC test at the 97.5% level. The $tEWMA_R$ is the only ES model to fail the ER test, and the only approach to fail all three tests at the 97.5% level. No ES model failed the ER test at the 99% level. The nGARCH, tGARCH, $tEWMA_{ml}$, vwHS-EWMA, and vwHS-GARCH remained satisfactory from a VaR and ES backtesting perspective at both levels. The top-performing VaR model is the tGARCH at the 97.5% level and the vwHS-EWMA at the 99% level. At both levels, the $tEWMA_{ml}$ is the top-performing ES model. Its corresponding VaR model is however close to being rejected by the average UC test. Looking at Table 19 it can be concluded that at the 97.5% level, all approaches except the $nEWMA_R$, $nEWMA_{ml}$, and $tEWMA_R$ are improvements over the plain HS. At the 99% level, no approach stands out as an obvious improvement. The t-distribution again yields the lowest statistic at both confidence levels for both the VaR and ES models.

q=0.975		UC ∑LR$_{UC}$	∑∑LR$_{UC}$	CC ∑LR$_{CC}$	∑∑LR$_{CC}$	ER -2∑ln(p$_i$)	∑∑X$_{ER}$
Normal	GARCH	3.51		5.36		15.43	
	EWMA$_R$	37.67	52.14	47.48	70.86	9.14	21.20
	EWMA$_{ml}$	10.95		18.09		5.78	
t-dist.	GARCH	$\boxed{2.13}$		$\boxed{3.61}$		6.66	
	EWMA$_R$	28.96	$\boxed{41.60}$	40.00	$\boxed{49.90}$	24.71	$\boxed{11.49}$
	EWMA$_{ml}$	10.51		6.29		$\boxed{4.82}$	
HS	Plain	12.62		12.57		5.07	
	vwEWMA	2.65		8.59		7.08	
	vwGARCH	3.06		4.31		11.81	

Table 19: Post COVID-19 Outbreak average performance statistics, 97.5% level. Time period: 01-10-2020 to 31-12-2021. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

q=0.99		UC ∑LR$_{UC}$	∑∑LR$_{UC}$	CC ∑LR$_{CC}$	∑∑LR$_{CC}$	ER -2∑ln(p$_i$)	∑∑X$_{ER}$
Normal	GARCH	7.95		8.77		14.12	
	EWMA$_R$	62.67	98.74	66.48	104.27	13.52	17.60
	EWMA$_{ml}$	28.09		28.60		3.47	
t-dist.	GARCH	4.56		4.72		9.20	
	EWMA$_R$	40.23	$\boxed{55.45}$	52.79	$\boxed{63.10}$	18.30	$\boxed{11.36}$
	EWMA$_{ml}$	10.66		5.59		$\boxed{2.16}$	
HS	Plain	6.27		6.37		3.32	
	vwEWMA	$\boxed{2.89}$		$\boxed{3.41}$		6.78	
	vwGARCH	4.04		4.28		8.58	

Table 20: Post COVID-19 Outbreak average performance statistics, 99% level. Time period: 01-10-2020 to 31-12-2021. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 11.1 for the UC and CC test and 18.31 for the ER test.

4.4.1 Portfolio Level VaR Backtesting Results

Looking at tables 21 and 22, the nEWMA$_R$, which failed the average UC and CC tests, failed both tests for all individual portfolios except the interest rate portfolio. The tEWMA$_R$, which failed the average VaR tests, had an inaccurate coverage and produced clustered exceptions for the commodity and trading book portfolio at the 97.5% level, and the equity, commodity, and trading book portfolio at the 99% level. The above results showed that the plain HS failed the average UC and CC tests at the 97.5% level. Looking at tables 21 and 22, this resulted from the model failing the 97.5% UC test for the interest rate portfolio and high CC test statistics for the equity and interest rate portfolio. The plain HS is outperformed by the vwHS-EWMA, vwHS-GARCH, and tGARCH for almost all portfolios, as indicated by the lower test statistics.

	Normal			Student t			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	1.47	13.14	0.08	1.47	4.84	0.18	0.08	1.61	1.27
Equity	0.16	5.51	2.98	0.10	4.81	2.87	4.04	0.10	0.49
Interest rate	1.47	0.71	2.48	0.18	4.07	2.48	6.84	0.91	1.27
Commodity	0.04	9.29	2.41	0.31	8.52	2.71	0.83	0.03	0.03
TB	0.38	9.02	2.99	0.07	6.72	2.27	0.83	0.01	0.01
q = 0.99									
FX	0.02	19.18	1.85	0.02	5.72	0.58	0.02	0.80	0.01
Equity	2.06	16.84	5.29	2.02	9.84	0.01	2.02	0.01	2.01
Interest rate	0.58	2.87	0.80	2.19	5.86	0.02	2.19	0.58	0.01
Commodity	1.15	18.32	9.91	0.00	7.94	5.98	0.33	1.15	0.32
TB	4.14	5.49	10.25	0.33	10.87	4.08	1.71	0.36	1.70

Table 21: Post COVID-19 Outbreak UC test statistics. Test statistics marked in gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 5.02 for the 97.5% models, and 6.63 for the 99% models.

	Normal			Student t			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	1.62	13.75	0.59	1.62	6.29	0.49	0.59	2.53	1.43
Equity	0.69	8.53	4.29	0.42	4.88	1.99	4.09	0.42	0.72
Interest rate	1.62	2.15	4.30	0.49	5.98	1.30	5.86	4.92	1.43
Commodity	0.48	12.81	3.42	0.56	13.12	1.10	1.01	0.37	0.37
TB	0.95	10.24	5.41	0.52	9.73	1.42	1.01	0.36	0.36
q = 0.99									
FX	0.08	19.19	2.07	0.08	7.11	0.60	0.08	0.95	0.07
Equity	2.29	18.56	5.71	2.03	11.89	0.07	2.03	0.07	2.01
Interest rate	0.60	3.38	0.95	2.19	8.37	0.08	2.19	0.60	0.07
Commodity	1.32	20.15	9.76	0.06	11.54	3.54	0.36	1.32	0.43
TB	4.48	5.62	10.11	0.36	13.89	1.30	1.72	0.47	1.70

Table 22: Post COVID-19 Outbreak CC test statistics. Test statistics marked in light gray indicate rejection of H0: "the model is correct". The best statistics are marked by a box. The critical values are 7.38 for the 97.5% models, and 9.21 for the 99% models.

The vwHS-EWMA, which passed all tests in tables 19 and 20, obtained the lowest UC and CC statistics for the equity, commodity, and trading book portfolio at the 97.5% level, and the lowest UC and CC statistics for the equity portfolio at the 99% level. The vwHS-GARCH, which also passed all three average tests, obtained the lowest UC and CC statistics for the commodity and trading book portfolio at the 97.5% level, and the lowest statistics for the FX and interest rate portfolio at the 99% level.

4.2.2 Portfolio Level ES Backtesting Results

In tables 19 and 20, only the 97.5% $tEWMA_R$ failed the average ER test. Looking at Table 23, this is a result of the 97.5% $tEWMA_R$ failing the ER test for the FX and interest rate portfolios. At the 99% level, the $tEWMA_{ml}$ was the only model to fail the ER test for a portfolio, with that being the FX portfolio. Though the vwHS-EWMA and vwHS-GARCH VaR models performed better for many of the portfolios than the VaR HS, the plain HS ES model outperforms the ES vwHS-EWMA and the wvHS-GARCH for all portfolios except the equity portfolio at the 97.5% level.

	Normal			Student t			HS		
	GARCH	$EWMA_R$	$EWMA_{ml}$	GARCH	$EWMA_R$	$EWMA_{ml}$	Plain	vw EWMA	vw GARCH
q = 0.975									
FX	0.09	0.06	0.25	0.19	0.02	0.21	$\boxed{0.64}$	0.50	0.19
Equity	0.34	$\boxed{0.97}$	$\boxed{0.97}$	0.88	0.28	0.98	0.48	0.56	0.62
Interest rate	0.73	0.58	0.74	$\boxed{1.00}$	0.01	0.88	0.64	0.90	0.19
Commodity	0.10	0.32	0.33	0.24	0.32	0.51	$\boxed{0.64}$	0.24	0.21
TB	0.20	$\boxed{0.96}$	0.94	0.89	0.16	0.97	0.63	0.48	0.58
q = 0.99									
FX	0.09	0.03	0.29	0.15	0.09	0.00	0.56	0.48	0.21
Equity	0.54	0.88	0.99	$\boxed{1.00}$	0.32	0.87	$\boxed{1.00}$	0.45	$\boxed{1.00}$
Interest rate	0.60	0.48	0.86	$\boxed{1.00}$	0.07	0.71	$\boxed{1.00}$	$\boxed{1.00}$	0.21
Commodity	0.06	0.11	$\boxed{0.72}$	0.11	0.22	0.59	0.34	0.30	0.31
TB	0.49	0.83	0.99	0.61	0.24	0.93	$\boxed{1.00}$	0.52	$\boxed{1.00}$

Table 23: Post COVID-19 Outbreak ER p-values. P-values marked in gray indicate rejection of H0: "the model is correct". The best p-values are marked by a box.

4.5 Jarque-Bera Test and Estimated Parameters

Jarque-Bera is a test used to determine whether a given dataset has skewness and kurtosis that matches normality. The Null hypothesis is that the data fits the description of a normal distribution. Looking at Table 24, the null hypothesis is rejected for all portfolios, during all four backtesting periods. This explains why the t-innovated models generally performed better than the models assuming normal innovations.

	Financial Crisis		Bull Market		COVID-19 Outbreak		Post COVID-19	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
Equity	55.59	8.49E-13	583.25	2.20E-16	109.62	2.20E-16	62.63	2.51E-14
FX	3045.60	2.20E-16	16570563	2.20E-16	20107	2.20E-16	4529	2.20E-16
Interest Rate	602.30	2.20E-16	59606.00	2.20E-16	982.83	2.20E-16	73.28	2.20E-16
Commodity	278.24	2.20E-16	699.94	2.20E-16	239.70	2.20E-16	325.98	2.20E-16
TB	57.03	4.14E-13	569.07	2.20E-16	95.90	2.20E-16	80.874	2.20E-16

Table 24: Jarque-Bera test for normality.

Table 25 state the estimated nGARCH coefficients for the trading book portfolio before, during, and after each of the four backtesting periods. Looking at the parameters, the estimated coefficient on the lagged squared residual, α, rose from 0.06 before the financial crisis to 0.11 throughout. This increase of α implies that the nGARCH model would have underestimated the impact of recent and sudden market volatility. The same holds for the bull market period where α rose from 0.04 to 0.10, and during the COVID-19 outbreak, when α rose from 0.10 to 0.21 in just two months before decreasing slightly, to 0.15, after the outbreak period. In line with Abboud et al. (2021), the parameters in Table 25 indicate structural market shifts during the backtesting periods. This suggests that the market conditions were driving these exceptions, rather than model misspecification alone. The estimated parameters for the other models and portfolios, which can be found in Appendix A, showed similar results.

	Date	ω	α	β	μ
	01.12.2007	1.36E-05	0.06	0.85	0
	01.07.2009	1.49E-05	0.11	0.88	0
Estimated nGARCH	01.06.2015	5.01E-06	0.04	0.92	0
parameters	20.02.2020	6.17E-06	0.10	0.81	0
	20.04.2020	6.70E-06	0.21	0.77	0
	01.10.2020	6.68E-06	0.15	0.83	0
	31.12.2021	1.66E-06	7.40E-13	0.98	0

Table 25: Mle estimated nGARCH parameters, trading book portfolio.

5 Discussion and Further Research

5.1 The plain HS

The results show that the predominant simulation approach among banks, the plain HS, only remained satisfactory from a VaR and ES backtesting perspective at the 99% level during the post COVID-19 backtesting period. This makes it one of the worst performers, together with the nGARCH, $nEWMA_R$, and $tEWMA_R$. The results further show that the tGARCH and vwHS-EWMA are indeed capable of addressing a wider variety of market conditions and thereby improve risk management compared to the widely used plain HS.

This might not be too surprising as the plain HS with a rolling window of 250 days by construction would struggle to keep up with sudden volatility spikes during the stressed periods. What might seem more surprising is that even some of the approaches that are designed to account for time-varying volatility failed to accurately reflect the risk of loss in many cases. This can be partly explained by the estimated nGARCH parameters in Table 25, which suggested that structural market changes likely contributed to the large number of exceptions incurred by some of the volatility modeling approaches. Model misspecification is another possible explanation, which will be returned to later. The finding of possible structural changes driving these backtesting exceptions underlines the importance of regulatory tools allowed by the FRTB. That is, flexibilities in assigning the backtesting multiplier and in allowing banks to continue to use IMA even if they fail several backtests during extraordinary circumstances are important tools to safeguard against capital procyclicality.

Though the plain HS ES model only failed the average performance 99% ES backtest during the financial crisis, its average performance ER statistic was rather high in several other cases. Moreover, during three of the four periods, the plain HS ES failed at least one backtest at the portfolio level, all of which were associated with the equity and trading book portfolio. During the financial crisis, the plain HS ES model failed at the 97.5% level for both these portfolios. The model was rejected for the equity portfolio at the 99% level during the bull market period. During the COVID-19 outbreak, the model was rejected for both portfolios at the 99% level. Yamai and Yoshiba's (2002d) finding that even ES has tail risk under extreme value distributions, could indicate that the plain HS ES model inaccurately measured the risk of loss due to these portfolios having more fat-tailed properties. However, the descriptive statistics in Table 3 stated that the equity and trading book portfolio did not have the highest kurtosis, skewness, or range among the portfolios in any of the periods. Yet, the combination of these estimates could have impacted the estimations. For instance, the equity and trading book portfolios had the highest measured sd during the financial crisis. In addition, the equity portfolio had the second-highest left skew, and the trading book portfolio had the second-highest range during this period. It is not obvious exactly which characteristics of the equity and trading book portfolio led to the plain HS ES model performing worse for these portfolios. This finding does however align with Hansen and Lunde (2001) as it suggests that the distributional risk properties of different portfolios affect the efficiency of models and that different models do not fit different datasets equally well. In practice, this implies that one must be careful when interpreting outcomes from various models. This further upholds Yamai and Yoshiba's (2002d) argument that financial risk management should not depend entirely on VaR or ES, as both risk measures could suffer from inaccurate risk estimates under certain conditions. A combined approach is likely to be more sophisticated than either alone.

5.2 Most Accurate VaR Model: vwHS-EWMA

With the implementation of the FRTB, banks will be obliged to use the 97.5% ES for capital purposes together with backtests on the corresponding 97.5% and 99% VaR. The only VaR model that passed the average performance UC and CC tests during the four periods was the vwHS-EWMA. According to the average performance CC tests, the model only produced clustered exceptions during the COVID-19 outbreak at the 99% level. In all other cases, the model satisfied the requirements of having approximately the expected number of VaR exceedances as well as the exceedances being independent of each other.

However, the vwHS-EWMA ES model was close to being rejected during the financial crisis and the COVID-19 outbreak while its VaR model performed well. Furthermore, the $nEWMA_{ml}$ and $tEWMA_{ml}$ were the overall best performing ES models while their corresponding VaR models performed poorly during all backtesting periods. This implies that when implementing the FRTB, banks will face a trade-off between selecting an accurate ES model at the cost of less accurate VaR model; and selecting a VaR model with good backtesting properties at the cost of a less accurate ES model. As banks must backtest their ES models based on VaR, this could potentially discourage banks from implementing the most accurate ES models, which could result in wrong capital numbers.

The portfolio level results reveal that the high average performance statistic of the ES vwHS-EWMA during the financial crisis follows from the low, though not rejected, p-value the model obtained for the FX and commodity portfolio. During the COVID-19 outbreak, the high average performance statistic of the ES vwHS-EWMA follows from the model obtaining low, though not rejected, p-values for the interest rate portfolio at the 97.5% level. At the 99% level, the model was rejected for the interest rate portfolio and obtained low p-values for the equity and commodity portfolios. The descriptive statistics in Table 3 stated that the FX portfolio had the highest kurtosis, and the commodity portfolio had the highest range during the financial crisis. During the COVID-19 outbreak, the commodity portfolio had the highest sd and range among the portfolios, while the interest rate portfolio had the next highest range. For the FX, interest rate, and commodity portfolio, the poor performance of the ES vwHS-EWMA can hence be explained by the characteristics of these portfolios. For the equity portfolio, the descriptive statistics do not point towards a certain characteristic but could be due to a combination of factors.

Another possible explanation for why the ES vwHS-EWMA model performed worse than its corresponding VaR model is Yamai and Yoshiba's (2002b) finding that ES estimates may not be as accurate as estimates of VaR when the datasets have less than 1000 data points. In the analysis, a rolling window of the past 250 trading days was used to estimate VaR and ES. It cannot be ruled out that this might have affected the results. Furthermore, Yamai and Yoshiba's (2002b) finding of ES having larger estimation errors at higher confidence levels could explain why the ES vwHS-EWMA failed for more portfolios at the 99% level than at the 97.5% level.

5.3 Top Performers: vwHS-EWMA and tGARCH

Regardless of its ES model obtaining low p-values in some cases, the vwHS-EWMA is not only the top-performing VaR model but also one of the two overall top performers during the four periods. The other top performer is the tGARCH. That is, in most of the cases, these models were satisfactory from a VaR backtesting perspective and an ES backtesting perspective. During the bull market period, the vwHS-EMWA VaR model was rejected by the average performance UC and CC tests at the 99% level. During the COVID-19 outbreak at the 99% level, the ES model was rejected while its VaR model had a correct coverage but produced clustered exceptions. In the other cases, the model managed to stay satisfactory from a VaR and ES backtesting perspective. The tGARCH had its VaR model rejected during the COVID-19 outbreak at both confidence levels while its corresponding ES model passed the backtests during all four periods at both confidence levels. In fact, even when looking at the portfolio level during the four periods, the ES tGARCH model did not fail a single ER test at any confidence level for any of the individual portfolios. This makes the tGARCH one of the top-performing ES models as well as one of the overall top performers. The only other ES model that did not fail a single ES backtest, neither overall nor at the portfolio level, was the $nEWMA_{ml}$. However, as previously stated, the $nEWMA_{ml}$ VaR model did yield poor results and would hence not be a favorable choice for a bank that will need to evaluate its ES model based on the corresponding VaR model.

5.4 Worst Performers: plain HS, nGARCH, $nEWMA_R$, $tEWMA_R$

The models with the highest frequency of failed VaR and ES backtests during the four periods were the plain HS, nGARCH, $nEWMA_R$, and $tEWMA_R$. As previously discussed, the fact that the volatility modeling nGARCH, $nEWMA_R$, and $tEWMA_R$ ended up as the worst performers could be due to the changing structure of the market during the backtesting periods. Moreover, both the nGARCH and $EWMA_R$ models could potentially be subject to different sources of misspecification. For instance, the specification of the variance equation and the distribution used to build the maximum estimation may be wrong, and the standardized residuals may not be i.i.d. As indicated by the Jarque-Bera test in Table 24 and descriptive statistics in Table 3, the t-distribution is a better fit for all datasets than the normal distribution. This explains why the two of the four worst performers belong to the parametric models which assume a normal distribution. Contrary to the Jarque-Bera results, the normal distribution was a better fit among the VaR models during the financial crisis, according to the $\sum\sum LR_{UC}$ and $\sum\sum LR_{CC}$ statistics. A surprising result, as one would expect the t-distribution to yield better results during market stress. The result however owes to the poor performance of the $tEWMA_R$ VaR model during this period, which could have suffered from misspecification. Another possible explanation as to why the $nEWMA_R$ and $tEWMA_R$ represent two of the worst performers is how the models had to be specified in R due to the limitations of the `rugarch` package. As all parameters could not be fixed the $EWMA_R$ had to assume a non-zero mean when λ was set to a fixed value. This could have affected the results as the descriptive statistics in Table 3 clearly stated a mean of zero, or very close to zero, for all portfolios during all periods. It cannot be ruled out that the $nEWMA_R$ and $tEWMA_R$ would have performed better, had the mean not been estimated by mle.

5.5 Further Research

Only long positions, i.e., left tail risk, were considered in the analysis. VaR and ES resulting from short positions, i.e., right tail risk, were not considered. As the findings of this paper support Hansen and Lunde's (2001) conclusion that different models do not fit different datasets equally well, one might reasonably assume that models that fit well to left tail risk may not work equally well for right tail risk. Extending this study by considering both long and short positions is hence encouraged.

This study considered only two stylized facts of financial returns: fat tails and volatility clustering. Another observed characteristic that was not considered is the leverage effect. I.e., the volatility of financial returns data tends to be higher when previous returns have been negative. The asymmetric cousins of the GARCH model, like the EGARCH and GJR-GARCH which account for the leverage effect, were hence not considered. The results further suggest that accounting for the excess kurtosis typically observed in financial returns does improve the accuracy of the parametric models compared to the Gaussian distribution. Modeling the underlying loss distribution with various leptokurtic properties or estimating the degrees of freedom by mle may hence be of interest. Furthermore, as the descriptive statistics showed that the data exhibited high skewness during the COVID-19 pandemic, applying the skewed t-distribution is another option. An extension of this study with the left-out models and specifications is left for further research.

As this thesis considered the VaR-based backtesting constraint of the FRTB when interpreting the results, another important issue regarding ES for capital purposes has not been properly addressed. That is, considering the regulatory requirements banks are constrained by, there are good arguments against using volatility responsive approaches. In particular, Laurent and Firouzi (2017) found that vwHS-EWMA has good backtesting properties but at the cost of increased capital numbers under the FRTB due to ES being stress calibrated. Laurent and Firouzi argue that this could potentially incentivize banks to stick to the less accurate HS approach rather than implementing the more accurate vwHS-EWMA. To further investigate the practical implications of the FRTB, it would hence be a valuable exercise to extend this study by following a similar approach to that of Laurent and Firouzi (2017) and comparing the estimated capital requirements of the two top performers in this thesis, the tGARCH and vwHS-EWMA, to plain HS under different market conditions.

6 Conclusion

With the implementation of the FRTB, banks using IMA will be obliged to calculate their market risk capital requirements based on ES, rather than VaR. However, backtesting will still be VaR-based. As the FRTB is designed to address the issues arising from the financial crisis, this thesis investigated whether volatility modeling of VaR and ES can improve banks' risk management compared to the prevalent HS approach under various market conditions. The accuracy of corresponding VaR and ES models was further compared to inform whether VaR-based backtesting requirements might incentivize banks to choose certain model specifications.

Three major implications can be drawn from the analysis. Firstly, the predominant plain HS, which was one of the worst-performing approaches, is outperformed by vwHS using an EWMA filter and GARCH(1,1) applying t-innovations with four degrees of freedom under all market conditions. However, even some of the approaches that are designed to account for time-varying volatility suffered failed backtests in many cases. The finding of possible structural market changes being the main driver of these backtesting exceptions underlines the importance of regulatory tools allowed by the FRTB. That is, during extraordinary circumstances, regulatory flexibilities in assigning the backtesting multiplier and in allowing banks to continue to use IMA even if they fail several backtests are important tools to safeguard against capital procyclicality.

Secondly, even though the ES models performed better than their corresponding VaR models in most cases, there were some cases where the ES models performed worse than their corresponding VaR models. While there does not seem to be a straightforward criterion for selecting proper models, the results suggest that the distributional risk properties of different portfolios affect the efficiency of models and that both VaR and ES could suffer from inaccurate risk estimates under certain conditions. In practice, this implies that one must be careful when interpreting outcomes from various models and that financial risk management should not depend on either risk measure alone. A combined approach is likely to be more sophisticated.

Thirdly, while the transition from VAR to ES for capital calculation ensures more conservative risk estimates and reduces capital procyclicality, the results suggest that VaR-based backtesting might provide wrong incentives. Though the vwHS-EWMA was the overall top-performing VaR model, its corresponding ES model obtained rather high test statistics in some cases. Furthermore, the $nEWMA_{ml}$ and $tEWMA_{ml}$ were the overall best performing ES models while their corresponding VaR models performed poorly in many cases. The results from this paper hence imply that when implementing the FRTB, banks will face a trade-off between selecting an accurate ES model at the cost of a less accurate VaR model; and selecting a VaR model with good backtesting properties at the cost of a less accurate ES model. This implies that VaR-based backtesting could potentially discourage banks from implementing the best ES models, which could result in wrong capital numbers.

References

Abboud, Alice, Chris Anderson, Aaron L. Game, Diana Iercosan, Hulusi Inanoglu, and David K. Lynch. 2021. "Banks' Backtesting Exceptions during the COVID-19 Crash: Causes and Consequences." *FEDS Notes.* July. Accessed January 2020. https://www.federalreserve.gov/econres/notes/feds-notes/banks-backtesting-exceptions-during-the-covid-19-crash-causes-and-consequences-20210708.htm.

Artzner, Philippe, Eber Jean-Marc, Freddy Delbaen, and David D. Heath. 1999. "Coherent Measures of Risk." *Mathematical Finance 9*, July: 203-228. DOI:10.1111/1467-9965.00068.

BCBS. 2013. "Consultative Document. Fundamental review of the trading book: A revised market risk framework." Basel, Switzerland: Bank for International Settlements.

BCBS. 2006. "International Convergence of Capital Measurement and Capital Standards: A Revised Framework." Basel, Switzerland: Bank for International Settlements.

—. 2022. "MAR. Calculation of RWA for Market Risk. MAR12. Definition of Trading Desk. ." Bank for International Settlements. Version effective as of 01 Jan 2023.

—. 2011. "Messages from the academic literature on risk measurement for the trading book. Working Paper No. 19." CH-4002 Basel, Switzerland. ISSN: 1561-8854: Bank for International Settlements, January 31.

—. 2020. "Press release: Governors and Heads of Supervision announce deferral of Basel III implementation to increase operational capacity of banks and supervisors to respond to Covid-19." *bis.org.* March 27. Accessed March 2022. https://www.bis.org/press/p200327.htm.

BCBS. 2009. *Revisions to the Basel II market risk framework.* Basel, Switzerland, July.

BCBS. 1996. *Supervisory Framework for the Use of "Backtesting" in Conjunction With the Internal Models Approach to Market Risk Capital Requirements.* Basel, Switzerland, January.

—. 2019. "The market risk framework in brief." January. Accessed February 2022. https://www.bis.org/bcbs/publ/d457_inbrief.pdf.

Bera, Anil K., and Matthew L. Higgins. 1993. "ARCH Models: Properties, Estimation and Testing." *Journal of Economic Surveys 7(4)*, 305-366. https://doi.org/10.1111/j.1467-6419.1993.tb00170.x.

Bollerslev, Tim. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics Volume 31, no. 3*, 307-327. https://doi.org/10.1016/0304-4076(86)90063-1.

Cao, Jin. 2022. *The Economics of Banking.* Abingdon, Oxford: Routledge, forthcoming. ISBN: 978-0-367-40571-7. DOI: 10.4324/9780429356773.

Danske Bank. 2020. *Risk Management 2020.* Pilar 3, Danske Bank. URL: https://danskebank.com/-/media/danske-bank-com/file-cloud/2021/2/risk-management-2020.pdf?rev=25603689101af84815de20ba0edf3.

Deloitte. 2020. "In defence of VaR. Risk measurement and backtesting in times of crisis." June. Accessed January 2022. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/risk/deloitte-uk-risk-in-defence-of-var.pdf.

DNB. 2020. *Risk and Capital Management Report 2020.* Pilar 3, DNB. URL: https://www.ir.dnb.no/sites/default/files/pr/202103107750-1.pdf?ts=1629356842. Accessed February 2022. https://www.ir.dnb.no/sites/default/files/pr/202103107750-1.pdf?ts=1629356842.

Dowd, Kevin. 2005. *Measuring Market Risk.* The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons. ISBN: 978-0470013038.

Edwards, Stefan McKinnon, Baptiste Auguie, Shaun Jackman, Hadley Wickham , and Winston Chang. 2020. "lemon: Freshing Up your 'ggplot2' Plots. Version 0.4.5." https://CRAN.R-project.org/package=lemon.

European Banking Authority. 2017. "EBA Report: Resukts from the 2016 Market Risk Benchmarking Exercise."

Galanos, Alexios. 2022. "rugarch: Univariate GARCH models. Version 1.4-7." https://CRAN.R-project.org/package=rugarch. Accessed January 2022. https://bitbucket.org/alexiosg.

Ghalanos, Alexios. 2022. "Introduction to the rugarch package (Version 1.4-3)." February. Accessed March 2022. https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf.

Hansen, Peter R., and Asger Lunde. 2001. "A Forecast Comparison of Volatility Models: Does anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics 20(7)*, January: 873-889. DOI:10.2139/ssrn.264571.

Hendricks, Darryll. 1996. "Evaluation of Value-at-Risk Models Using Historical Data." *Economic Policy Review, Vol. 2, No. 1*, April: 39-67. DOI: http://dx.doi.org/10.2139/ssrn.1028807.

Hull, John C. 2012. *Risk management and financial institutions.* Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN: 978-1-119-44816-7.

Hull, John, and Alan White. 2014. "Hull and White on the pros and cons of expected shortfall ." *Risk.net.* October. Accessed February 2022. https://www.risk.net/risk-management/market-risk/2375185/hull-and-white-on-the-pros-and-cons-of-expected-shortfall.

Hurst, Simon R., and Eckhard Platen. 1997. "The Marginal Distributions of Returns and Volatility." *Lecture Notes-Monograph Series 31*, 301–314. http://www.jstor.org/stable/4355986.

Karlsson, Johannes Solheim, and Henning Zakrisson. 2016. "Practical estimation of Value at Risk and Expected Shortfall: Are complex methods really necessary?" [Bachelor Thesis]. Lund University. Department of Economics , May 30.

Komsta, Lukasz, and Frederick Novomestky. 2005. "moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests. Version 0.14.1." https://CRAN.R-project.org/package=moments.

Laurent, Jean-Paul , and Hassan Omidi Firouzi. 2017. "Market Risk and Volatility Weighted Historical Simulation After Basel III." December. DOI:10.2139/ssrn.3094463.

Letmathe, Sebastian. 2022. "quarks: Simple Methods for Calculating Value at Risk and Expected Shortfall. Version 1.1.0." https://CRAN.R-project.org/package=quarks.

Markowitz, Harry M. , and Nilufer Usmen. 1996a. "The Likelihood of Various Stock Market Return Distributions, Part 1: Principles of inference." *Journal of Risk and Uncertainty 13, no. 3*, 207-219. https://doi.org/10.1007/BF00056153.

Markowitz, Harry M., and Nilufer Usmen. 1996b. "The Likelihood of Various Stock Market Return Distributions, Part 2: Empirical Results." *Journal of Risk and Uncertainty 13, no. 3*, 221–247. http://www.jstor.org/stable/41760831.

McNeil, Alexander J., and Rüdiger Frey. 2000. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach." *Journal of Empirical Finance 7(3-4)*, 271–300. DOI:10.1016/S0927-5398(00)00012-8.

McNeil, Alexander, Rüdiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management.* Princeton University Press. ISBN: 0691122555.

Nordea. 2020. *Capital and Risk Management Report 2020.* Pilar 3, Nordea. URL: https://www.nordea.com/en/doc/nordea-group-capital-and-risk-management-report-2020.pdf. https://www.nordea.com/en/doc/nordea-group-capital-and-risk-management-report-2020.pdf.

Norges Bank. 2019. "Activity in the Norwegian foreign exchange and derivatives markets in April 2019." URL: https://www.norges-bank.no/contentassets/6f9e54cfbfa54b22986aaf10dbf7b436/mms_bis_triennial_survey_2019.pdf?v=09/16/2019140703&ft=.pdf&v=09/16/2019140703&ft=.pdf.

O'Brien, James, and Paweł J. Szerszen. 2017. "An evaluation of bank measures for market risk before, during and after the financial crisis." *Journal of Banking and Finance 80*, 215–234. DOI: 10.1016/j.jbankfin.2017.03.002.

Ooms, Jeroen. 2021. "writexl: Export Data Frames to Excel 'xlsx' Format. Version 1.4.0." https://cran.r-project.org/web/packages/writexl/writexl.pdf, September.

Peterson, Brian G., and Peter Carl. 2020. "PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis R package 2.04." https://cran.r-project.org/web/packages/PerformanceAnalytics/PerformanceAnalytics.pdf, June.

Platen, Eckhard, and Renata Sidorowicz. 2007. "Empirical Evidence on Student-t Log-Returns of Diversified World Stock Indices." *Quantitative Finance Research Centre*, March: 1-23. DOI: 10.1080/15598608.2008.10411873.

RiskMetrics. 1996. *RiskMetrics - Technical Document (4).* J.P.Morgan/Reuters. URL: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Ryan, Jeffrey A., Joshua M. Ulrich, Ross Bennett, and Corwin Joy. 2020. "xts: eXtensible Time Series. Version 0.12.1." https://cran.r-project.org/web/packages/xts/xts.pdf, September.

Sharma, Meera. 2012. "The Historical Simulation Method for Value-at-Risk: A Research Based Evaluation of the Industry Favorite." ISBN: 1556-5068. DOI: 10.2139/ssrn.2042594.

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. 2021. "lubridate: Make Dealing with Dates a Little Easier. Version 1.8.0." https://CRAN.R-project.org/package=lubridate, July.

Wickham, Hadley. 2009. "ggplot2: elegant graphics for data analysis. Version 3.3.6." https://CRAN.R-project.org/package=ggplot2: Springer New York.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software, 4(43)*, 1686.

Wigglesworth, Robin. 2020. "US stocks' record bull run brought to abrupt end by coronavirus." *FinancialTimes.com.* March. Accessed February 2022. https://www.ft.com/content/6b987f46-644f-11ea-b3f3-fe4680ea68b5.

Wilke, Claus O. 2020. "cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. Version 1.1.1." https://cran.r-project.org/web/packages/cowplot/vignettes/introduction.html, December.

Yamai, Yasuhiro, and Toshinao Yoshiba. 2002d. "Comparative analyses of expected shortfall and value-at-risk under market stress." *Monetary and Economic Studies 20*, Bank of Japan. 181-237.

—. 2002b. "Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization." *Monetary and Economic Studies 20*, 87-122.

Youngman, Peter. 2009. "Procyclicality and Value at Risk." *Financial System Review*, June: 51-54.

Zhang, Y., and S. Nadarajah. 2017. "A review of backtesting for value at risk." *Communications in Statistics - Theory and Methods*, 1-24. https://doi.org/10.1080/03610926.2017.1361984.

Appendix A: Descriptive statistics & Plots

i. Portfolio Price Series

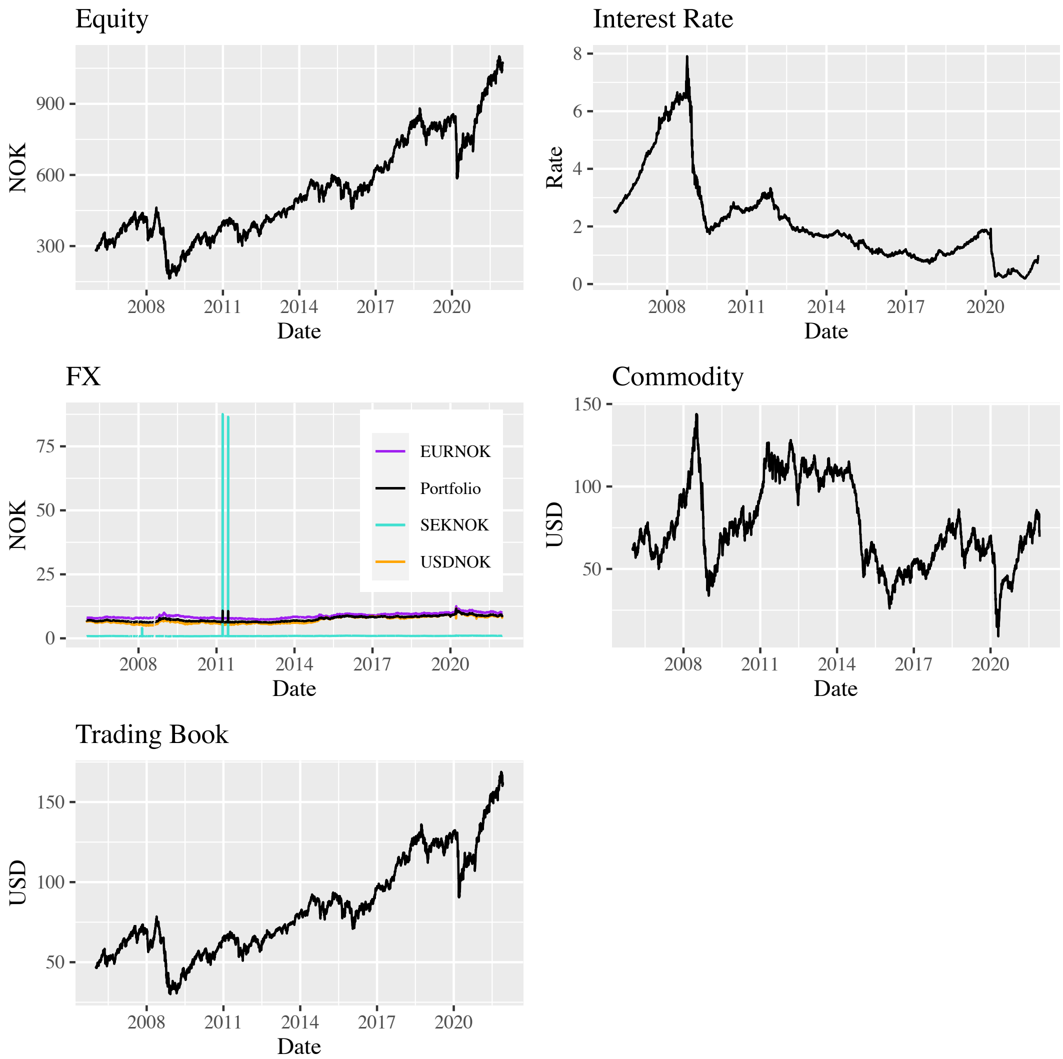

Figure 4: Portfolio price series.

ii. Pearson Correlation Coefficients

	NIBOR3M	BOND10Y	EURNOK	USDNOK	SEKNOK	OBX	Brent
NIBOR3M	1	0.84	-0.48	-0.48	-0.50	-0.55	0.37
BOND10Y	0.84	1	-0.65	-0.78	-0.59	-0.73	0.36
EURNOK	-0.48	-0.65	1	0.92	0.62	0.57	-0.62
USDNOK	-0.48	-0.78	0.92	1	0.63	0.67	-0.67
SEKNOK	-0.50	-0.59	0.62	0.63	1	0.58	-0.41
OBX	-0.55	-0.73	0.57	0.67	0.58	1	-0.28
Brent	0.37	0.36	-0.62	-0.67	-0.41	-0.28	1

Table 26: Pearson correlation coefficients.

BOND10Y denotes Norway 10-year Government Bond Yield, sourced from Investing.com.

iii. Histogram of Log Returns

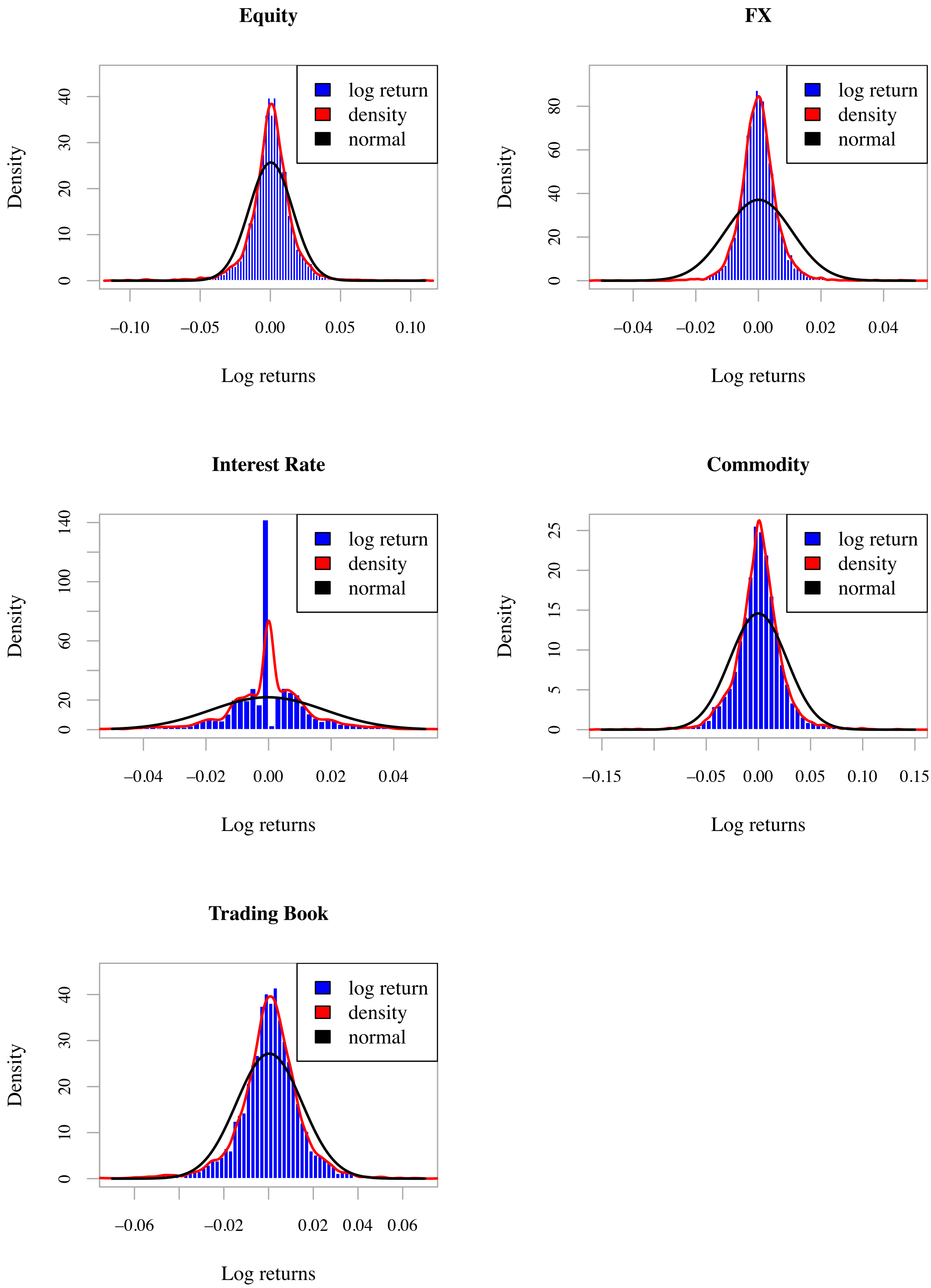

Figure 5: Histogram of log returns.

iv. Estimated GARCH & EWMA Parameters

	Date	ω	α	β	μ
nGARCH	01.12.2007	1.84E-05	0.09	0.81	-
	01.07.2009	2.41E-05	0.14	0.85	-
	01.06.2015	7.31E-05	0.29	0.10	-
	20.02.2020	6.02E-06	0.09	0.83	-
	20.04.2020	5.25E-06	0.16	0.82	-
	01.10.2020	7.04E-06	0.15	0.83	-
	31.12.2021	9.40E-08	2.24E-08	0.99	-
tGARCH	01.12.2007	2.15E-05	0.16	0.80	-
	01.07.2009	5.94E-05	0.13	0.87	-
	01.06.2015	7.28E-06	0.08	0.88	-
	20.02.2020	6.34E-06	0.14	0.83	-
	20.04.2020	5.50E-06	0.16	0.84	-
	01.10.2020	4.15E-06	0.14	0.86	-
	31.12.2021	6.71E-05	0.22	0.21	-
$nEWMA_{ml}$	01.12.2007	-	0.05	0.95	-
	01.07.2009	-	0.13	0.87	-
	01.06.2015	-	0.04	0.96	-
	20.02.2020	-	0.04	0.96	-
	20.04.2020	-	0.11	0.89	-
	01.10.2020	-	0.11	0.89	-
	31.12.2021	-	0.00	1.00	-
$tEWMA_{ml}$	01.12.2007	-	0.04	0.96	-
	01.07.2009	-	0.14	0.86	-
	01.06.2015	-	0.11	0.89	-
	20.02.2020	-	0.07	0.93	-
	20.04.2020		0.11	0.89	-
	01.10.2020	-	0.02	0.98	-
	31.12.2021	-	0.02	0.98	-
$nEWMA_{R}$	01.12.2007	-	-	-	-6.90E-04
	01.07.2009	-	-	-	3.81E-05
	01.06.2015	-	-	-	-1.53E-03
	20.02.2020	-	-	-	2.97E-03
	20.04.2020	-	-	-	1.63E-03
	01.10.2020	-	-	-	-3.07E-03
	31.12.2021	-	-	-	-3.15E-03
$tEWMA_{R}$	01.12.2007	-	-	-	1.72E-03
	01.07.2009	-	-	-	-1.26E-04
	01.06.2015	-	-	-	3.07E-04
	20.02.2020	-	-	-	9.79E-05
	20.04.2020	-	-	-	4.39E-05
	01.10.2020	-	-	-	1.68E-03
	31.12.2021	-	-	-	7.47E-04

Table 27: Equity portfolio mle estimated GARCH and EWMA parameters.

	Date	ω	α	β	μ
nGARCH	01.12.2007	3.05E-07	0.05	0.94	-
	01.07.2009	2.69E-06	0.20	0.80	-
	01.06.2015	6.93E-07	0.09	0.91	-
	20.02.2020	2.43E-05	0.11	1.10E-07	-
	20.04.2020	2.99E-05	0.95	0.05	-
	01.10.2020	5.56E-05	0.78	0.02	-
	31.12.2021	2.83E-05	0.54	0.03	-
tGARCH	01.12.2007	5.24E-07	0.04	0.94	-
	01.07.2009	5.36E-06	0.15	0.85	-
	01.06.2015	1.09E-06	0.10	0.90	-
	20.02.2020	1.89E-05	0.15	0.01	-
	20.04.2020	1.40E-05	0.81	0.15	-
	01.10.2020	3.87E-05	0.59	0.12	-
	31.12.2021	8.11E-07	2.99E-07	9.80E-01	-
$nEWMA_{ml}$	01.12.2007	-	0.04	0.96	-
	01.07.2009	-	0.15	0.85	-
	01.06.2015	-	0.05	0.95	-
	20.02.2020	-	0.10	0.90	-
	20.04.2020	-	0.26	0.74	-
	01.10.2020	-	0.09	0.91	-
	31.12.2021	-	0.06	0.94	-
$tEWMA_{ml}$	01.12.2007	-	0.05	0.93	-
	01.07.2009	-	0.14	0.84	-
	01.06.2015	-	0.03	0.95	-
	20.02.2020	-	0.13	0.84	-
	20.04.2020	-	0.24	0.73	-
	01.10.2020	-	0.37	0.60	-
	31.12.2021	-	4.10E-08	1.00	-
$nEWMA_R$	01.12.2007	-	-	-	2.44E-03
	01.07.2009	-	-	-	-9.68E-05
	01.06.2015	-	-	-	-2.51E-03
	20.02.2020	-	-	-	3.74E-03
	20.04.2020	-	-	-	1.44E-03
	01.10.2020	-	-	-	2.62E-03
	31.12.2021	-	-	-	1.39E-03
$tEWMA_R$	01.12.2007	-	-	-	-5.45E-04
	01.07.2009	-	-	-	3.35E-04
	01.06.2015	-	-	-	3.41E-04
	20.02.2020	-	-	-	3.87E-04
	20.04.2020	-	-	-	1.57E-04
	01.10.2020	-	-	-	-9.65E-05
	31.12.2021	-	-	-	-2.29E-04

Table 28: FX portfolio mle estimated GARCH and EWMA parameters.

	Date	ω	α	β	μ
nGARCH	01.12.2007	2.51E-06	0.06	0.99	-
	01.07.2009	2.97E-05	0,06	0.91	-
	01.06.2015	3.15E-05	0.54	0.43	-
	20.02.2020	2.47E-07	6.85E-09	0.99	-
	20.04.2020	8.69E-06	0.25	0.75	-
	01.10.2020	4.95E-06	0.24	0.76	-
	31.12.2021	3.89E-07	1.09E-08	1.00	-
tGARCH	01.12.2007	1.09E-06	0.07	0.91	-
	01.07.2009	4.06E-06	0.12	0.88	-
	01.06.2015	9.07E-06	0.02	0.91	-
	20.02.2020	1.06E-07	3.16E-08	1.00	-
	20.04.2020	1.17E-05	0.28	0.72	-
	01.10.2020	1.57E-05	0.35	0.65	-
	31.12.2021	4.36E-04	0.21	0.16	-
$nEWMA_{ml}$	01.12.2007	-	0.05	0.95	-
	01.07.2009	-	0.03	0.97	-
	01.06.2015	-	0.28	0.72	-
	20.02.2020	-	0.05	0.95	-
	20.04.2020	-	0.19	0.81	-
	01.10.2020	-	0.22	0.78	-
	31.12.2021	-	4.16E-08	1	-
$tEWMA_{ml}$	01.12.2007	-	0.07	0.94	-
	01.07.2009	-	0.03	0.96	-
	01.06.2015	-	0.29	0.70	-
	20.02.2020	-	0.04	0.96	-
	20.04.2020	-	0.20	0.82	-
	01.10.2020	-	0.23	0.79	-
	31.12.2021	-	4.28E-08	1	-
$nEWMA_R$	01.12.2007	-	-	-	-3.58E-03
	01.07.2009	-	-	-	-1.77E-02
	01.06.2015	-	-	-	-3.74E-03
	20.02.2020	-	-	-	2.58E-03
	20.04.2020	-	-	-	1.12E-03
	01.10.2020	-	-	-	-1.32E-02
	31.12.2021	-	-	-	1.82E-02
$tEWMA_R$	01.12.2007	-	-	-	1.31E-03
	01.07.2009	-	-	-	-3.14E-04
	01.06.2015	-	-	-	-2.43E-03
	20.02.2020	-	-	-	1.39E-03
	20.04.2020	-	-	-	1.13E-03
	01.10.2020	-	-	-	-2.62E-03
	31.12.2021	-	-	-	3.23E-03

Table 29: Interest rate portfolio mle estimated GARCH and EWMA parameters.

	Date	ω	α	β	μ
nGARCH	01.12.2007	2.14E-06	8.01E-10	0.99	-
	01.07.2009	1.87E-05	0.06	0.93	-
	01.06.2015	3.34E-06	0.08	0.92	-
	20.02.2020	2.82E-05	0.07	0.87	-
	20.04.2020	3.77E-05	0.19	0.81	-
	01.10.2020	3.76E-05	0.24	0.76	-
	31.12.2021	3.42E-05	0.11	0.83	-
tGARCH	01.12.2007	3.34E-05	1.00E-07	0.92	-
	01.07.2009	2.23E-05	6.47E-02	0.93	-
	01.06.2015	3.55E-06	0.08	0.92	-
	20.02.2020	2.12E-05	0.06	0.91	-
	20.04.2020	3.12E-05	0.25	0.51	-
	01.10.2020	4.44E-05	0.20	0.78	-
	31.12.2021	1.10E-04	0.22	0.58	-
$nEWMA_{ml}$	01.12.2007	-	1.64E-08	1	-
	01.07.2009	-	0.06	0.94	-
	01.06.2015	-	0.07	0.93	-
	20.02.2020	-	0.04	0.96	-
	20.04.2020	-	0.17	0.91	-
	01.10.2020	-	0.19	0.81	-
	31.12.2021	-	0.11	0.89	-
$tEWMA_{ml}$	01.12.2007	-	5.00E-02	1	-
	01.07.2009	-	0.04	0.93	-
	01.06.2015	-	0.08	0.92	-
	20.02.2020	-	0.06	0.95	-
	20.04.2020	-	0.19	0.90	-
	01.10.2020	-	0.21	0.81	-
	31.12.2021	-	0.12	0.88	-
$nEWMA_R$	01.12.2007	-	-	-	5.16E-03
	01.07.2009	-	-	-	-4.31E-02
	01.06.2015	-	-	-	-7.24E-03
	20.02.2020	-	-	-	1.11E-02
	20.04.2020	-	-	-	4.21E-02
	01.10.2020	-	-	-	8.75E-04
	31.12.2021	-	-	-	1.05E-03
$tEWMA_R$	01.12.2007	-	-	-	2.73E-03
	01.07.2009	-	-	-	1.28E-04
	01.06.2015	-	-	-	-1.16E-03
	20.02.2020	-	-	-	1.25E-03
	20.04.2020	-	-	-	2.45E-03
	01.10.2020	-	-	-	3.95E-03
	31.12.2021	-	-	-	3.04E-03

Table 30: Commodity portfolio mle estimated GARCH and EWMA parameters.

	Date	ω	α	β	μ
nGARCH	01.12.2007	1.36E-05	0.06	0.85	-
	01.07.2009	1.49E-05	0.11	0.88	-
	01.06.2015	5.01E-06	0.04	0.92	-
	20.02.2020	6.17E-06	0.10	0.81	-
	20.04.2020	6.70E-06	0.21	0.77	-
	01.10.2020	6.68E-06	0.15	0.83	-
	31.12.2021	1.66E-06	7.40E-13	0.98	-
tGARCH	01.12.2007	1.53E-05	0.12	0.84	-
	01.07.2009	3.80E-05	0.13	0.88	-
	01.06.2015	4.64E-06	0.05	0.92	-
	20.02.2020	6.26E-06	0.15	0.82	-
	20.04.2020	7.33E-06	0.21	0.79	-
	01.10.2020	3.95E-06	0.14	0.86	-
	31.12.2021	5.49E-05	0.27	0.27	-
$nEWMA_{ml}$	01.12.2007	-	0.04	0.96	-
	01.07.2009	-	0.10	0.90	-
	01.06.2015	-	0.03	0.97	-
	20.02.2020	-	0.05	0.95	-
	20.04.2020	-	0.16	0.88	-
	01.10.2020	-	0.11	0.89	-
	31.12.2021	-	3.03E-09	1	-
$tEWMA_{ml}$	01.12.2007	-	0.05	0.95	-
	01.07.2009	-	0.08	0.92	-
	01.06.2015	-	0.03	0.97	-
	20.02.2020	-	0.03	0.97	-
	20.04.2020	-	0.12	0.88	-
	01.10.2020	-	0.10	0.90	-
	31.12.2021	-	3.71E-08	1	-
$nEWMA_R$	01.12.2007	-	-	-	-1.55E-04
	01.07.2009	-	-	-	-1.07E-03
	01.06.2015	-	-	-	-8.79E-04
	20.02.2020	-	-	-	2.71E-03
	20.04.2020	-	-	-	-0.0041921
	01.10.2020	-	-	-	-4.45E-03
	31.12.2021	-	-	-	-3.26E-03
$tEWMA_R$	01.12.2007	-	-	-	1.93E-03
	01.07.2009	-	-	-	-2.48E-03
	01.06.2015	-	-	-	3.30E-04
	20.02.2020	-	-	-	3.98E-04
	20.04.2020	-	-	-	2.80E-05
	01.10.2020	-	-	-	4.01E-04
	31.12.2021	-	-	-	1.13E-03

Table 31: Trading book portfolio mle estimated GARCH and EWMA parameters.

v. Time Series of VaR & ES Estimates

a. Foreign Exchange Portfolio

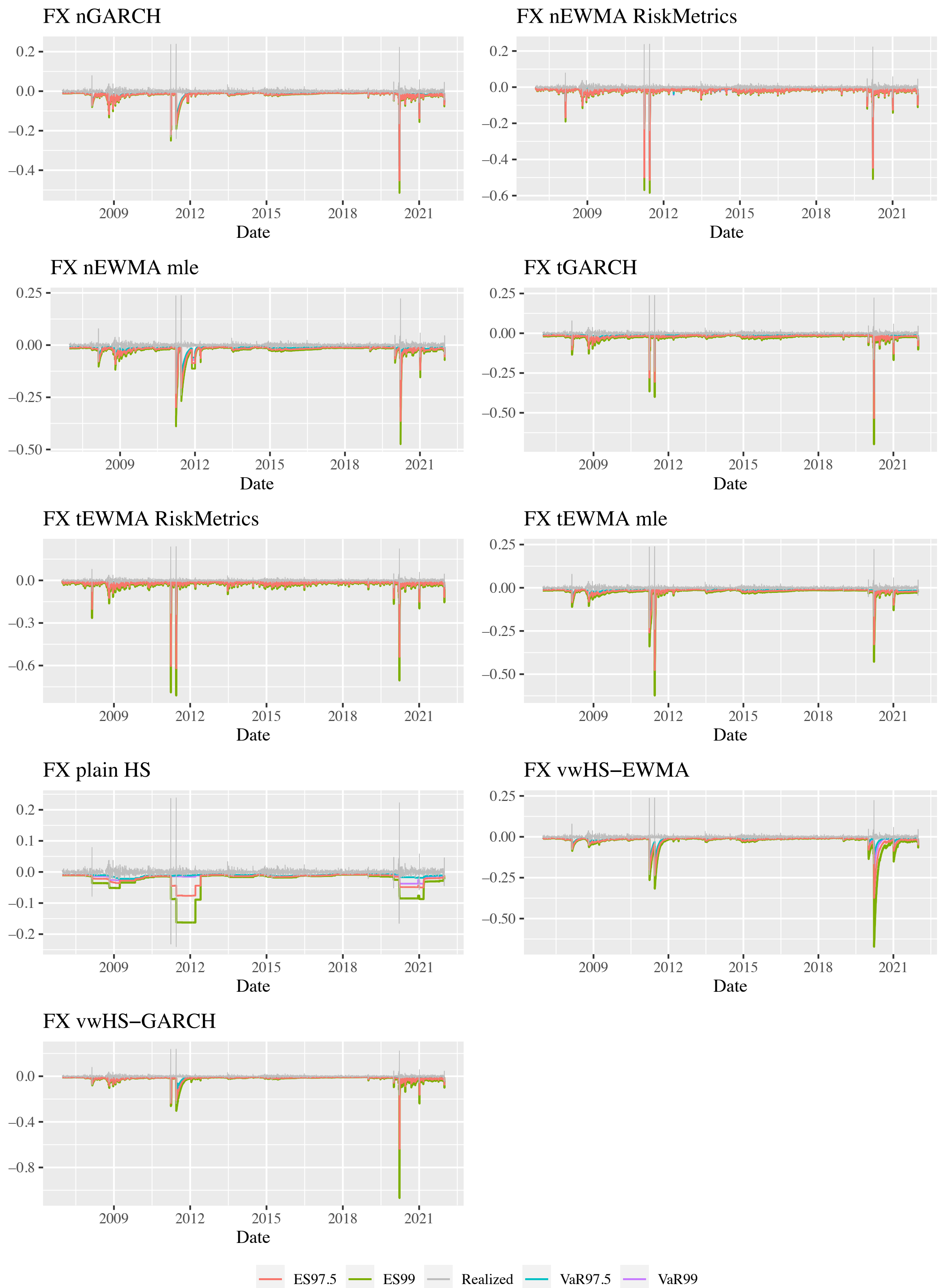

Figure 6: Foreign Exchange VaR and ES estimates.

b. Equity Portfolio

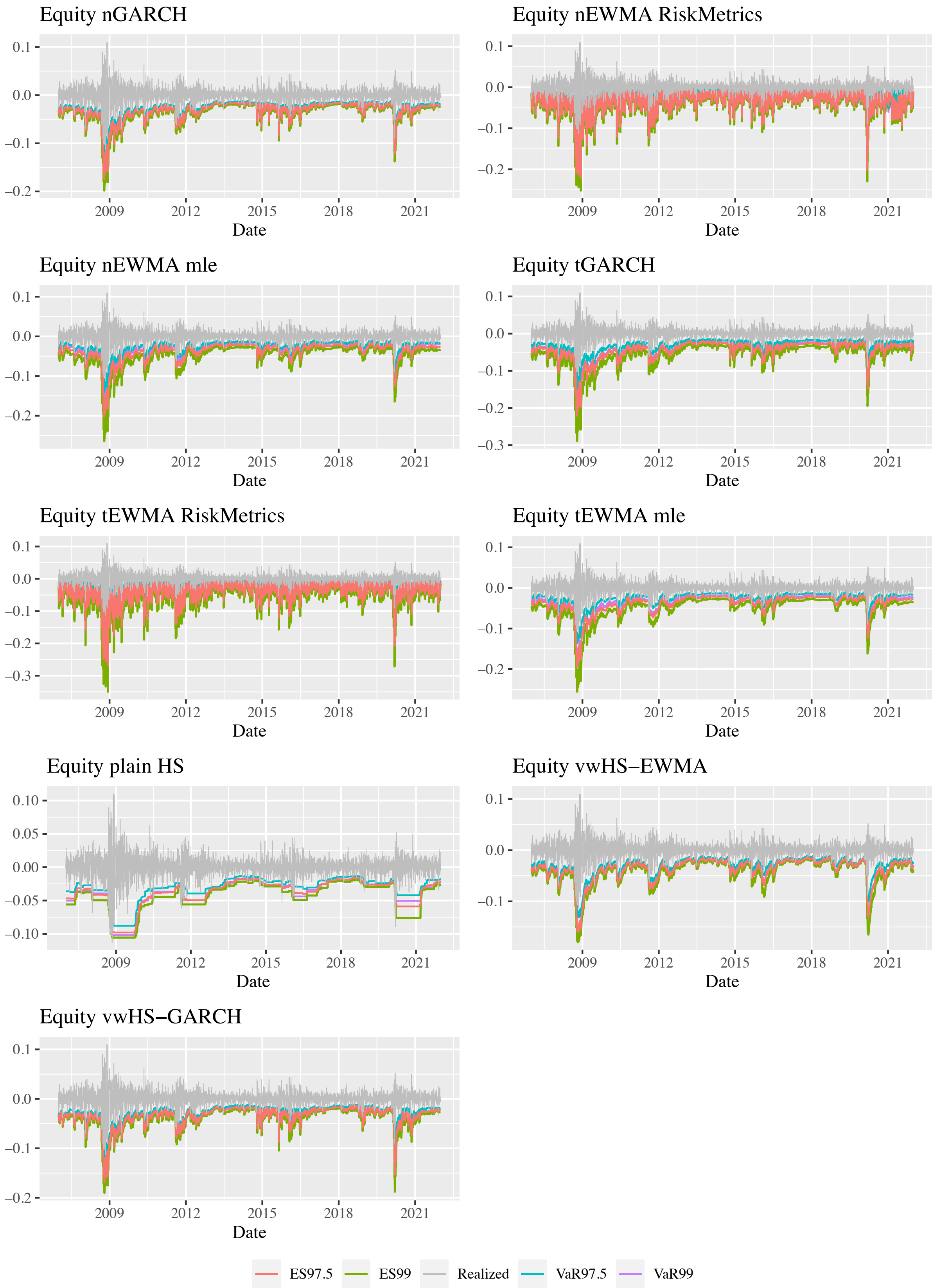

Figure 7: Equity VaR and ES estimates.

c. Commodity Portfolio

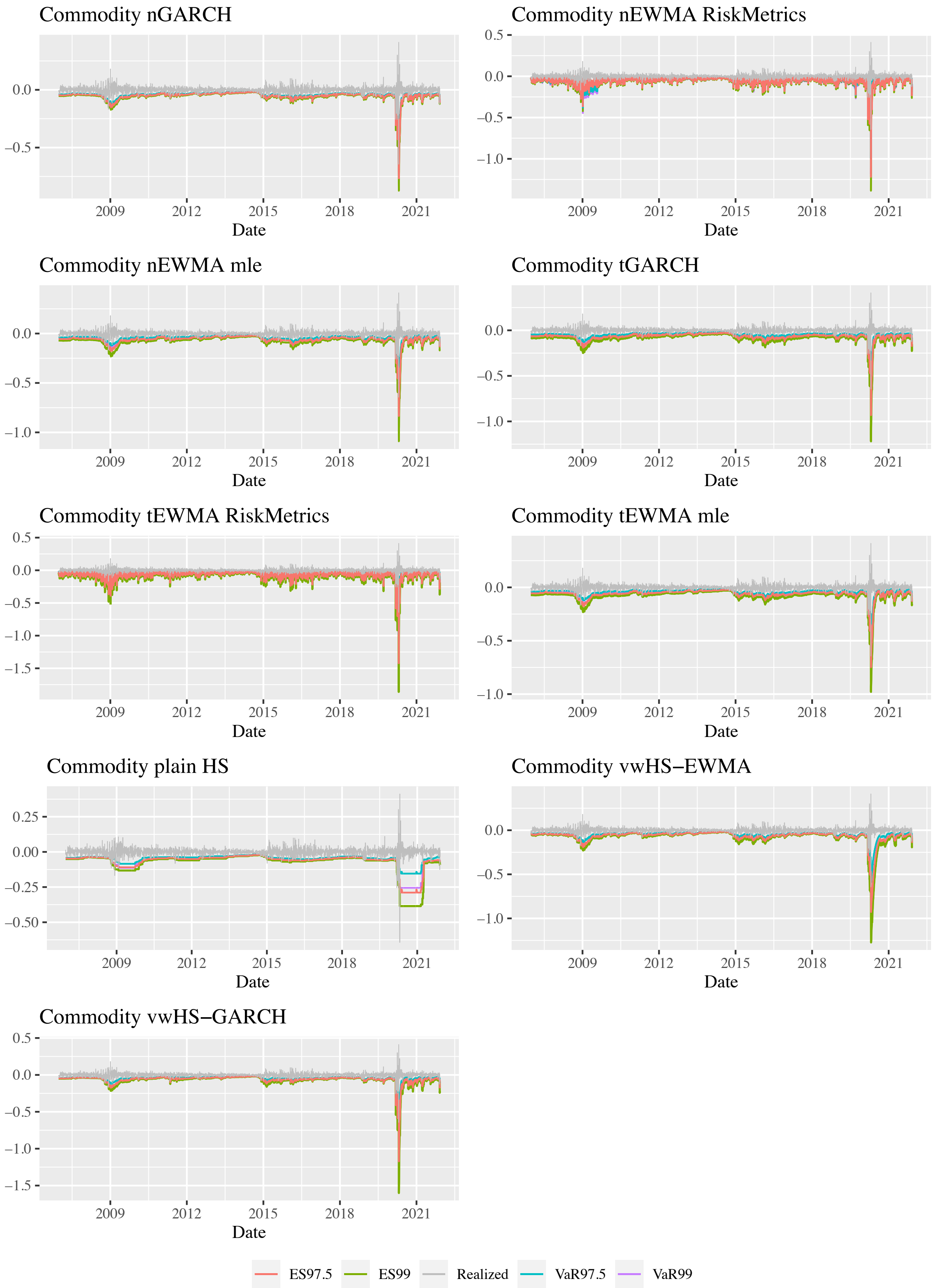

Figure 8: Commodity VaR and ES estimates.

d. Interest Rate Portfolio

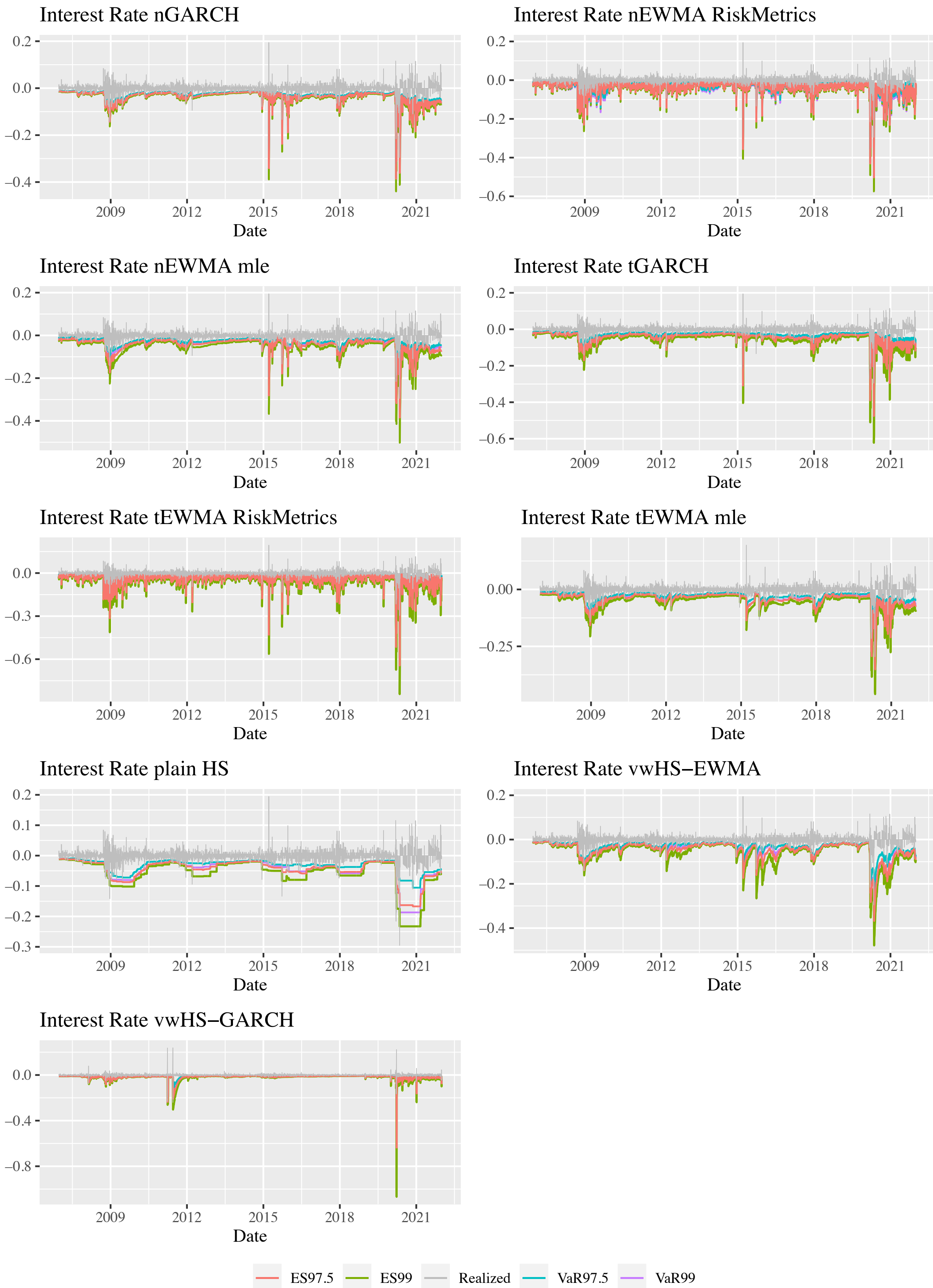

Figure 9: Interest Rate VaR and ES estimates.

e. Trading Book Portfolio

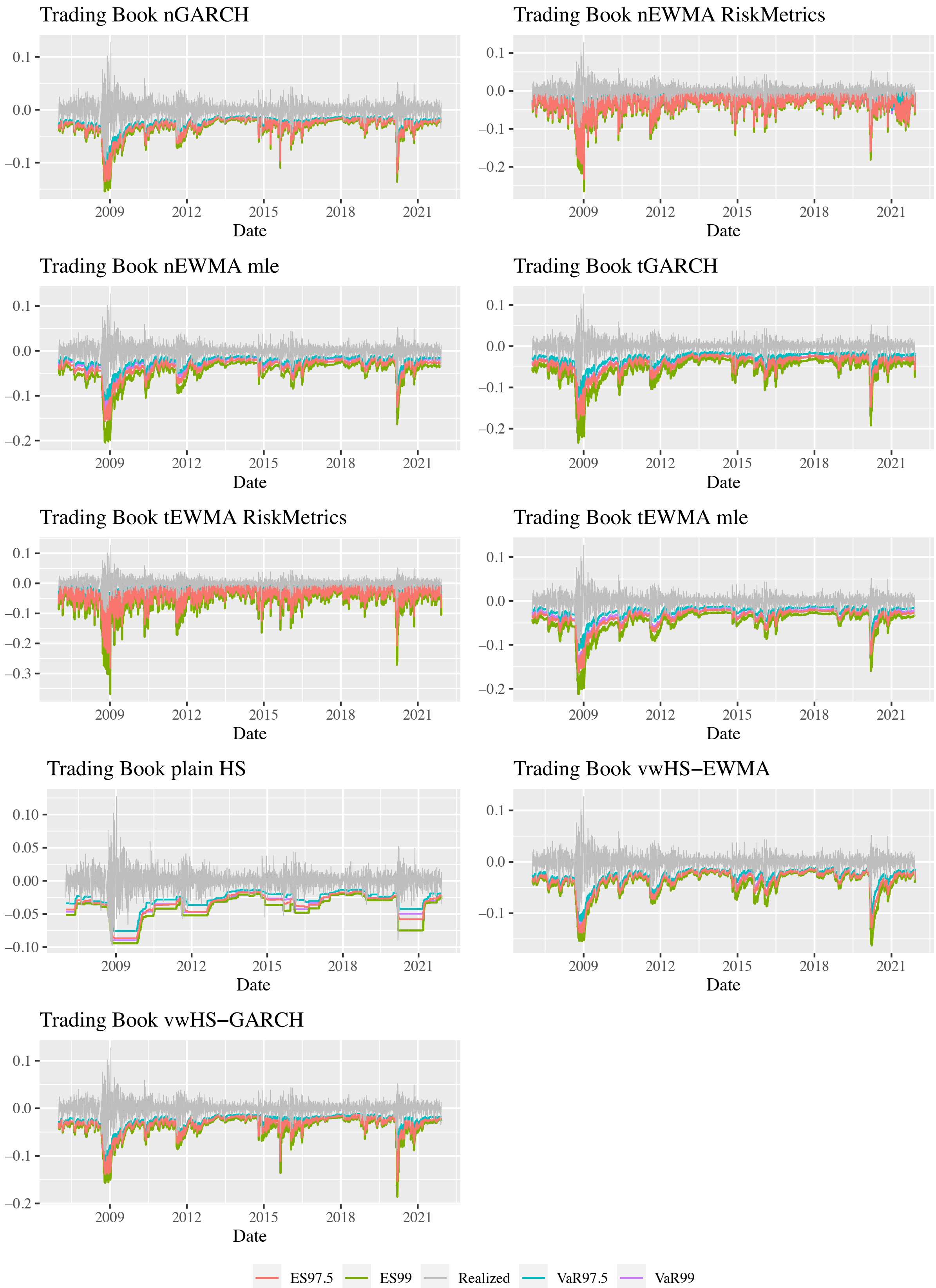

Figure 10: Trading Book VaR and ES estimates.

Appendix B: R-script

i. Implemented Functions

```
################## PREPARATIONS ####################
library(tidyverse)
library(ggplot2)
library(xts)
library(lubridate)
library(quarks)
library(rugarch)
library(PerformanceAnalytics)
library(moments)
library(writexl)
library(lemon)
library(cowplot)


######## HS rolling window forecast function ########

HS.method <- function(y, alpha = 0.01, WE = 250){
  # y: data frame of returns, ordered by date
  # alpha: alpha to be used for VaR and ES - Default 1%
  # WE: Estimation window for the forecast - Default 250 days
  x <- y$log_return
  n <- length(x)

  # Initialize empty VaR and ES vectors
  VaR <- rep(NA, n)
  ES <- rep(NA, n)

  # loop for the forecast
  for (i in 1:(n-WE)){

    # Sort returns for the given estimation window
    xs <- sort(x[i:(i+WE-1)])

    # Obtain the quantile position
    quant <- ceiling(alpha * length(xs))

    # Allocate forecasts in the vectors
    VaR[i+WE] <- xs[quant]
    ES[i+WE] <- mean(xs[1:quant])
  }

# drop first WE observations
  VaR <- VaR[-1:-WE]
  ES <- ES[-1:-WE]
  dates <- y$dato[-1:-WE]
  actual <- y$log_return[-1:-WE]

  # combine vectors to data frame
  hs <- data.frame(dates, VaR, ES, actual)
}


########### Parametric VaR and ES functions ##########

# VaR for a normal distribution
```

```
var.nor <- function(mean, sd, quant=c(0.975, 0.99)){
  var <- mean + sd * qnorm(p=quant)
  return(var)
}

# VaR for a Student t distribution
var.std <- function(mean, sd, quant=c(0.975, 0.99), df){
  scaling.factor <- sqrt((df-2) / df)
  var <- mean + sd * (scaling.factor * qt(p = quant, df = df) )
  return(var)
}

# ES for a normal distribution
es.nor <- function(mean, sd, quant=c(0.975, 0.99)){
  es <- mean + sd * (dnorm(x=qnorm(p=quant)) / (1-quant))
  return(es)
}

# ES for a Student t distribution
es.std <- function(mean, sd, quant=c(0.975, 0.99), df){
  scaling.factor <- sqrt((df-2)/df)
  factor1 <- dt(x=qt(p=quant, df=df), df=df) / (1-quant)
  factor2 <- (df + (qt(p=quant, df=df))^2 ) / (df-1)
  es <- mean + sd * scaling.factor * factor1 * factor2
  return(es)
}

############## Parametric ES estimation ##################

ForecastES <- function(forecastframe, quant = c(0.975, 0.99),distribution){
  # forecastframe: A ugarchroll data frame object
  # quant: quantiles to be used for ES estimates
  # distribution: distribution to be used for ES

  # number of observations
  n <- length(forecastframe$mean)

  # Initialize empty ES vectors
  ES97.5 <- rep(NA, n)
  ES99 <- rep(NA, n)

  if (distribution == "norm"){
    for (i in 1:n){
      mean <- forecastframe$mean[i]
      sd <- forecastframe$sd[i]
      ES97.5[i] <- -es.nor(mean, sd, quant = 0.975)
      ES99[i] <- -es.nor(mean, sd, quant = 0.99)
    }}
  if (distribution == "std"){
    for (i in 1:n){
      mean <- forecastframe$mean[i]
      sd <- forecastframe$sd[i]
      ES97.5[i] <- -es.std(mean, sd, quant = 0.975, df)
      ES99[i] <- -es.std(mean, sd, quant = 0.99, df)
    }}

  ES <- data.frame(ES97.5, ES99)
  return(ES)
}

############## Function for data handling ##################
```

```
# converts GARCH and EWMA ugarchroll objects to data frame
getData <- function(forecastmodel){
  mean <- forecastmodel@forecast[["density"]][["Mu"]]
  sd <- forecastmodel@forecast[["density"]][["Sigma"]]
  VaR99 <- forecastmodel@forecast[["VaR"]][["alpha(1%)"]]
  VaR97.5 <- forecastmodel@forecast[["VaR"]][["alpha(3%)"]]
  actual <- forecastmodel@forecast[["VaR"]][["realized"]]

  data <- data.frame(mean, sd, VaR99, VaR97.5, actual)
  return(data)
}


######## Functions to run backtests on all estimates #######
RunTests <- function(Subset, alpha, method, period){

  if (alpha == 0.025){
    VaRtest <- data.frame(VaRTest(alpha = 0.025, as.numeric(Subset$actual),
      as.numeric(Subset$VaR97.5), conf.level = 0.975))

    set.seed(250)
    EStest <- data.frame(ESTest(alpha = 0.025, as.numeric(Subset$actual),
                                as.numeric(Subset$ES97.5),
      as.numeric(Subset$VaR97.5), conf.level = 0.975,
                                boot = TRUE, n.boot = 10000))
      }

  else if (alpha == 0.01){
    VaRtest <- data.frame(VaRTest(alpha = 0.01, as.numeric(Subset$actual),
      as.numeric(Subset$VaR99), conf.level = 0.99))

    set.seed(250)
    EStest <- data.frame(ESTest(alpha = 0.01, as.numeric(Subset$actual),
      as.numeric(Subset$ES99), as.numeric(Subset$VaR99), conf.level = 0.99,
                                boot = TRUE, n.boot = 10000))

  }

  results <- data.frame(VaRtest, EStest)
  results <- data.frame(alpha = alpha, results)
  results <- data.frame(method = method, results)
  results <- data.frame(period = period, results)
  return(results)
}

EstimateSubset <- function(Subset, method, period) {
  estimates97.5 <- RunTests(Subset, 0.025, method, period)
  estimates99 <- RunTests(Subset, 0.01, method, period)
  E <- merge(estimates97.5, estimates99, all = TRUE)
  return(E)
}


EstimateData <- function(DataFrameEstimates, method){
  DataFrameEstimates <- DataFrameEstimates %>% na.omit()
  # subset data
  Financial.Crisis <- DataFrameEstimates %>%
    subset(dato >= "2007-12-01" & dato <="2009-06-30")

  Bull.Market <- DataFrameEstimates %>%
    subset(dato >= "2009-07-01" & dato <="2020-02-19")
```

```
  # COVID-19 outbreak
  Covid19.c <- DataFrameEstimates %>%
    subset(dato >= "2020-02-20" & dato <="2020-09-30")

  # post COVID-19 outbreak
  post.Covid19.c <- DataFrameEstimates %>%
    subset(dato >= "2020-10-01" & dato <="2021-12-31")

  Financial.Crisis <- EstimateSubset(Financial.Crisis, method=method,
      period = "Financial.Crisis")

  Bull.Market <- EstimateSubset(Bull.Market, method=method, period =
      "Bull.Market")

  Covid19.c <- EstimateSubset(Covid19.c, method=method, period =
      "Covid19.c")

  post.Covid19.c <- EstimateSubset(post.Covid19.c, method=method, period =
      "post.Covid19.c")

  results <- Reduce(function(...) merge(..., all = TRUE),
      list(Financial.Crisis, Bull.Market, Covid19.c, post.Covid19.c))
  return(results)
}


makeEstimates <- function(nGARCH, nEWMA_R, nEWMA_ml, tGARCH, tEWMA_R,
tEWMA_ml, HS, vwHSewma, vwHSgarch){
  # takes in all data frames of estimates for a trading desk

  nGARCH <- EstimateData(nGARCH, method = "nGARCH")
  nEWMA_R <- EstimateData(nEWMA_R, method = "nEWMA_R")
  nEWMA_ml <- EstimateData(nEWMA_ml, method = "nEWMA_ml")
  tGARCH <- EstimateData(tGARCH, method = "tGARCH")
  tEWMA_R <- EstimateData(tEWMA_R, method = "tEWMA_R")
  tEWMA_ml <- EstimateData(tEWMA_ml, method = "tEWMA_ml")
  HS <- EstimateData(HS, method = "HS")
  vwHSewma <- EstimateData(vwHSewma, method = "vwHSewma")
  vwHSgarch <- EstimateData(vwHSgarch, method = "vwHSgarch")

  estimates <- Reduce(function(...) merge(..., all = TRUE), list(nGARCH,
nEWMA_R, nEWMA_ml, tGARCH, tEWMA_R, tEWMA_ml, HS, vwHSewma, vwHSgarch))
  return(estimates)
}

######## Functions for descriptive statistics #########

Descriptives <- function(Subset, period){
  Period <- period
  n <- length(Subset$log_return)
  mean <- mean(Subset$log_return, na.rm = TRUE)
  sd <- sd(Subset$log_return)
  kurtosis <- kurtosis(Subset$log_return)
  skewness <- skewness(Subset$log_return)
  min <- min(Subset$log_return)
  max <- max(Subset$log_return)

  descriptives <- data.frame(Period, n, mean, sd, kurtosis, skewness, min,
      max)
  return(descriptives)
```

```
}


EstimateDescriptives <- function(Portfolio){
  Portfolio <- Portfolio %>% na.omit()
  # subset data
  Financial.Crisis <- Portfolio %>%
    subset(dato >= "2007-12-01" & dato <="2009-06-30")

  Bull.Market <- Portfolio %>%
    subset(dato >= "2009-07-01" & dato <="2020-02-19")

  # COVID-19 outbreak
  Covid19.c <- Portfolio %>%
    subset(dato >= "2020-02-20" & dato <="2020-09-30")

  # post COVID-19 outbreak
  post.Covid19.c <- Portfolio %>%
    subset(dato >= "2020-10-01" & dato <="2021-12-31")

  Financial.Crisis <- Descriptives(Financial.Crisis, "Financial.Crisis")
  Bull.Market <- Descriptives(Bull.Market, "Bull.Market")
  Covid19.c <- Descriptives(Covid19.c, "Covid19.c")
  post.Covid19.c <- Descriptives(post.Covid19.c, "post.Covid19.c")

  results <- Reduce(function(...) merge(..., all = TRUE),
      list(Financial.Crisis, Bull.Market, Covid19.c, post.Covid19.c))
  return(results)
}

EstimateAllDescriptives <- function(E, IR, FX, C, TB){

  E <- EstimateDescriptives(E)
  E <- E %>% mutate(Portfolio = rep("Equity", 4))

  IR <- EstimateDescriptives(IR)
  IR <- IR %>% mutate(Portfolio = rep("Interest Rate", 4))

  FX <- EstimateDescriptives(FX)
  FX <- FX %>% mutate(Portfolio = rep("FX", 4))

  C <- EstimateDescriptives(C)
  C <- C %>% mutate(Portfolio = rep("Commodity", 4))

  TB <- EstimateDescriptives(TB)
  TB <- TB %>% mutate(Portfolio = rep("Trading Book", 4))

  descriptives <- Reduce(function(...) merge(..., all = TRUE), list(E, IR,
      FX, C, TB))
  return(descriptives)
}
```

ii. Parametric Model Specifications

```
# GARCH(1,1) normal distribution
GARCH.spec.norm <- ugarchspec(variance.model = list(model="sGARCH",
      garchOrder = c(1,1)), mean.model= list(armaOrder = c(0,0),
      include.mean = FALSE), distribution.model = "norm")
```

```
# GARCH(1,1) student t distribution
GARCH.spec.std <- ugarchspec(variance.model = list(model="sGARCH",
      garchOrder = c(1,1)),mean.model= list(armaOrder = c(0,0),
      include.mean = FALSE), distribution.model = "std", fixed.pars =
      list(shape=df))

# RiskMetrics EWMA normal distribution
EWMA.spec.norm <- ugarchspec(variance.model = list(model = "iGARCH",
      garchOrder = c(1,1)), mean.model=list(armaOrder=c(0,0),
      include.mean=TRUE),  distribution.model = "norm", fixed.pars =
      list(alpha1=0.6, omega = 0)) # omega=0, alpha=1-λ

# RiskMetrics EWMA student t distribution, 4 df
EWMA.spec.std <- ugarchspec(variance.model = list(model = "iGARCH",
      garchOrder = c(1,1)), mean.model=list(armaOrder=c(0,0),
      include.mean=TRUE),  distribution.model = "std", fixed.pars =
      list(alpha1=0.6, shape = df, omega = 0)) # omega=0, alpha=1-λ

# EWMA mle estimation normal distribution
EWMA.spec.norm.x <- ugarchspec(variance.model = list(model = "iGARCH",
      garchOrder = c(1,1)), mean.model=list(armaOrder=c(0,0),
      include.mean=FALSE), distribution.model = "norm", fixed.pars =
      list(omega=0))

# EWMA mle estimation student t distribution, 4 df
EWMA.spec.std.x <- ugarchspec(variance.model = list(model = "iGARCH",
      garchOrder = c(1,1)), mean.model=list(armaOrder=c(0,0),
      include.mean=FALSE), distribution.model = "std", fixed.pars =
      list(omega=0, shape = df))
```

iii. Analysis

This section includes the importation and forecast estimation of the OBX dataset only, while it includes the backtest of all portfolio estimates. Complete script is available upon request.

```
######################### Variable definitions #########################
WE <- 250 # estimation window
df <- 4 # degrees freedom for t-distribution
ctrl = list(tol=1e-7, delta=1e-9) # solver for parametric forecasts
dec = "."

################ IMPORT DATA ################
setwd("~/Downloads")

# OBX price series retrieved from Investing.com 2006-2017
OBX <- read.csv("Oslo OBX Historical Data-2.csv", dec =dec,
      stringsAsFactors = FALSE)

OBX[OBX == "null"] <- NA

OBX <- OBX %>%
  select(Date, Price) %>%
  mutate(Price = as.numeric(gsub(",", "", Price))) %>%
  mutate(Date = as.Date(Date, "%b%d,%Y")) %>%
  rename(dato = Date) %>%
  arrange(dato) %>%
  na.omit()
```

```
# generate log returns column, xts object, and vector
OBX$log_returns <- log(OBX$Price) - log(lag(OBX$Price, k = 1))

OBX.log_return <- xts(OBX$log_return, order.by = OBX$dato)
OBX.log_return <- OBX.log_return[-1]
colnames(OBX.log_return) <- "Log return"

OBX.log_returns <- OBX$log_returns

#################### Plain HS ####################
# 99% VaR and ES
OBXX <- OBX %>% select(dato, log_returns)
OBX.VaRES_99 <- HS.method(OBXX, alpha = 0.01, WE = 250)

# 97.5% VaR and ES
OBX.VaRES_97.5 <- HS.method(OBXX, alpha = 0.025, WE = 250)

actual_HS <- OBX$log_returns[-1:-WE]

OBX.HS.estimates <- data.frame(OBX.VaRES_97.5, OBX.VaRES_99, actual_HS)

############# vwHS EWMA #############
OBX.vwHS.ewma.975 <- rollcast(OBX.log_return, p = 0.975,
                             model = "EWMA", method = "vwhs", lambda =
                              0.94, nout = 3766, nwin = 250)

dates.vwhs <- OBX$dato[-1:-WE]
OBX.vwHS.ewma.975.data <- data.frame(dato = dates.vwhs, VaR97.5=-
      OBX.vwHS.ewma.975[["VaR"]], ES97.5=-OBX.vwHS.ewma.975[["ES"]],
            actual = OBX.vwHS.ewma.975[["xout"]])

OBX.vwHS.ewma.99 <- rollcast(OBX.log_return, p = 0.99,
                            model = "EWMA", method = "vwhs", lambda = 0.94,
                            nout = 3766, nwin = 250)

OBX.vwHS.ewma.99.data <- data.frame(VaR99=-OBX.vwHS.ewma.99[["VaR"]],
      ES99=-OBX.vwHS.ewma.99[["ES"]])

OBX.vwHS.ewma.data <- data.frame(OBX.vwHS.ewma.975.data,
      OBX.vwHS.ewma.99.data)

############## vwHS GARCH ##############

OBX.vwHS.garch.975 <- rollcast(OBX.log_returns, p = 0.975, model = "GARCH",
      method = "vwhs", nout = 3766, nwin = 250)

OBX.vwHS.garch.975.data <- data.frame(dato=dates.vwhs,
      VaR97.5=-OBX.vwHS.garch.975[["VaR"]],
      ES97.5=-OBX.vwHS.garch.975[["ES"]],
      actual=OBX.vwHS.garch.975[["xout"]])

OBX.vwHS.garch.99 <- rollcast(OBX.log_returns, p = 0.99,
                             model = "GARCH", method = "vwhs",
                             nout = 3766, nwin = 250)

OBX.vwHS.garch.99.data <- data.frame(VaR99=-OBX.vwHS.garch.99[["VaR"]],
      ES99=-OBX.vwHS.garch.99[["ES"]])

OBX.vwHS.garch.data <- data.frame(OBX.vwHS.garch.975.data,
      OBX.vwHS.garch.99.data)
```

```
########### nGARCH ###########

OBX.GARCH.norm <- ugarchroll(GARCH.spec.norm, OBX.log_return, n.start=250,
                        refit.every = 1,
                        refit.window= "moving", solver= "hybrid",
                       calculate.VaR =TRUE, VaR.alpha=c(0.01,0.025),
                       keep.coef=TRUE, solver.control =ctrl,
                        fit.control = list(scale=1))

# dataframe with mean, sd, VaR forecasts and realized
OBX.data.GARCH.norm <- getData(OBX.GARCH.norm)

# ES forecasts
OBX.ES.GARCH.norm <- ForecastES(OBX.data.GARCH.norm, quant = c(0.975,
      0.99), distribution = "norm")

# dates from rollcast object
VaRgarch <- data.frame(OBX.GARCH.norm@forecast[["VaR"]])
datesGarch <- as.data.frame(dato=as.Date(rownames(VaRgarch)))

# merge to one data frame
OBX.data.GARCH.norm <- data.frame(datesGarch, OBX.data.GARCH.norm,
      OBX.ES.GARCH.norm)

############## tGARCH ##############

OBX.GARCH.std <- ugarchroll(GARCH.spec.std, OBX.log_return, n.start=250,
                       refit.every = 1,
                       refit.window= "moving", solver= "hybrid",
                        calculate.VaR =TRUE,
                       VaR.alpha=c(0.01,0.025), keep.coef=TRUE,
                        solver.control =ctrl,
                         fit.control = list(scale=1))

OBX.data.GARCH.std <- getData(OBX.GARCH.std)

OBX.ES.GARCH.std <- ForecastES(OBX.data.GARCH.std, quant = c(0.975, 0.99),
      distribution = "std")

VaRgarch.std <- data.frame(OBX.GARCH.std@forecast[["VaR"]])
datesGarch.std <- as.data.frame(dato=as.Date(rownames(VaRgarch.std)))

OBX.data.GARCH.std <- data.frame(datesGarch.std, OBX.data.GARCH.std,
      OBX.ES.GARCH.std)

############# nEWMA R #############

OBX.EWMA.norm <- ugarchroll(EWMA.spec.norm, OBX.log_return, n.start=250,
                        refit.every = 1,
                        refit.window= "moving", solver= "hybrid",
                       calculate.VaR =TRUE,
                        VaR.alpha=c(0.01,0.025), keep.coef=TRUE,
                       solver.control =ctrl,
                        fit.control = list(scale=1))

OBX.data.EWMA.norm <- getData(OBX.EWMA.norm)

OBX.ES.EWMA.norm <- ForecastES(OBX.data.EWMA.norm, quant = c(0.975, 0.99),
      distribution = "norm")
```

```
VaREwma.norm <- data.frame(OBX.EWMA.norm@forecast[["VaR"]])
datesEwma.norm <- as.data.frame(dato=as.Date(rownames(VaREwma.norm)))

OBX.data.EWMA.norm <- data.frame(datesEwma.norm, OBX.data.EWMA.norm,
      OBX.ES.EWMA.norm)

########### tEWMA R ###########

OBX.EWMA.std <- ugarchroll(EWMA.spec.std, OBX.log_return, n.start=250,
                      refit.every = 1,
                      refit.window= "moving", solver= "hybrid",
                       calculate.VaR =TRUE, VaR.alpha=c(0.01,0.025),
                       keep.coef=TRUE, fit.control = list(list(scale=1)),
                       solver.control =ctrl)

OBX.data.EWMA.std <- getData(OBX.EWMA.std)

OBX.ES.EWMA.std <- ForecastES(OBX.data.EWMA.std, quant = c(0.975, 0.99),
      distribution = "std")

VaREwma.std <- data.frame(OBX.EWMA.std@forecast[["VaR"]])
datesEwma.std <- as.data.frame(dato=as.Date(rownames(VaREwma.std)))

OBX.data.EWMA.std <- data.frame(datesEwma.std, OBX.data.EWMA.std,
      OBX.ES.EWMA.std)

############ nEWMA ml ############
OBX.EWMA.norm.x <- ugarchroll(EWMA.spec.norm.x, OBX.log_return,
      n.start=250, refit.every = 1, refit.window= "moving",
      solver= "hybrid", calculate.VaR =TRUE,
      VaR.alpha=c(0.01,0.025), keep.coef=TRUE, solver.control =ctrl,
      fit.control = list(scale=1))

OBX.data.EWMA.norm.x <- getData(OBX.EWMA.norm.x)

OBX.ES.EWMA.norm.x <- ForecastES(OBX.data.EWMA.norm.x, quant = c(0.975,
      0.99), distribution = "std")

VaREwma.x <- data.frame(OBX.EWMA.norm.x@forecast[["VaR"]])
datesEwma.x <- as.data.frame(dato=as.Date(rownames(VaREwma.x)))

OBX.data.EWMA.norm.x <- data.frame(datesEwma.x, OBX.data.EWMA.norm.x,
      OBX.ES.EWMA.norm.x)

############ tEWMA ml ############

OBX.EWMA.std.x <- ugarchroll(EWMA.spec.std.x, OBX.log_return, n.start=250,
                       refit.every = 1,
                       refit.window= "moving", solver= "hybrid",
                       calculate.VaR =TRUE,
                       VaR.alpha=c(0.01,0.025), keep.coef=TRUE,
                       solver.control = ctrl, fit.control = list(scale=1))

OBX.data.EWMA.std.x <- getData(OBX.EWMA.std.x)

OBX.ES.EWMA.std.x <- ForecastES(OBX.data.EWMA.std.x, quant = c(0.975,
      0.99), distribution = "std")

VaRgarch.std.x <- data.frame(OBX.EWMA.std.x@forecast[["VaR"]])
datesGarch.x <- as.data.frame(dato=as.Date(rownames(VaRgarch.std.x)))
```

```
OBX.data.EWMA.std.x <- data.frame(datesGarch.x, OBX.data.EWMA.std.x,
OBX.ES.EWMA.std.x)

########### Backtest of all model-generated forecasts ###########

TB.results <- makeEstimates(TB.data.GARCH.norm, TB.data.EWMA.norm,
      TB.data.EWMA.norm.x,
      TB.data.GARCH.std, TB.data.EWMA.std, TB.data.EWMA.std.x,
      TB.HS.estimates, TB.vwHS.ewma.data, TB.vwHS.garch.data)


OBX.results <- makeEstimates(OBX.data.GARCH.norm, OBX.data.EWMA.norm,
      OBX.data.EWMA.norm.x,
      OBX.data.GARCH.std, OBX.data.EWMA.std, OBX.data.EWMA.std.x,
      OBX.HS.estimates, OBX.vwHS.ewma.data, OBX.vwHS.garch.data)

C.results <- makeEstimates(C.data.GARCH.norm, C.data.EWMA.norm,
      C.data.EWMA.norm.x,
      C.data.GARCH.std, C.data.EWMA.std, C.data.EWMA.std.x,
      C.HS.estimates, C.vwHS.ewma.data, C.vwHS.garch.data)

FX.results <- makeEstimates(FX.data.GARCH.norm, FX.data.EWMA.norm,
      FX.data.EWMA.norm.x,
      FX.data.GARCH.std, FX.data.EWMA.std, FX.data.EWMA.std.x,
      FX.HS.estimates, FX.vwHS.ewma.data, FX.vwHS.garch.data)

IR.results <- makeEstimates(R.data.GARCH.norm, R.data.EWMA.norm,
      R.data.EWMA.norm.x,
      R.data.GARCH.std, R.data.EWMA.std, R.data.EWMA.std.x,
      R.HS.estimates, R.vwHS.ewma.data, R.vwHS.garch.data)

# Write to excel
write_xlsx(list(Equity=OBX.results, IR=IR.results, FX=FX.results,
      C=C.results, TB=TB.results),"/path/filename.xlsx")
```

iv. Plots

```
setwd("~/Path/Example")

mean_1 <- 0
sd_1 <- 1

VaR99 <- qnorm(0.99, mean = mean_1, sd = sd_1) %>% round(digits = 2)
ES99 <- ESnorm(0.99, mu = mean_1, sd = sd_1) %>% round(digits = 2)

# 99% VaR
pdf("99VaR.pdf", width = 7, height = 5)
ggplot(data.frame(x=c(-4, 4)), aes(x=x)) +
      stat_function(fun = dnorm) +
      stat_function(fun = dnorm, xlim = c(VaR99, 4), geom = "area",
      fill="blue", alpha=0.4) +
      annotate("label", x = VaR99, y = -0.025, label = "99% VaR = 2.3",
      size = 5) +
      stat_function(fun = dnorm, xlim = c(-4, VaR99), geom = "area",
      fill="green", alpha=0.4) +
      theme(legend.position='none', text=element_text(family="serif", size
      = 20)) +
```

```
      geom_segment(aes(x = VaR99, y = 0.0, yend = c(-0.01, 0.025), xend =
      VaR99)) +
      geom_text(aes(x=2.7, y=0.02, label = "a = 1%", size=5, hjust = -0.1,
      vjust = -0.2, check_overlap = TRUE, family="serif"))+
      geom_text(aes(x=0.2, y=0.02, label = "1 - a = 99%", size=5, hjust = -
      0.1, vjust = -0.2, check_overlap = TRUE, family="serif"))+
      scale_y_continuous(name = "Density") +
      scale_x_continuous(name = "Loss")
dev.off()


# 99% VaR and 99% ES
pdf("99VaRES.pdf", width = 7, height = 5)
ggplot(data.frame(x=c(-4, 4)), aes(x=x)) +
      stat_function(fun = dnorm) +
      stat_function(fun = dnorm, xlim = c(-4, VaR99), geom = "area",
      fill="blue", alpha=0.2) +
      stat_function(fun = dnorm, xlim = c(-4, ES99), geom = "area",
      fill="green", alpha=0.2) +
      geom_segment(aes(x = VaR99, y = 0.0, yend = c(-0.0, 0.1), xend =
      VaR99), color = "blue") +
      geom_segment(aes(x = ES99, y = 0.0, yend = c(-0.0, 0.05), xend =
      ES99), color = "green") +
      geom_text(aes(x=2, y=0.1, label = "99% VaR = 2.3", size=5, hjust = -
      0.1, vjust = -0.2, check_overlap = TRUE, family="serif"))+
      geom_text(aes(x=2.25, y=0.05, label = "99% ES = 2.67", size=5, hjust
      = -0.1, vjust = -0.2, check_overlap = TRUE, family="serif"))+
      theme(legend.position='none', text=element_text(family="serif", size
      = 20)) +
      scale_y_continuous(name = "Density") +
      scale_x_continuous(name = "Loss")
dev.off()


# Price series
OBXP <- OBX %>% ggplot() +
  geom_line(aes(x = dato, y = Price, group = 1)) +
  labs(x = "Date", y = "NOK", title= "Equity") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

FXP <- FX %>% ggplot() +
  geom_line(aes(x = dato, y = EURNOK, group = 1, color = "EURNOK")) +
  geom_line(aes(x = dato, y = USDNOK, group = 1, color = "USDNOK")) +
  geom_line(aes(x = dato, y = SEKNOK, group = 1, color = "SEKNOK")) +
  geom_line(aes(x = dato, y = Portfolio, group = 1, color = "Portfolio")) +
  scale_color_manual(values=c("purple", "black", "turquoise", "orange")) +
  labs(x = "Date", y = "NOK", title= "FX") +
  theme(legend.title = element_blank(), legend.position = c(0.8, 0.6),
        legend.text = element_text(size = 8),
text=element_text(family="serif")) +
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

IRP <- NIBOR %>% ggplot() +
  geom_line(aes(x = dato, y = Price, group = 1)) +
  labs(x = "Date", y = "Rate", title= "Interest Rate") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")
```

```
CP <- Brent %>% ggplot() +
  geom_line(aes(x = dato, y = Price, group = 1)) +
  labs(x = "Date", y = "USD", title= "Commodity") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

TBP <- TB %>% ggplot() +
  geom_line(aes(x = dato, y = return, group = 1)) +
  labs(x = "Date", y = "USD", title= "Trading Book") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

pdf(file="PriceSeries.pdf")
plot_grid(OBXP, IRP, FXP, CP, TBP, ncol = 2, nrow = 3, label_fontfamily =
      "serif")
dev.off()

# log returns series
OBXP <- OBX %>% ggplot() +
  geom_line(aes(x = dato, y = log_return, group = 1),color="dark orange") +
  labs(x = "Date", y = "log return", title= "Equity") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

FXP <- CURRENCY %>% ggplot() +
  geom_line(aes(x = dato, y = log_return, group = 1),color="dark orange") +
  labs(x = "Date", y = "log return", title= "FX") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

IRP <- NIBOR %>% ggplot() +
  geom_line(aes(x = dato, y = log_return, group = 1),color="dark orange")+
  labs(x = "Date", y = "Log return", title= "Interest Rate") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

CP <- Brent %>% ggplot() +
  geom_line(aes(x = dato, y = log_return, group = 1),color="dark orange") +
  labs(x = "Date", y = "Log return", title= "Commodity") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

TBP <- TB %>% ggplot() +
  geom_line(aes(x = dato, y = log_return, group = 1),color="dark orange") +
  labs(x = "Date", y = "Log return", title= "Trading Book") +
  theme(legend.title = element_blank(), text=element_text(family="serif"))
+
  scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

pdf(file="Log_ret.pdf")
plot_grid(OBXP, IRP, FXP, CP, TBP, ncol = 2, nrow = 3, label_fontfamily =
      "serif")
dev.off()

# Histogram of log returns
```

```
pdf(file="histogram.pdf", width =8, height = 11)
par(mfrow=c(3,2))
par(family = "serif", cex = 1)
chart.Histogram(OBX.log_return, methods = c("add.density", "add.normal"),
      colorset = c("blue", "red", "black"), main="Equity", xlab =
      "Log returns", ylim=c(0,45))
legend("topright", legend = c("log return", "density", "normal"), fill =
      c("blue", "red", "black"))

chart.Histogram(FX.log_return, methods = c("add.density", "add.normal"),
      colorset = c("blue", "red", "black"), main="FX", xlab = "Log
      returns", ylim=c(0,95), xlim = c(-0.05,0.05))
legend("topright", legend = c("log return", "density", "normal"), fill =
      c("blue", "red", "black"))

chart.Histogram(R.log_return, methods = c("add.density", "add.normal"),
      colorset = c("blue", "red", "black"), main="Interest Rate", xlab =
      "Log returns", ylim=c(0,140), xlim = c(-0.05, 0.05))
legend("topright", legend = c("log return", "density", "normal"), fill =
      c("blue", "red", "black"))

chart.Histogram(Brent.log_return, methods = c("add.density", "add.normal"),
      colorset = c("blue", "red", "black"), main="Commodity", xlab = "Log
      returns", ylim = c(0,26), xlim = c(-0.15,0.15))
legend("topright", legend = c("log return", "density", "normal"), fill =
      c("blue", "red", "black"))

chart.Histogram(TB.log_return, methods = c("add.density", "add.normal"),
      colorset = c("blue", "red", "black"), main="Trading Book", xlab =
      "Log returns", xlim = c(-0.07, 0.07), ylim = c(0, 45))
legend("topright", legend = c("log return", "density", "normal"), fill =
      c("blue", "red", "black"))
dev.off()


# VaR & ES estimates time series. Example: OBX portfolio
nGARCH <- OBX.data.GARCH.norm %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity nGARCH", y = "VaR, ES") +
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

nEWMAR <- OBX.data.EWMA.norm %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
```

```
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity nEWMA RiskMetrics ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

nEWMAml <- OBX.data.EWMA.norm.x %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity nEWMAml mle ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

tGARCH <- OBX.data.GARCH.std %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity tGARCH", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

tEWMAR <- OBX.data.EWMA.std %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity tEWMA RiskMetrics ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

tEWMAml <- OBX.data.EWMA.std.x %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
```

```
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity tEWMA mle ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

HSP <- OBX.HS.estimates %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= "Equity plain HS ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

vwHSewmaP <- OBX.vwHS.ewma.data %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= " Equity vwHS-EWMA ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")

vwHSgarchP <- OBX.vwHS.garch.data %>% ggplot() +
      geom_line(aes(x = dato, y = VaR99, group = 1, color = "VaR99"), size
      = 0.5) + geom_line(aes(x = dato, y = ES99, group = 1, color =
      "ES99"), size = 0.6) +
      geom_line(aes(x = dato, y = VaR97.5, group = 1, color = "VaR97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = ES97.5, group = 1, color = "ES97.5"),
      size = 0.5) +
      geom_line(aes(x = dato, y = actual, group = 1, color = "Realized"),
      color = "gray", size = 0.2) +
      theme(legend.title = element_blank(),
      text=element_text(family="serif"),axis.title.y = element_blank()) +
      labs(x = "Date", title= " Equity vwHS-GARCH ", y = "VaR, ES")+
      scale_x_date(date_breaks = "3 years" , date_labels = "%Y")


pdf(file="OBXplot.pdf", width = 8, height = 11)
grid_arrange_shared_legend(nGARCH, nEWMAR, nEWMAml, tGARCH, tEWMAR,
      tEWMAml, HSP, vwHSewmaP, vwHSgarchP, ncol = 2, nrow = 5, position =
      "bottom")
dev.off()
```