SCHOOL OF ECONOMICS AND MANAGEMENT

LUND UNIVERSITY

Master's Programme in Finance

Forecasting Value-at-Risk and Expected Shortfall

A comparison of non- and parametric methods for crude oil amidst extreme volatility

by

Otto Colbin

Yugam Sharma

NEKN02
Master's Thesis (15 credits ECTS)
June 2023
Supervisor: Anders Vilhelmsson

This page intentionally left blank.

Abstract

Practitioners primarily utilise nonparametric methods when estimating Value-at-Risk (VaR) and Expected Shortfall (ES) for computing capital requirements. However, various researchers assert that there are issues with those estimates, particularly amidst periods of market turmoil. Academia produces novel parametric methods to estimate extreme risk measures to address these deficiencies. Nevertheless, empirical findings of the discrepancies in the performance between non- and parametric estimation methods are inconclusive. Various authors discover that the nonparametric methodologies display superior backtesting results, while others demonstrate contrary results. Therefore, this thesis contrasts the backtesting performance of non- and parametric estimation methods for VaR and ES in the context of crude oil amidst a significant geopolitical event which had major implications for West Texas Intermediate (WTI) and Europe Brent: the COVID-19 pandemic. We contrast the performance of the nonparametric BHS, AWHS, and VWHS methods with the parametric Gaussian, Student's t, and conditional EVT methods. Our backtesting results demonstrate that the conditional EVT is the superior method for estimating VaR, whilst the Student's t-distribution displays the most rigorous performance in estimating ES. These results are robust across WTI and Brent crude oil and for the duration of the backtesting period. Hence, we recommend that practitioners utilise parametric methods for estimating measures of extreme risks for crude oil.

Keywords: Value-at-Risk (VaR); Expected Shortfall (ES); Nonparametric estimation methods; Parametric estimation methods; Crude oil.

Acknowledgement

We would like to extend our deepest and most sincere gratitude to our supervisor, Anders Vilhelmsson, who provided valuable feedback and guidance for this Master's thesis.

We would like to acknowledge the unwavering support from family and friends throughout our academic endeavours. Thank you!

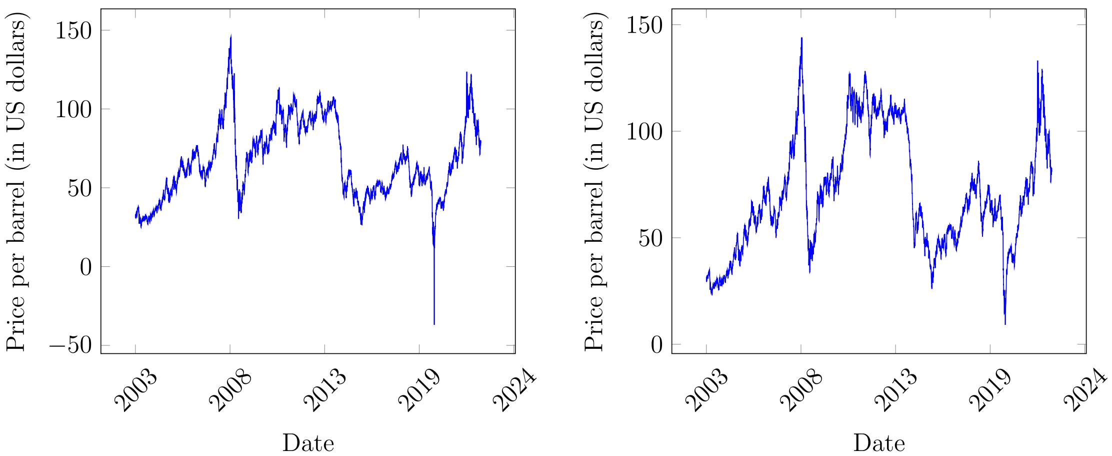
Contents

List of Figures

List of Tables

1 Introduction

Financial institutions and other practitioners ought to adopt effective estimation methods for calculating the capital requirements for market risk. Previous research on the methodology that financial institutions utilise to compute Value-at-Risk (VaR) and Expected Shortfall (ES) demonstrate that they rely on nonparametric methods such as historical simulation (Pérignon and Smith, 2010). Conversely, academia produces novel empirical findings on superior methods for estimating VaR and ES for various asset classes and predominantly accentuates the parametric estimation methods such as conditional Extreme Value Theory (EVT) (e.g., Marimoutou, Raggad and Trabelsi, 2009; Youssef, Belkacem and Mokni, 2015). Numerous studies corroborate that financial institutions are prone to produce biased estimates of VaR and ES (e.g., Pérignon, Deng and Wang, 2008; Campbell and Smith, 2022). Campbell and Smith (2022) study Australian banks and observe that practitioners are inclined to overstate estimates in calm periods and underestimate them during market turbulence, inducing potential self-inflicted solvency issues. Berkowitz and O'Brien (2002) further demonstrate that commercial banks' VaR methods perform less effectively than simple time-series models. Given these findings, it appears counterintuitive that practitioners do not embrace the vanguard of academic literature and rely on parametric estimation methods. In particular, practitioners should exert prudence when managing risk for financial assets susceptible to sudden and substantial fluctuations in volatility.



(a) Daily Spot Price of WTI *(b) Daily Spot Price of Brent*

Figure 1.1: Spot price series of WTI and Brent crude oil from 2003 to 2022.

Crude oil is a commodity renowned for its high level of volatility when compared to stock indices such as S&P500 or FTSE100. The recent COVID-19 pandemic precipitates a further increase in the volatility of crude oil, posing a challenge to implementing estimates of VaR and ES. Figure (1.1) illustrates the spot prices of two eminent commodities, West Texas Intermediate (WTI) crude oil and Europe Brent crude oil, from 2003 to 2022. As evident from Figure (1.1a), WTI crude oil experienced a brief period with negative prices in 2020 at the onset of the pandemic, an unprecedented event. Given the importance of crude oil as a commodity for banks and financial institutions and commodities' significant role as a risk factor for the calculation of capital requirements in the Fundamental Review of the Trading Book (FRTB) (Hull, 2018), it is more critical than ever to generate accurate estimates of VaR and ES concerning crude oil.

Considering the discordance between the estimation approaches of scholars and practitioners, where the former stresses parametric techniques and the latter favours nonparametric methods, the primary aim of this thesis is to investigate whether parametric methods for calculating VaR and ES outperform nonparametric methods for the volatile commodity of crude oil. An application of suboptimal estimation techniques potentially impels practitioners, such as banks, to adopt inefficient capital requirements for market risk and, thus, endure significant and tangible repercussions. That is, a feeble implementation methodology escalates the risk of insolvency and begets adverse externalities for the real economy. We utilise three nonparametric and three parametric methods to scrutinise the discrepancies in performance between their application to the formidable commodity of crude oil. The backtesting period, from 2016 to 2022, encompasses a significant event for crude oil - the COVID-19 pandemic.

Previous studies yield inconclusive results when evaluating nonparametric and parametric estimation methods regarding their relative performance. Brooks and Persand (2002) and Sadorsky (2006) ascertain that nonparametric methods excel over parametric methods in various backtests. Furthermore, they conclude that simple estimation methods are superior to complex methodologies, bolstering the case for implementing historical simulation by financial institutions. On the other hand, Marimoutou, Raggad and Trabelsi (2009) utilise conditional Extreme Value Theory (EVT), a more recent application in risk management, and discover a substantial augmentation in the accuracy of VaR and ES estimates for commodities such as WTI and Brent crude oil. Our findings demonstrate that parametric methods transcend nonparametric methods for VaR and ES. More specifically, we find that conditional EVT is optimal for estimating VaR at the 95% and 99% confidence levels, whilst the Student's t-distribution is the more efficient method for estimating ES at both confidence levels. However, the volatility-weighted historical simulation (VWHS) exhibits a sturdy performance in estimating VaR and ES at both confidence levels, which is robust across commodities. The efficacy of basic historical simulation (BHS) and age-weighted historical simulation (AWHS) deteriorates after a marked and sudden change in the risk level of the underlying commodity, corroborated in each of the backtests. Overall, the most efficient estimates of VaR and ES emanate from parametric methodologies. Therefore, we urge practitioners to espouse these techniques, especially regarding commodities exhibiting equivalent features to crude oil.

The thesis is structured as follows: Section 2 provides a comprehensive review of the essential background, encompassing the risk measures of VaR and ES, including their respective advantages and disadvantages, and previous findings. Section 3 describes the commodities of interest and presents summary statistics. Section

4 delineates the models and backtests we utilise in this thesis and evaluates the limitations. Moreover, we discuss our approach to account for and attenuate these constraints. Section 5 presents the results of the backtests and compares these to previous literature. Finally, Section 6 concludes the thesis and provides recommendations to practitioners and academics.

2 Literature Review & Background

2.1 Market Risk and Basel Capital Requirements

Market risk is a primary concern for practitioners because it pertains to the implications they endure due to altering market conditions (Hull, 2018). There is a possibility, albeit dubious, that financial institutions are rendered insolvent because of adverse shocks caused by, *inter alia*, market turmoil. Bankruptcies concerning financial institutions are disruptive and have economy-wide repercussions as it inhibits them from, amongst other things, acting as financial intermediaries for the real economy. Therefore, the Basel Committee on Banking Supervision (BCBS) enforce restrictive policies to curtail the probability of bankruptcies among financial institutions and preserve the serenity and durability of the macroeconomy. Adherence to these regulations requires financial institutions to develop and pursue adequate risk management. A constituent part of the risk management of banks and financial institutions is their compliance with capital requirements, which enables them to absorb adverse shocks and remain solvent. Under the Internal Models Approach (IMA), contrary to the Standardised Approach, financial institutions retain discretion over the model for calculating the aggregate capital carried on their balance sheets if they satisfy various stringent restrictions (Basel Committee on Banking Supervision, 2023a). More specifically, the ultimate capital requirement imposed on financial institutions emanates partially from the estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) on their Trading Book (cf. Section 2.3.1-2.3.2 for an overview of these risk measures). As a regulating body, the BCBS prefers banks to utilise a model that begets few violations (i.e., exceedances over VaR) to ensure the continued solvency of banks.

The IMA outlines various requirements to regulate and control market risk amongst financial institutions. The BCBS requires banks to compute risk estimation daily using an in-sample period of at least one year at the 99% confidence level in the IMA (Hull, 2018, p. 357). The banks are also subject to the local supervising authority's version of the IMA in line with the country's exposure. Although they possess discretionary power over the model for estimating VaR and ES under the IMA, supervising authorities demand the implementation of backtests and stress-tests to ensure that banks retain sufficient capital to withstand extended periods of market turmoil (Basel Committee on Banking Supervision, 2023a, 2023c). The capital requirements bestowed on banks' aggregate capital restrict the availability of funds for investment activities. However, it facilitates rigorous risk management and, consequently, abates the likelihood of bankruptcy in a recession.

The regulatory framework currently being implemented is the Basel IV Accord. This iteration of the Basel regulations recognises that VaR is not an adequate measure of the extreme risk. Therefore, the computation of capital requirements concerning market risk will pivot towards ES (cf. Section 2.3.1-2.3.2) (Basel Committee on Banking Supervision, 2023b). Also, the Basel Committee on Banking Supervision (2023b), under the FRTB, delineates several market risk factors that banks must appraise. The risk factors constitute aspects including foreign exchange, commodity, interest rate, credit spread and equity (Hull, 2018). Hence, a solid forecasting scheme for extreme risk measures heavily concerns the crude oil markets. Stress-tests and liquidity are equivalently vital constituents of risk management for financial institutions. However, in the interest of time, these are beyond the scope of the thesis and not further considered.

2.2 Coherency

What distinctive attributes of a risk measure compel it to excel over candidate risk measures? This conundrum has been the topic of vigorous debate and puzzled academia. A widely acknowledged explanation emanates from the concept of coherency (Szegö, 2002; Tasche, 2002; Hull, 2018), initially developed by Artzner *et al.* (1999). According to the authors, a coherent risk measure should fulfil four requirements for all loss distributions: monotonicity, sub-additivity, positive homogeneity, and translation invariance. Hence, this section briefly elucidates the properties of a 'desirable' risk measure, as proposed by Artzner *et al.* (1999).

Firstly, a monotonic risk measure states that for all portfolios A and B, if the losses of portfolio A are weakly less than the losses of portfolio B for every state of nature, then $R(L_A) \leq R(L_B)$ where $R(\cdot)$ denotes a risk measure and L_i denotes the stochastic loss of the ith portfolio (Szegö, 2002). This requirement is a desirable property of a risk measure since if an asset consistently delivers inferior outcomes than another asset in absolute terms, it should intuitively be deemed riskier. Monotonicity is analogous to first-order stochastic dominance (FSD), a prevalent concept when ranking miscellaneous assets, such as stocks, bonds, and commodities. An asset FSD another asset if and only if it produces weakly better outcomes than the other assets in all states of nature. Consequently, its cumulative distribution function (CDF) is always located to the right and below the other asset's CDF.

Furthermore, Markowitz (1952) is the pioneer to formally accentuate that an investor cannot be made worse off through diversification (Elton *et al.*, 2014). An adequate risk measure ought to recognise the benefits of diversification, which Artzner *et al.* (1999) denote as sub-additivity. Mathematically, for all portfolios A and B, a sub-additive risk measure states that $R(L_A + L_B) \leq R(L_A) + R(L_B)$. That is, the risk of the aggregated portfolio cannot exceed the sum of the risk of the two sub-portfolios for sub-additivity to be satisfied.

The third property of a coherent risk measure is labelled positive homogeneity. A positive homogeneous risk measure states that the risk of a portfolio scales in concordance with its size, $R(kL_A) = kR(L_A)$ for $k > 0$ (Artzner *et al.*, 1999). Intuitively, scaling up the quantity of an investment made in a position should scale the measure by an equivalent factor, given that there are no diversification benefits stemming from investing more in the same asset. However, positive homogeneity is the most controversial requirement as it neglects liquidity risks associated with the size of a specific position (Acerbi and Scandolo, 2008). The likelihood of a fire sale, or the necessity of selling an asset at a discount, increases proportionally with the size of the position. Therefore, the risk inherent in holdings does not increase linearly with the investment size but instead exhibits a convex relationship with the underlying risk. As aforementioned in Section (2.1), measuring liquidity risks is beyond the scope of this thesis and will not be further contemplated.

Lastly, translation invariance is commonly conceptualised as the normalisation of a risk measure. That is, a risk measure that abides by the translation invariance requirement implies that for all portfolios A, $R(L_A - m) = R(L_A) - m$ (Artzner *et al.*, 1999). The constant, m, is typically inferred as money, or other forms of liquid funds, added to a portfolio. Translation invariance dictates that adding an amount of money decreases the overall risk but does not affect the inherent riskiness of the risk-bearing components of the portfolio. Hence, the risk measure is "invariant" to the "translation" (i.e., translation referring to a movement of the underlying probability distribution). This requirement normalises the risk measure because it must be denoted in monetary terms to consider m as money or, more generally, liquid funds.

After outlining a theoretical foundation of what determines the desirability of a risk measure, we continue this thesis with a holistic overview of two extreme risk measures - Value-at-Risk and Expected Shortfall.

2.3 Measures of Extreme Risk

2.3.1 Value-at-Risk (VaR)

For a stochastic loss, L, Value-at-Risk (VaR) at the α per cent confidence level is defined as the minimum loss (i.e., quantile of the underlying loss distribution) satisfying (Pflug, 2000):

$$VaR_\alpha : \min\{\ell : Pr(L > \ell) \leq 1 - \alpha\} \tag{2.1}$$

where ℓ denotes a realised loss or a quantile of the pertinent loss distribution. Regarding a continuous distribution, the mathematical definition of VaR becomes (Hull, 2018):

$$VaR_\alpha : \inf\{\ell : Pr(L > \ell) \leq 1 - \alpha\}$$

$$\implies Pr(L > VaR_\alpha) = 1 - \alpha \tag{2.2}$$

The difference between Equation (2.1) and (2.2) is that for a discrete loss distribution, it may not be feasible to locate a quantile on the loss distribution that begets a probability mass to be precisely $(1-\alpha)$ to the right of VaR (Hull, 2018), whilst that is continually the case for a continuous loss distribution. Nevertheless, Equation (2.1) nests the definition of VaR for both a continuous and discrete loss distribution for this specific reason.

There are numerous reasons for the endorsement of VaR from the perspective of academics and practitioners. The risk measure conveniently encompasses the risk of an economic agent's total exposure to a specific position in one monetary unit (Beder, 1995; Manganelli and Engle, 2001). Therefore, the susceptibility to risk factors is communicated efficiently and effortlessly through VaR to various internal and external stakeholders. The risk measure summarises all positions for every asset class because VaR emanates from the underlying probability distribution as a quantile (Acerbi, Nordio and Sirtori, 2001; Dowd, 2002). Additionally, the focus on the downside risk of VaR is intuitively appealing as it is congruent with practitioners' and academics' conceptualisation of risk. Due to these reasons, *inter alia*, VaR has established itself as a widely accepted and extensively used risk measure.

However, various academics and regulatory institutions deem VaR an inadequate measure of extreme risk. A frequently levied critique towards the risk measure is that it is agnostic about losses exceeding VaR (Acerbi and Tasche, 2002; Yamai and Yoshiba, 2005). VaR does not fully describe tail risk, or downside risk, which is the purpose of measuring extreme risk and may induce adverse incentives for banks and traders working under VaR restrictions (see e.g., Hull, 2018, p. 273). Basak and Shapiro (2001) implement a conceptual analysis of VaR in a utility optimisation framework, concluding that risk managers operating under VaR restrictions optimally insure themselves in intermediate states of nature rather than extreme losses. Therefore, regulatory policies applied to VaR might inadvertently induce periods of market turmoil (Dowd, 2002, p. 12). Another limitation of VaR is that it is not a coherent risk measure (Artzner *et al.*, 1999; Acerbi and Tasche, 2002; Szegö, 2002). Specifically, VaR does not satisfy the sub-additivity requirement for all loss distributions (cf. Section 2.2), implying that the risk measure occasionally discourages diversification with adverse concentration risk as a consequence. Beder (1995) accentuates the practical limitations of VaR because of, amongst other things, its dependency on stringent assumptions of the underlying loss distribution and a myriad of implementation methodologies. Therefore, the risk measure entails a nontrivial model risk (Marshall and Siegel, 1997), which requires consideration when employing VaR. Numerous researchers delineate specific frameworks to manage this model risk in practice (see e.g., Kerkhof, Melenberg and Schumacher, 2010), but further deliberation of these frameworks for model risk is beyond the scope of this thesis. Due to many of these issues, academics and regulators have partially diverted their attention towards risk measures with more favourable theoretical properties.

2.3.2 Expected Shortfall (ES)

Expected Shortfall (ES)[1] is defined as the average VaR for all confidence levels x in the interval $\alpha \leq x \leq 1$ (Hull, 2018). Mathematically, we denote ES as:

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_x dx \tag{2.3}$$

where α denotes the confidence level. Equation (2.3) dictates that an average is taken over an infinite number of VaR which is computationally tedious and, in practice, Riemann's integrals are often used for simplification. McNeil, Frey and Embrechts' (2015) dual-representation of ES, originally derived by Rockafeller and Uryasev (1999), provides further insight into this risk measure. For a continuous loss distribution, this representation of ES is:

[1]Some authors in the literature refer to Expected Shortfall as Conditional Value-at-Risk (CVaR). Due to consistency, this thesis utilises the former term throughout the paper.

$$
\begin{aligned}
ES_\alpha &= \frac{\mathbb{E}[L \cdot I]}{1-\alpha} = \mathbb{E}\left[L \mid L > VaR_\alpha\right] \\
I &= \begin{cases} 1 & \text{for } L > VaR_\alpha \\ 0 & \text{for } L \leq VaR_\alpha \end{cases}
\end{aligned}
\tag{2.4}
$$

where I is a dummy variable and $\mathbb{E}[\cdot]$ is the expectation operator. Equation (2.4) states that ES is, by definition, the expected loss conditional on it exceeding VaR (Hull, 2018, p. 274).

The advantages of the risk measure become palpable following the definition of ES. In contrast to VaR, ES describes the entire tail risk of a loss distribution. Consequently, ES does not discourage diversification or, in other words, it satisfies the sub-additivity condition (cf. Acerbi and Tasche (2002, pp. 1501-1502) for formal proof). ES is a coherent risk measure as it reconciles all of the requirements for coherency. These appealing properties have prompted the Basel Committee on Banking Supervision (2023b) to revise the Fundamental Review of the Trading Book (FRTB) concerning the computation of the capital requirements for market risk in Basel IV by substituting VaR for ES.

ES rectifies many of the critiques levied at VaR, albeit there are issues to consider for this risk measure. Yamai and Yoshiba (2005) ascertain that ES endures considerable estimation error. In contrast to VaR, ES demands a greater sample size to achieve accurate estimates. The implication is that ES is more onerous to implement in practice. Furthermore, academics have contested the backtestability of ES as it lacks elicitability (see e.g., Lambert, Pennock and Shoham, 2008; Gneiting, 2011). Nevertheless, authors such as Acerbi and Szekely (2014) refute the importance of elicitability for backtesting, who develop numerous backtests for ES (cf. Section 4.5.3). We put ourselves 'on the shoulders of giants' in this thesis and utilise the backtests as described by Acerbi and Szekely (2014), neglecting the intellectual debate on the significance of elicitability.

2.3.3 Estimation Methods and Backtests

There are commonly two distinct techniques to implement the estimation of VaR and ES in practice: non- and parametric estimation methods.[2] Nonparametric methods unequivocally rely on the empirical loss distribution when estimating, for instance, one-day-ahead forecasts and do not specify a distributional form *a priori*. Due to the flexibility of the assumptions and relative ease of computation, *inter alia*, the nonparametric methods are widely implemented in practical contexts (Pérignon and Smith, 2010). Contrarily, the parametric estimation methods of VaR and ES differ because they explicitly assume that the loss distribution follows a known and well-defined theoretical continuous probability distribution (Hull, 2018). These methods permit academia more flexibility to delineate intricate methodologies to develop robust and accurate estimates of VaR and ES. However, neither implementation procedure can adequately capture latent risks, such as the two classes of climate risk: transition and disaster risk. The nonparametric and parametric methods derive their estimates from empirical data, which does not fully acknowledge those risk

[2]Some estimation methods are referred to as semi-parametric if it is nonparametric but implemented with, amongst other things, GARCH-type models.

factors. Therefore, there is a nontrivial possibility of underestimation concerning climate risk. We do not consider climate risk further as it is beyond the scope of the thesis. Thus, the various procedures differ primarily on the statistical sophistication required to attain the estimates for the measures of extreme risk. Nevertheless, a formal backtest of the estimations for the point and interval forecasts of VaR and ES, respectively, determines the desirability of a particular methodology.

According to Jorion (2007, p. 139), the rationale for backtests concerning the implementation of VaR and ES is to determine if an estimation method adheres to the expected losses given the size of a backtesting sample and confidence level. A desirable estimation method ought to yield the correct number of violations in the backtest, where a violation is defined as a loss exceeding VaR (Hull, 2018). Backtests for VaR generally aim to assess whether a method produces accurate point forecasts by comparing the actual and projected violations. Therefore, tests of binary classification are sufficient. The most common of these models is the Kupiec (1995) test for unconditional coverage (Halilbegovic and Vehabovic, 2016), which basically provides the foundation for the Basel Traffic Light, i.e., the backtest required in the Basel regulations for capital requirements regarding market risk (Hull, 2018). These backtesting methodologies permit academics and practitioners to determine if the VaR level is appropriately estimated. However, only implementing such tests fails to detect bunching of exceedances, *videlicet*, violations of VaR appearing in clusters (Hull, 2018, pp. 287-288). Bunching begets the probability of a violation to elevate considerably during market turmoil, rendering financial institutions susceptible to solvency issues. Complementary tests to uncover estimation methods of VaR and ES that beget bunching are required. Prominent backtests that consider these issues are the Christoffersen (1998) and the Engle and Manganelli (2004) tests. The former essentially investigates the number of non- and violations occurring in succession, whilst the latter backtest utilises a linear regression model to test for dependencies in the time series of binary variables. Consideration of dependencies of violations over time is neglected in the backtests of the Basel regulation (Hull, 2018). Nevertheless, an estimation method that captures the dynamics of the model ensures that an adequate VaR level is maintained even during market turmoil.

Backtesting ES poses a conundrum because of its definition as an average VaR for all confidence levels in $\alpha \leq x \leq 1$ (Hull, 2018). The implication is that any backtest is, in principle, a test of the entire right tail of the loss distribution. Backtests concerning ES are a more recent phenomenon and their structure varies widely. For instance, Acerbi and Szekely (2014) introduce a set of backtest derived from Equation (2.4). On the other hand, Costanzino and Curran (2018) develop another model for backtesting ES utilising a dummy variable that measures the severity of a VaR violation. Academia performs numerous attempts to create robust backtests concerning ES, albeit Basel IV implements its backtest on VaR (Basel Committee on Banking Supervision, 2023c). Therefore, there is no transparent standard for backtesting estimates regarding ES. The nonparametric and parametric estimation methods and the backtests we implement in this thesis are delineated in Section 4.

2.4 Stylised Facts of Financial Time Series

Empirical research of returns concerning various asset classes discovers consistent characteristics of the time series. These findings are denoted collectively as the stylised facts of financial time series. For instance, empirical results exhibit robust evidence of distinct attributes such as sharp peaks and heavy tails (i.e., leptokurtosis) regarding the distributional form of asset returns (Pagan, 1996; Cont, 2001). Acute peakedness pertains to the propensity of asset returns to congregate around the mean in frequency, whilst fat tails indicate a higher likelihood of extreme outcomes compared to a Gaussian distribution. Additionally, Pagan (1996) and Cont (2001) assert that skewness accompanies leptokurtosis, which further deviates asset returns from a Gaussian distribution. It is discernible by the observations that considerable losses are more frequent than equal gains, i.e., gain-loss asymmetry. Time-varying volatility is another prevailing stylised fact innate to financial time series (Cont, 2001). Returns of financial assets exhibit periods of high and low volatility clusters which tend to persist. Volatility clustering indicates that returns in subsequent periods are not independent. Despite the general consensus that asset returns are a random walk process, time-varying volatility reveals the presence of nonlinear dependence in the time series. Additionally, time-varying volatility in financial time series frequently exhibits leverage effects. Black (1976), credited with the initial development of the concept, asserts that stock price returns and future volatility display a negative causal relationship. That is, negative shocks impact the future volatility of share price returns relatively more than positive shocks, i.e., negative returns at time t increase volatility for time $t + 1$, and vice versa.

Evidence of these stylised facts is also present in crude oil commodities. Ebrahimi and Pirrong (2018) observe that portfolios with high exposure to kurtosis in the oil market tend to assume negative returns. However, when controlled for the kurtosis, the variance in returns becomes insignificant. They further find that skewness only displays a nontrivial impact on portfolios with exposure to the crude oil market in the sub-period but not in the full-time series, whereas kurtosis is significant in all periods. Additionally, the tendency for volatility clustering to persist for long periods, especially in crude oil, indicates a long-memory process in returns (Choi and Hammoudeh, 2009). Wang and Liu (2010) find that large fluctuations in the crude oil markets tend to be highly unstable, whereas small fluctuations often persist and can be forecasted. Moreover, Kristoufek (2014) demonstrate that leverage effects in volatility clustering are highly prevalent for the crude oil commodities of WTI and Brent. Chen and Mu (2021) further provide evidence of a distinct disparity between futures and spot prices. Returns derived from the former exhibit "standard" leverage effects, whilst returns computed from the latter display "inverse" leverage effects. The phenomenon refers to a positive correlation between the returns and future volatility, i.e., when the spot price of a commodity increases, the conditional volatility also increases. Inverse leverage effects can be explained by the theory of storage, first conceptualised by Working (1949), where a low inventory increases the risk of a supply shortage and consequently increases the price and volatility of a commodity, and vice versa. Hence, equivalent to the time series of various financial assets, crude oil commodities exhibit nuanced evidence of the stylised facts.

This section has presented numerous prevailing statistical commonalities in the time series of financial assets and, specifically, crude oil commodities. The subsequent section develops further on the empirical findings on nonparametric and parametric estimation methods for VaR and ES.

2.5 Previous Research and Empirical Findings

The primitive endeavours in the risk management literature at documenting the empirical findings on nonparametric and parametric estimation methods for VaR find trivial discrepancies in their respective performance (e.g., Beder, 1995; Hendricks, 1996; Pritsker, 1997). Hendricks (1996) compares the efficacy of historical simulation methodologies and Monte Carlo simulation with models of volatility based on foreign exchange portfolios. The author discovers that all models attain their objectives satisfactorily. More recently, authors such as Angelidis, Benos and Degiannakis (2007) and Abad and Benito (2013) present evidence demonstrating that nonparametric methodologies render nontrivial imprecisions in their estimates. Abad and Benito (2013) study various stock indices and uncover that methods for estimating VaR emanating from historical simulation display an inferior performance amidst market turmoil, whilst parametric methods implemented with nonlinear GARCH-type models excel. Contrarily, Sadorsky (2006) utilises a bundle of nonparametric and parametric methods to estimate VaR for future contracts regarding commodities associated with unleaded gasoline, natural gas, heating oil, and crude oil. The findings illustrate that the general backtesting performance of nonparametric estimation methods transcends those of the parametric methods (Sadorsky, 2006). Brooks and Persand (2002) observe that nonparametric methods outperform their parametric rivals, including Extreme Value Theory (EVT), for six broad indices in the U.K. and the U.S., concluding that simple estimation methods are superior to the more complex methodologies. It is apparent that the empirical literature yields inconclusive answers as to whether the parametric estimation methods are more desirable than nonparametric methods. Regardless, the parametric methods demonstrate a conspicuous predominance in the academic literature and warrant a closer review of the empirical findings concerning crude oil.

Accurate volatility forecasting is accentuated in the literature as crucial for the computation of VaR and ES due to extreme risk levels and clustering of volatility in crude oil (Aloui and Mabrouk, 2010; Youssef, Belkacem and Mokni, 2015). Numerous previous researchers find that the FIAPARCH model is the optimal model for volatility forecasting to exploit when estimating VaR and ES for crude oil commodities due to its ability to capture long-memory processes and leverage effects (see e.g., Aloui and Mabrouk, 2010; Chkili, Hammoudeh and Nguyen, 2014; Youssef, Belkacem and Mokni, 2015). Specifically, Aloui and Mabrouk (2010) contrast three separate long-memory GARCH-type models (i.e., FIGARCH, FIAPARCH, and HYGARCH) on, *inter alia*, WTI and Brent and conclude that the FIAPARCH displays the most robust performance when estimating VaR and ES. However, the underlying distributional assumption is susceptible to leptokurtosis and skewness. Therefore, the authors suggest that a skewed Student's t-distribution is required to yield accurate risk forecasts. On the other hand, Wei, Wang and Huang (2010) find that no specific GARCH-type specification is superior for the forecasting ability of crude oil commodities as long as a nonlinear specification is provided to capture the leverage effects, especially for shorter-term horizon forecasts (Liu *et al.*, 2022).

A significant development in the parametric estimation of extreme risk is Extreme Value Theory (EVT). Accordingly, numerous endeavours have been made in the literature to apply it to commodities relating to crude oil. Authors such as Marimoutou, Raggad and Trabelsi (2009) and Youssef, Belkacem and Mokni (2015) uncover that conditional EVT, a method introduced by McNeil and Frey (2000), is the optimal parametric method when estimating VaR and ES for WTI and Brent crude oil. Marimoutou, Raggad and Trabelsi (2009) implement a simple AR(1)-GARCH(1,1) model for the two crude oil commodities from 1987 to 2007 and apply EVT on the standardised GARCH residuals. The findings demonstrate a significant improvement from more simple estimation methods. Contrarily, Chiu, Chuang and Lai (2010) find that the conditional EVT is inferior to the more elementary volatility-weighted historical simulation for the same commodities in the sub-period 2000-2007. Hence, the question naturally arises of whether there is an outperformance of the more complex methods which depend on advanced statistical theory. Nevertheless, conditional EVT has been an eminent development in the estimation of extreme risk also for volatile commodities, and even the simple GARCH-type specifications display an adequate performance when utilised in conjunction with EVT.

3 Data

The data used for this thesis entails the daily spot prices of WTI and Brent crude oil. WTI, or the West Texas Intermediate, is available to the spot market through pipelines and delivered to Cushing, Oklahoma in the U.S., whereas Brent crude oil is extracted in the North Sea. WTI and Brent are focal reference points worldwide for crude oil prices (Lang and Auer, 2020), in addition to being the most widely researched and traded crude oil commodities. Therefore, these two commodities are the subjects of analysis in this thesis. The time period encompasses the daily spot price commencing January 2003 to December 2022, i.e., a 20-year horizon. The backtesting period starts in January 2016 and ends in December 2022. We gather the data for the spot price from the FRED database (see U.S. Energy Information Administration, 2023a, 2023b). The crude oil commodities exhibit extended periods of extreme volatility, which poses a challenge for all implementation methodologies for VaR and ES, regardless of their classification. Crucially, at the commencement of the COVID-19 pandemic (i.e., 20 April 2020), WTI traded at a negative price, an extraordinary event for commodities. According to Kearney (2020), the pandemic restricted travelling, causing an overcrowding of crude oil at storage facilities. This compelled suppliers to compensate buyers for alleviating the oversupply, i.e., causing negative prices. This phenomenon raises issues for the computation of returns because logarithms cannot be applied to negative numbers. Therefore, we opt to index the price series of WTI and employ the logarithms on the index series to calculate the continuously compounded returns (see Appendix A.1 for further explanation). We compute the time series of returns as:

$$r_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right) = \ln P_{i,t} - \ln P_{i,t-1} \tag{3.1}$$

where $P_{i,t}$ and $r_{i,t}$ denote the spot price and the continuously compounded daily return for the ith commodity at time t, respectively.

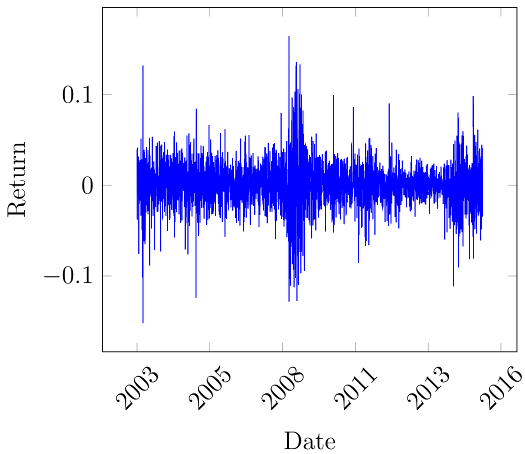
Table (3.1) presents summary statistics, Ljung-Box test for autocorrelation and the test statistic for the Jarque and Bera (1980) test for a Gaussian distribution.

	Mean	STD	Min	Max	Skew	Kurt	LB-Q(20)	JB Statistic
WTI	0.000045800	0.0241154	-0.151909	0.164137	-0.0755907	4.7198114	65.36239***	3035.51612***
Brent	0.000057212	0.0213847	-0.1683201	0.1812974	0.0196234	5.1149835	44.02818***	3592.18106***

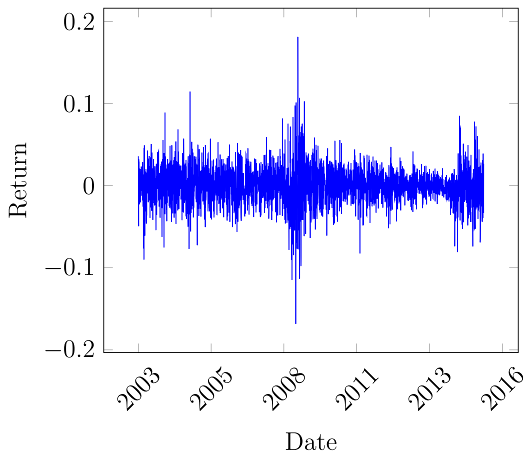
Table 3.1: Summary statistics of the sample data from estimation period. *, **, and *** denote statistical significance at the 0.1, 0.05, and 0.01 level, respectively. LB and JB denote the Ljung-Box test and Jarque-Bera test, respectively.

We notice that the mean of both return series is approximately zero for WTI and Brent crude oil commodities. The return series exhibits evidence of both asymmetry and leptokurtosis, as displayed by the skewness and kurtosis estimate. Specifically, the Jarque and Bera (1980) test rejects the null hypothesis of a Gaussian distribution because of a p-value that is essentially zero for both commodities. Congruent with the stylised facts of financial time series (cf. Section 2.4) and previous research on similar assets, our sample data exhibits robust evidence of heavy tails. Moreover, the Ljung-Box test demonstrates proof of autocorrelation up to the lag of order 20. This provides issues for implementing methods such as Extreme Value Theory (EVT) as it assumes an i.i.d. series (Marimoutou, Raggad and Trabelsi, 2009). However, Appendix (B.1) presents, amongst other things, the Ljung-Box test of the standardised residual series of a GARCH-type model, which concludes no autocorrelation. Therefore, we can apply EVT on the standardised GARCH residuals.

Figure (3.1) visualises the daily return series for the two commodities throughout the estimation period.



(a) Daily Return of WTI



(b) Daily Return of Brent

Figure 3.1: Visualisation of the daily return series of WTI and Brent crude oil for the estimation period.

Figure (3.1) demonstrates that the return series of the commodities centres around the mean through the estimation period. The stylised fact of volatility clustering for financial time series (cf. Section 2.4) is evident in our data from Figure (3.1). There are extended periods where the volatility is relatively high and long periods where the volatility is relatively low, which is valid for both commodities. The Great Recession of 2007/2008 is also nuanced for WTI and Brent crude oil during the estimation period with extreme returns. A visual inspection of Figure (3.1) indicates the presence of GARCH-type effects in line with previous research and signals the need for a volatility model. Nevertheless, we implement a formal test for ARCH effects in the return series, which generates a LaGrange multiplier statistic of 639.4746 ($p < 0.01$) and 445.6150 ($p < 0.01$) for WTI and Brent, respectively. Hence, we soundly conclude that the return series exhibits volatility clustering.

The estimates of parameters in time-series models depend on the absence of a unit root in the underlying data. Therefore, we implement two tests for stationarity: the augmented Dickey and Fuller (1979) (ADF) test and the Kwiatkowski *et al.* (1992) (KPSS) test. The intuition for implementing two tests is because of the low power of the ADF test, which has the non-stationarity under the null hypothesis (Brooks, 2019, pp. 451-452). Hence, we complement the ADF test with another test that positions the stationarity condition under the null hypothesis, i.e., the KPSS test. The results from these tests are presented in Table (3.2).

	ADF	KPSS
WTI	-15.2155810***	0.3593333*
Brent	-13.4613351***	0.4369624*

Table 3.2: Test for stationarity for the return series of WTI and Brent in the estimation period. *, **, and *** denote statistical significance at the 0.1, 0.05, and 0.01 level, respectively.

The results from the stationarity tests present some contradictory conclusions. Specifically, we reject the null hypothesis of non-stationarity at the 0.01 level of significance for both commodities. However, the KPSS test rejects the null of stationarity at the 0.1 significance level for WTI and Brent. Nevertheless, since the time series in Figure (3.1) is centred around the mean and the extremely low p-value of the ADF test, we conclude that the return series for both crude oil commodities are stationary enough for our purposes in this thesis.

We extrapolate the proper theoretical distribution that best describes the sample data by plotting a Quantile-Quantile (Q-Q) plot of the commodities. A Q-Q plot compares the quantiles of an empirical distribution with the quantiles of a known and well-defined theoretical distribution (Marden, 2004, p. 606). Intuitively, a perfect fit between the observed and theoretical distribution would result in the scatters being located precisely on the 45-degree line. Figure (3.2) plots the quantiles of the return series of the commodities jointly with the theoretical quantiles of the Gaussian distribution (cf. Section 4.3.1 for an overview of the Gaussian distribution).

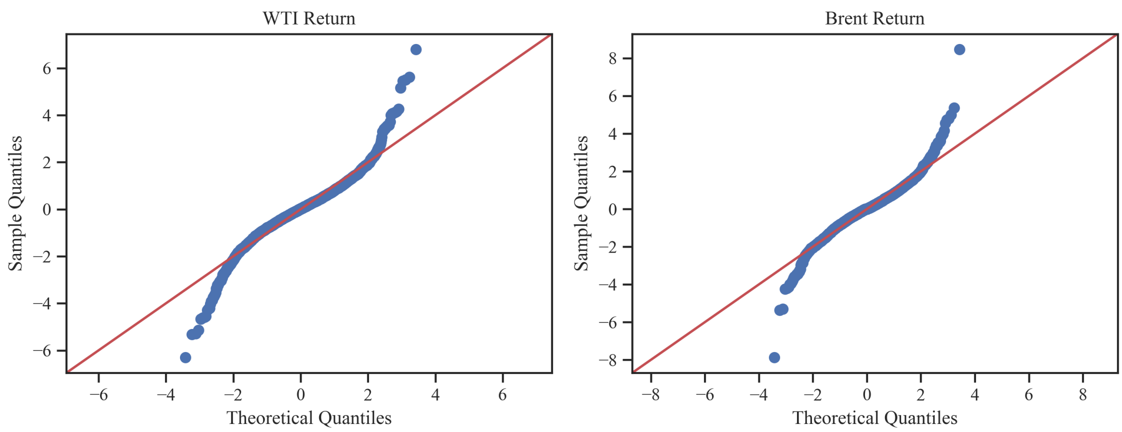


Figure 3.2: Q-Q plot of the empirical return distribution together with the theoretical Gaussian distribution.

As can be inferred from Figure (3.2), the nonlinearity of the Q-Q plot suggests that there is a poor fit between the empirical return series for the estimation period and the theoretical Gaussian distribution, providing support for the conclusion from the Jarque and Bera (1980) test. More specifically, the leptokurtosis of the return series for the two commodities, as formally shown in Table (3.1), is nuanced in Figure (3.2) by the scatters diverging from the 45-degree line at more extreme quantiles. We further demonstrate the fit between the Gaussian distribution[1] and the observed

[1] We estimate the parameter for fitting the theoretical distribution upon the empirical frequency distribution by Maximum Likelihood.

return distribution by superimposing the theoretical distribution on a histogram of the returns, as demonstrated in Figure (3.3).

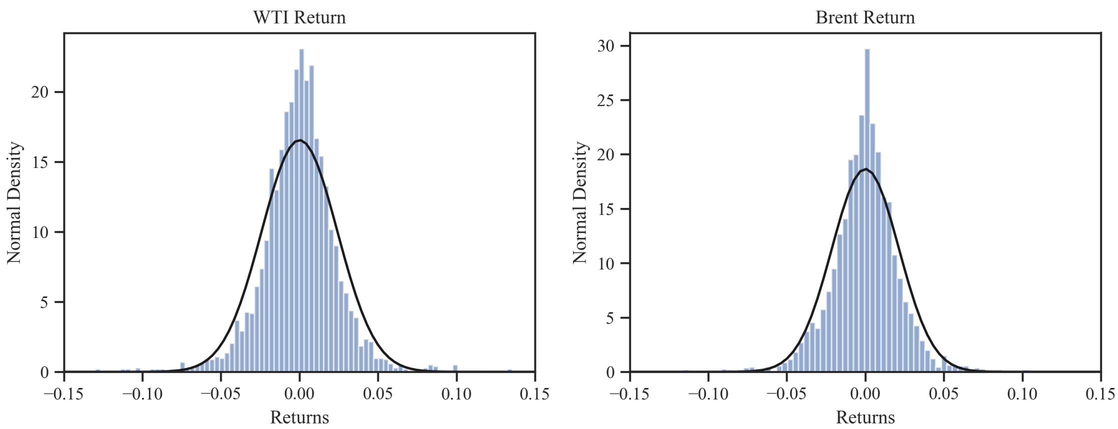


Figure 3.3: Histogram of the return series with a superimposed theoretical Gaussian distribution.

The Gaussian distribution does not depict a suitable fit for the empirical return distribution of either commodity. This result distinctly corroborates previous research and the empirical findings of, for instance, Youssef, Belkacem and Mokni (2015) of an EVT approach for estimating VaR and ES as this approach does not explicitly assume a distribution *a priori* and concentrates on the tails (Hull, 2018). Nevertheless, we investigate how well the empirical return distributions fit the Student's t-distribution (cf. Section 4.3.2 for an overview of the Student's t-distribution). Figure (3.4) illustrates the Q-Q plots of the commodities in conjunction with the theoretical quantiles of the Student's t-distribution.

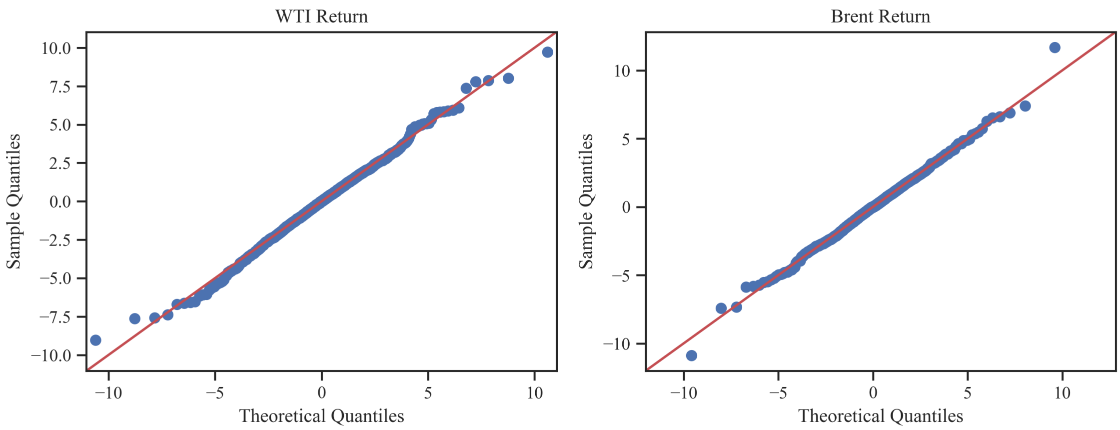


Figure 3.4: Q-Q plot of the empirical return distribution together with the theoretical quantiles of the Student's t-distribution.

Comparing Figure (3.2) and (3.4), it is indisputable that the Student's t-distribution presents a refinement to the Gaussian distribution. Contrary to the Gaussian distribution, the scatters at the extreme ends of the Q-Q plot for both commodities are more adjacent to the 45-degree line. It is consistent with our expectations that this theoretical distribution demonstrates a more suitable fit for our data given the leptokurtosis presented in Table (3.1). This result emerges because the Student's t-distribution with low degrees of freedom has relatively higher kurtosis and is more qualified to capture certain aspects of the stylised facts of financial time series. Figure (3.5) illustrates a histogram of the empirical return distribution alongside the theoretical Student's t-distribution.

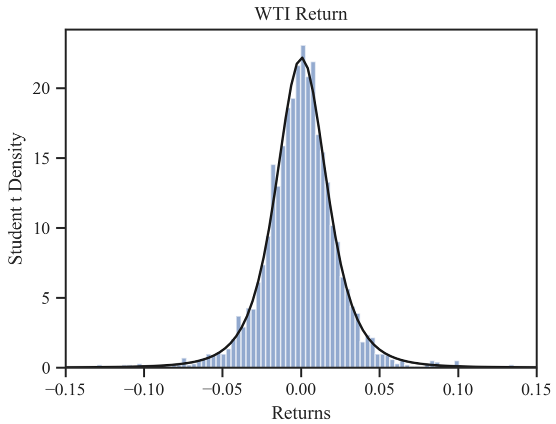


Figure 3.5: Histogram of the return series with a superimposed theoretical Student's t-distribution.

It is discernible that the Student's t-distribution better captures both the leptokurtosis and intermediate parts of the empirical distributions of the daily returns of WTI and Brent crude oil. However, given that we do not uncover a perfect fit even for this theoretical distribution, it further supports the intuition of fitting the generalised Pareto distribution of the EVT approach (cf. section 4.3.3 for an overview of the EVT). A Q-Q plot of the return series jointly with the generalised Pareto distribution from the EVT method presents an intriguing observation, as illustrated by Figure (3.6).

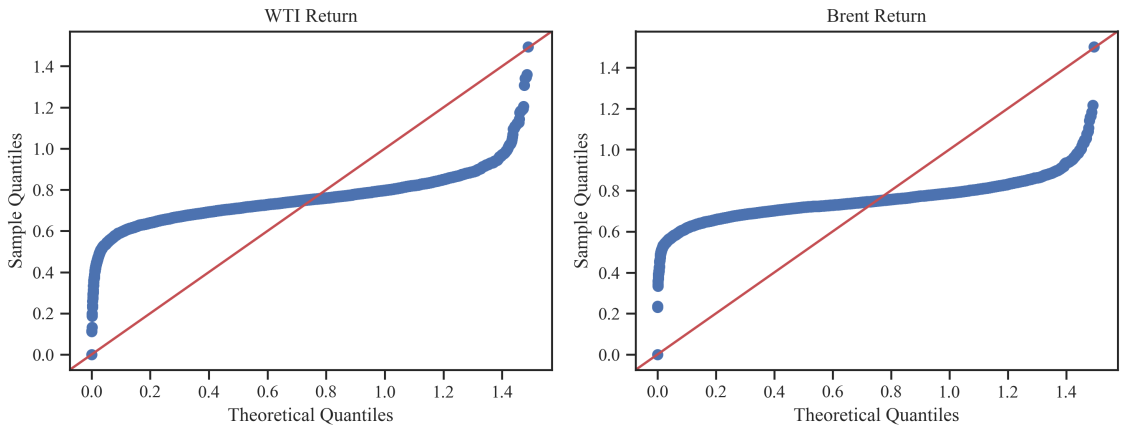


Figure 3.6: Q-Q plot of the empirical return distribution together with the theoretical quantiles of the generalised Pareto distribution.

As depicted by Figure (3.6), the fit of the generalised Pareto distribution is suboptimal for the intermediate parts of the empirical distribution for both WTI and Brent crude oil. Nevertheless, the most extreme quantiles lie virtually precisely on the 45-degree line, suggesting a suitable fit for these quantiles of the distribution and that the generalised Pareto distribution is highly applicable for VaR and ES estimation.

This section has presented plots and summary statistics based on the daily returns. The remainder of the thesis employs the loss distribution. That is, we multiply the daily continuously compounded returns by negative one, as demonstrated by:

$$\ell_{i,t} = r_{i,t} \cdot (-1)$$

where $\ell_{i,t}$ denotes the loss observed at time t for the ith commodity. This thesis assumes that we hold a long position of 100 units of each commodity i, implying that we multiply each loss by the factor 100.

4 Methodology

4.1 Out-of-Sample Forecasting

A dataset used for forecasting requires an in-sample and an out-of-sample component. We can conceptualise the timeline as an event study. In the event, the in-sample is akin to the estimation period, i.e., the data utilised to derive a model and estimate its parameters to yield forecasts. The out-of-sample is analogous to the event window of the event study, i.e., the data used to test the model's goodness of fit. Given our intent of conducting one-step-ahead forecasts, it is not feasible to observe and control the model's fit contra the realisation at $t+1$ at time t. Therefore, we allot a portion of the previous dataset between an in-sample (estimation) set and an out-of-sample (test) set, where we use the fitted values from the in-sample to determine the suitability of the model on the out-of-sample to locate the optimal model for forecasting.

A rolling window alludes to a fixed length of an in-sample period where the start and end date increase by one with each observation (Brooks, 2019). This technique is particularly effective when the statistical properties of a dataset progressively change. That is, the first data point is replaced from the window by the incremental observation and, consequently, alters the descriptive statistics of the sample data. Hence, it possesses the capability to capture the dynamics of the data. In contrast, an expanding window keeps the initial estimation date fixed and incrementally adds more recent observations to the sample period (Brooks, 2019). This window incorporates more data points and can generate more accurate parameter estimates vis-à-vis a rolling window.

This thesis utilises a rolling window for out-of-sample forecasting instead of an expanding window despite the greater accuracy in the parameter estimates. A rolling window accommodates a standardised sample size to compute descriptive statistics, which renders the results comparable. In contrast, an expanding window is unable to trace the dynamics of the data as rigorously as its rolling counterpart due to its recursive qualities. Therefore, we opt for a window size of 500 observations for the nonparametric and 1000 observations for the parametric estimation window to produce one-step-ahead forecasts. The rationale for implementing a smaller window size for the nonparametric methods is that most of them effectively assume a constant variance for the duration of the window size. Contrarily, we implement the parametric estimation methods with a GARCH-type model, which introduces time dependence in the volatility level. Therefore, we can utilise a larger window size for the parametric methods and generate more consistent estimates of the parameters.

4.2 Nonparametric Methods

As aforementioned in Section 2.3.3, nonparametric estimation methods do not specify a distributional form *a priori* and depend on the empirical loss distribution when estimating one-day-ahead VaR and ES forecasts. We proceed by delineating the three nonparametric methods implemented in this thesis.

4.2.1 Basic Historical Simulation (BHS)

Basic historical simulation (BHS) is a straightforward nonparametric estimation method. The BHS is implemented directly on the empirical loss distribution without modification by electing the pertinent loss as its estimate for time t. For a sample size of M and a confidence level of α, we expect $(1-\alpha)\cdot M$ losses to be in excess of VaR. Accordingly, the estimate for VaR is simply the $(1-\alpha)\cdot M+1$ largest loss. Mathematically, we describe the BHS as:

$$Pr(L > \ell^s_k) \leq 1-\alpha$$

$$Pr(L > \ell^s_{k+1}) > 1-\alpha$$

$$\implies VaR_{\alpha,t} = \ell^s_k \tag{4.1}$$

where ℓ^s denotes the sorted loss distribution in descending order. Since the probability of losing more than ℓ^s_{k+1} is greater than the desired level of $1-\alpha$, we select the smallest loss that begets the probability of losing more than VaR being less than or equal to $1-\alpha$, i.e., the kth largest loss. Furthermore, a prevailing implementation scheme of the estimate of ES using BHS is the arithmetic average of the $k-1$ losses, which are greater than the VaR estimate from the empirical loss distribution. However, this interpretation of ES is not theoretically valid because the observed losses emanate from a discrete distribution. A common justification of the approach is the implicit assumption that the sample data is drawn randomly from some underlying but unknown continuous loss distribution. We utilise this somewhat arbitrary interpretation of ES in this thesis. Therefore, our approach for deriving the estimate for ES is given by:

$$ES_{\alpha,t} = \frac{1}{k-1}\sum_{i=1}^{k-1} \ell^s_i \tag{4.2}$$

where ℓ^s once again denotes the sorted loss distribution in descending order.

Although the simplicity of the BHS is intuitively appealing, there are various limitations that require attention when applying it in practice. Boudoukh, Richardson and Whitelaw (1998) accentuate two specific problems related to the procedure of locating extreme quantiles when the data availability is scarce and that the method essentially assumes an identical and independent distribution. The latter issue can be rectified by simply implementing the BHS with a rolling window, which introduces some degree of time dependence. For our purposes, the simplicity of the BHS makes it an exemplary model to employ as a benchmark. Ideally, a more complex methodology to generate estimates for VaR and ES should at least outperform the BHS.

4.2.2 Age Weighted Historical Simulation (AWHS)

The age-weighted historical simulation (AWHS) was first introduced by Boudoukh, Richardson and Whitelaw (1998) and is a nonparametric estimation method which accentuates the intuition that newer observations are more pertinent for one-day-ahead estimation of VaR and ES (Hull, 2018, p. 301). Recent observations are allocated a superior weighting than older ones. Conceptually, the BHS is nested as a special case of the AWHS where all observations are assigned an equal weight. This is formally proved by utilising l'Hôpital for evaluating the limit where the exponential decay factor, λ, of AWHS tends towards one, albeit not demonstrated in this thesis. Moreover, the weight of the first observation (i.e., the newest loss) in the AWHS is defined as:

$$w_T = \frac{1-\lambda}{1-\lambda^T} \tag{4.3}$$

where $\lambda < 1$. Older observations are then recursively multiplied by the exponential decay factor as:

$$w_t = \lambda^{T-t} \cdot w_T, \ \ t = T-1, T-2, ..., 2, 1 \tag{4.4}$$

To estimate VaR using AWHS, the losses are sorted in descending order according to the size of the loss and the weights are then summed until we find $Pr(L > \ell_k^s) \leq 1-\alpha$ and $Pr(L > \ell_{k+1}^s) > 1-\alpha$ where the kth largest loss is taken as the VaR estimate for time t. To estimate ES, the same procedure and assumptions are applied as in the case for the BHS. As for the selection of the exponential decay factor, we specify λ as 0.995 for technical reasons of estimating VaR and ES at both the 95% and the 99% confidence level.

4.2.3 Volatility Weighted Historical Simulation (VWHS)

The volatility-weighted historical simulation (VWHS) emphasises that estimates of VaR and ES should reflect volatility clustering (Hull and White, 1998; Hull, 2018). More specifically, estimates of VaR and ES ought to account for turbulent market conditions with elevated risk levels. Many authors in the literature classify VWHS as a semi-parametric method, as GARCH-type models are frequently utilised in the implementation. Nevertheless, we divert from this classification in this thesis for a more unambiguous distinction between the parametric and nonparametric methods. Therefore, we implement the VWHS by utilising the EWMA volatility forecasting scheme, which in turn can be conceptualised as a simplified GARCH(1,1) model (cf. Section 4.4.2), to preserve this method as fully nonparametric. One technique for implementing the VWHS is to scale the observed sample losses by the ratio of the volatility on day t and the one-day-ahead forecasted volatility on day $T+1$. Mathematically, we describe the scaled losses, ℓ_t^*, as:

$$\ell_1^* = \frac{\sigma_{T+1}}{\sigma_1} \cdot \ell_1$$

$$\ell_2^* = \frac{\sigma_{T+1}}{\sigma_2} \cdot \ell_2$$

$$\vdots$$

$$\ell_T^* = \frac{\sigma_{T+1}}{\sigma_T} \cdot \ell_T$$

To implement the estimation of VaR at time t, BHS (cf. Section 4.2.1) is applied to the rescaled losses. Similarly, the estimate of ES at time t is computed in the same manner as for the BHS but for the rescaled losses. The VWHS is a straightforward and efficient approach to accommodate the stylised fact of time-varying volatility in the estimates of VaR and ES. Finally, as the nonparametric methods have been extensively covered, we divert our attention to a different approach towards estimating VaR and ES, namely, the parametric methods.

4.3 Parametric Methods

In Section 2.3.3, we discuss that the parametric estimation methods of VaR and ES differ from the nonparametric in that they explicitly assume that the loss distribution follows a known and well-defined theoretical probability distribution. We proceed by providing an overview of three different parametric estimation methods, which we implement in this thesis.

4.3.1 Gaussian Distribution

A standard benchmark model for parametric methods is the Gaussian distribution, also known as the normal distribution, which is a bell-shaped continuous distribution. The distribution is fully described by two parameters, and the probability density function (pdf) is given by:

$$L \sim \Phi(\mu, \sigma) \implies f(\ell) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ell-\mu}{\sigma}\right)^2\right] \tag{4.5}$$

where μ denotes the location parameter (i.e., the mean) and σ denotes the scale parameter (i.e., the standard deviation). Given that VaR is the α quantile of the loss distribution (Hull, 2018), the formula for VaR and ES based on the Gaussian distribution is:

$$VaR_{\alpha,t} = \mu + \sigma \cdot z_\alpha \tag{4.6}$$

$$ES_{\alpha,t} = \mu + \sigma \cdot \frac{f_{std}(z_\alpha)}{1-\alpha}, \text{ where } f_{std}(z_\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z_\alpha^2\right] \tag{4.7}$$

where z_α denotes the α quantile of the standardised normal distribution. It is straightforward to incorporate time-varying volatility in these estimates by substituting σ by σ_t, i.e., the conditional volatility estimate at time t. Moreover, we estimate the parameters of the Gaussian distribution and other parametric methods in each rolling window by Maximum Likelihood, which is an estimation method that selects the parameter estimates to maximise the likelihood that the observed sample is collected from the relevant distribution.

4.3.2 Student's t-distribution

The Student's t-distribution possesses a similar distributional form to the Gaussian distribution but consists of three parameters. Therefore, the pdf is:

$$L \sim ST(\mu, \sigma^*, \nu) \Longrightarrow f(\ell) = \frac{\Gamma[(\nu+1)/2]}{\sigma^* \sqrt{\nu\pi}\Gamma(\nu/2)} \left[1 + \frac{1}{\nu}\left(\frac{\ell - \mu}{\sigma^*}\right)^2\right]^{-(\nu+1)/2} \tag{4.8}$$

where Γ denotes the gamma function, ν denotes the degrees of freedom, σ^* denotes the scale parameter, and μ denotes the location parameter. Notice that μ is equivalent to the location parameter of the Gaussian distribution and σ^* is similar to the σ (i.e., the scale parameter of the Gaussian distribution), nevertheless, not equivalent. There is a defined relationship between σ and σ^* if and only if the degrees of freedom is greater than two. The relationship is captured by:

$$\sigma^* = \sqrt{\frac{\nu - 2}{\nu}} \cdot \sigma \tag{4.9}$$

As briefly discussed in Section 3, the Student's t-distribution with low degrees of freedom has a higher kurtosis than the Gaussian distribution. Therefore, the Student's t-distribution is better suited for capturing some of the stylised facts of financial time series, such as leptokurtosis. The closed-form equations for VaR and ES of the Student's t-distribution are given by:

$$VaR_{\alpha,t} = \mu + \sigma^* \cdot t_{\alpha,\nu} = \mu + \sqrt{\frac{\nu - 2}{\nu}} \cdot \sigma \cdot t_{\alpha,\nu} \tag{4.10}$$

$$\begin{aligned} ES_{\alpha,t} &= \mu + \sigma^* \cdot \frac{f_{std}(t_{\alpha,\nu})}{1 - \alpha} \left(\frac{\nu + t_{\alpha,\nu}^2}{\nu - 1}\right), \\ \text{where } f_{std} &= \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left[1 + \frac{1}{\nu} t_{\alpha,\nu}^2\right]^{-(\nu+1)/2} \end{aligned} \tag{4.11}$$

where $t_{\alpha,\nu}$ denotes the α quantile of a standardised Student's t-distribution with ν degrees of freedom. Analogous to the Gaussian distribution, it is straightforward to incorporate time-varying volatility utilising the conditional estimate instead of the unconditional volatility.

A deficiency of the Student's t-distribution is that it is not closed under convolution. Specifically, if $x \sim ST$ and $y \sim ST$, the sum of the two random variables x and y cannot follow a Student's t-distribution. Consequently, if continuously compounded daily returns follow a Student's t-distribution, weekly returns cannot pursue the same distributional form. The Gaussian distribution, on the other hand, is closed under convolution, which makes it a conducive distributional assumption when developing theory. Nevertheless, this limitation of the Student's t-distribution is negligible for our purposes since we implement one-day-ahead forecasts of VaR and ES for the backtesting period and are not interested in the multi-period return.

We restrict the degrees of freedom in the Maximum Likelihood estimate to be strictly greater than 2. This restriction is because the relationship between the scale parameter of the Student's t-distribution and the standard deviation in Equation (4.9) is stipulated if and only if this restriction holds.

4.3.3 Extreme Value Theory - Peaks over Threshold (POT)

The Extreme Value Theory (EVT) introduces an alternative approach to implement a parametric estimation of VaR and ES, with authors such as Gnedenko (1943) credited with the early development. Peaks over Threshold (POT) has become the favoured implementation of EVT in finance and, unlike the former parametric estimation methods presented, it does not explicitly pre-specify an underlying distribution, f, of the losses and instead endeavours to model the tails. Hence, EVT alleviates the adverse model risk inherent in implementation methodologies for the parametric estimation methods. It is straightforward to locate VaR by taking the inverse of the CDF, F, of this unspecified distribution as:

$$F(VaR_\alpha) = \alpha \Longrightarrow VaR_\alpha = F^{-1}(\alpha) \tag{4.12}$$

To find the inverse of the F, we define two separate events, A and B, where B denotes the event of some loss being greater than some threshold, u, and A denotes the event where the loss is smaller than $u + y$ where $y = \ell - u$. We specify the conditional probability of event A given event B as:

$$\left.\begin{array}{ll} A: & L \leq u + y \\ B: & L > u \end{array}\right\} \Longrightarrow Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

The conditional cumulative distribution function of A given B, $F_u(y)$, is described as:

$$F_u(y) = Pr(A \mid B) = Pr(L \leq u + y \mid L > u) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

$$F_u(\ell - u) = \frac{F(\ell) - F(u)}{1 - F(u)} \Longleftrightarrow F(\ell) = (1 - F(u))\, F_u(\ell - u) + F(u)$$

Furthermore, Balkema and de Haan (1974) and Pickands (1975) formally prove that if the defined threshold is sufficiently large, then the conditional cumulative distribution function can be approximated by the generalised Pareto distribution (GPD) (Marimoutou, Raggad and Trabelsi, 2009). We mathematically capture the approximation by the cumulative GPD of $F_u(\ell - u)$ as:

$$F_u(\ell - u) \approx G_{\xi,\beta}(\ell - u) = \begin{cases} 1 - \left(1 + \xi \frac{\ell - u}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{\ell - u}{\beta}\right), & \xi = 0 \end{cases} \tag{4.13}$$

where ξ is the parameter related to the right tail of the loss distribution and β is the parameter associated with the volatility of the (unspecified) underlying distribution (Marimoutou, Raggad and Trabelsi, 2009). The estimation of the parameters is implemented via the first derivative of Equation (4.13) and applying Maximum Likelihood (Hull, 2018, p. 308). However, there is a pragmatic trade-off to consider in the EVT method. Maximum Likelihood is an asymptotic estimator, implying that a large sample is required to derive an adequate estimate which argues for selecting a low threshold, u. On the other hand, the underlying EVT theory asserts that we can approximate with the GPD if and only if the threshold is 'sufficiently' large. Therefore, we utilise two separate thresholds in this thesis. When estimating VaR and ES at the 95% confidence level, we select the threshold as the loss at the 93^{rd} quantile. For the estimation at the 99% confidence level, we define the threshold as the loss at the 97^{th} quantile of the sample of losses. The rationale is that we desire to retain an adjacent threshold for each confidence level without discarding too many observations. The equations for estimating VaR and ES using the EVT method are given by:

$$VaR_{\alpha,t} = \begin{cases} u + \frac{\beta}{\xi}\left[\left(\frac{N}{N_u}(1-\alpha)\right)^{-\xi} - 1\right],\ \xi \neq 0 \\ u - \beta \ln\left(\frac{N}{N_u}(1-\alpha)\right),\ \xi = 0 \end{cases} \tag{4.14}$$

$$ES_{\alpha,t} = \begin{cases} \frac{VaR_{\alpha,t}+\beta-u\cdot\xi}{1-\xi},\ \xi \neq 0 \\ VaR_{\alpha,t}+\beta,\ \xi = 0 \end{cases} \tag{4.15}$$

where N_u denotes the number of large losses (i.e., losses above the threshold).

McNeil and Frey (2000) introduce the idea of conditional volatility in the EVT method to incorporate the stylised fact of time-varying volatility, referred to as the conditional EVT. Youssef, Belkacem and Mokni (2015, p. 103) outline the conditional EVT methodology in three steps: (i) fit a volatility model to estimate the conditional mean and standard deviation on the losses by Maximum Likelihood; (ii) standardise the residuals from the volatility model to make the assumption that they are white noise and apply the EVT method on the standardised residuals to compute the quantile at the germane confidence level; and (iii) use the estimated conditional mean and standard deviation with the results from step (ii) to get the estimates for the conditional EVT. Mathematically, we describe the conditional EVT by:

$$VaR_{\alpha,t+1} = \mu_{t+1} + \sigma_{t+1} \cdot VaR(q_\alpha) \tag{4.16}$$

$$ES_{\alpha,t+1} = \mu_{t+1} + \sigma_{t+1} \cdot ES(q_\alpha) \tag{4.17}$$

where $VaR(q_\alpha)$ and $ES(q_\alpha)$ denote the VaR and ES estimates, respectively, of the standardised residuals from the volatility model, μ_{t+1} denotes the conditional mean at time $t+1$ and σ_{t+1} denotes the conditional volatility at time $t+1$. We utilise the conditional EVT method in this thesis, consistent with previous research (cf. Section 2.5).

4.4 Volatility Forecasting

As discussed in the stylised facts of financial time series (cf. Section 2.4), volatility clustering suggests the presence of autocorrelation and nonlinear dependence in crude oil returns, i.e., violating the assumptions of conventional econometric models. To account for such violations, Engle (1982) proposes a class of models called Autoregressive Conditional Heteroscedasticity (ARCH) models. Suppose η_t is the stochastic innovation that impacts asset returns and is conditional on some past information, Ω_{t-1}, with some unspecified continuous distribution, f:

$$r_t = \mu + \eta_t$$

$$\eta_t \mid \Omega_{t-1} \sim f(0, \sigma_t^2)$$

The ARCH(q) model, where q refers to the lagged values of η_t, follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2$$

Bollerslev (1986) generalises the ARCH(q) process to a GARCH(p, q) model, where p denotes the order of autoregressive terms, σ_{t-j}^2, to facilitate the model to depend on past variances. The GARCH model is specified by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

Bollserslev (1986) also suggests that a GARCH(1,1) is highly effective in most volatility forecasting instances despite being the simplest form of the model with one lagged value of the squared stochastic innovation process, η_{t-1}^2, and one autoregressive term, σ_{t-1}^2:

$$\sigma_t^2 = \omega + \alpha \eta_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{4.18}$$

where the α parameter reflects the sensitivity of volatility to new innovations and the β parameter relates to the persistence of the effects of past shocks and defines the smoothness of the variance series.

Subsequently, we consider the mean equation utilised in the volatility models and present alternative specifications for a GARCH-type model and how they differ from the simple GARCH(1,1) model.

4.4.1 Conditional Mean

Autoregressive Moving Average (ARMA) model is often employed in practice and academia to derive and forecast the conditional mean. The ARMA model is specified as:

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{4.19}$$

where p and q are the orders of the AR and MA processes, respectively (Brooks, 2019). The AR and MA processes are commonly determined by observing the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) in correlograms (i.e., the ACF and PACF plots) (Tsay, 2005).

Figures (4.1) and (4.2) demonstrate a lack of clear AR or MA process for both WTI and Brent crude oil. However, ACF and PACF plots can be cumbersome to deduce accurate information from and, therefore, we employ the pmdarima package in Python to find the AR and MA orders to WTI and Brent. We observe an information criterion to determine the correct specifications for the ARMA orders, namely, Akaike's (1974) Information Criterion (AIC) and Schwarz' (1978) Bayesian

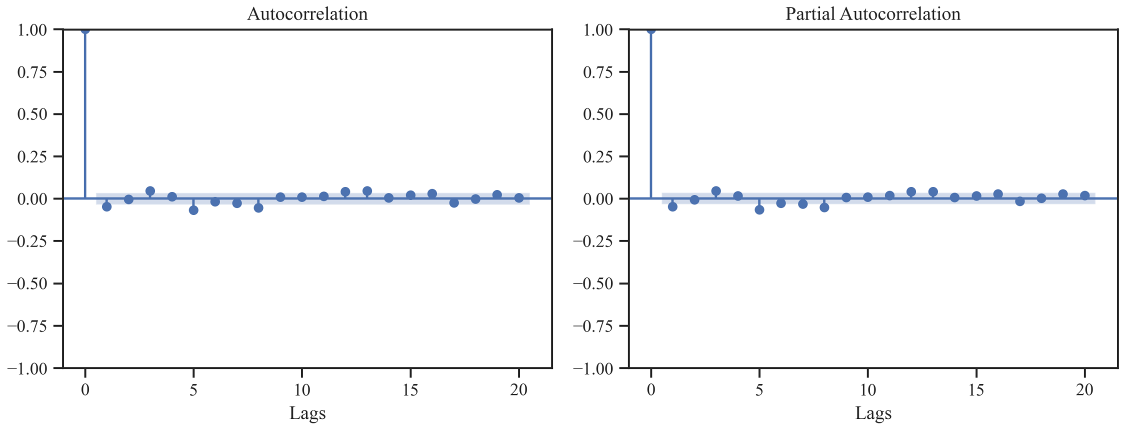


Figure 4.1: ACF and PACF plots for WTI crude oil based on returns from estimation period.

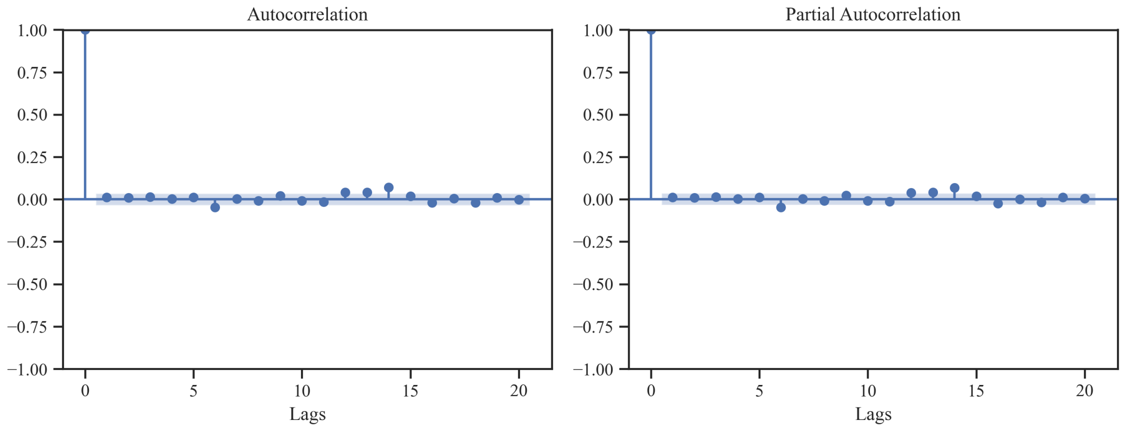


Figure 4.2: ACF and PACF plots for Brent crude oil based on returns from estimation period.

Information Criterion (BIC). According to Chakrabarti and Ghosh (2011, pp. 599-600) and Brooks (2019, pp. 275-276), the AIC tends to opt for more complex models than the BIC when selecting parameters to fit a specific dataset. That is, the AIC has a tendency to induce over-fitting, which implies modelling the randomness of the data. The model picked by the AIC may perform exemplary on in-sample data but ineffectively on out-of-sample data, whereas the BIC tends to select a model better suited for generalisation purposes. We denote the AIC and the BIC by:

$$AIC = \ln \hat{\sigma}^2 + \frac{2(p+q+1)}{T}$$
$$BIC = \ln \hat{\sigma}^2 + \frac{p+q+1}{T} \ln T$$

where p and q are simply the orders from the ARMA model, $\hat{\sigma}^2$ denotes the estimated residual variance, and T denotes the number of observations (Brooks, 2019).

Searching for the optimal ARMA model for the data from the estimation period of daily returns with the BIC and the AIC results in two different model specifications. The BIC returns a model without any AR or MA orders for both WTI and

Brent, which is in line with our expectations from Figure (4.1) and (4.2). Contrarily, the AIC criterion returns an ARMA(4,4) and an ARMA(0,0) for WTI and Brent, respectively. An ARMA model of these dimensions raises concerns for over-fitting given the size of our rolling window and supports the arguments of Chakrabarti and Ghosh (2011) and Brooks (2019) of the AIC choosing a model that is overly complex. Therefore, we assume that a constant mean model best describes the daily continuously compounded returns, consistent with the BIC estimates. We utilise these results for the volatility models considered in this thesis.

4.4.2 Exponentially Weighted Moving Average (EWMA)

Although GARCH-type models are strongly preferred in academia, practitioners often opt for more simplistic models for forecasting volatility. Specifically, the Exponentially Weighted Moving Average (EWMA) model is widely employed in practice and is a model that captures time-varying volatility. Hull and White (1998) utilise the EWMA model when developing their version of the VWHS. Interestingly, the EWMA model can be interpreted as a simplified GARCH(1,1) with parameters chosen in a specific manner. The EWMA model is described as:

$$\sigma_t^2 = (1-\lambda)\eta_{t-1}^2 + \lambda\sigma_{t-1}^2 \tag{4.20}$$

where σ_t^2 denotes the conditional variance at time t, η_t denotes the innovation at time t, and λ denotes the exponential decay factor. EWMA places higher weights on recent observations and lesser weights on older observations to capture the effects of the latest shocks while exponentially decaying the impact of preceding changes in volatility. The resemblance of a GARCH(1,1) is that EWMA puts the constant to zero and the coefficients as a function of the exponential decay factor. Hence, the EWMA model does not require any estimation, as the exponential decay factor is endowed with a value arbitrarily.

The simplicity of the EWMA model serves as a double-edged sword. The advantage of not having to estimate the parameters accompanies the cost of the exponential decay factor endowed with a suboptimal value. Nevertheless, RiskMetrics (1996) establishes $\lambda = 0.94$ for daily volatility estimates, which Hull (2018, p. 226) notes as the parameter value producing the closest estimates to the realised volatility rate. We embrace this practice in this thesis and put the exponential decay factor equal to 0.94. As aforementioned in Section 4.2.3, we utilise the EWMA model in conjunction with the VWHS because no parameter estimates are required and, thus, the VWHS remains entirely nonparametric.

4.4.3 GJR-GARCH

Consistent with the stylised facts of financial time series, Wei, Wang and Huang (2010) evaluate the leverage effects present in the time series of crude oil to better forecast VaR and ES. They corroborate that the financial time series is nonlinear and, therefore, a nonlinear GARCH specification is required to capture these leverage effects. The simple GARCH(1,1) specification is a symmetric model and does not account for these properties. Therefore, Glosten, Jagannathan and Runkle (1993) introduce a GARCH specification that captures the asymmetry in the effects of positive and negative shocks. The GJR-GARCH(1,1) model is given by:

$$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1})\eta_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{4.21}$$

where the γ parameter reflects the magnitude and direction of the leverage effects with the binary variable, I_{t-1}, which follows the condition:

$$I_{t-1} = \begin{cases} 1, & \text{if } \eta_{t-1} < 0 \\ 0, & \text{if } \eta_{t-1} \geq 0 \end{cases}$$

Under conventional leverage effects, adverse shocks define the binary variable as one and, therefore, increase the conditional variance in the following observation. However, as discussed in Section 2.4, spot prices of crude oil encounter inverse leverage effects where positive shocks accelerate volatility more than adverse shocks (Chen and Mu, 2021). After implementing the model for our data during the estimation period with Maximum Likelihood estimation, the γ parameter takes the expected negative sign, given the inverse leverage effects.

There are other specifications of the nonlinear GARCH model that capture leverage effects. Ding, Granger and Engle (1993) propose the Asymmetry Power ARCH (APARCH) model, which is a substitute for the GJR-GARCH and used in an equivalent manner for the purpose of capturing leverage effects. As discussed in Section 2.4, crude oil markets follow a long-memory process and the extension to the APARCH model, the Fractionally Integrated APARCH (FIAPARCH) model (Tse, 1998), captures the asymmetry along with the long-memory process for volatility forecasting. Yet, Wei, Wang, and Huang (2010) find that the GARCH specification is irrelevant as long as the specification is nonlinear for crude oil commodities concerning shorter-term horizon forecasts. Therefore, we disregard the GARCH-type specifications that model the long-memory processes in this thesis as we implement one-step-ahead forecasts.

In implementing this volatility forecasting model, we incorporate it within the Gaussian and Student's t-distribution for VaR and ES. Concerning the implementation of the conditional EVT, we adhere to the method by Marimoutou, Raggad and Trabelsi (2009) and utilise a simple GARCH(1,1). For the distributional assumptions of the innovations of the GARCH-type models, the parametric methods use their respective distributions except for the conditional EVT method, which utilises the Gaussian distribution.

4.5 Backtesting

Section 2.3.3 discusses that the best estimation method ought to yield the correct number of violations given the confidence level, α, and sample size, M. There are a myriad of backtests available to determine if the method performs adequately. Nonetheless, this thesis implements three predominant backtests for the backtesting period: the Kupiec (1995) test, the Christoffersen (1998) test, and the Acerbi and Szekely (2014) Test 2.

4.5.1 The Kupiec (1995) Test

Comparing the one-day-ahead estimate of VaR with the observed loss at time t, ℓ_t, denoting losses exceeding VaR with a 1 and 0 otherwise, we introduce a Bernoulli distributed random variable (Brooks, 2019, p. 113). From elementary statistics, the sum of a Bernoulli distributed variable over the backtesting sample is a binomial distributed random variable. Hence, the Kupiec (1995) test is a binomial test given by:

$$Pr(X \leq x) = \sum_{i=1}^{x} \binom{M}{i} p_0^{M-i} p_1^i$$

where $\binom{M}{i}$ denotes the binomial coefficient and p_1 equals the frequency of violations under the null hypothesis with p_0 being the complement. That is, p_1 under a correct method is equal to $1 - \alpha$. The Kupiec (1995) test counts the number of violations over a specified period and compares it to the expected number of violations given a correct VaR level (Gordy and McNeil, 2020). The simple implementation and interpretation of the Kupiec (1995) test make it a prevalent backtest in practical settings.

It is possible to investigate the Kupiec (1995) test utilising a likelihood ratio test, in which case, it is common to refer to it as a test for unconditional coverage. That is,

$$\begin{cases} H_0: & \pi_1 = p_1 = 1 - \alpha \\ H_1: & \pi_1 \neq p_1 = 1 - \alpha \end{cases}$$

$$LR_{uc} = -2(\ln \mathcal{L}_0 - \ln \mathcal{L}_1) = -2(\ln p_0^{n_0} p_1^{n_1} - \ln \pi_0^{n_0} \pi_1^{n_1}) \sim \chi^2(1) \tag{4.22}$$

where $\mathcal{L}_0$ and $\mathcal{L}_1$ denote the likelihood functions for the null and the alternative hypothesis, respectively, and π_0 and π_1 denote the observed probabilities of a non- and violation, respectively. Given that we only impose one restriction, this two-sided likelihood ratio test statistic follows a χ^2 distribution with one degree of freedom under the null hypothesis (Youssef, Belkacem and Mokni, 2015).

The Kupiec (1995) test encounters backlash from authors such as Escanciano and Pei (2012), providing evidence that the test yields inconsistencies in its endeavour to discover suboptimal nonparametric estimation methods of VaR forecasts. Nevertheless, we employ the Kupiec (1995) test due to its straightforward interpretation and extensive usage in practice.

4.5.2 The Christoffersen (1998) Test of Independence and Conditional Coverage

In contrast to the Kupiec (1995) test, the Christoffersen (1998) test is agnostic towards the number of violations and models the stochastic process of the Bernoulli distributed random variables as a two-state Markov Chain. The Christoffersen (1998) test inquires whether the probability of a violation differs, contingent on the observation of a violation at a previous time or not. The conditional probabilities of interest are denoted as $\pi_{11} = Pr(s_1|s_1)$ and $\pi_{01} = Pr(s_1|s_0)$ where s_i denotes state i and the violation is coded as a 1. A model that captures the dynamics of the data properly should not have a first-order Markov Chain (i.e., violations ought to be independently distributed over time) and, therefore, the null hypothesis, H_0, states that $\pi_{11} = \pi_{01} = \pi_1$. That is, the conditional probability of observing a violation is equal to its unconditional probability. The Christoffersen (1998) test is a two-sided likelihood ratio test following a χ^2-distribution with one degree of freedom under the null hypothesis denoted as:

$$\begin{cases} H_0: & \pi_{11} = \pi_{01} = \pi_1 \\ H_1: & \pi_{11} \neq \pi_{01} \neq \pi_1 \end{cases}$$

$$LR_{ind} = -2(\ln \mathcal{L}_0 - \ln \mathcal{L}_1) = -2(\ln \pi_0^{n_0}\pi_1^{n_1} - \ln \pi_{00}^{n_{00}}\pi_{01}^{n_{01}}\pi_{10}^{n_{10}}\pi_{11}^{n_{11}}) \sim \chi^2(1) \quad (4.23)$$

where LR_{ind} denotes the likelihood ratio of the independence test statistic, $\mathcal{L}_0$ and $\mathcal{L}_1$ denote the likelihood functions of the null and alternative hypothesis, respectively. The benefit of this test is that dependencies of violations over time mean that the probability of a violation, given a violation state, is greater than what is desirable. Thus, the user of a model with dependencies of violations over time endures a greater risk of imposing large losses amidst turbulent times than the expected level.

The Christoffersen (1998) test of independence may be merged with a Kupiec (1995) type test for unconditional coverage, in which case, a test for conditional coverage is implemented. The null hypothesis of the conditional coverage test dictates that the probability of a violation at time t is equal to $1 - \alpha$. The actual test is a likelihood ratio test following a χ^2 distribution with two degrees of freedom under the null hypothesis (Aloui and Mabrouk, 2010). Mathematically, we summarise the conditional coverage test as:

$$\begin{cases} H_0: & Pr(\text{violation}_t|\Omega_{t-1}) = 1 - \alpha \\ H_1: & Pr(\text{violation}_t|\Omega_{t-1}) \neq 1 - \alpha \end{cases}$$

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \quad (4.24)$$

where Ω_{t-1} denotes the information available at time $t-1$ (i.e., whether a violation is observed or not at time $t-1$), α denotes the confidence level and LR_{cc} denotes the likelihood ratio of the conditional coverage test.

There are limitations to the Christoffersen (1998) test as it models only first-order dependencies in the time series of Bernoulli-distributed random variables. The implication is that more complex dependencies remain undetected, for instance, a violation every Monday. More sophisticated backtests (e.g., Engle and Manganelli, 2004) amend this limitation considerably. Nevertheless, we utilise the Christoffersen (1998) test to be able to test for conditional coverage and produce results comparable to previous authors who deploy the same test (see e.g., Marimoutou, Raggad and Trabelsi, 2009; Aloui and Mabrouk, 2010; Chiu, Chuang and Lai, 2010; Abad and Benito, 2013). Hence, the Christoffersen (1998) test of independence and conditional coverage are simple models to test for dependencies in the time series of non- and violations.

4.5.3 The Acerbi and Szekely (2014) Test 2

Numerous backtesting models for ES have recently been developed, including models by authors such as Costanzino and Curran (2018), Du and Escanciano (2017), and

Gordy and McNeil (2020). However, this thesis utilises the backtest developed by Acerbi and Szekely (2014) because of its transparent connection to the mathematical definition of ES.[1] The authors employ the dual-representation of ES in McNeil, Frey and Embrechts (2015) for a continuous loss distribution, which is given by:

$$ES_{\alpha,t} = \frac{\mathbb{E}[L_t \cdot I_t]}{1-\alpha} = \mathbb{E}\left[L_t \mid L_t > VaR_{\alpha,t}\right], \ \forall t$$

$$\iff -\frac{\mathbb{E}[L_t \cdot I_t]/(1-\alpha)}{ES_{\alpha,t}} + 1 = 0, \ \forall t$$

where I_t is a dummy variable taking the value 1 if a VaR violation and 0 otherwise and $\mathbb{E}[\cdot]$ denotes the expectation operator. The test statistic, Z_2, is implemented by utilising the sample data of size M as:

$$Z_2 = -\frac{1}{M}\sum_{t=1}^{M} \frac{\ell_t \cdot I_t/(1-\alpha)}{ES^P_{\alpha,t}} + 1 \tag{4.25}$$

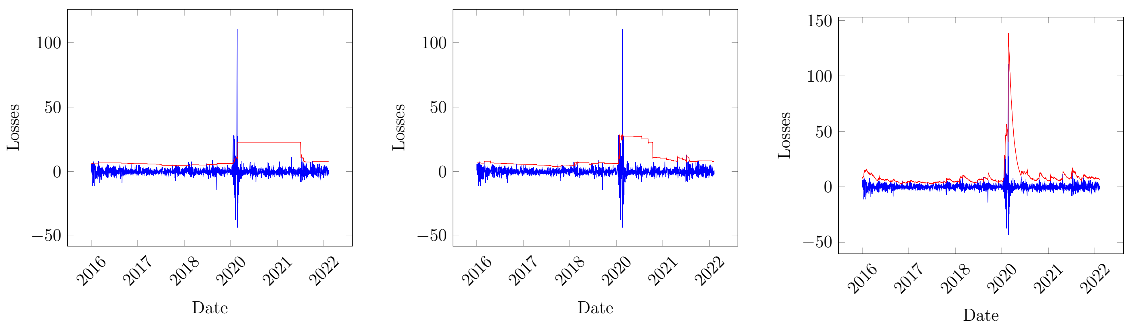
where $ES^P_{\alpha,t}$ denotes the estimated ES at the α confidence level at time t and ℓ_t denotes a realised loss at time t. Under the null hypothesis, the expected value of the test statistic, $\mathbb{E}[Z_2]$, is equal to zero, whilst the null hypothesis states that the expected value is less than zero (Acerbi and Szekely, 2014). Hence, this is a one-sided test for underestimation of ES.

A considerable disadvantage of the Acerbi and Szekely (2014) test in comparison to the previous tests of VaR is that the asymptotic distribution of the test statistic under the null hypothesis is not known (cf. Costanzino and Curran, 2018). Therefore, a Monte Carlo simulation is necessary to retrieve the critical values of the test statistic. Furthermore, a simple manipulation of Equation (4.25) reveals that the Acerbi and Szekely (2014) Test 2 is susceptible to the magnitude of the losses and the number of violations. The authors develop another test that is solely sensitive to the extent of the losses (i.e., Test 1). Contrary to Test 1, the critical values under Test 2 are more stable across different loss distributions (Acerbi and Szekely, 2014) and, therefore, this thesis utilises the latter test statistic and extracts the critical values from the original article. This introduces some stringent limitations and restrictions to our thesis. By employing the critical values directly from the Acerbi and Szekely (2014) paper, we are required to outline the backtesting sample according to the specification of the authors. Nevertheless, these limitations do not have major consequences for the generalisations of our results since the Acerbi and Szekely (2014) test is developed according to the Basel requirements. For uniformity between the backtesting schemes of VaR and ES, we backtest on approximately 250 observations of trading days for an annual backtesting sample. We apply this procedure for each calendar year, allowing us to investigate how a particular method performs between different years and risk levels.

[1] Any of the aforementioned models can be used with no change in credibility as a model is yet to be formally recognised by the Basel Committee on Banking Supervision.

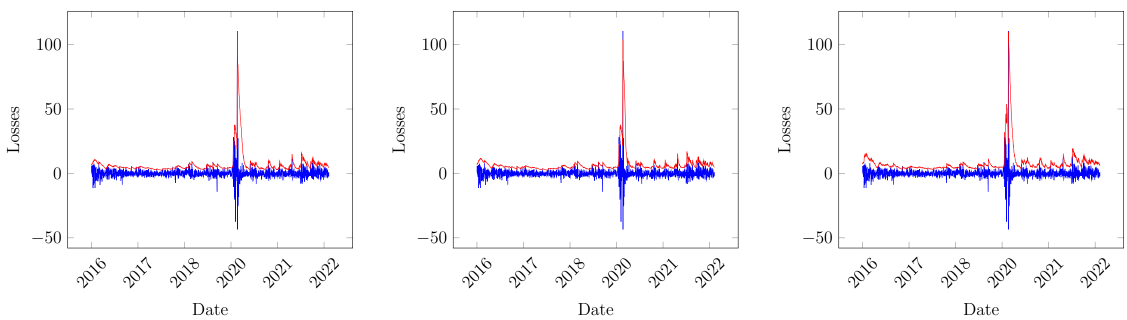
5 Results and Analysis

Implementing the models yields evidence that the different implementation methodologies generate vastly different estimates for VaR and ES, especially amidst market turmoil and post-pandemic. Appendix (C.1) to (D.2) presents an ocular overview of all models. However, an interesting comparison is observed by contrasting the different models at the 99% confidence level.



(a) *BHS: VaR at* $\alpha = 0.99$ (b) *AWHS: VaR at* $\alpha = 0.99$ (c) *VWHS: VaR at* $\alpha = 0.99$

Figure 5.1: Nonparametric estimation methods for VaR at the 99th quantile for WTI for the backtesting period.



(a) *Gaussian: VaR at* $\alpha = 0.99$ (b) *t: VaR at* $\alpha = 0.99$ (c) *Cond. EVT: VaR at* $\alpha = 0.99$

Figure 5.2: Parametric estimation methods for VaR at the 99th quantile for WTI for the backtesting period.

Figure (5.1) contributes to the inference that the simple nonparametric methods of BHS and AWHS seem to heavily overestimate VaR in the direct period after the pandemic, requiring an extended period to readjust to more appropriate levels. Nevertheless, the VWHS appears to be a simple but effective method to capture the dynamics of the data. Figure (5.2), on the other hand, showcases that the parametric methods do not generate any obvious over- or underestimation of VaR. The formal results of the backtests for VaR at the 95% and 99% confidence levels concerning WTI are rendered in Table (5.1).

		Confidence level: 95%			Confidence level: 99%		
	Year	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
BHS	2016	0.6367	0.4519	0.6740	0.7327	0.8264	0.9209
	2017	0.0042	0.6865	0.0154	0.0250	1.0000	0.0811
	2018	0.0761	0.0686	0.0395	0.0583	0.5561	0.1401
	2019	0.4812	0.7130	0.7293	0.3805	0.6865	0.6275
	2020	0.0006	0.0018	0.0000	0.0003	0.3720	0.0011
	2021	0.0133	0.0736	0.0094	0.0247	1.0000	0.0802
	2022	0.0815	0.5632	0.1855	0.7630	0.7555	0.9103
AWHS	2016	0.4363	0.3720	0.4958	0.2732	0.8994	0.5443
	2017	0.0366	0.5570	0.0947	0.0250	1.0000	0.0811
	2018	0.0761	0.0686	0.0395	0.0583	0.5561	0.1401
	2019	0.6571	0.4552	0.6856	0.7419	0.8257	0.9245
	2020	0.0260	0.0004	0.0002	0.3880	0.0415	0.0863
	2021	0.0805	0.1671	0.0836	0.2756	0.8992	0.5475
	2022	0.3370	0.7157	0.5902	0.7373	0.8261	0.9227
VWHS	2016	0.1553	0.4418	0.2710	0.0244	1.0000	0.0794
	2017	0.8853	0.2146	0.4582	0.1619	0.6203	0.3325
	2018	0.0230	0.0437	0.0099	0.3767	0.6859	0.6235
	2019	0.4529	0.0488	0.1083	0.7419	0.8257	0.9245
	2020	0.4363	0.3720	0.4958	0.1662	0.0732	0.0771
	2021	0.6799	0.1825	0.3777	0.7630	0.7555	0.9103
	2022	0.8726	0.2526	0.5131	0.7630	0.7555	0.9103
Gaussian	2016	0.4363	0.3398	0.4683	0.0244	1.0000	0.0794
	2017	0.2860	0.3869	0.3893	0.2781	0.8990	0.5508
	2018	0.6674	0.2916	0.5229	0.7466	0.8254	0.9263
	2019	0.1632	0.4400	0.2808	0.7419	0.8257	0.9245
	2020	0.2738	0.2963	0.3185	0.3880	0.0415	0.0863
	2021	0.0010	0.7555	0.0041	0.0247	1.0000	0.0802
	2022	0.0010	0.7555	0.0041	0.0247	1.0000	0.0802
Student's t	2016	0.6367	0.2947	0.5165	0.0244	1.0000	0.0794
	2017	0.2860	0.3869	0.3893	0.2781	0.8990	0.5508
	2018	0.1292	0.0848	0.0716	0.2806	0.8988	0.5542
	2019	0.6571	0.0741	0.1839	0.7419	0.8257	0.9245
	2020	0.2261	0.1033	0.1276	0.3880	0.0415	0.0863
	2021	0.0805	0.4976	0.1725	0.0247	1.0000	0.0802
	2022	0.4446	0.3388	0.4725	0.0247	1.0000	0.0802
C. EVT	2016	0.9084	0.2166	0.4630	0.2732	0.8994	0.5443
	2017	0.8853	0.2146	0.4582	0.1619	0.6203	0.3325
	2018	0.1292	0.0848	0.0716	0.3767	0.6859	0.6235
	2019	0.8853	0.1484	0.3482	0.7419	0.8257	0.9245
	2020	0.5001	0.0065	0.0197	0.0614	0.1149	0.0502
	2021	0.4446	0.3388	0.4725	0.7630	0.7555	0.9103
	2022	0.8969	0.2156	0.4606	0.7373	0.8261	0.9227

Table 5.1: Backtesting results of p-value for VaR regarding non- and parametric methods for WTI crude oil. A method is rejected for a given year if it produces a p-value less than 5%

At the 95% confidence level for WTI crude oil, the BHS produces a correct number of violations for the majority of the years, evident from the unconditional coverage test. However, amid and post-2020, the performance of the BHS deteriorates, where we observe considerably more violations than expected. In 2020, we reject the hypothesis of independent violations and retrieve a misspecified conditional coverage. It also produces a misspecified conditional coverage for 2021, however, we do not reject the hypothesis of independent violations. The AWHS produces similar results at the 95% confidence level, although, marginally improved. On the other hand, the VWHS generates only one rejection at the 95% confidence level for all three tests, specifically, in 2018. This means that the VWHS performs reasonably well on the unconditional and conditional coverage test mid- and post-2020. At the 99% confidence level, the same tendencies are prevalent for all three nonparametric methods. However, an interesting and unexpected observation is that the

BHS performs reasonably well on the independence test. The only time we reject independence of violations for the BHS is at the 95% confidence level during 2020. Nevertheless, this is presumably due to more complex dependencies that are not captured by a simple Christoffersen (1998) test of independence.

Investigating the backtesting results of VaR for WTI for the parametric estimation methods, as expected, the Gaussian assumption performs the worst at both the 95% and 99% confidence levels for the unconditional coverage test. The Student's t-distribution, on the other hand, performs very well in the unconditional and conditional coverage tests at the 95% confidence level. At the 99% confidence level, the performance of the Student's t-distribution deteriorates post-2020, possibly due to an overestimation of the unconditional coverage. Conditional EVT does not reject the unconditional coverage at all for both confidence levels. However, it does reject independence and conditional coverage at the 95% confidence level in 2020. This does not carry over at the 99% confidence level. Nevertheless, it is in accordance with our expectations that the conditional EVT displays a superior performance at very high quantiles, considering that it is a method developed for modelling extreme risk.

Comparing the nonparametric with the parametric backtesting results for VaR of WTI crude oil at the 99% confidence level, the AWHS and the VWHS perform slightly better than the Gaussian and Student's t-distribution for all three backtests. Yet, they perform marginally worse than the conditional EVT regarding the unconditional and conditional coverage and the independence test. It is interesting to note that the AWHS and the VWHS seem to handle 2020 and post-2020 equally well as the conditional EVT at the 99% confidence level. At the 95% confidence level, on the other hand, the AWHS displays an inferior performance in 2020 for all three backtests. The VWHS still performs well and is a solid contender for conditional EVT. Overall, the different backtests demonstrate that a parametric estimation method performs best (i.e., the conditional EVT), but it is noteworthy that the VWHS displays a robust performance at both confidence levels and even transcends some of the parametric methods. Therefore, the initial result indicates that practitioners should employ conditional EVT to estimate VaR, whilst the VWHS appears to be an uncomplicated substitute that yields robust results.

Table (5.2) provides the results of the same backtests for the VaR of Brent to investigate if the same results also hold for this crude oil commodity.

Regarding the unconditional and conditional coverage at the 95% confidence level in Table (5.2), the BHS is rejected for the majority of the years, failing to yield a correct estimate of VaR. It performs marginally better for the unconditional and conditional coverage at the 99% confidence level but is unsuccessful in managing the extreme change in volatility at the commencement of the pandemic. The AWHS provides an improvement to the BHS at both confidence levels but still displays a suboptimal performance amid the pandemic in 2020. At the 99% confidence level for 2020, it fails to produce the correct conditional coverage when the risk of the underlying change drastically. Similar to WTI, the VWHS for Brent crude oil appears to capture the dynamics of the data and yields a reasonable VaR level in most of the years at both confidence levels. The conditional coverage test is only rejected once, videlicet, at the 95% confidence level in 2018. Therefore, the VWHS replicates its effectiveness for Brent.

Analogous to WTI, the Gaussian distribution performs reasonably well on the

		Confidence level: 95%			Confidence level: 99%		
	Year	LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
BHS	2016	0.3682	0.7185	0.6251	0.7829	0.7574	0.9179
	2017	0.0008	0.7578	0.0034	0.0233	1.0000	0.0763
	2018	0.3446	0.7165	0.5990	0.0614	0.5586	0.1465
	2019	0.3844	0.0694	0.1318	0.4071	0.6906	0.6552
	2020	0.0038	0.0157	0.0008	0.0000	0.1031	0.0000
	2021	0.0038	0.6883	0.0139	0.0241	1.0000	0.0787
	2022	0.0135	0.6661	0.0432	0.7327	0.8264	0.9209
AWHS	2016	0.7235	0.6688	0.8572	0.7190	0.8274	0.9153
	2017	0.0302	0.5617	0.0807	0.0233	1.0000	0.0763
	2018	0.0843	0.5660	0.1911	0.0614	0.5586	0.1465
	2019	0.5876	0.0702	0.1675	0.7928	0.7583	0.9214
	2020	0.0930	0.1817	0.1000	0.0211	0.1642	0.0266
	2021	0.0333	0.5594	0.0875	0.2708	0.8996	0.5410
	2022	0.5001	0.7122	0.7442	0.7327	0.8264	0.9209
VWHS	2016	0.0718	0.5011	0.1578	0.0236	1.0000	0.0771
	2017	0.5972	0.2987	0.5068	0.7145	0.8278	0.9134
	2018	0.0135	0.6661	0.0432	0.3880	0.6877	0.6355
	2019	0.2452	0.2903	0.2911	0.7928	0.7583	0.9214
	2020	0.8278	0.5273	0.7998	0.3995	0.0410	0.0869
	2021	0.4282	0.3408	0.4641	0.7730	0.7564	0.9142
	2022	0.3446	0.7165	0.5990	0.7327	0.8264	0.9209
Gaussian	2016	0.1440	0.4446	0.2567	0.0236	1.0000	0.0771
	2017	0.4045	0.3662	0.4697	0.0233	1.0000	0.0763
	2018	0.5001	0.7122	0.7442	0.7680	0.7560	0.9123
	2019	0.2452	0.2903	0.2911	0.7100	0.8281	0.9115
	2020	0.4123	0.3676	0.4762	0.7829	0.0191	0.0619
	2021	0.0038	0.6883	0.0139	0.0241	1.0000	0.0787
	2022	0.0039	0.6877	0.0144	0.0244	1.0000	0.0794
Student's t	2016	0.2564	0.3918	0.3639	0.0236	1.0000	0.0771
	2017	0.4045	0.3437	0.4513	0.0233	1.0000	0.0763
	2018	0.0843	0.5321	0.1853	0.7327	0.8264	0.9209
	2019	0.3968	0.3648	0.4632	0.2613	0.9004	0.5279
	2020	0.0535	0.5902	0.1342	0.3995	0.0410	0.0869
	2021	0.1514	0.4428	0.2662	0.0241	1.0000	0.0787
	2022	0.0344	0.5586	0.0898	0.0244	1.0000	0.0794
C. EVT	2016	0.8278	0.5273	0.7998	0.2660	0.9000	0.5344
	2017	0.4045	0.3437	0.4513	0.7145	0.8278	0.9134
	2018	0.2261	0.1060	0.1302	0.1662	0.6217	0.3396
	2019	0.8058	0.0113	0.0393	0.4071	0.6906	0.6552
	2020	0.0535	0.5902	0.1342	0.0646	0.1134	0.0518
	2021	0.2679	0.3899	0.3740	0.7730	0.7564	0.9142
	2022	0.6907	0.6727	0.8450	0.2732	0.8994	0.5443

Table 5.2: Backtesting results of p-value for VaR regarding non- and parametric methods for Brent crude oil. A method is rejected for a given year if it produces a p-value less than 5%

unconditional coverage for Brent at the 95% and 99% confidence levels pre-pandemic, which is consistent with our expectations given a relatively stable period. However, the performance of the Gaussian distribution deteriorates in 2021 and 2022 and even generates the wrong conditional coverage at the 95% confidence level. The Student's t-distribution exhibits similar tendencies as for WTI at the lower confidence level. A noteworthy observation regarding the unconditional coverage for Brent is that we reject the Student's t-distribution in the majority of the years at the higher confidence level, whilst the other two backtests remain conducive. On the other hand, conditional EVT performs extraordinarily well for Brent, especially at a higher confidence level. More specifically, at the 99% confidence level, we do not reject any of the three tests applied to VaR. It also controls the risk effectively when the underlying risk level changes dramatically at the commencement of the pandemic in 2020. Therefore, it is apparent that the conditional EVT is well-suited for estimating

VaR for crude oil at high confidence levels, corresponding to results for WTI.

Comparing the nonparametric and parametric estimation methods for Brent, the Gaussian and Student's t-distribution underperform the BHS and the AWHS at the 99% confidence level regarding the unconditional coverage but marginally transcend them regarding the conditional coverage test. At the lower confidence level, the Gaussian and Student's t-distribution are superior, at least at the margin. The conditional EVT displays a dominant performance at both confidence levels and all three tests. Nevertheless, the VWHS appear to be only marginally inferior to the conditional EVT at both confidence levels, which indicates that it is a robust alternative to the more preeminent parametric method for estimating VaR concerning crude oil commodities. All in all, the backtesting results of VaR for Brent are analogous to WTI. Therefore, our findings demonstrate that the conditional EVT is particularly conducive for estimating VaR concerning crude oil, which is robust for an extended backtesting period and across commodities.

The results of the Acerbi and Szekely (2014) Test 2 for the nonparametric and parametric estimation methods for both WTI and Brent are displayed in Table (5.3). Nonetheless, we reiterate that our procedure concerning the implementation of the Acerbi and Szekely (2014) Test 2 is a test for underestimation only, unlike the backtests applied to VaR.

The BHS is not rejected in the majority of the backtesting years at either confidence levels for the Acerbi and Szekely (2014) Test 2 for underestimation. However, we strongly reject a correct ES level in 2020 at both confidence levels. Hence, the BHS strongly underestimates ES as the risk abruptly increases for WTI and Brent crude oil. The AWHS displays a similar tendency to the BHS for both commodities and is strongly rejected when the underlying riskiness suddenly increases at the commencement of the pandemic. Nevertheless, the AWHS exhibits an improvement in comparison to the BHS. The VWHS provides only one rejection at the 95% confidence level for both WTI and Brent. However, there is evidence of underestimation generated by the VWHS at the higher confidence level, especially in 2020 for both crude oil commodities. None of the nonparametric estimation methods displays a robust performance for estimating ES for the duration of the backtesting period and across confidence levels. Therefore, computing ES with a nonparametric methodology is suboptimal and may induce severe underestimation amidst market turmoil.

As can be inferred from Table (5.3), we do not reject the Student t-distribution at all for the backtesting period for either WTI or Brent crude oil. The Gaussian distribution displays an equivalent performance with a few exceptions at the commencement of the pandemic in 2020. Given the backtesting results of VaR, the worst overall performance in the category of parametric estimation methods for ES is unexpectedly the conditional EVT. It underestimates ES in 2020 for WTI and Brent crude oil at both confidence levels. Overall, the robust performance of the Student's t-distribution suggests that the parametric estimation methods are especially preeminent for a volatile commodity, such as Brent and WTI crude oil.

Comparing the nonparametric and parametric estimation methods for ES, it is apparent that the Student's t-distribution and Gaussian distribution transcend all the nonparametric estimation methodologies. A noteworthy observation is that the VWHS and conditional EVT produce a comparable backtesting performance across WTI and Brent crude oil for the backtest concerning ES. This finding further

		Confidence level: 95%		Confidence level: 99%	
	Year	Brent	WTI	Brent	WTI
BHS	2016	-0.2413	0.1335	-0.0043	0.3352
	2017	0.7895	0.7167	1.0000	1.0000
	2018	-0.4356	-0.6897	-1.5909	-1.7572
	2019	-0.2355	-0.1873	-0.5170	-0.5678
	2020	-2.1784	-2.5733	-5.8216	-5.6590
	2021	0.7891	0.8078	1.0000	1.0000
	2022	-0.4561	-0.1909	0.4182	0.1868
AWHS	2016	-0.0264	0.2645	0.3365	0.6340
	2017	0.5873	0.5747	1.0000	1.0000
	2018	-0.6119	-0.6838	-1.4422	-1.6395
	2019	0.1266	0.1078	-0.1077	0.1715
	2020	-1.3764	-1.6789	-2.9271	-2.7784
	2021	0.6747	0.7050	0.7528	0.8005
	2022	-0.0512	0.0165	0.3668	0.5003
VWHS	2016	0.5089	0.4922	1.0000	1.0000
	2017	0.1462	-0.1614	0.2540	-1.0573
	2018	-0.8123	-0.6653	-0.8389	-0.6155
	2019	0.2496	0.1330	-0.1178	-0.0266
	2020	-0.1662	-0.1361	-1.1422	-1.4335
	2021	0.1117	0.0398	-0.2541	0.0350
	2022	-0.0836	0.1090	0.3333	0.1666
Gaussian	2016	0.4131	0.2957	1.0000	1.0000
	2017	0.2666	0.2718	1.0000	0.6087
	2018	-0.1448	0.1086	-0.1383	0.2384
	2019	0.2977	0.3674	0.2562	0.2303
	2020	0.1636	0.1908	-0.2469	-0.7068
	2021	0.7070	0.7795	1.0000	1.0000
	2022	0.7221	0.7913	1.0000	1.0000
Student's t	2016	0.3769	0.2719	1.0000	1.0000
	2017	0.3022	0.3004	1.0000	0.6558
	2018	-0.3382	-0.2968	0.3184	0.6387
	2019	0.2849	0.2034	0.6705	0.3097
	2020	-0.4214	-0.2615	-0.4145	-0.4082
	2021	0.4722	0.5610	1.0000	1.0000
	2022	0.6091	0.3981	1.0000	1.0000
C. EVT	2016	0.1866	0.1400	0.7045	0.7088
	2017	0.2595	-0.0639	0.3675	-0.6242
	2018	-0.4032	-0.4058	-1.0223	-0.6493
	2019	0.0157	-0.0663	-0.5449	-0.0805
	2020	-1.0243	-0.7473	-2.8966	-2.6213
	2021	0.1447	0.1905	-0.3831	-0.0637
	2022	-0.0057	0.1728	0.5320	0.5035

Table 5.3: Backtesting results of the test statistic of the Acerbi and Szekely (2014) Test 2. For simplicity, a method is rejected at the 5% level of significance in a given year if the test statistic is less than -0.70. NB: the test statistic equals exactly one whenever no violations are observed in a given year.

supports the proposal that the VWHS is an uncomplicated substitute for conditional EVT. Nevertheless, as evident from the Acerbi and Szekely (2014) Test 2, the Student's t-distribution generates the most rigorous ES estimates across confidence levels and crude oil commodities. Therefore, this finding further supports that the parametric estimation methods excel the nonparametric methods concerning crude oil, which exhibits periods of extreme and nuanced volatility levels.

Subsequent to presenting the results from our backtests on VaR and ES, how do our findings compare to previous research? Regarding the VaR estimates for the parametric estimation methods, in line with Marimoutou, Raggad and Trabelsi (2009) and Youssef, Belkacem and Mokni (2015), we find that the conditional EVT based on a simple GARCH(1,1) model performs extraordinarily well for both commodities throughout the backtesting period. In contrast to Chiu, Chuang and Lai

(2010), our findings demonstrate that the VWHS slightly underperforms the conditional EVT, which is robust across crude oil commodities and for the duration of the backtesting period. Interestingly, the nonparametric estimation methods regarding the AWHS and the VWHS display a more robust performance than the Student's t-distribution at the 99% confidence level for unconditional coverage. The AWHS and the VWHS swiftly readjust themselves to the appropriate risk level in 2021 and 2022, as far as VaR is concerned, whilst the Student's t-distribution does not adjust adequately. Future research can apply a comparison of the skewed Student's t-distribution as this might better describe the underlying data than an ordinary Student's t-distribution and generate more robust estimates for VaR (see e.g., Aloui and Mabrouk, 2010). Overall, our findings for VaR contrast with authors such as Brooks and Persand (2002) and Sadorsky (2006) since there is an advantage for the parametric estimation methods given the strong results of the conditional EVT.

The conclusions from our backtest of ES contrast with those of the VaR backtests and, equivalently, contrast with previous research. We find that the Student's t-distribution display the most robust performance across commodities and for the duration of the backtesting period. This finding is possibly due to the leptokurtosis of the Student's t-distribution with low degrees of freedom, which is especially prevalent mid- and post-2020. In contrast to Marimoutou, Raggad and Trabelsi (2009), we find evidence that the conditional EVT intermittently underestimates ES, which is overstated around the commencement of the pandemic. Hence, our findings demonstrate that the backtesting performance of the conditional EVT is equivalent to the VWHS, except that the VWHS appears to marginally excel over the conditional EVT at the 95% confidence level in 2020. These results are more consistent with Chiu, Chuang and Lai (2010), who find similar performance tendencies between the VWHS and the conditional EVT. However, the results of the backtesting performance regarding conditional EVT for ES are plausibly due to the volatility model and not inherent to the model itself. Future research can investigate these results for other commodities and periods and compare the performance of different GARCH-type models applied to EVT. In contrast to Brooks and Persand (2002) and Sadorsky (2006), all parametric estimation methods for ES transcend the nonparametric BHS and AWHS for the duration of the backtesting period, indicating that these methods are suboptimal for estimating ES concerning a volatile commodity such as crude oil. However, Boudoukh, Richardson and Whitelaw (1998) champion a relatively small weight of old observations relative to new ones. A smaller value of the exponential decay factor might have induced a more robust result for the AWHS. Future research may consider providing a similar comparison but with a smaller exponential decay factor for the AWHS. Additionally, it is vital to remember that our implementation of the Acerbi and Szekely (2014) test does not consider the overestimation of ES. Future research should amend the backtest to permit a two-sided test to provide a more robust conclusion of the performance of ES. Overall, our findings for ES are consistent with authors such as Abad and Benito (2013) since the parametric estimation methods yield better estimates during both stable and turbulent periods.

It is ambiguous whether our findings translate to other periods. For instance, Chiu, Chuang and Lai (2010) find that nonparametric methods excel over parametric peers for a sub-period of the horizon utilised by Marimoutou, Raggad and Trabelsi (2009), who make an equivalent conclusion to ours. Therefore, it is inconclusive whether similar findings transfer to shorter periods. Nevertheless, as we exploit a relatively extensive backtesting period in this thesis, we expect that the parametric estimation methods transcend nonparametric over an extended period. This suggestion is apparent by the finding that conditional EVT performs extremely well pre-, mid-, and post-pandemic, whilst most of the nonparametric estimation methods struggle considerably at the commencement of the pandemic. Previous research on parametric estimation methods, specifically, conditional EVT (see e.g., Marimoutou, Raggad and Trabelsi, 2009; Youssef, Belkacem and Mokni, 2015), indicates that our results translate to securities which exhibit similar features to crude oil. We expect the parametric estimation methods to be particularly conducive for asset classes prone to sudden and extreme changes in the underlying risk levels, such as cryptocurrency. Nevertheless, since our results demonstrate that nonparametric estimation methods display an adequate performance pre-pandemic for the majority of the backtests, we expect that differences between the estimation methods are negligible for asset classes which are stable over extended periods of time, such as Treasury bonds.

To sum up, we find that a parametric method displays superior performance to the nonparametric estimation methods for VaR and ES. This finding is especially apparent for ES. More specifically, our findings contradict the conclusion of Brooks and Persand (2002) that uncomplicated estimation methods are superior to more complex methodologies. We find that the conditional EVT is the most robust estimation method for VaR, consistent across WTI and Brent. Concomitantly, the Student's t-distribution markedly outperforms the nonparametric methods concerning the estimation of ES. The VWHS appears to be the dominant methodology in the category of nonparametric estimation methods and transcends the BHS and AWHS. However, our findings suggest that practitioners ought to implement parametric methods to estimate the risk of commodities that exhibit equivalent features to crude oil. Practitioners generate biased estimates of VaR and ES that induce periods of solvency issues by implementing inferior estimation methods. Our findings demonstrate that parametric estimation methods are least likely to induce misspecification of the risk level.

6 Conclusion

In this thesis, we contrast the performance of nonparametric and parametric estimation methods of VaR and ES for crude oil commodities amidst a challenging backtesting period with major geopolitical events such as the COVID-19 pandemic. WTI displays a negative price due to the COVID-19 pandemic, which is extraordinary for a commodity. The nonparametric estimation methods entail the basic historical simulation, age-weighted historical simulation, and volatility-weighted historical simulation. The parametric estimation methods include the Gaussian distribution, the Student's t-distribution, and the conditional EVT. Our findings demonstrate that parametric estimation methods transcend those of the nonparametric methods in backtests for unconditional and conditional coverage, independence of violations and ES. This finding is especially prevalent mid- and post-2020. Specifically, the conditional EVT is the most robust estimation method for VaR, while the Student's t-distribution is superior for estimating ES. The BHS and AWHS display significant issues in dealing with drastic changes in the risk of the underlying commodity. The VWHS captures these changes in a more satisfactory manner in comparison to its nonparametric peers and outperforms the Student's t-distribution and the Gaussian distribution in estimating VaR at both confidence levels post-2020. Albeit the nonparametric estimation methods are inferior to more sophisticated parametric methods, the VWHS is an uncomplicated substitute of conditional EVT that captures the dynamics of the underlying data and yields adequate estimates of VaR at both confidence levels. The results for ES differ due to the extremely robust performance of the Student's t-distribution for the duration of the backtesting period. This result further supports the finding that parametric estimation methods are generally superior to nonparametric methodologies. These results are crucial given the amendments to Basel IV, in which the derivation of capital requirement calculations emanates from ES instead of VaR.

Following our findings, we recommend that practitioners seize their reliance on nonparametric methods (i.e., historical simulation) and adopt the parametric methods championed by academia. That is, a challenging commodity requires a sophisticated estimation method and elementary nonparametric estimation methods are not equipped to deal with sudden changes in risk. Such is evident from the independence test in 2020 at the 95% confidence level, simple nonparametric methods provide dependent violations which are detrimental to the ability of financial institutions to remain solvent during market turmoil. Hence, the parametric estimation methods are more robust across commodities and backtesting years for both VaR and ES and should, consequently, be the preferred estimation method among practitioners.

Reference List

Abad, P. and Benito, S. (2013). A detailed comparison of value at risk estimates, *Mathematics and Computers in Simulation*, vol. 94, pp. 258-276.

Acerbi, C., Nordio, C. and Sirtori, C. (2001). Expected Shortfall as a Tool for Financial Risk Management [pdf]. Available at: https://doi.org/10.48550/arXiv.cond-mat/0102304 [Accessed 5 April 2023].

Acerbi, C. and Scandolo, G. (2008). Liquidity risk theory and coherent measures of risk, *Quantitative Finance*, vol. 8, no. 7, pp. 681-692.

Acerbi, C. and Szekely, B. (2014). Backtesting Expected Shortfall: Introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall [pdf]. Available at: https://dlu-umich.github.io/docs/Research_Insight_Backtesting_Expected_Shortfall_December_2014.pdf [Accessed 28 March 2023].

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall, *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1487-1503.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723.

Aloui, C. and Mabrouk, S. (2010). Value-at-Risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models, *Energy Policy*, vol. 38, no. 5, pp. 2326-2339.

Angelidies, T., Benos, A. and Degiannakis, S. (2007). A robust VaR model under different time periods and weighting schemes, *Review of Quantitative Finance and Accounting*, vol. 28, no. 2, pp. 187-201.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent Measures of Risk, *Mathematical Finance*, vol. 9, no. 3, pp. 203-228.

Balkema, A.A. and de Haan, L. (1974). Residual Life Time at Great Age, *The Annals of Probability*, vol. 2, no. 5, pp. 792-804.

Basak, S. and Shapiro, A. (2001). Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices, *The Review of Financial Studies*, vol. 14, no. 2, pp. 371-405.

Basel Committee on Banking Supervision. (2023a). MAR30 - Internal models approach: general provisions [pdf], Available at: https://www.bis.org/basel_framework /chapter/MAR/30.htm [Accessed 25 April 2023].

Basel Committee on Banking Supervision. (2023b). MAR33 - Internal models approach: capital requirements calculation [pdf], Available at: https://www.bis.org/ basel_framework/chapter/MAR/33.htm [Accessed 25 April 2023].

Basel Committee on Banking Supervision. (2023c). MAR99 - Guidance of use of the internal models approach [pdf], Available at: https://www.bis.org/basel_framework/ chapter/MAR/99.htm [Accessed 25 April 2023].

Beder, T.S. (1995). VaR: Seductive but Dangerous, *Financial Analysts Journal*, vol. 51, no. 5, pp. 12-24.

Berkowitz, J. and O'Brien, J. (2002). How Accurate are Value-at-Risk Models at Commercial Banks?, *The Journal of Finance*, vol. 57, no. 3, pp. 1093-1111.

Black, F. (1976). Studies of Stock Price Volatility Changes, Proceedings of the 1976 Meeting of the Business and Economics Statistics Section, American Statistical Association.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, vol. 31, no. 3, pp. 307-327.

Boudoukh, J., Richardson, M.P and Whitelaw, R. (1998). The best of both worlds: a hybrid approach to calculating Value at Risk, SSRN Scholarly Paper, 51420, Available Online: https://papers.ssrn.com/abstract=51420 [Accessed 2 April 2023].

Brooks, C. (2019). Introductory econometrics for finance, 4th edn, Cambridge: Cambridge University Press.

Brooks, C. and Persand, G. (2002). Model Choice and Value-at-Risk Performance, *Financial Analysts Journal*, vol. 58, no. 5, pp. 87-97.

Campbell, A. and Smith, D.R. (2022). An empirical investigation of the quality of Value-at-Risk disclosure in Australia, *Accounting & Finance*, vol. 62, no. 1, pp. 469-491.

Chakrabarti, A. and Ghosh, J.K. (2011). AIC, BIC and Recent Advances in Model Selection, in P.S. Bandyopadhyay and M.R. Forster (eds), *Philosophy of Statistics*, Amsterdam: North-Holland, pp. 583-605. Available Online: https://www.sciencedirect.com/science/article/pii/B9780444518620500186 [Accessed 3 April 2023].

Chen, Y.F. and Mu, X. (2021). Asymmetric volatility in commodity markets, *Journal of Commodity Markets*, vol. 22, pp. 1-19.

Chiu, Y.C., Chuang, I.Y. and Lai, J.Y. (2010). The performance of composite forecast models of Value-at-Risk in the energy market, *Energy Economics*, vol. 32, no. 2, pp. 423-431.

Chkili, W., Hammoudeh, S. and Nguyen, D.K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long-memory, *Energy Economics*, vol. 41, pp. 1-18.

Choi, K. and Hammoudeh, S. (2009). Long-Memory in Oil and Refined Products Markets, *The Energy Journal*, vol. 30, no. 2, pp. 97-116.

Christoffersen, P.F. (1998). Evaluating Interval Forecasts, *International Economic Review*, vol. 39, no. 4, pp. 841-862.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance*, vol. 1, no. 2, pp. 223-236.

Costanzino, N. and Curran, M. (2018). A Simple Traffic Light Approach to Backtesting Expected Shortfall, *Risks*, vol. 6, no. 1, p. 2.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427-431.

Ding, Z., Granger, C.W.J and Engle, R.F. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, vol. 1, no. 1, pp. 83-106.

Dowd, K. (2002). Measuring market risk, Chichester: Wiley Finance Series.

Du, Z. and Escanciano, J.C. (2017). Backtesting Expected Shortfall: Accounting for Tail Risk, *Management Science*, vol. 63, no. 4, pp. 940-958.

Ebrahimi, N. and Pirrong, C. (2018). The Risk of Skewness and Kurtosis in Oil Market and the Cross-Section of Stock Returns, SSRN Scholarly Paper, 3168191, Available Online: https://papers.ssrn.com/abstract=3168191 [Accessed 2 April 2023].

Elton, E.J., Gruber, M.J., Brown, S.J. and Goetzmann, W.N. (2014). Modern portfolio theory and investment analysis, 9th edn, Hoboken: Wiley.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, vol. 50, no. 4, pp. 987-1007.

Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and & Economic Statistics*, vol. 22, no. 4, pp. 367-381.

Escanciano, J.C. and Pei, P. (2012). Pitfalls in backtesting Historical Simulation VAR models, *Journal of Banking & Finance*, vol. 36, no. 8, pp. 2233-2244.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance*, vol. 48, no. 5, pp. 1779-1801.

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, *Annals of Mathematics*, vol. 44, no. 3, pp. 423-453.

Gneiting, T. (2011). Making and Evaluating Point Forecasts, *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 746-762.

Gordy, M.B. and McNeil, A.J. (2020). Spectral backtests of forecast distributions with application to risk management, *Journal of Banking & Finance*, vol. 116, pp. 1-9.

Halilbegovic, S. and Vehabovic, M. (2016). Backtesting Value at Risk Forecast: The Case of Kupiec POF-Test, *European Journal of Economic Studies*, vol. 17, no. 3., pp. 393-404.

Hendricks, D. (1996). Evaluation of Value-at-Risk Models using Historical Data, *Economic Policy Review*, vol. 2, pp. 39-70.

Hull, J. (2018). Risk Management and Financial Institutions, 5th edn, Hoboken: Wiley.

Hull, J. and White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk, *Journal of Risk* [Preprint]. Available Online: https://www.risk.net/node/2161156 [Accessed 2 April 2023].

Jarque, C.M. and Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, vol. 6, no. 3, pp. 255-259.

Jorion, P. (2007). Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edn, New York: McGraw-Hill.

Kearney, L. (2020). Explainer: What is a negative crude future and does it mean anything for consumers?, Reuters, 21 April, Available Online: https://www.reuters.com/article/us-global-oil-crash-explainer-idUSKBN22301M [Accessed 9 May 2023].

Kerkhof, J., Melenberg, B. and Schumacher, H. (2010). Model risk and capital reserves, *Journal of Banking & Finance*, vol. 34, no. 1, pp. 267-279.

Kristoufek, L. (2014). Leverage effect in energy futures, *Energy Economics*, vol. 45, pp. 1-9.

Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models, *The Journal of Derivatives*, vol. 3, no. 2, pp. 73-84.

Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics*, vol. 54, no. 1, pp. 159-178.

Lambert, N., Pennock, D.M. and Shoham, Y. (2008). Eliciting properties of probability distributions: the highlights, *ACM SIGecom Exchanges*, vol. 7, no. 3, pp. 1-5.

Lang, K. and Auer, B.R. (2020). The economic and financial properties of crude oil: A review, *North American Journal of Economics and Finance*, vol. 52, pp. 1-45.

Liu, L., Geng, Q., Zhang, Y. and Wang, Y. (2022). Investor's perspective on forecasting crude oil return volatility: where do we stand today?, *Journal of Management Science and Engineering*, vol. 7, no. 3, pp. 423-438.

Manganelli, S. and Engle, R.F. (2001). Value at Risk Models in Finance, working paper series, no. 75, European Central Bank (ECB), Available Online: https://papers.ssrn.com/abstract=356220 [Accessed 5 April 2023].

Marden, J.I. (2004). Positions and QQ Plots, *Statistical Science*, vol. 19, no. 4, pp. 606-614.

Marimoutou, V., Raggad, B. and Trabelsi, A. (2009). Extreme Value Theory and Value at Risk: Application to oil market, *Energy Economics*, vol. 31, no. 4, pp. 519-530.

Markowitz, H. (1952). Portfolio Selection, *The Journal of Finance*, vol. 7, no. 1, pp. 77-91.

Marshall, C. and Siegel, M. (1997). Value at Risk: Implementing a Risk Measurement Standard, *The Journal of Derivatives*, vol. 4, no. 3, pp. 91-111.

McNeil, A.J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance*, vol. 7, no. 3, pp. 271-300.

McNeil, A.J., Frey, R. and Embrechts, P. (2015). Quantitative risk management: concepts, techniques and tools, revised edn, Princeton: Princeton University Press.

Pagan, A. (1996). The econometrics of financial markets, *Journal of Empirical Finance*, vol. 3, no. 1, pp. 15-102.

Pérignon, C., Deng, Z.Y. and Wang, Z.J. (2008). Do banks overstate their Value-at-Risk?, *Journal of Banking & Finance*, vol. 32, no. 5, pp. 783-794.

Pérignon, C. and Smith, D.R. (2010). The level and quality of Value-at-Risk disclosure by commercial banks, *Journal of Banking & Finance*, vol. 34, no. 2, pp. 362-377.

Pflug, G.C. (2000). Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk, in S.P. Uryasev (ed.), *Probabilistic Constrained Optimization: Methodology and Applications*, Boston: Springer US, pp. 272-281. Available Online: https://doi.org/10.1007/978-1-4757-3150-7_15 [Accessed 1 April 2023].

Pickands, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, vol. 3, no. 1, pp. 119-131.

Pritsker, M. (1997). Evaluating Value at Risk Methodologies: Accuracy versus Computational Time, *Journal of Financial Services Research*, vol. 12, no. 2, pp. 201-242.

RiskMetrics. (1996). RiskMetrics Technical Document [pdf], 4th edn. Available at: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a [Accessed 20 April 2023].

Rockafellar, R.T. and Uryasev, S. (1999). Optimization of conditional value-at-risk, *The Journal of Risk*, vol. 2, no. 3, pp. 21-41.

Sadorsky, P. (2006). Modelling and forecasting petroleum futures volatility, *Energy Economics*, vol. 28, no. 4, pp. 467-488.

Schwarz, G. (1978). Estimating the Dimension of a Model, *The Annals of Statistics*, vol. 6, no. 2, pp. 461-464.

Szegö, G. (2002). Measures of risk, *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1253-1272.

Tasche, D. (2002). Expected shortfall and beyond, *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1519-1533.

Tsay, R.S. (2005). Analysis of financial time series, 2nd edn, Hoboken: Wiley.

Tse, Y.K. (1998). The Conditional Heteroscedasticity of the Yen-Dollar Exchange Rate, *Journal of Applied Econometrics*, vol. 13, no. 1, pp. 49-55.

U.S. Energy Information Administration. (2023a). Crude Oil Prices: Brent - Europe, *FRED, Federal Reserve Bank of St. Louis*, Available Online: https://fred.stlouisfed.org/series/DCOILBRENTEU [Accessed 28 March 2023].

U.S. Energy Information Administration. (2023b). Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma, *FRED, Federal Reserve Bank of St. Louis*, Available Online: https://fred.stlouisfed.org/series/DCOILWTICO [Accessed 28 March 2023].

Wang, Y. and Liu, L. (2010). Is WTI Crude Oil Market Becoming Weakly Efficient over Time?: New Evidence from Multiscale Analysis Based on Detrended Fluctuation Analysis, *Energy Economics*, vol. 32, no. 6, pp. 987-992.

Wei, Y., Wang, Y. and Huang, D. (2010). Forecasting Crude Oil Market Volatility: Further Evidence Using GARCH-Class Models, *Energy Economics*, vol. 32, no. 6, pp. 1477-1484.

Working, H. (1949). The Theory of Price of Storage, *The American Economic Review*, vol. 39, no. 6, pp. 1254-1262.

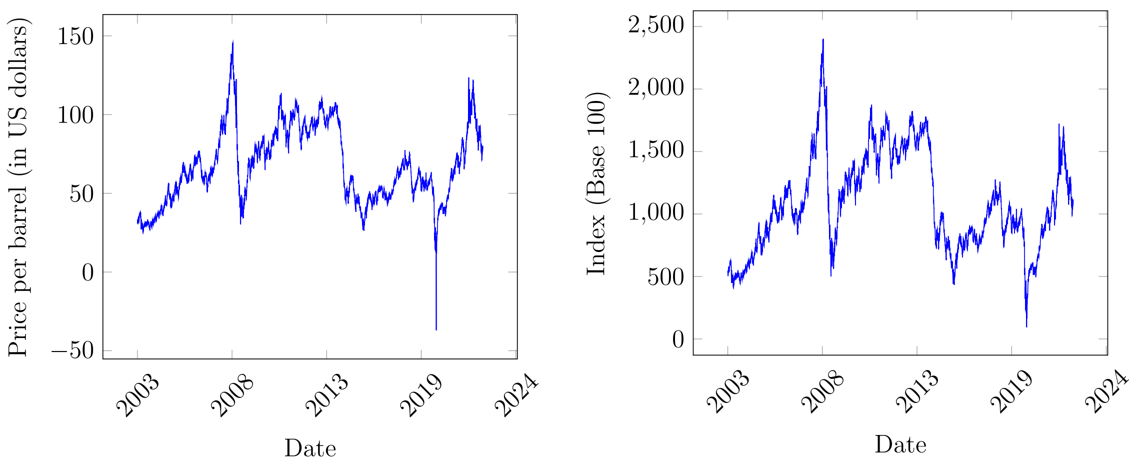
Yamai, Y. and Yoshiba, T. (2005). Value-at-Risk versus Expected Shortfall: A Practical Perspective, *Journal of Banking & Finance*, vol. 29, no. 4, pp. 997-1015.

Youssef, M., Belkacem, L. and Mokni, K. (2015). Value-at-Risk Estimation of Energy Commodities: A Long-Memory GARCH-EVT Approach, *Energy Economics*, vol. 51, pp. 99-110.

Appendix A WTI

A.1 Indexing of WTI

Since the spot price of WTI crude oil turns negative in 2020, it is not possible to apply logarithms to calculate continuously compounded returns. Therefore, to remedy this issue, we index the price series by putting the base of 100 on April 20, 2020, which is the data where the price is at its lowest. Subsequently, we replicate the shape of the series by multiplying the simple returns to the index. The result of this manipulation is showcased in Figure (A.1).



(a) Daily Spot Price of WTI Crude Oil

(b) Index of WTI Crude Oil

Figure A.1: Spot price and index of WTI crude oil from 2003 to 2022.

As can be seen in Figure (A.1b), the general shape and dynamics of the price series is the same in the index series. However, one small difference is that there is not as nuanced drop in the index in comparison to the price series in 2020. Nevertheless, this is a necessary manipulation to be able to apply logarithms and the extreme volatility of the commodity is still salvaged in the index. Therefore, we believe that the manipulation will not have an adverse effect on the generalisability of our results.

Appendix B Time Series Effects

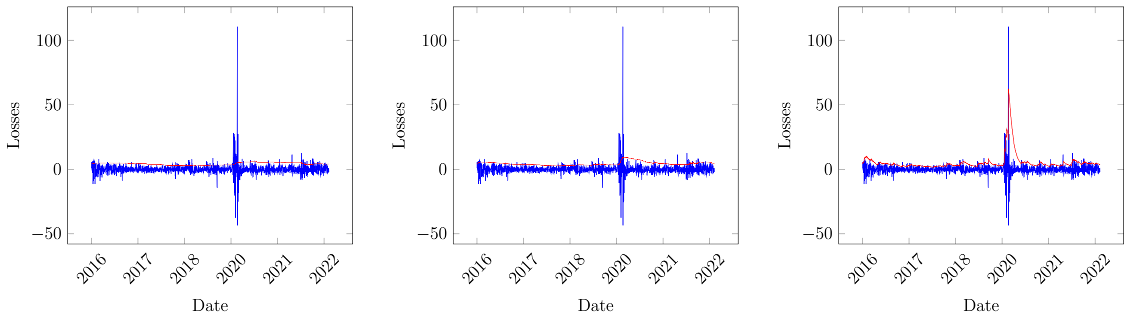
B.1 Tests for Time Series Effects in GARCH Residuals

	LB-Q(20)	LB p-value	LM	LM p-value
WTI	15.30036	0.75897	21.07347	0.39282
Brent	19.33405	0.50022	13.25061	0.86637

Table B.1: Tests for autocorrelation (i.e., the Ljung-Box Test) and ARCH effects (i.e., the LM Test) in the standardised residuals of a GARCH(1,1) for the estimation period.

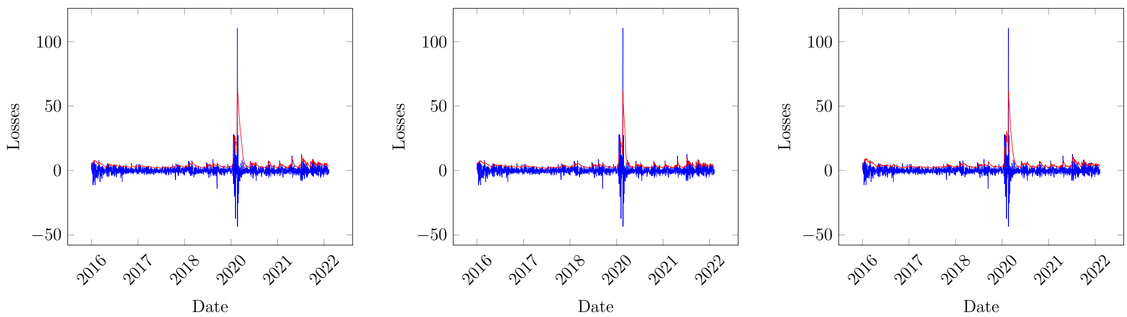
Appendix C VaR Plots

C.1 VaR Plots for WTI



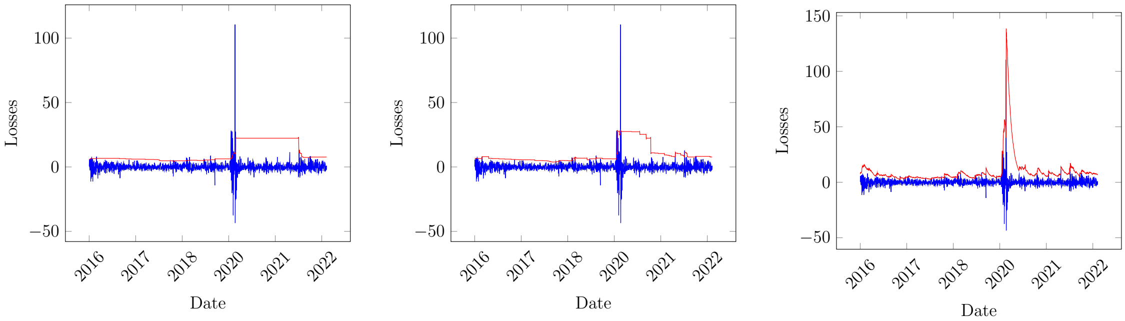
(a) BHS: VaR at $\alpha = 0.95$ *(b) AWHS: VaR at* $\alpha = 0.95$ *(c) VWHS: VaR at* $\alpha = 0.95$

Figure C.1: Nonparametric estimation methods for VaR at the 95th quantile for WTI for the backtesting period



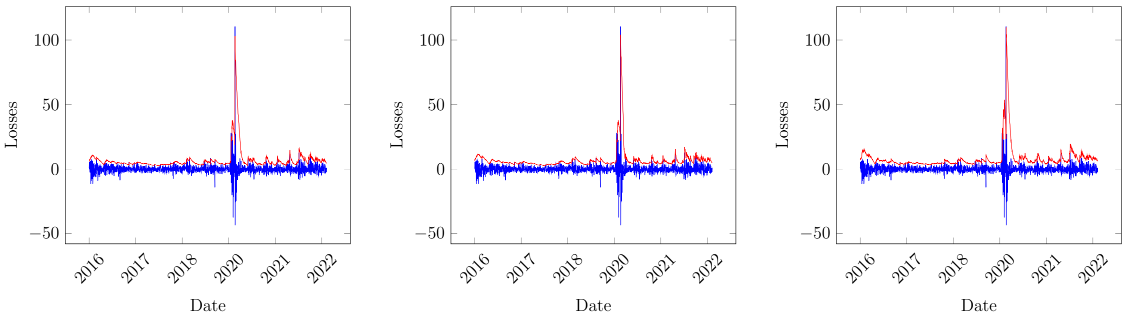
(a) Gaussian: VaR at $\alpha = 0.95$ *(b) t: VaR at* $\alpha = 0.95$ *(c) Cond. EVT: VaR at* $\alpha = 0.95$

Figure C.2: Parametric estimation methods for VaR at the 95th quantile for WTI for the backtesting period



(a) BHS: VaR at $\alpha = 0.99$ *(b) AWHS: VaR at* $\alpha = 0.99$ *(c) VWHS: VaR at* $\alpha = 0.99$

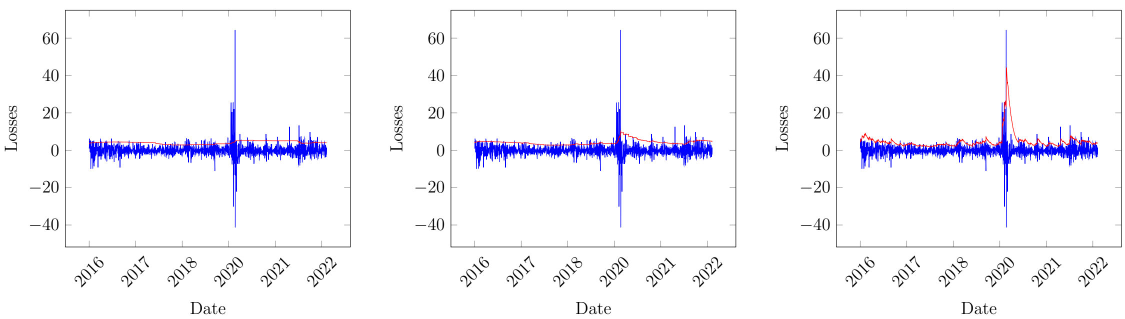
Figure C.3: Nonparametric estimation methods for VaR at the 99th quantile for WTI for the backtesting period.



(a) Gaussian: VaR at $\alpha = 0.99$ *(b) t: VaR at* $\alpha = 0.99$ *(c) Cond. EVT: VaR at* $\alpha = 0.99$

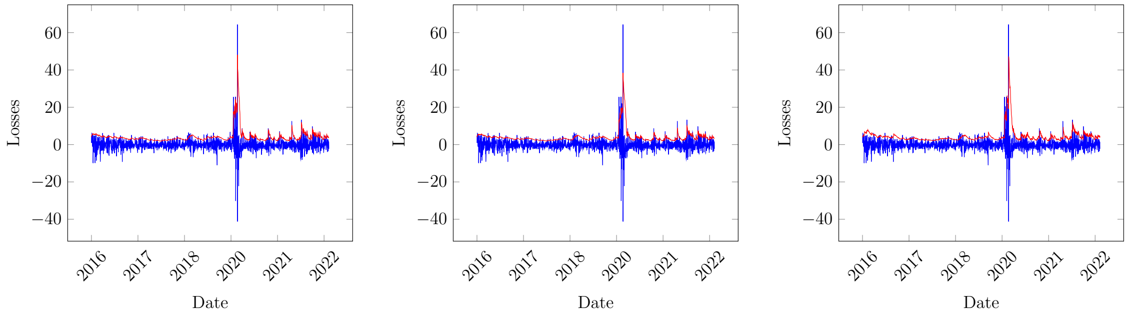
Figure C.4: Parametric estimation methods for VaR at the 99th quantile for WTI for the backtesting period.

C.2 VaR Plots for Brent



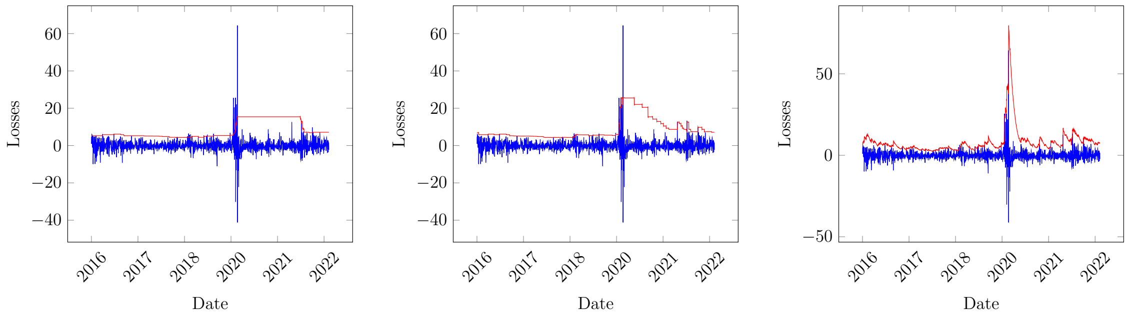
(a) BHS: VaR at $\alpha = 0.95$ *(b) AWHS: VaR at* $\alpha = 0.95$ *(c) VWHS: VaR at* $\alpha = 0.95$

Figure C.5: Nonparametric estimation methods for VaR at the 95th quantile for Brent for the backtesting period.



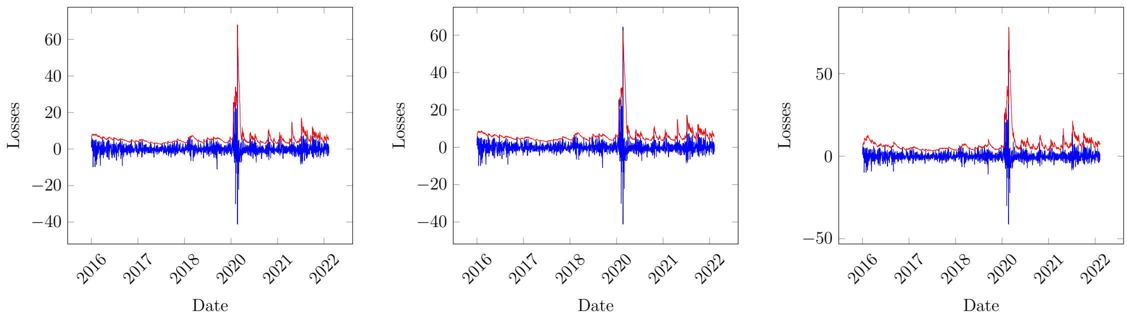
(a) Gaussian: VaR at $\alpha = 0.95$ *(b) t: VaR at* $\alpha = 0.95$ *(c) Cond. EVT: VaR at* $\alpha = 0.95$

Figure C.6: Parametric estimation methods for VaR at the 95th quantile for Brent for the backtesting period.



(a) BHS: VaR at $\alpha = 0.99$ *(b) AWHS: VaR at* $\alpha = 0.99$ *(c) VWHS: VaR at* $\alpha = 0.99$

Figure C.7: Nonparametric estimation methods for VaR at the 99th quantile for Brent for the backtesting period.

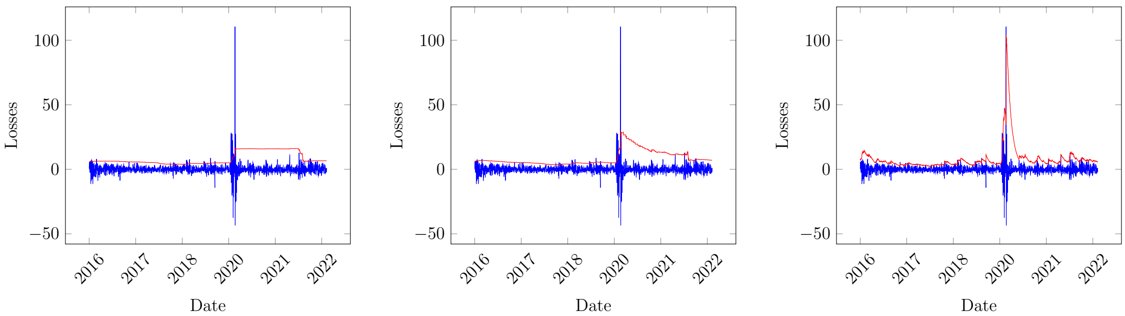


(a) Gaussian: VaR at $\alpha = 0.99$ *(b) t: VaR at* $\alpha = 0.99$ *(c) Cond. EVT: VaR at* $\alpha = 0.99$

Figure C.8: Parametric estimation methods for VaR at the 99th quantile for Brent for the backtesting period.

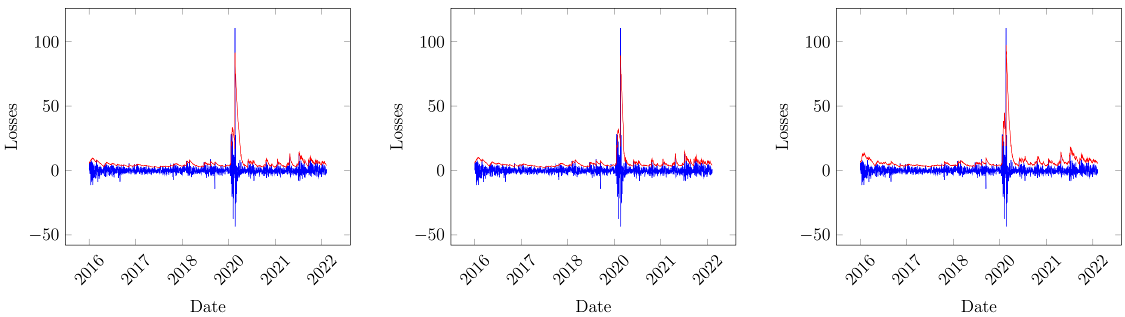
Appendix D ES Plots

D.1 ES Plots for WTI



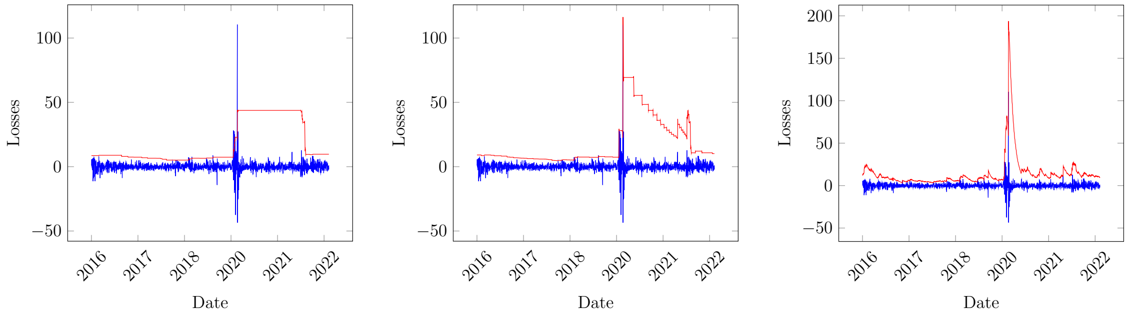
(a) BHS: ES at $\alpha = 0.95$ *(b) AWHS: ES at* $\alpha = 0.95$ *(c) VWHS: ES at* $\alpha = 0.95$

Figure D.1: Nonparametric estimation methods for ES at the 95th quantile for WTI for the backtesting period.



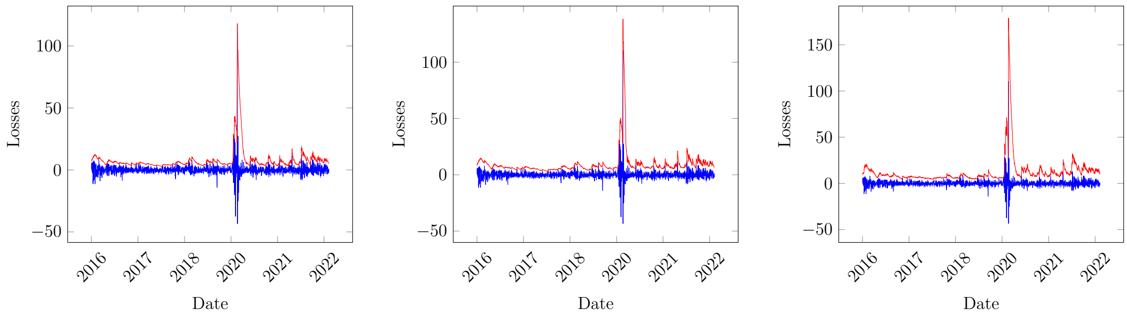
(a) Gaussian: ES at $\alpha = 0.95$ *(b) t: ES at* $\alpha = 0.95$ *(c) Cond. EVT: ES at* $\alpha = 0.95$

Figure D.2: Parametric estimation methods for ES at the 95th quantile for WTI for the backtesting period.



(a) BHS: ES at $\alpha = 0.99$ *(b) AWHS: ES at* $\alpha = 0.99$ *(c) VWHS: ES at* $\alpha = 0.99$

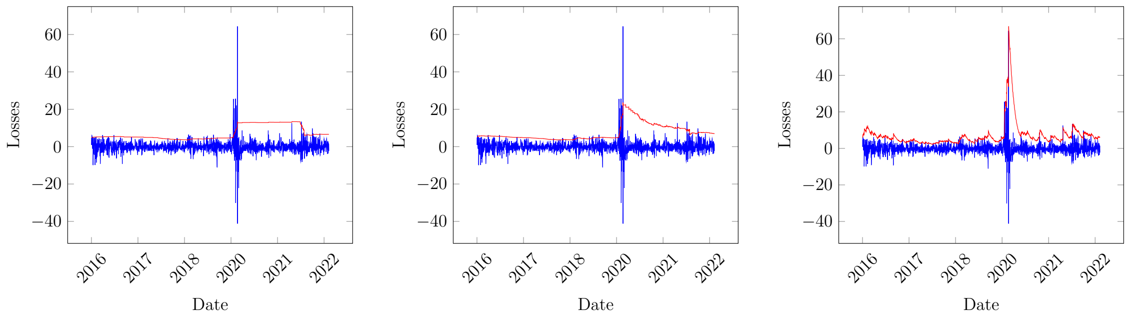
Figure D.3: Nonparametric estimation methods for ES at the 99th quantile for WTI for the backtesting period.



(a) Gaussian: ES at $\alpha = 0.99$ *(b) t: ES at* $\alpha = 0.99$ *(c) Cond. EVT: ES at* $\alpha = 0.99$

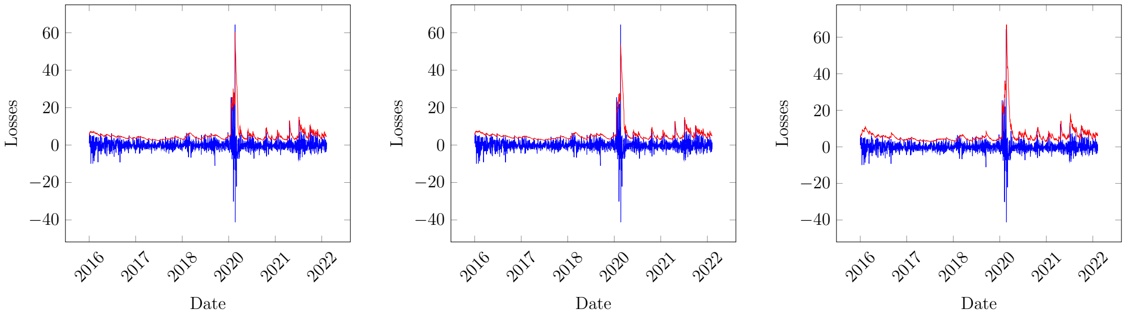
Figure D.4: Parametric estimation methods for ES at the 99th quantile for WTI for the backtesting period.

D.2 ES Plots for Brent



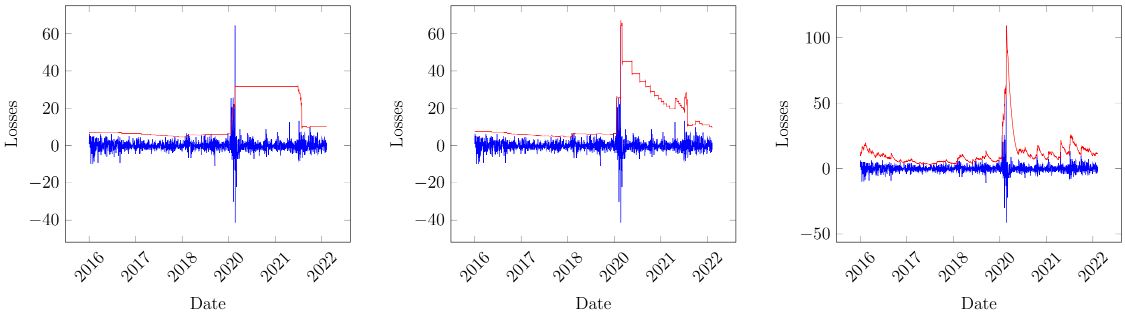
(a) BHS: ES at $\alpha = 0.95$ *(b) AWHS: ES at* $\alpha = 0.95$ *(c) VWHS: ES at* $\alpha = 0.95$

Figure D.5: Nonparametric estimation methods for ES at the 95th quantile for Brent for the backtesting period.



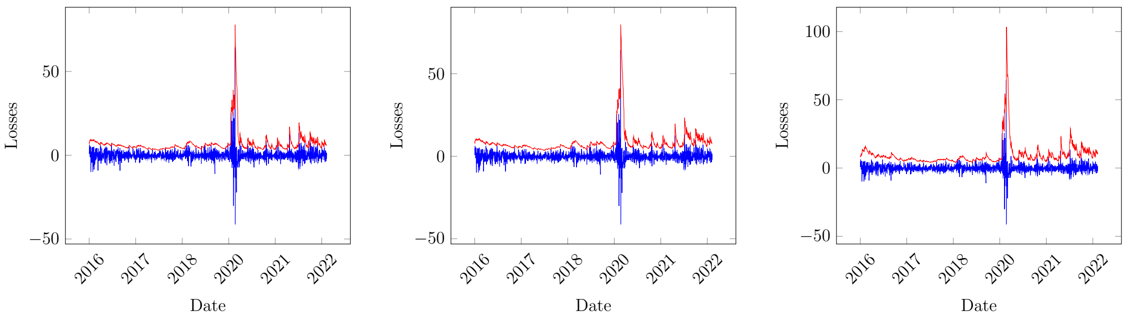
(a) Gaussian: ES at $\alpha = 0.95$ *(b) t: ES at* $\alpha = 0.95$ *(c) Cond. EVT: ES at* $\alpha = 0.95$

Figure D.6: Parametric estimation methods for ES at the 95th quantile for Brent for the backtesting period.



(a) BHS: ES at $\alpha = 0.99$ *(b) AWHS: ES at* $\alpha = 0.99$ *(c) VWHS: ES at* $\alpha = 0.99$

Figure D.7: Nonparametric estimation methods for ES at the 99th quantile for Brent for the backtesting period.



(a) Gaussian: ES at $\alpha = 0.99$ *(b) t: ES at* $\alpha = 0.99$ *(c) Cond. EVT: ES at* $\alpha = 0.99$

Figure D.8: Parametric estimation methods for ES at the 99th quantile for Brent for the backtesting period.