Risk Forecasting of Electricity Futures: A Comparative Study of VaR and ES Forecasting Models on German Electricity Futures

Christopher Laloi Dybdahl[a], Syver Verlo Nes[a]

[a]*Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim, Norway*

ARTICLE INFO

Keywords:
Value at risk
Expected shortfall
Electricity market
Electricity futures
Risk management
Forecasting

ABSTRACT

Electricity futures are essential tools for hedging risks in the increasingly volatile electricity markets, caused by renewable energy growth and geopolitical uncertainties. Risk measures like Value-at-Risk (VaR) and Expected Shortfall (ES) are crucial for understanding and mitigating the risks of these hedged positions. However, joint forecasting of these metrics remains underexplored for electricity futures, especially across contracts with varying return and volatility dynamics. This study evaluates 31 models for joint VaR and ES forecasting on German electricity futures traded on the European Energy Exchange. The contracts analyzed include peak- and base-load front contracts for daily, monthly, quarterly and yearly delivery. Models evaluated include Filtered Historical Simulation (FHS), Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, Extreme Value Theory (EVT) hybrids, Generalized Autoregressive Score (GAS) models, and Conditional Autoregressive Value at Risk (CAViaR) models, with Historical Simulation (HS) serving as the benchmark. Forecasting performance is evaluated using the AL loss function, various backtests, and the Model Confidence Set framework. Results reveal no single model excels universally, with performance varying between contracts and market conditions. Certain FHS and CAViaR specifications demonstrate strong performance across all contracts, while the HS benchmark underperforms for longer durations. These findings highlight the necessity of tailoring model selection to the unique characteristics of each futures contract. This research bridges gaps in the literature by addressing joint VaR and ES forecasting for electricity futures, offering actionable insights for market participants. Additionally, it adds to the foundation for future studies exploring risk assessment methodologies.

1. Introduction

The transition to a sustainable energy system has significantly heightened the demand for robust risk management frameworks among participants in the electricity ecosystem. The growing reliance on electricity, particularly from renewable energy sources, coupled with increasing geopolitical instability, has intensified electricity price uncertainty. This uncertainty arises from factors such as greater challenges in matching production with consumption and the heightened influence of inherently unpredictable weather conditions. These dynamics create significant challenges for electricity market participants, making effective risk management critical to addressing market and operational uncertainty, optimizing capital allocation, and providing financial stability. Electricity futures have emerged as key tools for hedging, offering a means to enhance predictability. However, trading these instruments introduces additional complexities, including ensuring sufficient liquidity to be able to meet margin calls. In striving to navigate these complexities, electricity market participants are increasingly looking to adopt risk management practices inspired by trends in the banking sector, such as rigorous backtesting procedures. Accurately forecasting the risk associated with futures-hedging positions enables market participants to proactively adapt to changing environments and effectively mitigate these challenges.

This study focuses on evaluating models that forecast the widely recognized risk metrics Value-at-Risk (VaR) and Expected Shortfall (ES) by applying them to futures traded on the European Energy Exchange (EEX). The EEX offers high liquidity, enabling cost-efficient trading, and a wide variety of futures contracts that support hedging diverse types of risks. Building on insights from a Norwegian electricity producer that trades German electricity futures for hedging purposes and is transitioning its risk assessment processes in-house, this research examines eight distinct contracts. Each contract presents unique return characteristics, such as varying levels of volatility and leverage effects, posing significant challenges for accurately forecasting the risk of the entire hedging position. Despite the critical importance of reliable forecasts, there is no consensus on which models are best suited for different contracts or market conditions. While prior studies have extensively explored risk metrics in financial markets, the heterogeneity of electricity futures contracts introduces complexities that remain insufficiently examined. Addressing this gap provides a valuable opportunity to enhance understanding and deliver actionable insights for risk management across diverse energy commodities.

The objective of this study is to identify effective forecasting models for VaR and ES that perform well across the diverse return characteristics exhibited by electricity futures. By accounting for these variations, the study aims to provide electricity producers with actionable insights to enhance their risk management practices. We formulate our research question as follows: *Which univariate joint VaR and ES forecasting models are best-performing for different return characteristics*. The research focuses on contracts with varying durations—ranging from daily to yearly and

christdy@stud.ntnu.no (C.L. Dybdahl); syvervn@stud.ntnu.no (S.V. Nes)

underlying products classified as peak or baseload, addressing the unique return environments and volatility patterns associated with these instruments. Ultimately, the study seeks to advance understanding of how these models can improve risk forecasting, helping electricity producers better manage uncertainty and align their hedging strategies with their risk exposures.

To achieve this objective, we evaluate univariate forecasting models for joint VaR and ES, prioritizing intrinsic return characteristics of electricity futures by excluding exogenous variables. The contracts analyzed include peak- and base-load front contracts for daily, monthly, quarterly and yearly delivery. The price-data for each contract spans from April 9, 2019, to October 31, 2024, covering 1,239 log-return observations. A diverse set of relevant forecasting models was selected based on a bibliometric review of 272 studies published in the last five years. These models include Filtered Historical Simulation (FHS), Generalized Autoregressive Conditional Heteroskedasticity (GARCH) type models, Extreme Value Theory (EVT) hybrids, Generalized Autoregressive Score (GAS) models, and Conditional Autoregressive Value at Risk (CAViaR) models, with Historical Simulation (HS) included as a benchmark to provide additional nuance to the analysis.

The forecasting out-of-sample accuracy of these models was assessed using the Asymmetric Laplace (AL) log score loss function, ensuring evaluation of joint VaR and ES performance. Backtesting frameworks applied include unconditional coverage (UC), conditional coverage (CC), dynamic quantile (DQ), and VDT tests for VaR, as well as ER and ESR tests for ES. Additionally, the Model Confidence Set (MCS) test was employed to rank models based on performance. This methodology emphasizes univariate models to limit scope, ensure computational efficiency, and focus on intrinsic characteristics of futures contracts, as portfolio returns can often be aggregated into a single time series. Exogenous variables were deliberately excluded to concentrate on intrinsic return dynamics, ensuring that the findings are robust across diverse market conditions.

The main findings of this study can be summarized as follows: First, no single model consistently outperforms others across all contracts and conditions, emphasizing the need for model selection tailored to the specific characteristics of the futures contract, such as duration and underlying product. Second, some configurations of FHS and CAViaR models demonstrate strong accuracy and are well-specified, with the FHS models performing best at the 2.5% quantile and the CAViaR model showing strong performance at the 5% quantile. Notably, these models demonstrate reliability during stressed periods, drawing parallels to banking-inspired approaches like stressed VaR. Third, the traditional model HS underperforms particularly for longer-duration contracts, underscoring the potential benefits of adopting more advanced modeling techniques. These findings reveal the contextual strengths and limitations of widely used risk forecasting models, offering a foundation for further research while providing valuable insights to inform and refine risk management practices in electricity futures markets.

This article contributes to the literature on risk forecasting by incorporating recent data from periods of heightened market turbulence, such as the volatile energy markets of 2021–2022. Furthermore, it identifies trends from recent bibliometric analysis, employing relevant models cited in the literature, and addressing the limited research on joint VaR and ES forecast evaluation in electricity futures markets. The study systematically assesses widely used models by evaluating their performance across diverse market conditions, with a focus on robust backtesting procedures. This emphasis reflects the growing aspiration of electricity market participants to match the rigorous risk management standards of banking institutions, where backtesting and the inclusion of ES are increasingly vital for capturing tail risks.

By addressing the unique return characteristics and risk dynamics of electricity futures, this work bridges theoretical advancements and practical applications, providing actionable insights to enhance risk management and laying a foundation for further research in energy markets.

The remainder of this article is structured as follows: Section 2 provides the background for the study, emphasizing the importance of risk forecasting in energy markets and the unique characteristics of electricity futures. Section 3 presents a literature review, identifying recent trends, existing gaps, and key advancements in VaR and ES forecasting models. Section 4 introduces the forecasting models evaluated in this study. Section 5 describes the backtesting framework employed to assess the predictive performance and adequacy of these models. Section 6 outlines the data used in the analysis, with a focus on the German electricity futures market and its specific characteristics. Section 7 discusses the results, comparing model performance across different return environments, including base and peak contracts, various durations, and volatility regimes. Finally, Section 8 concludes the article by summarizing the key findings, highlighting practical implications, and offering directions for future research.

2. Background

2.1. Value at risk (VaR) and Expected Shortfall (ES)

Most companies have an interest in limiting their financial losses. To do this, it is beneficial to measure the risks involved in their operation. For instance, supervised entities, such as banks or insurance companies, need good risk measures to maintain sufficient capital buffers and meet regulatory standards. This ensures they have the reserves necessary to absorb losses, such as paying out deposits or claims. Many of these entities follow the recommendations on risk management issued by the Basel Committee on Banking Supervision (BCBS). In 1996, BCBS issued the Basel II Accord, where they recommended the use of VaR as risk measure (BIS, 2006). VaR dates back to the 20th century, but the use of VaR today is mostly credited to JP Morgan, which in 1994 published RiskMetrics (J.P. Morgan/Reuters, 1996), a technical document on how they estimated VaR for tradable assets, leading to VaR becoming a popular choice for other financial institutions (Adamko et al., 2015; Manh, 2017). In simple terms, VaR is a single value representing the potential loss, over a specified time period, given a certain confidence level. It has the advantage of being easy to understand, and relatively simple to implement. However, the drawbacks of VaR became evident during the 2008 financial crisis.

VaR does not account for extreme losses beyond the given confidence level, which in 2008 led to banks underestimating the market turndown, leading to insufficient capital reserves (Halbleib and Pohlmeier, 2012). Moreover, it does not reward diversification, as it is not sub-additive (Tasche, 2002). The Basel committee acknowledged these drawbacks, and in 2012 they published the Basel III accord, replacing VaR with ES as the preferred risk measure for calculating capital requirements (Zaevski and Nedeltchev, 2023). ES accounts for extreme losses by calculating the average of potential losses that exceed the set VaR confidence level (Saralees Nadarajah and Chan, 2014). Despite the drawbacks of VaR, the risk measure is still used in practice, and is often written about in the literature. Hence, understanding both VaR and ES is important for banks and insurance companies.

Furthermore, companies exposed to volatile markets also widely apply VaR and ES. For instance, financial institutions engaged in speculative stock market activities must quantify the risks of their positions, as large movements in held positions can result in substantial losses. In addition, asset managers use risk measures in dynamic portfolio allocation strategies, such as risk targeting, to allocate assets in order to maintain a preferred level of risk (Happersberger et al., 2020). VaR and ES applications also reach beyond financial institutions. The risk measures are used, for example, in medicine, engineering, supply chain management and energy (Filippi et al., 2020).

Another area where the use of risk measures has become increasingly important is for companies holding hedged positions through derivatives such as futures and options. The electricity market, in particular, has experienced significant growth in trading volumes for such contracts. This expansion began following the transition to liberalized electricity markets in the 1990s, with the UK and Norway pioneering the shift toward competitive market structures (Pietz, 2009). Today, electricty stakeholders can hold large positions in futures contracts. Risk associated with price movements in these contracts facilitate the need for risk measures in order to avoid potential downside. However, the unique characteristics of electricity prices and associated derivatives, such as high volatility and seasonality, make forecasting VaR and ES particularly challenging. Furthermore, the growing penetration of renewable electricity generation and significant geopolitical events, such as the Russian invasion of Ukraine, have introduced additional complexities, emphasizing the need for studies that evaluate and compare models under these evolving market dynamics. In the next section, we will take a deeper look at the risk exposure of electricity producers, how they can use futures to hedge this risk, before we adress the importance of robust risk measures such as VaR and ES. While the focus is on electricity producers, the insights can be readily applied to other energy stakeholders facing similar challenges.

2.2. Electricity producers - Risks, hedging and the role of risk measures

The electricity market is distinct due to its inherent characteristics, including the non-storability of electricity, the need for real-time supply-demand balancing, and the volatile electricity prices. Electricity prices exhibit unique dynamics such as volatility clustering, frequent spikes, negative prices during specific hours, and strong seasonal patterns (Knittel and Roberts, 2005). These price movements are increasingly influenced by weather-dependent renewable electricity production, making the market's risk profile more complex and less predictable (May et al., 2019). For electricity producers, electricity prices pose significant risks, as their volatile and unpredictable nature directly impacts their revenue and operational planning. Nuclear power plants for instance, face high shutdown and startup costs, which make them particularly susceptible to losses during periods of negative electricity prices (Westgaard et al., 2018).

Electricity producers also face additional risks. Traditional producers such as those relying on gas and coal for electricity generation face exposure to fuel and policy risks. For instance, gas-based producers are vulnerable to fluctuations in gas prices, and policy risk related to carbon taxes.

Renewable electricity producers, on the other hand, are less exposed to production costs and policy risk, but more exposed to volume risk. Their production volumes are largely dependent on weather conditions, such as rainfall for hydropower, wind for wind power, or sunshine for solar power. An extreme example to substantiate this is the 30% drop in solar electricity production reported by some companies in California in 1992 due to the Mount Pinatubo volcanic eruption (Molineaux and Ineichen, 1996; Michalsky et al., 1994).

Electricity producers can mitigate these risks in several ways. One is by diversifying their energy portfolios, investing in multiple methods of electricity generation. This allows the producer to leverage the unique advantages of different technologies, such as the storability of energy from certain sources and the low-cost production offered by others. Another key instrument for mitigating said risks, include the trading of options, futures, and other derivatives, which enable producers to hedge against unfavorable outcomes. Future contracts in particular, are popular for their high liquidity, which provides a more stable hedge against price volatility (ter Haar, 2010). Additionally, the availability of various contracts with different underlying assets, such as for instance carbon credit futures, enables producers to hedge not only against price risk but also against specific risks. Carbon credit futures for instance, can be used to hedge against risk associated with policy uncertainty connected to carbon emission costs. Electricity futures contracts in particular, are traded on established exchanges such as the European Energy Exchange (EEX) and NASDAQ Commodities (Cartesan and Penzo, 2020).

The broad range of different futures contracts enable energy producers and other stakeholders to manage risks by allowing them to construct portfolios that align with their specific risk profiles. For instance, the gas-based electricity producer exposed to fluctuating gas prices and carbon taxes used in the example above, can hedge its risk by buying gas futures and carbon credit futures, stabilizing input- and carbon emission costs. Furthermore the same producer could buy electricity futures to directly hedge against the volatile electricity prices, resulting in a more predictable income. Similarly, a renewable energy producer facing volume risks due to unpredictable weather patterns could use weather-related futures to hedge against fluctuations in production (May et al., 2019). Using various futures contracts can thus help the producer hedge against multiple risks. However, there is a significant trade-of between trading highly liquid futures and trading less liquid contracts that better match the risk profile of the producer (May et al., 2019).

Higher liquidity leads to smaller bid-ask spreads, which lowers transaction costs and improves hedging by allowing participants to execute trades more efficiently and cost-effectively (DNV, 2020; Marcato and Ward, 2007). Given this trade-off, the optimal mix of futures used for hedging by an energy producer often consists of a limited number of products with sufficient liquidity, rather than a large array of illiquid futures to match their specific risk profile (May et al., 2019). A position where the risk a producer seeks to hedge does not perfectly align with the futures contracts used, is often referred to as a “dirty hedge” (May et al., 2019). A prominent example of dirty hedging is that electricity stakeholders across Europe, including the Norwegian electricity producer motivating this study, frequently use German electricity futures for hedging. German electricity futures are the most liquid in Europe, making them desirable hedging tools even for foreign stakeholders (EEX). While Norwegian and German electricity prices are often correlated due to interconnected grids and shared influences like gas prices, local weather conditions and transmission constraints can lead to significant price deviations. Despite this mismatch, the liquidity of German electricity futures make them a preferred choice of derivatives for hedging.

In general, the front year electricity futures contracts with one year delivery of baseload and peakload are the most liquid, followed by the front month contracts (May et al., 2019). This makes these products popular hedging tools, creating a lock-in effect where stakeholders continue to trade these, further providing liquidity (May et al., 2019). Many studies also focus on these contracts when evaluating models for forecasting VaR and ES. Although these contracts are the most popular, quarterly and daily delivery contracts for baseload and peakload futures also exist. The recent growth in trading volumes on exchanges like EEX[1], results in increased liquidity for for these traded contracts as well. Given the popularity of the yearly and monthly duration contracts, it is plausible that quarterly and daily contracts could gain traction in the future as their liquidity improves, making them important to study. Futures contracts are an effective tool for hedging risks, such as price volatility, but holding large positions in these contracts can introduce significant risks of their own.

Consider an electricity producer holding a large position in a front year baseload futures contract. If the price of the futures drops, the producer faces monetary losses, which remain unrealized until a margin call forces the producer to deposit additional funds to maintain their position. This can strain liquidity and divert capital from other operational needs. If the producer fails to meet the margin call, their position also may be liquidated at a loss. Moreover, the producer might face opportunity costs, as capital tied up in the futures market could otherwise be deployed elsewhere. These risks underscore the importance of carefully managing futures positions, as the consequences of unwanted price movements can result in financial problems for the producer.

To understand and measure the riskiness of these hedged positions is in other words crucial, allowing the electricity producers to take preventive actions to mitigate potential losses and tied up capital. Risk measures, such as VaR, play an important role in managing these positions. By predicting VaR and ES for their held futures positions, the electricity producer gains insight into the level of risk associated with their position, allowing them to align their strategies according to their risk tolerance.

There have been many studies comparing the performance of VaR and ES models on the stock market, as well as on electricity prices. However, fewer studies have been done on futures contracts, and more specifically German electricity futures contracts. The performance of models for forcasting VaR and ES, differ significantly from market to market, and between assets with different return dynamics (Amaro et al., 2022). Hence, studies comparing models on different markets and assets are important. For instance, returns of German front-day electricity futures contracts display positive skewness and higher volatility, whereas the front-year contracts exhibit negative skewness and lower volatility. One model therefore might perform well for one contract but poorly for the other, underscoring the importance of testing several models across different contracts to identify the best fit. This helps the producer choose what models to implement, or which to direct their focus at when conducting further research.

Furthermore, an understanding of how to backtest VaR and ES models is crucial for electricity producers with in-house risk forecasting, as wrongly assessed models could result in biased results. Open-source code would further help facilitate easy and effective implementation of these tests, providing electricity producers with a valuable framework

[1] Trading volume on the EEX Group Power Markets grew by 36% from 2022 to 2023 (EEX, 2024b).

for testing new models against their current risk management strategies.

Evaluating models on new data is further valuable, as it enables testing their ability to capture specific events and evolving market dynamics that previous studies may not have addressed. For example, the increased penetration of renewable electricity production has shifted the factors affecting electricity prices. These prices have traditionally been driven by demand, but are now increasingly affected by weather, which makes the prices less predictable (May et al., 2019). Additionally, recent events, such as the Russian invasion of Ukraine and the subsequent Russian halt of gas through the Nord Stream 1 pipeline, have led to record-high and volatile electricity prices, impacting futures markets as well (EU Council, 2024). Testing and evaluating models with data that include these dynamics provides producers with critical insights into which models perform well during extreme periods.

This background section has explored the role of futures contracts as a critical tool for electricity producers to hedge against the volatility and unpredictability of electricity prices, providing more stable revenue streams. German electricity futures have gained popularity among electricity producers across Europe due to their high liquidity, making them a reliable choice for hedging. Among these, year and month baseload and peakload contracts are the most liquid, though the evolving market suggests that daily and quarterly contracts could become increasingly viable hedging tools in the future.

The risks associated with holding hedged positions were also discussed, underscoring the need for effective risk management and the important role of risk measures in mitigating these risks. The prominence of VaR as a risk measure was examined, alongside its limitations, which have led to the adoption of ES as a more comprehensive alternative, while highlighting the value of using both measures jointly. Lastly, the necessity of comparing models across different contracts and incorporating new data was observed, as model performance can vary significantly with differing return dynamics and shifting market conditions.

3. Literature Review

A lot has been written in literature about forecasting VaR or ES for stock and electricity prices. However, significantly less has been written about VaR and ES forecasting for electricity futures, more specifically German electricity futures, apart from some exceptions (Westgaard et al., 2019; Amaro et al., 2022; Peña et al., 2020, e.g.). This section is a review of the literature on existing models and backtesting methods for forecasting VaR and ES for the last five years. Our objective is threefold: first, to analyze the overarching trends in the literature; second, to derive suitable backtesting methods based on these trends; and finally, to identify promising models to implement for our analysis.

There are in principle two approaches to writing a literature review: performing a literature query to identify studies that meet some specified requirements or utilizing existing reviews on the subject (Weron, 2014). This section combines both approaches and is divided into their respective parts. In the literature query analysis, we first discuss general trends and then select a subset of articles whose presented models are further considered for implementation. In the part on existing systematic literature reviews, we highlight the identified trends and present the models deemed superior.

3.1. Bibliometric Query

The query analysis followed the methodology of Weron (2014) and used the Scopus database for its comprehensive coverage and user-friendly interface. A constructed query[2] identified recent articles on forecasting VaR and ES published from 2019 to October 17, 2024, yielding 384 results.

After manually reviewing the articles, 38 were removed as they focused on Vector Autoregression Models, which share the same commonly used abbreviation as value at risk. This narrowed the dataset to 344 relevant articles, including journal articles and reviews.

3.1.1. Trends

The resulting articles from the query were published by 185 distinct journals, where Figure 1 depicts the five most popular ones. The most popular journal was the Journal Of Forecasting with 26 articles, followed by Finance and Research Letters and the International Journal Of Forecasting with 16 and 14 articles, respectively. Figure 2 shows the five most common subject areas identified in the search. The leading subject area is Economics, Econometrics and Finance, followed by Business, Management and Accounting, with Mathematics ranking third. Note that some articles are categorized into multiple subject areas.

Figure 3 illustrates the distribution of articles from the query search by publication year, segmented into those focusing solely on VaR and those addressing both VaR and ES. The total number of annual publications show a relatively steady growth, with a notable increase in articles discussing both VaR and ES. The growing inclusion of ES in research may indicate an increasing recognition among scholars of its advantages over VaR, elaborated on in the background section. Despite the weaknesses of VaR, it remains widely used, and studies investigating ES often include the forecasting of VaR as well. There are several reasons for this. One reason is related to the disadvantage that ES is not elicitable, as proven by Gneiting and Raftery (2007). A statistical function, such as VaR or ES, is elicitable if it is the unique minimizer of a strictly consistent scoring function. This makes it challenging to use ES alone in a loss function to create forecasts or optimize risk models. However, Fissler

[2]Constructed query for the bibliometric analysis: "TITLE-ABS(("forecast*" OR "predict*" OR "estimating") W/2 (("value at risk" OR "value-at-risk" OR "VaR") OR ("Expected shortfall" OR "cVaR" OR ("conditional" AND ("value-at-risk" OR "value at risk"))))) AND KEY(("value at risk" OR "value-at-risk" OR "VaR") OR ("Expected shortfall" OR "ES" OR "cVaR" OR ("conditional" AND ("value-at-risk" OR "value at risk")))) AND PUBYEAR > 2018 AND (LIMIT-TO(SRCTYPE,"j")) AND (LIMIT-TO(LANGUAGE,"English"))"

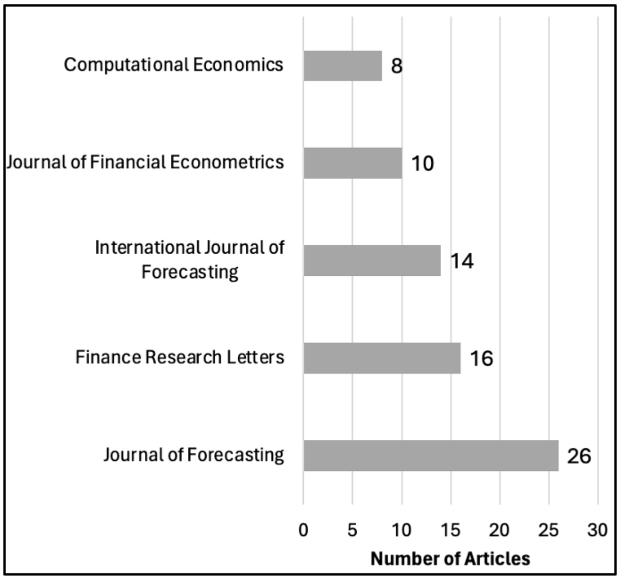

Figure 1: The top five most popular journals from the bibliometric query search, out of a total of 188 different journals.

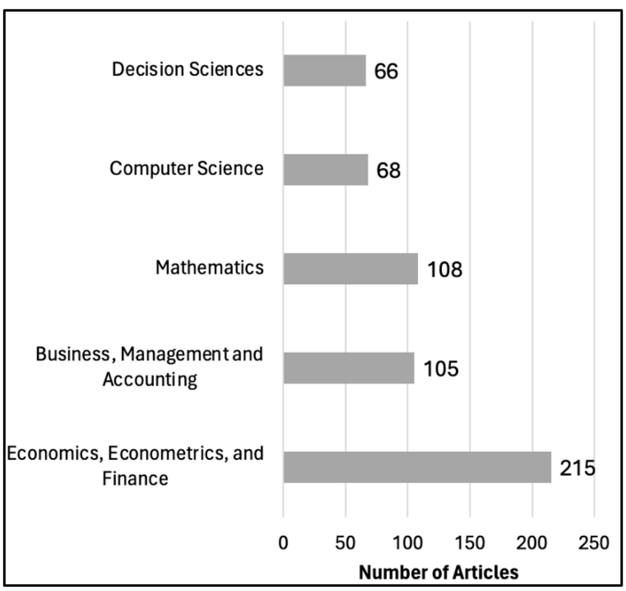

Figure 2: The top five most popular subject areas from the bibliometric query search.

and Ziegel (2016) found that the pair Var and ES is jointly elicitable, i.e. that there exists a scoring function for which VaR and ES together constitute the minimizer, making the estimation of VaR nevertheless desirable.

Furthermore, the increasing trend of studies including ES benefits this study by providing insights into the methods that perform best for ES forecasting. Equally important, it allows us to examine the backtesting approaches employed to evaluate these models, ensuring a comprehensive understanding of their effectiveness and best practices.

While the initial exploration includes a broad range of articles related to VaR and ES, specific attention is given to those that assess the actual performance of these measures. From the manual survey, we identified less relevant articles as those that do not introduce new methods for forecasting VaR or ES (Siu, 2019; Tsoukala and Tsiotas, 2021; Tsukahara, 2024; Golodnikov et al., 2019, e.g.) or that primarily use VaR or ES for other applications (Población and Serna, 2021; Lee and Kramer, 2023; Behera et al., 2023; Zhao et al., 2020, e.g.). These less relevant articles total 48, leaving 296 articles. Out of these, 24 require subscriptions that we do not have access to, further reducing the set to 272 articles.

3.1.2. Backtesting Methods

The 272 resulting set of articles revealed a significant diversity in backtesting methods employed. Table 1 and 2 illustrates the five most common backtesting approaches used to assess the adequacy of VaR models and joint VaR and ES models, respectively. Methods such as the Unconditional Coverage (UC) test by Kupiec (1995), Conditional Coverage (CC) test by Christoffersen (1998), and Dynamic Quantile (DQ) by Engle and Manganelli (2004) are the most widely used for VaR. For joint VaR and ES forecasts, the Exceedance Residual (ER) by McNeil and Frey (2000), Expected Shortfall Regression (ESR) by Bayer and Dimitriadis (2020), Conditional Calibration (CoC) by Nolde and Ziegel (2017), and Conditional- and Unconditional Coverage of Expected Shortfall tests (CCE and UCE) by Du and Escanciano (2017) were the most frequent, as displayed in Table 2. These methods are hypothesis-based tests applied to individual models to evaluate their adequacy. However, they are limited in that they do not provide a definitive answer about which models are statistically optimal within the specific study.

In contrast to adequacy tests, comparative backtesting methods focus on ranking models based on their performance computed with a selected loss functions. Table 3 illustrates the five most common comparative backtesting approaches identified in the articles. The most popular of these was the Model Confidence Set (MCS) by Hansen et al. (2011) which identifies a Superior Set of Models (SSM), followed by the Diebold-Mariano test by Diebold and Mariano (1995), and the V1 and V2 tests (VT) by McNeil et al. (2005). Regarding the employed loss functions, the Quantile Loss for VaR by González-Rivera et al. (2004) and the Fissler-Ziegel class of loss functions by Fissler and Ziegel (2016) for joint VaR and ES forecasts, were the most popular.

We return to a more comprehensive discussion on the differences between these tests and loss functions in Section 5.

3.1.3. Model Selection

This section explains the process of how we ultimately narrowed down from 272 articles to 41 articles selected for model comparison and consideration.

Comparing studies on forecasting is difficult because they are often done on different data, they use different backtesting and error measures, and often have different code implementations (Lago et al., 2021). These differences

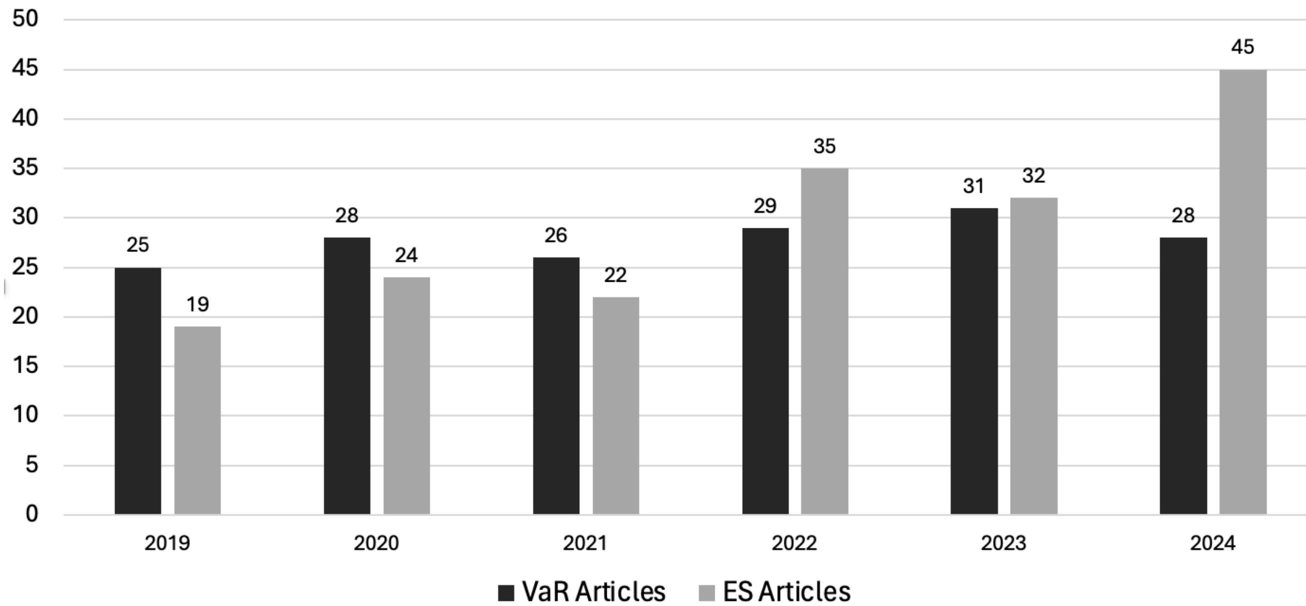

Figure 3: This bar chart illustrates the number of articles obtained from the bibliometric search by publishing year, categorized into two groups: articles focusing exclusively on VaR and articles that also include ES. The black bars represent VaR-only articles, while the gray bars correspond to articles that consider both VaR and ES.

Table 1
The top five most popular backtesting methods for VaR from the bibliometric query analysis. The "-" indicates that no specific author or publication is attributed to the development of the corresponding test.

Name	**Abbreviation**	**Proposed by**
Unconditional Coverage	UC	Kupiec (1995)
Conditional Coverage	CC	Christoffersen (1998)
Dynamic Quantile VaR	DQ	Engle and Manganelli (2004)
Actual Exceedance Rate	AE	-
Independence	IND	Christoffersen (1998)

make it hard to follow the progress in the field. Lago et al. (2021) discuss common pitfalls of studies related to electricity forecasting models that make them hard to compare. The pitfalls can be summarized as the following:

- Comparing advanced models with too simple benchmark models.
- Too short testing periods, or limiting test periods to on-week periods.

Table 2
The top five most popular hypothesis backtesting methods for joint VaR and ES from the bibliometric query analysis.

Name	**Abbreviation**	**Proposed by**
Exceedance Residual	ER	McNeil and Frey (2000)
Expected Shortfall Regression	ESR	Bayer and Dimitriadis (2020)
Conditional Calibration	CoC	Nolde and Ziegel (2017)
Unconditional Backtest of ES	CCE	Du and Escanciano (2017)
Conditional Backtest of ES	UCE	Du and Escanciano (2017)

Table 3
The top five most popular comparative backtesting methods from the bibliometric query analysis.

Name	Abbreviation	Proposed by
Model Confidence Set	MCS	Hansen et al. (2011)
Diebold-Mariano	DM	Diebold and Mariano (1995)
V1 and V2 Test	VT	McNeil et al. (2005)
Pairwise Weak Dominance Test	ECPA	Giacomini and White (2006)
Superior Predictive Ability	SPA	Hansen (2005)

- Lack of details to reproduce the results in the study (dataset, train/test split, model parameters)
- Only testing models for one dependent variable.
- Combination forecasts are not generically assessed.

One of the pitfalls frequently exhibited by the resulting articles from the query, is not specifying the model parameters. However, since we will apply the models to new data, the model parameters will need to be tuned accordingly, making the requirement insignificant. The pitfalls identified by Lago et al. (2021) highlight important requirements for comparing forecasting models across studies. However, when dealing with VaR and ES forecasting, additional requirements should be considered.

One of the key aspects of forecasting VaR and ES is backtesting (Qihui Su and Qin, 2021; Argyropoulos and Panopoulou, 2019). Since this study aims to identify suitable models for forecasting VaR and ES, the backtesting methods employed should be considered when comparing models across different studies. Although the literature on risk models is extensive, research on model ranking and selection that aligns with the Basel III framework remains limited (Owusu Junior et al., 2022). For instance, two popular tests identified in the bibliometric search, the Superior Predictive Ability (SPA) test by Hansen (2005) and the Pairwise Weak Dominance (ECPA) test by Giacomini and White (2006), lack the use of a consistent scoring function. In contrast, the MCS test referenced earlier provides a robust framework for determining, with a certain confidence level, which models are statistically indistinguishable from the best-performing model, regardless of the specific loss function employed (Hansen et al., 2011). This is achieved through the integration of a DM test applied across all competing models. Given the frequent use of the MCS test in our bibliometric query and its clear conclusions regarding model performance, we have chosen to prioritize studies that incorporate the MCS test as a requirement for comparison. Filtering out the ones not using MCS to evaluate the models, leaves us with 53 articles.

Taking into account the pitfalls proposed by Lago et al. (2021) and discussed in the previously in this section narrow the search down further. The remaining articles avoid most of these pitfalls, the exception being only applying models to no more than one dependent variable, where 12 of the articles are being excluded (Štefan Lyócsa et al., 2024; Lu et al., 2023; Naimoli et al., 2022; Pourrezaee and Hajizadeh, 2024; Trabelsi and Tiwari, 2023, e.g.). This narrows the selection down to 41 articles.

Out of the these articles, several of the studies include models that have specific implementation variations which we will, for the purpose of comparison between classes of models, not incorporate. Some variations include, for instance, the use of realized measures or high-frequency data as independent variables (Wang et al., 2023; Gerlach and Wang, 2020, 2022; Wu and Cai, 2024; Naimoli and Storti, 2021; Lazar and Xue, 2020; Pourrezaee and Hajizadeh, 2024; Ma et al., 2024; Antwi et al., 2020). We acknowledge the potential of incorporating realized measures or high-frequency data, and highlight it as a key area for future research, particularly in the context of improving VaR and ES forecasting for electricity futures. Furthermore, we find that some articles only employ multivariate approaches or combine forecasts from multiple models, which do not align with our objective (Leccadito et al., 2024; Fülle and Herwartz, 2024; Saidane, 2024; Marchese et al., 2020). Multivariate models fall beyond the scope of this study as they primarily aim to forecast multiple assets simultaneously, leveraging the relationships between them. These models could be explored in future research to forecast the risk dynamics of a portfolio comprising different futures. While some studies highlight combination forecast models as yielding superior results (Taylor, 2020), we will not implement them, as our focus is on evaluating models that perform well independently. Despite some of the models not being directly relevant for implementation, we can still draw some insights from the resulting articles.

The majority of the resulting articles find that well-performing models often include various implementations of GARCH models (Le, 2024; Naimoli and Storti, 2021; Kuang, 2022; Taylor, 2022, 2020; Fiszeder et al., 2024; Rice et al., 2020; Fantazzini, 2024; Garcia-Jorcano and Novales, 2021; Owusu Junior et al., 2022; Fu et al., 2023; Amaro and Pinho, 2022; Diks and Fang, 2020; Băra et al., 2024; Chen et al., 2024), CAViaR models (Taylor, 2020, 2022; Wang et al., 2023; Gerlach and Wang, 2020), GAS models (Wu and Cai, 2024; Trucíos and Taylor, 2023; Lazar and Xue, 2020; Iqbal et al., 2023; Owusu Junior et al., 2022; Liu et al., 2020; Song et al., 2021; Song and Li, 2023, 2022) and EVT models (Candia and Herrera, 2024; Taylor, 2020; Tong et al., 2024). It is also worth noting that for models relying on distributional assumptions (e.g. models employing GARCH and GAS), distributions with asymmetric properties and/or heavy tails, such as the skewed Student-t or generalized Student-t, are the most common. The subsequent section, which delves into the review of existing literature reviews, also shows that these models are widely utilized and demonstrate strong performance.

3.2. Review of Systematic Literature Reviews

A literature review is a focused summary of existing research in a particular field, aimed at capturing the current state of knowledge and identifying gaps where further research could be beneficial (Rowley and Slack, 2004). In this context, analyzing recent literature reviews on VaR or ES is essential to understand advancements in the field. Unlike the bibliometric review, which focused on analyzing articles proposing new forecasting methods for VaR or ES, this section examines literature reviews that compare multiple established forecasting methods for these risk measures. The literature reviews discussed in this section are obtained in the same way as the articles in the section above, by constructing a query[3] to search on the Scopus database. The search is limited to articles from the year 2019 up to the time of search (17.10.2024), and yielded 36 articles, where 12 could be considered literature reviews of either VaR or ES. Despite the shortcomings of VaR in comparison with ES, only 2 of these articles compare ES models.

Shayya et al. (2023) offer a systematic review of VaR modeling from 1996 to 2017, examining shifts in model popularity before and after the 2007–2009 financial crisis. Their findings show that, following the crisis, publications on VaR models increased in general. However, ARCH/GARCH models and copula-based methods gained the most traction, surpassing traditional approaches such as Variance-Covariance (VC), HS, and Monte Carlo Simulation. The literature reviewed in this section further illustrates that GARCH models continue to be popular.

For instance, Serrano Bautista and Núñez Mora (2021) compares the performance of several CaViar implementations to 12 variations of GARCH-type models with different distributional assumptions (GARCH, EGARCH, and GJR-GARCH under normal, skewed-normal, Student-t, and skewed-Student-t distributions). Their study shows that CaViaR with the Indirect Garch- and Asymmetric- configurations (henceforth IGARCH and AS) are the most accurate in out-of-sample predictions across all tested confidence levels, beating the GARCH models. Despite the findings of Serrano Bautista and Núñez Mora (2021), other studies find that simple GARCH models outperform other well-known models in some cases.

For instance, Buczyński and Chlebus (2020) highlight the robustness of GARCH models. Their study evaluates several well-known VaR forecasting models, namely, GARCH, CAViaR, and hybrid approaches such as EVT-GARCH and FHS. These models were empirically tested using stock indices data across varying levels of volatility. The study found that GARCH (1,1) with a Student-t distribution performs consistently well across periods with different volatility. Furthermore, they found that FHS with a skewed normal distribution works reasonably well in stable markets, but decline in accuracy under high volatility.

The results from a comparison of various VaR models on different futures by Amaro et al. (2022) further demonstrate that increasing the complexity of GARCH does not necessarily lead to more accurate forecasts. The authors state that selecting an appropriate distribution is often more critical for accuracy than increasing model complexity. For instance, they conclude that for oil and soy futures, probability distributions with multiple parameters (e.g., skewed and heavy-tailed distributions) are better suited than simpler ones, like the normal distribution. In contrast, normal and asymmetric normal distributions provide the best fit for gas futures, underscoring the importance of comparing GARCH models across various probability distributions to enhance forecast accuracy.

Further supporting the relevance of GARCH models, Jongadsayakul (2021) examine different VaR estimation methods on the Thailand stock market and find that asymmetric GARCH models, specifically TARCH (1,1) and EGARCH (1,1), outperform symmetric GARCH and volatility-weighted HS approaches. Together, these studies highlight the versatility and adaptability of GARCH models under various market conditions, with particular emphasis on the importance of distribution selection and model asymmetry for accurate forecasting.

Another approach often written about in literature for forecasting VaR and ES is EVT. EVT is concerned with modeling the tail dynamics of distributions by only focusing on selected extreme observations McNeil and Frey (2000).

Tong et al. (2024) compares EVT models combined with different GARCH implentations for predicting VaR on four major Chinese stock indices. Out-of-sample testing of 24 models shows that the heavy-tailed EVT hybrids (std-EVT, sstd-EVT) outperform light-tailed versions (norm-EVT, snorm-EVT) in extreme lower tail VaR forecasts (1% and 0.5%). The authors further find that the choice of distribution is found to be more influential on lower tail accuracy than the choice of volatility model, while upper tail forecasts show less variability across methods. Lastly, Amaro et al. (2022) finds that EVT-GARCH with CAViaR perform well, underscoring the robustness of the EVT family of models.

Candia and Herrera (2024) also study EVT-based models for predicting VaR and ES in the stock market. The article concludes that model performance varies depending on the evaluation method employed and the specific risk

[3]Constructed query for obtaining literature reviews: SEARCH:TITLE(("value at risk" OR "value-at-risk" OR "VaR") OR ("Expected shortfall" OR "ES" OR (("conditional" OR "tail") AND ("risk" OR "expectation")))) AND KEY(("value at risk" OR "value-at-risk") OR ("Expected shortfall" OR (("conditional" OR "tail") AND ("risk" OR "expectation")))) AND TITLE("review" OR "compar*" OR "survey") AND (PUBYEAR > 2018) AND (LIMIT-TO (LANGUAGE,"English"))

measure being forecasted. The overall best model based on backtests was the SPOT-β model. For VaR specifically, the best models were the BDT-POT and MPOT, while for ES, SPOT-β was the best performer, followed by NPOT. Using another evaluation method, namely strictly consistent scoring functions for VaR alone and ES jointly with VaR, the article finds that BDT-POT, MSM-POT, and MPOT had the most accurate forecasts. Common for these models is that they have regime-switching mechanisms that allow for the model to adapt to different states in a time series (often volatility)(Candia and Herrera, 2024). The study further highlights the importance of employing multiple backtests, as results can differ significantly between them.

Gupta and Sisir (2023) further finds that regime-switching models perform well in forecasting VaR and ES. The study evaluates the effectiveness of a regime-switching model, specifically the Markov regime-switching GJR-GARCH, against a single-regime GJR-GARCH model within the Australian energy stock market. For out-of-sample VaR forecasts, the DQ test shows the MS-GJR-GARCH model performs better, whereas the CC test indicates similar results for both models.

Although Quantile Regression (QR) relies on exogenous variables and is therefore outside the scope of our study, it remains a widely recognized method for forecasting VaR and ES that should be mentioned. Westgaard et al. (2019) compares HS, a parametric model, and a QR model for forecasting VaR across nine European energy futures (crude oil, coal, electricity and natural gas) on data from 2007 to 2017. The QR model, enhanced with an Exponentially Weighted Moving Average (EWMA) to capture volatility clustering, proves the most accurate. The authors also state that, despite this accuracy, the QR model is not without limitations, as the occurrences of VaR violations exhibit dependence. However, given the limited research on VaR forecasting for power futures, especially German electricity futures, the findings of this article are particularly interesting.

Combining forecasts from different models also falls beyond the scope of this study. However, it is an intriguing approach that should be considered in future research. Trucíos and Taylor (2023) experiment with combining forecasts from multiple models, and finds that it does not enhance accuracy beyond the best individual models when applied to Bitcoin and Ethereum data. Furthermore, they observe that no single model consistently excels for Bitcoin, but that the GAS model performs best for Ethereum. In contrast, Happersberger et al. (2020) find that combination forecasts of several different models (e.g. HS, EVT, Copula, GAS) outperformed the individual models, underscoring the potential value of combination forecasts.

Also worth noting in the name of future research is the findings of Müller and Righi (2024), that assess various multivariate models for VaR forecasting in the Brazilian stock market. The authors find that skewed Student's t copulas are the best in terms of minimizing realized loss, while DCC, R-Vine, and D-Vine models perform best under realized cost evaluation.

3.3. Review Conclusion

The analysis of systematic literature reviews in the previous section highlight well performing models in forecasting VaR and ES. Besides some multivariate models (e.g. Copula and DCC) and QR models, which fall outside the scope of this study, it is evident that GARCH, EVT hybrids, FHS, and variants of CAViaR have been widely recognized and demonstrated robust performance. An additional key insight from this analysis is the reported effectiveness of models that incorporate heavy-tails or asymmetric effects, either by construction or through distributional assumptions. These traits have shown significant improvements in the forecasting accuracy of VaR and ES metrics.

These results align closely with the findings of the bibliometric query analysis discussed in Section 3.1, which identified GARCH, CAViaR, hybrid EVT and GAS models, as the most frequently used. However, a distinct contrast is that the bibliometric query analysis found that realized volatility measures are commonly employed and popular among researchers. Furthermore, it is worth highlighting that both the bibliometric analysis and the survey of existing literature reviews indicate a relative scarcity of machine learning approaches specifically focused on VaR and ES. This gap underscores an opportunity for future studies to explore the potential of machine learning methodologies in this domain.

Both the bibliometric analysis and the review of systematic literature reviews reveal consistent trends regarding the performance of different VaR and ES models. Furthermore, they emphasize the importance of exploring diverse model configurations and distributions, as these factors significantly impact performance. Based on these insights, our study applies a broad selection of approaches to thoroughly investigate the effects of these variations on forecasting accuracy.

The class of models chosen in our study can be found in Table 4, and consist of the HS model, FHS with EWMA (FHS-EWMA) and GARCH (FHS-GARCH) as filters, standalone GARCH, EVT-GARCH, GAS, and CAViaR. The normal distribution (norm), Student's t-Distribution (std) and Skewed Student's t-Distribution (sstd) are each implemented for the models assuming distributions, namely the FHS-GARCH, GARCH, EVT-GARCH, and GAS.

In addition to model selection, the bibliometric analysis provided valuable insights into the most frequently employed backtesting methods for VaR (Table 1), joint VaR and ES (Table 2), and comparative evaluation (Table 3). These findings provide valuable insight into the popularity of various methods, which can reflect their ease of implementation, widespread recognition, and comprehensibility. By selecting the most commonly used methods for our backtesting framework, we ensure that the approach is not only well-understood, but also facilitate comparability and alignment with established practices in the field. Our study adopts most of the backtesting methods mentioned , which for evaluating VaR include the CC, UC, and DQ tests. For ES, the selected methods comprise the CoC, ER, and ESR tests. Additionally, the MCS test is employed for model comparison where the AL loss function is applied for performance evaluation of joint VaR and ES forecasts. In addition to the widely used VaR backtests identified in the literature review, we have incorporated the VaR Duration Test (VDT), also known as the Continuous Weibull test, proposed by Christoffersen and Pelletier (2004). This addition enhances the backtesting of VaR violation clustering by incorporating a duration-based approach, complementing the other VaR tests, such as first-order dependence tests (CC) and regression-based tests (DQ). The inclusion of VDT provides a more comprehensive assessment of model performance by integrating all the three typical approaches to independence testing (Berger and Moys, 2021).

4. VaR and ES Forecasting Models

In this section, we present and discuss the the methodology and selected forecasting models employed in our study to forecast day-ahead VaR and ES, along with the backtesting methods and loss functions chosen for evaluation.

We implemented a rolling horizon forecast methodology with a window size of 250 (roughly a year of data), resulting in 989 observations for out-of-sample day-ahead forecasting. The window size of 250 meets Basel's requirement for a minimum of one year of historical observations (BIS, 2006). This window size also aligns with the European Banking Authority's methodology, which employs a continuous 12-month period for calibrating stressed VaR (Authority, 2024). These day-ahead VaR and ES forecasts are evaluated against the actual returns for the corresponding days. To maintain computational feasibility for parameter estimation in parametric and semi-parametric models, the model parameters are refitted every 10 observations within the rolling window.

For consistency, we will henceforth denote VaR at confidence level α as the lower quantile of the return profit-and-loss (P&L) distribution, with ES representing the mean of losses that exceed this quantile. As a result, both VaR and ES will primarily take positive values. Forecasting these risk metrics is mainly done in three distinct approaches; non-parametric, semi-parametric, parametric.

Parametric models assume a specific distribution, such as the normal (norm), Student's t (std), or skewed Student's t (sstd). In these models, day-ahead volatility is typically estimated first, allowing VaR and ES to be computed analytically based on the assumed distribution.

Non-parametric models make no assumptions about the underlying distribution and rely solely on historical observations to forecast future values. These models are simple, computationally efficient, and intuitive, as they do not require parameter estimation. However, their performance is sensitive to hyperparameters, and they inherently assume that historical patterns will repeat in the future.

Semi-parametric models lie somewhere between parametric and non-parametric models by typically not assuming any distribution, but still using parameters which are to be estimated.

In the remaining part of this section, we will shortly discuss each of the models and which specification we employed. For a detailed introduction on each of these models, please refer to the respective articles cited.

4.1. GARCH-Type Models

GARCH models were introduced by Bollerslev in 1986 (Bollerslev, 1986). In the standard GARCH(p,q), conditional volatility is modeled after its lagged p values and q lagged residuals:

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

where σ_t^2 is the conditional variance, ϵ_t are the residuals, and $\omega, \alpha_i, \beta_j$ are model parameters. In addition to this model, we will employ the GJR-GARCH model introduced by GLOSTEN et al. (1993). Here, the equation for conditional volatility is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \left(\alpha_i + \gamma_i I(\epsilon_{t-i} < 0)\right) \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where I is the indicator function, σ_t^2 is the conditional variance, ϵ_t are the residuals, q and p are lag orders, and $\omega, \alpha_i, \lambda_i, \beta_j$ are model parameters. The model uses a threshold of zero in the indicator function to separate the impacts of past shocks. This allows the GJR-GARCH model to capture asymmetric effects.

In our implementation, whether it is standalone GARCH or hybrids, we incorporate the lags $q = 1$ and $p = 1$ as Bollerslev (1986) showed that the GARCH(1,1) works well for most applied situations. Moreover, across all uses, the model parameters are estimated as per previously mentioned, after each 10 observation. We will incorporate norm, std, and sstd version wherever they are employed. Also, each model which use of GARCH to model conditional volatility will have one one version using the normal GARCH and one using the GJR-GARCH. In our code, we we used the `rugarch` R package described in Galanos (2024), which is frequently updated. Hill and Mccullough (2020) mentions this as being the preferred package among several option (`tseries` and `fGarch`), due to the comprehensive documentation and flexibility in terms of optimization choices, forecast capabilities, and options in tuning specifications.

4.2. EVT-GARCH

Extreme Value Theory (EVT) captures rare events by focusing on the tail of returns, and is usually combined with a conditional volatility model, such as GARCH, to account for volatility clustering. In the Peaks-over-threshold (POT) framework within EVT, standardized residuals from the conditional volatility model exceeding a threshold u, is used to fit a Generalized Pareto Distribution (GPD). The choice of a threshold is a demanding task, as a high threshold could exclude most of the observations, while a low threshold reduces the precision. We will employ the common $t = 95\%$ quantile as threshold, recognizing that the selection of high quantile thresholds has been shown to have minimal influence on the computed VaR (Benito et al., 2023; Echaust and Just, 2020). For a more comprehensive discussion on threshold selection, see Echaust and Just (2020).

The standardized residuals $z_t = \frac{r_t}{\sigma_t}$ exceeding a $u = F(t)^{-1}$, where σ_t is obtained through a GARCH model, are used to fit a generalized extreme value (GEV) distribution. Specifically, it is found that under normal assumption, the GEV approximates to the GPD:

$$F(x) = 1 - \left(1 + \xi \frac{x-u}{\beta}\right)^{-1/\xi}, \quad \xi \neq 0,$$

where ξ is the shape parameter and β is the scale parameter.

Using the GPD and Pickands-Balkema-de Haan theorem, VaR at quantile α can be computed analytically as:

$$\mathrm{VaR}_{\mathrm{EVT}} = \begin{cases} u + \frac{\beta}{\xi}\left(\left(\frac{m}{n_u}c\right)^{-\xi} - 1\right), & \text{if } \xi \neq 0, \\ u - \beta \log\left(\frac{m}{n_u}c\right), & \text{if } \xi = 0, \end{cases}$$

where m is the window size and n_u is the number of exceedances.

Expected Shortfall is then derived as:

$$\mathrm{ES}_{\mathrm{EVT}} = \begin{cases} \frac{\mathrm{VaR}_{\mathrm{EVT}} + \beta - u\xi}{1-\xi}, & \text{if } \xi \neq 0, \\ \mathrm{VaR}_{\mathrm{EVT}} + \beta, & \text{if } \xi = 0. \end{cases}$$

In our implementation, both the GARCH and GPD parameters are estimated with a 10 day refitting frequency. In cases where the GPD fitting failed, we will fall back to a standard GARCH computation using the conditional volatility.

4.3. Generalized Autoregressive Score (GAS) Models

The GAS models, introduced by Creal et al. (2013), utilize the full density structure of the distribution F, rather than relying solely on the first and second moments Laporta et al. (2018). This specification allows for a better modeling of time-varying parameters. Here, the scaled score of the conditional density is leveraged to update parameters dynamically. The volatility equation for the GAS model is specified as:

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \delta s_t \frac{\partial \log F(r_t|\sigma_t^2)}{\partial \sigma_t^2},$$

where s_t is a scaling factor, and ω, β, and δ are model parameters.

For this model we employed the rolling forecast function from the `GAS` R package of Ardia et al. (2019), using norm, std, and sstd as distributions.

4.4. Historical Simulation (HS)

Historical Simulation is a widely used non-parametric method, attributed to its simplicity, and is commonly employed in the industry to compute VaR for regulatory purposes. In this approach, past returns are sampled from a selected window, and the quantile at a chosen confidence is calculated to forecast day-ahead VaR. The ES is then computed as the mean of returns exceeding the set VaR quantile.

The method's reliance on historical data introduces some limitations. It assumes that past market behavior is a reliable predictor of future risks, which can lead to inaccuracies in periods of structural market changes or heightened volatility. Additionally, the choice of hyperparameters, such as the window size and bootstrap size, significantly impacts its performance. A window that is too short may fail to capture a representative distribution of returns, while a window that is too long can dilute the impact of recent market conditions.

We employed a bootstrap size of 10,000 and utilized the R package `quarks` by Letmathe (2024) for this model.

4.5. Filtered Historical Simulation (FHS)

Filtered Historical Simulation improves upon HS by adjusting returns for time-varying volatility. The residuals are obtained by using a conditional volatility model, such as a GARCH or an Exponentially Weighted Moving Average (EWMA) model, which are then rescaled using the current volatility forecast from the same model. Compared to HS, this allows for conditional volatility adjustment, enhancing the response of risk estimates during periods of market stress.

In our study, we will apply a $\lambda = 0.94$ in the EWMA specification of FHS. This choice aligns with the parameterization popularized by J.P. Morgan's RiskMetrics framework, which found that $\lambda = 0.94$ effectively captures the persistence in daily financial return volatilities (J.P. Morgan/Reuters, 1996). Due to the fact that this is constant, there is no need to refit any parameters. This is in contrast to for when the GARCH is employed.

For FHS, we modified the `quarks` package to incorporate parameter refitting frequency when computing rolling forecasts of VaR and ES using GARCH as a filter. A bootstrap size of 10,000 was still employed in this approach.

4.6. Conditional Autoregressive Value at Risk (CAViaR) Models

CAViaR models were introduced by (Engle and Manganelli, 2004), and forecast VaR directly without requiring volatility estimation. Instead, the quantiles are derived from regression equations using lagged values of VaR and returns. Unlike QR, which uses exogenous variables to model the quantiles as the dependent variable, this approach relies solely on endogenous factors. They introduced four distinct specifications for updating VaR: Symmetric Absolute Value (SAV), Asymmetric Slope (AS), Indirect GARCH (IGARCH), and Adaptive model.

For the SAV model, VaR evolves as:

$$\mathrm{VaR}_t = \beta_1 + \beta_2 \mathrm{VaR}_{t-1} + \beta_3 |y_{t-1}|$$

The AS model extends this by incorporating asymmetry, separately accounting for positive and negative y_{t-1} values:

$$\mathrm{VaR}_t = \beta_1 + \beta_2 \mathrm{VaR}_{t-1} + \beta_3 \max(y_{t-1}, 0) + \beta_4 \min(y_{t-1}, 0)$$

In the IGARCH approach, VaR is computed via a square-root GARCH process, reflecting nonlinear adjustments:

$$\mathrm{VaR}_t = \left[(1 - 2 \cdot (c < 0.5)) \cdot \left(\beta_1 + \beta_2 \mathrm{VaR}_{t-1}^2 + \beta_3 y_{t-1}^2\right)\right]^{1/2}$$

Finally, the Adaptive model incorporates a feedback mechanism to adjust VaR dynamically:

$$\mathrm{VaR}_t = \mathrm{VaR}_{t-1} + \beta_1 \left(c - \frac{1}{1 + \exp(G \cdot (y_{t-1} - \mathrm{VaR}_{t-1}))}\right),$$

where G is an arbitrary constant, and β_i are model parameters for all four specifications. While G could be treated as a parameter to be estimated, the model prioritizes simplicity in its design. These parameters are estimated such that VaR and ES remains non-negative, for some specifications this is by construction.

Taylor (Taylor, 2019) expanded on this model to jointly estimate ES. He proposed two approaches; auto-regressive (AR) and multiplicative (MULT).

For the AR approach, ES is obtained by adding an adjustment term to VaR. The additive term x_t is modeled using an auto-regressive equation:

$$x_t = \begin{cases} \gamma_0 + \gamma_1 (VaR_{t-1} - y_{t-1}) + \gamma_2 x_{t-1}, & \text{if } y_{t-1} \leq VaR_{t-1}, \\ x_{t-1}, & \text{otherwise.} \end{cases}$$

$$\mathrm{ES}_t = VaR_t + x_t,$$

where γ_i are non-negative model parameters. x_t is by construction a non-negative number which is added to VaR, which ensures that ES is larger than VaR at any time.

For the MULT approach, ES is modeled as a multiple of the VaR where the multiplier is modeled after the parameter γ_0:

$$\mathrm{ES}_t = (1 + \exp(\gamma_0)) VaR_t.$$

The multiple, $(1 + \exp(\gamma_0)) \geq 0$, ensures that ES never is smaller than VaR. Due to its simplicity, the multiplicative approach is often preferred over the AR method, which requires recursive computations.

The CAViaR model was implemented in R with mathematical computations and recursive loops written in C++. The code is based on the original Matlab code by Simone Manganelli and Robert Engle[4]. Under the same convergence criteria applied in the optimization algorithms of the other models, the CAViaR models frequently failed to converge, particularly in the SAV, AS, and Adaptive specifications. Consequently, these models are excluded from the tables presented in Section \ref{sec:results}.

[4] The original code is available at `https://www.simonemanganelli.org/Simone/Research.html`

5. Backtesting of VaR and ES Models

The backtesting framework used in this study is designed to evaluate how well VaR and ES models predict financial risks. The backtesting process involves several adequacy tests that help identify if the model aligns with theoretical expectations, and comparative tests to examine statistically performing models.

5.1. Adequacy Tests for VaR Models

Adequacy tests for VaR focuses on two critical aspects: the correct hit rate and the independence of violations. The correct hit rate ensures that the model proportion of exceptions (when actual losses exceed the predicted VaR) align with the model's specified confidence level, while the independence of violations checks that these exceedances are not clustered in time but occur randomly as expected.

The following tests are implemented in this study:

Unconditional Coverage (UC) Test introduced by Kupiec (1995), assesses whether the observed frequency of VaR violations matches the expected frequency implied by the confidence level α. The likelihood ratio (LR) statistic for the test is:

$$\begin{aligned} LR_{\mathrm{UC}} = & -2 \ln \left[(1-\alpha)^{n-n_1} \alpha^{n_1}\right] \\ & + 2 \ln \left[\left(1 - \frac{n_1}{n}\right)^{n-n_1} \left(\frac{n_1}{n}\right)^{n_1}\right], \end{aligned}$$

where n_1 is the number of VaR exceedances, n is the total number of observations, and α represents the confidence level (e.g., $\alpha = 0.05$ for a 95% confidence interval).

The null hypothesis states that the observed hit rate equals the expected hit rate. Under the null, LR_{UC} follows a chi-squared distribution with one degree of freedom, $\chi^2(1)$. A failure to reject the null indicates that the model correctly predicts the frequency of VaR violations.

Conditional Coverage (CC) Test introduced by Christoffersen (1998), combines the UC test with a test for the independence of violations. The test is based on modeling the relationship between VaR violations using a first-order Markov chain, where the transition probability matrix is given by:

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},$$

with $\pi_{ij} = \Pr(I_t(p) = j \mid I_{t-1}(p) = i)$, representing the probability of a violation ($I_t(p) = 1$) or no violation ($I_t(p) = 0$) conditional on the state of the previous day.

The test statistic for independence, LR_{IND}, is calculated using a LR test as follows:

$$LR_{\mathrm{IND}} = 2 \ln \left(\frac{(1 - \hat{\pi}_{01})^{T_{00}} \hat{\pi}_{01}^{T_{01}} (1 - \hat{\pi}_{11})^{T_{10}} \hat{\pi}_{11}^{T_{11}}}{(1 - \hat{\pi})^{(T_{00} + T_{10})} \hat{\pi}^{(T_{01} + T_{11})}} \right),$$

Table 4
Overview of models with specifications, parameters, and distributions. Abbreviations: HS stands for Historical Simulation, FHS stands for Filtered Historical Simulation, GAS represents Generalized Autoregressive Score, and CAViaR refers to Conditional Autoregressive Value at Risk. GARCH specifications include standard GARCH (sGARCH) and GJR-GARCH (gjrGARCH). The entries in each column represent specific model characteristics, such as distributions or specifications, which are used in combination with other model components, including bootstrap size (B), smoothing parameter (λ), and thresholds, are also noted where applicable.

Model Class	Distribution	GARCH Specification	CAViaR ES Specification	Refitting Frequency	Hyperparameters
HS	–	–	–	–	$B = 10000$
FHS-EWMA	–	–	–	–	$B = 10000$ $\lambda = 0.94$
FHS-GARCH	Normal, Student's t, Skewed t	sGARCH, gjrGARCH	–	10	–
GARCH	Normal, Student's t, Skewed t	sGARCH, gjrGARCH	–	10	–
EVT-GARCH	Normal, Student's t, Skewed t	sGARCH, gjrGARCH	–	10	Threshold = 0.95
GAS	Normal, Student's t, Skewed t	–	–	10	–
CAViaR	–	SAV, AS, indirect-GARCH, Adaptive	AR, MULT	10	–

where $\hat{\pi}_{ij}$ is the estimated probability of transitioning from state i to state j, and T_{ij} represents the number of observed transitions from state i to state j. The null hypothesis assumes independence of violations, which implies that the transition probabilities are constant, and under this hypothesis, LR_{IND} follows a chi-squared distribution with one degree of freedom, $\chi^2(1)$.

Due to its focus on single-lag dependencies, the LR_{IND} may overlook more higher-order behaviors. To overcome these limitations, we explore alternative VaR backtesting methods.

Dynamic Quantile (DQ) Test by Engle and Manganelli (2004) tests both the correct hit rate and the independence of violations in a regression framework. The auxiliary regression is:

$$Hit_t^{\alpha} = \beta_0 + \sum_{i=1}^{L} \beta_i Hit_{t-i}^{\alpha} + \beta_{L+1} \text{VaR}_{t-1}(\alpha) + u_t,$$

where L is the lag order, β_i are regression parameters, and Hit_t^{α} represents whether an exceedance occurred at time t (1 if true, 0 otherwise), minus the expected rate α. The null hypothesis of $\beta_i = 0$ is assessed using a Wald test, where a rejection signifies a misspecified model. When backtesting, we employed a lag order of $L = 4$, consistent with the specification used in Engle and Manganelli (2004).

Advantages of the regression-based approaches include the possibility of incorporating hit rates with different lag orders and exogenous variables, which is not feasible in the UC framework. However, Berger and Moys (2021) find that duration based approaches often outperform their regression counterparts, a result which can be attributed to the no-memory property of the distribution under the null.

VaR Duration Test (VDT) introduced by Christoffersen and Pelletier (2004), evaluates the time intervals, or durations, between consecutive VaR exceedances to assess the adequacy of a VaR model. The test is based on the idea that a well-specified model should produce exceedances that occur randomly over time, with durations between exceedances consistent with an exponential distribution under the null hypothesis. The exponential distribution assumes independence between violations and a constant hazard rate, meaning the likelihood of an exceedance does not depend on how much time has passed since the last one.

Under the alternative hypothesis, where duration dependence or clustering is considered, the durations are modeled using a Weibull distribution, which is more flexible and allows for changing hazard rates. The Weibull distribution can capture patterns such as clustering (frequent short durations) or dispersion (long durations), which the exponential distribution cannot. If the observed durations deviate significantly from what is expected under the null, it suggests potential issues with the model, such as clustering of violations. Since its introduction, the VDT has inspired the development of other duration-based tests, which often differ in the distributions assumed.

5.2. Adequacy Tests for ES Models

For ES, adequacy tests ensure that the predicted ES values accurately reflect observed extreme losses. These methods are rooted in the joint elicitability of VaR and ES as described by Fissler and Ziegel (2016).

Exceedance Residual (ER) Test Proposed by McNeil and Frey (2000), this test evaluates whether the losses that exceed the VaR threshold are consistent with the predictions of the ES model. Specifically, it examines the behavior of the excess conditional shortfall, i.e. the difference between the actual observed losses and the ES when the VaR is exceeded. The null hypothesis states that this excess conditional shortfall is independent and identically distributed (i.i.d.) with a zero mean.

The test employs a one-sided t-test to investigate the alternative hypothesis that the excess shortfall has a mean greater than zero, indicating that the ES model systematically underestimates the average severity of losses beyond the VaR level. If the null hypothesis is rejected, it suggests that the ES predictions are insufficiently conservative and fail to capture the tail risk accurately.

To enhance the robustness of the results, a bootstrap procedure is used to calculate p-values. This approach mitigates potential biases arising from assumptions about the underlying distribution of the excess shortfall, ensuring a more reliable evaluation. The ER test is particularly valuable in assessing whether the ES model correctly reflects the expected magnitude of extreme losses, a critical component of risk management frameworks.

Conditional Calibration (CoC) Test by Nolde and Ziegel (2017) leverages the joint elicitability of VaR and ES to test their adequacy:

$$S(Q_t, \mathrm{ES}_t, y_t) = -\ln\left(\frac{\alpha - 1}{\mathrm{ES}_t}\right) - \frac{(y_t - Q_t)(\alpha - I(y_t \le Q_t))}{\alpha \mathrm{ES}_t},$$

where Q_t is the VaR at time t, ES_t is the ES at time t, y_t represents the realized return or observation at time t, α denotes the quantile, and I is the indicator function.

This test evaluates whether VaR and ES predictions jointly satisfy theoretical properties such as calibration and coherence.

Expected Shortfall Regression (ESR) Tests proposed by Bayer and Dimitriadis (2020), are the first strictly consistent backtests relying solely on ES forecasts and realized returns. The ESR backtest model the ES forecast as a linear function:

$$\mathrm{ES}_t = c_1 + c_2 \hat{e}_t,$$

where $\hat{e}_t$ is the ES forecast, and c_1 and c_2 are regression parameters. For correctly specified forecasts, $c_1 = 0$ and $c_2 = 1$, and these are tested using a Wald statistic.

The Auxiliary ESR (AESR) backtest incorporates auxiliary VaR forecasts $\hat{v}_t$ as explanatory variables, jointly evaluating VaR and ES in regression equations for the quantile and ES:

$$\mathrm{VaR}_t = b_1 + b_2 \hat{v}_t, \quad \mathrm{ES}_t = c_1 + c_2 \hat{e}_t.$$

However, it is limited by its reliance on VaR forecasts.

The Strict ESR (SESR) backtest removes the dependency on VaR by using $\hat{e}_t$ as the explanatory variable for both quantile and ES regressions:

$$\mathrm{VaR}_t = b_1 + b_2 \hat{e}_t, \quad \mathrm{ES}_t = c_1 + c_2 \hat{e}_t.$$

This allows for standalone ES evaluation, but it may suffer from quantile equation misspecification.

The Intercept ESR (IESR) backtest fixes the slope $c_2 = 1$ and tests only the intercept c_1:

$$\mathrm{ES}_t = c_1 + \hat{e}_t.$$

This approach enables one-sided tests, addressing regulatory needs to prevent underestimation of risks.

Simulation studies demonstrate that these backtests are well-sized and outperform existing methods such as ER and CoC, particularly under model misspecification. We employ the ER, CoC, and ESR backtests from the `esback` R package by Bayer and Dimitriadis (2023).

5.3. Comparative Tests

Comparative backtesting uses loss functions to evaluate and rank models, offering a robust framework for comparing the relative performance of competing models. These tests are particularly valuable for selecting a model when multiple models are correctly specified.

Model Confidence Set (MCS) developed by Hansen et al. (2011), provides a rigorous statistical framework for evaluating and selecting models based on their relative performance across a set of competing alternatives. Unlike traditional model selection criteria, which aim to identify a single "best" model, the MCS focuses on identifying a subset of models, Superior Set of Models (SSM) that are statistically indistinguishable from the best-performing model at a given confidence level.

The MCS methodology begins with a finite set of candidate models, and iteratively reduces this set to a subset SSM, containing the best models with a pre-specified confidence level $(1 - \alpha)$. This is achieved through a series of sequential hypothesis tests, which evaluate whether all models in the set perform equally well under a user-defined loss function. If the null hypothesis of equal performance is rejected, the model contributing most to the loss function is eliminated. This process continues until the null hypothesis can no longer be rejected, resulting in a final set of models that are considered statistically equivalent in terms of performance.

The MCS relies on bootstrap techniques to estimate critical values and p-values for the hypothesis tests. Importantly, asymptotically the best-performing model(s) are included in SSM with a probability of at least $1 - \alpha$.

One of the key strengths of the MCS framework that it can be applied in a wide range of contexts, including forecasting and regression analysis, using any user-defined loss function to evaluate model performance. Additionally, the size of the MCS depends on the informativeness of the data. When the data is highly informative, the MCS is likely to consist of fewer models, as the differences between models are more easily detected. Conversely, when the data is less informative, the MCS tends to be larger, reflecting the difficulty of distinguishing between competing models.

Unlike other typical model evaluation approaches, the MCS does not require a benchmark model to be specified. Instead, it evaluates all models in the candidate set simultaneously and eliminates only those that are statistically inferior. This benchmark-free approach makes the MCS particularly appealing in situations where the true model is unknown or not included in the set of candidates.

5.4. Loss Functions

Loss functions play a central role in comparative backtesting, providing a quantitative basis for evaluating the forecast accuracy of competing models. They are also helpful in optimizing model parameters. However, for loss functions to be effective, they must satisfy specific characteristics that ensure their suitability for these tasks.

A key concept for the design of effective loss functions is elicitability. In essence, strictly consistent scoring functions reward accurate predictions by ensuring that the correct forecast minimizes the scoring function, serving as a proxy for accuracy. This property makes them especially useful for evaluating forecast accuracy and optimizing model parameters.

For VaR, the Quantile Loss (QL) function introduced by González-Rivera et al. (2004) is commonly used:

$$QL_t(\alpha) = (\alpha - I(r_t < \text{VaR}_t(\alpha)))(r_t - \text{VaR}_t(\alpha)),$$

where I is the indicator function. This loss function penalizes underestimations more heavily than overestimations, reflecting the asymmetric nature of financial risk. While the Quantile Loss function is well-suited for evaluating VaR forecasts, our focus will primarily focus on another loss function that also incorporates insights about ES.

Regarding ES, Fissler and Ziegel (2016) showed that VaR and ES are jointly elicitable. Specifically, they derived the following general form of strictly consistent scoring functions for evaluating joint VaR and ES forecasts:

$$\begin{aligned}
&S(Q_t, ES_t, y_t) = \\
&\big(I(y_t \leq Q_t) - \alpha\big)G_1(Q_t) - I(y_t \leq Q_t)G_1(y_t) \\
&+ G_2(ES_t)\big(ES_t - Q_t + I(y_t \leq Q_t) \times (Q_t - y_t)/\alpha\big) \\
&- \zeta_2(ES_t) + a(y_t),
\end{aligned}$$

where G_1, G_2, ζ_2, and a are functions that must satisfy several conditions to ensure consistency. Some conditions include that G_1 must be an increasing function, and $G_2 = \zeta_2'$, implying that G_2 is the derivative of ζ_2. Additionally, ζ_2 must be an increasing and convex function.

These conditions allow for a range of alternative specifications. Taylor (2019) introduced the AL log score (henceforth denoted AL or joint loss) which, by design, corresponds to the negative of the Asymmetric Laplace log-likelihood. This was achieved by setting $G_1 = 0$, $G_2(x) = -1/x$, $\zeta_2(x) = -\ln(-x)$, and $a = 1 - \ln(1 - \alpha)$:

$$S(Q_t, ES_t, y_t) = -\ln\left(\frac{\alpha}{|ES_t|}\right) - \frac{(y_t - Q_t)(\alpha - I(y_t \leq Q_t))}{\alpha ES_t}.$$

The AL log score is particularly appealing because it belongs to the set of scoring functions proposed by Fissler and Ziegel (2016). Furthermore, its connection to the Asymmetric Laplace log-likelihood, a well-established method in quantile estimation literature, enhances its theoretical foundation and practical applicability as a robust loss function. In this study, we use the AL log score not only to evaluate the accuracy of joint VaR and ES forecasts, but also to optimize model parameters and compare competing models within the MCS framework.

6. Empirical study

6.1. Source

This study evaluates VaR and ES forecasting models using eight German power futures contracts traded on the EEX. More specifically, it is based on out-of-sample analysis of daily-close settlement prices for futures contracts on baseload and peak load electricity, obtained via the Refinitiv API (eikon.refinitiv.com). These contracts span various delivery periods, including daily, monthly, quarterly, and yearly durations, and are expressed in Euro (EUR). Further details on the contracts can be found in Table 5.

The datasets spans from April 9, 2019, to October 31, 2024, encompassing 1,240 daily price observations for each contract. We chose this period as it was the most recent and longest available overlap between the eight time series, ensuring the models are evaluated over the same period for each contract. Furthermore, this period encompasses both the Covid-19 outbreak and the full-scale Russian invasion of Ukraine, events that significantly disrupted global markets. Analyzing the models' performance over different periods enables us to assess their robustness across varying market conditions.

6.2. Description of the Futures Contracts

This section provides a more in depth explanation of the contracts used in the study. Table 5 provides details on the daily, monthly, quarterly, and yearly base and peak load front contracts for German electricity futures traded at the EEX exchange, and used in this study. Front contracts refer to futures contract that correspond to the nearest upcoming delivery period (Badáňová, 2015). The DEBQ contract for instance, is a contract on baseload delivery for the next upcoming quarter. The prices are in Euro (EUR) and are only for weekdays.

Table 5
An overview of the contracts used in the study. All series are sourced from the Refinitiv API. The RIC column represents the Refinitiv Identification Code corresponding to the continuation values of the futures mentioned.

RIC	Futures Name	Underlying Product	Country	Exchange	Currency	Periodicity
DBc1	EEX Phelix DE Base Daily Energy Future	Baseload	Germany	EEX	EUR	Daily
DPc1	EEX Phelix DE Peak Daily Energy Future	Peakload	Germany	EEX	EUR	Daily
DEBMc1	EEX Phelix DE Base Monthly Energy Future	Baseload	Germany	EEX	EUR	Monthly
DEPMc1	EEX Phelix DE Peak Monthly Energy Future	Peakload	Germany	EEX	EUR	Monthly
DEBQc1	EEX Phelix DE Base Quarterly Energy Future	Baseload	Germany	EEX	EUR	Quarterly
DEPQc1	EEX Phelix DE Peak Quarterly Energy Future	Peakload	Germany	EEX	EUR	Quarterly
DEBYc1	EEX Phelix DE Base Yearly Energy Future	Baseload	Germany	EEX	EUR	Yearly
DEPYc1	EEX Phelix DE Peak Yearly Energy Future	Peakload	Germany	EEX	EUR	Yearly

The RIC column in Table 5 stands for Refinitiv Identification Code, and is the unique code used for that time series in the Refinitiv API. It is included in the table to enable the readers of this study to find the data used. The c1 at the end of the RIC code indicates that the series are continuation series, meaning they have been adjusted for roll-over returns. Roll-over returns occur when there is a price shock to the futures contract, as a result of the expired contract being replaced by a new one (Westgaard et al., 2019). Hence, the return dynamics of a continuous series are different from those of a non-adjusted series for the same contract.

Peak load, denoted by a P in the RIC, refers to 12 hour constant power delivery from 8:00 AM to 8:00 PM on weekdays, for the whole delivery period (EEX, 2024a). Base load, on the other hand, is denoted by a B and represents the constant supply of power over all hours of the delivery period (EEX, 2024a). The DE notation indicates that the contracts are on the German power market. For German base and peak power futures, the prices are derived from the Physical Electricity Index (Phelix), which is based on the EPEX SPOT price (EEX, 2024a). Notably, the first two contracts in Table 5 (DBc1 and DPc1), while not explicitly marked with 'DE,' are also German power futures.

The prices of the contracts are derived from the Physical Electricity Index (Phelix), which is based on the EPEX SPOT price (EEX, 2024a).

6.3. Cleaning and Preprocessing

For the data cleaning process, little work was needed, as the Refinitiv API provided us with clean and preprocessed data. As mentioned in the section above, the series was already adjusted for roll-over returns. Furthermore, all the contracts trade on the same dates, so we did not need remove any trading days either. The length of each series was however different. We chose to include observation for the longest overlapping period of all time series, to be sure the models are evaluated over the same period for each contract. Furthermore, we use log returns for the forecasting, calculated as shown by equation (1).

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right), \tag{1}$$

6.4. Descriptive Statistics

Figure 4 and 5 shows the price plots of the baseload and peakload contracts respectively. In both figures we observe a significant increase in prices in the year 2022 to 2023, likely driven by the Russian invasion of Ukraine. A large spike is also evident for both the peak and baseload futures prices on August 31, 2022, the date when Russian gas giant Gazprom halted gas deliveries through Nord Stream 1. In Figure 4 we further observe a spike in the daily baseload contract (DBc1) price at the end of the observed period.

Figure 6 compares the development of baseload and peakload prices, on contracts with the same delivery period. We observe that the prices of the peakload contracts (in yellow), is mostly higher than its baseload counterpart, with the exception of the yearly contracts. This price difference can be explained by the fact that electricity prices tend to be higher during during peak hours due to increased demand, resulting in higher prices for the futures contracts.

Figure 7 provides the plot of the log returns for all eight contracts. The plots show that the returns exhibit no directional trend, and varies across a constant mean. Furthermore,

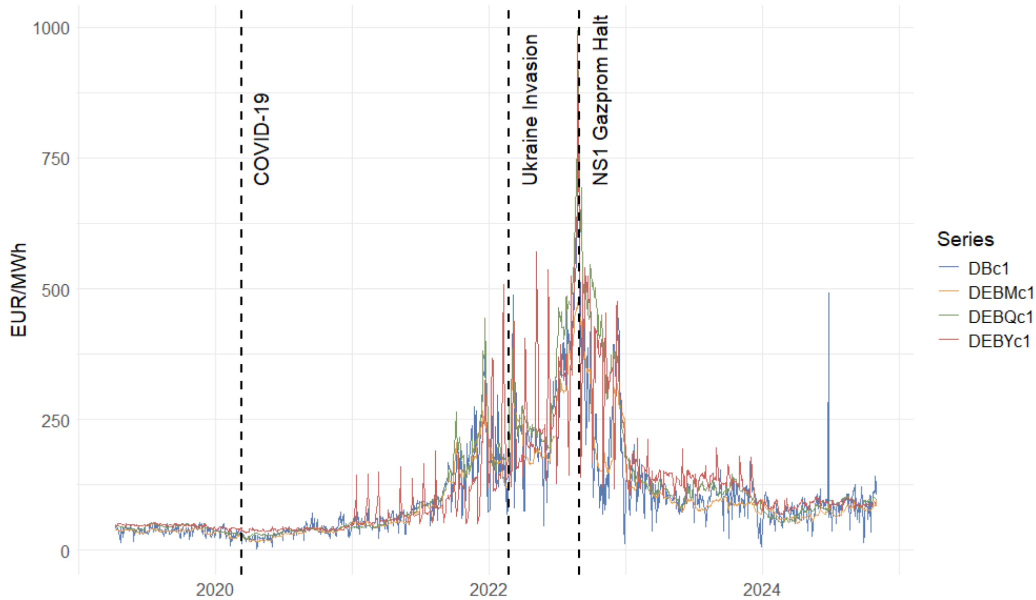

Figure 4: Development of the prices for the baseload futures with different delivery periods (DBc1, DEBMc1, DEBQc1, DEBYc1). Key events such as the COVID-19 outbreak, the Russian invasion of Ukraine, and the halt of the Nord Stream 1 gas pipeline are marked to illustrate their impact on price trends.

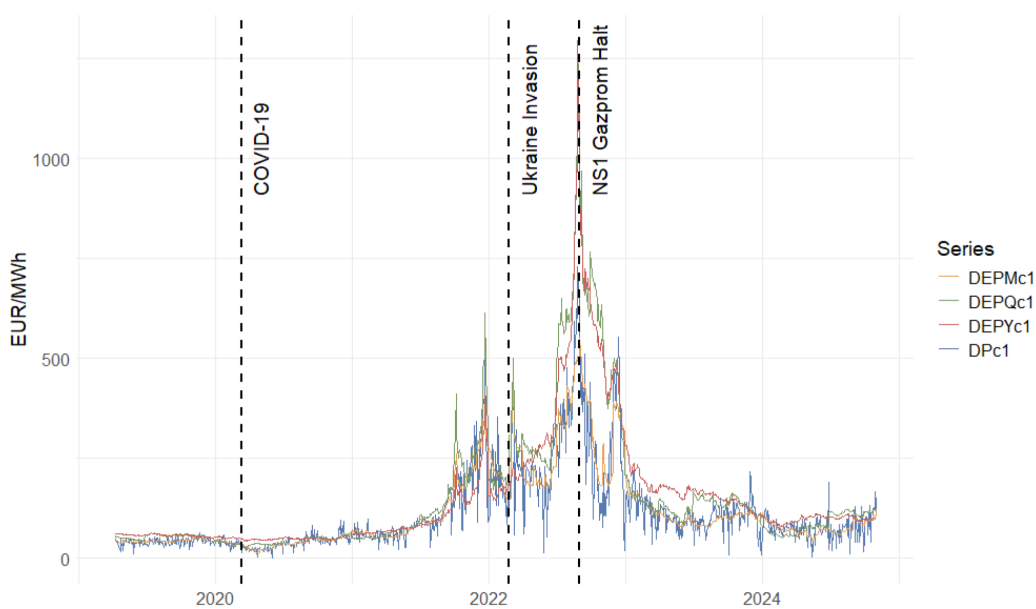

Figure 5: Development of the prices for the peakload futures with different delivery periods (DPc1, DEPMc1, DEPQc1, DEPYc1). Key events such as the COVID-19 outbreak, the Russian invasion of Ukraine, and the halt of the Nord Stream 1 gas pipeline are marked to illustrate their impact on price trends.

the contracts with longer delivery periods (e.g. DEBQc1, DEPQc1, DEBYc1, DEPYc1) have a notable increase in volatility and larger spikes around the start of 2022. A possible explanation for this could be increased longterm energy uncertainty as a result of the Russian invasion of Ukraine at the 24th of February 2022.

Table 6, which display the descriptive statistics of the log returns, confirms what we see in Figure 7, as the mean of the returns are zero across all contracts. The standard deviation, however, varies. The daily contracts are by far the most volatile contracts, with DPc1_return being the most volatile, with a standard deviation of 0.51, followed by DBc1_return with a standard deviation of 0.33. The volatility decrease slightly with the length of the delivery period, where the yearly contracts are the least volatile. We can observe a similar pattern by looking at the range of the returns.

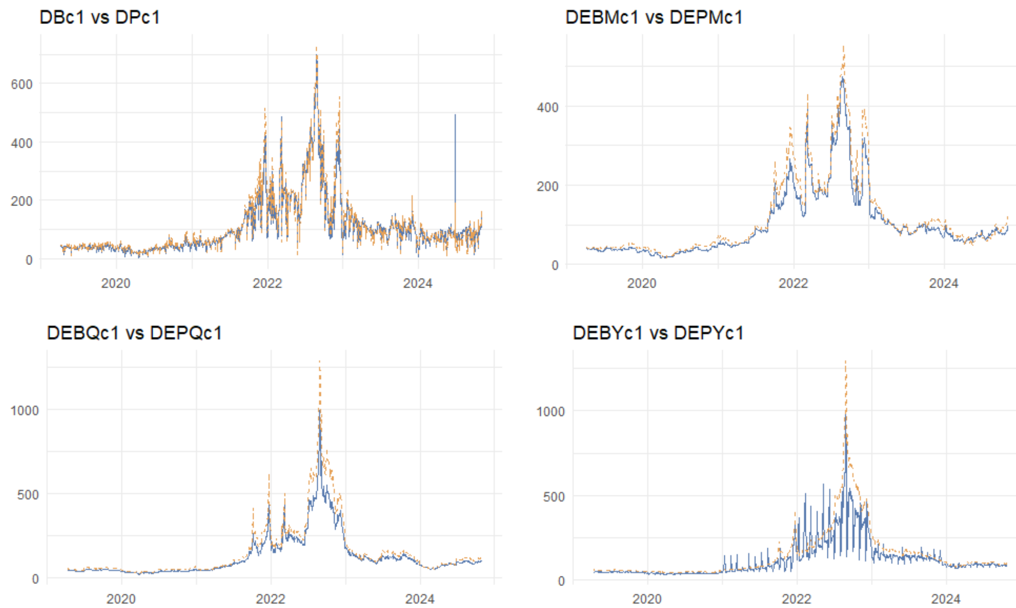

Figure 6: Prices in EUR/MWh of baseload (blue) and peakload (yellow) contracts with same delivery period.

Table 7 further displays the annualized standard deviation of the different contracts each year. A notable observation is the heightened volatility in longer-duration contracts during 2021 and 2022, which may signal market stress as these contracts reflect expected prices over extended periods.

From Table 6 we observe that the daily, monthly and quarterly contracts display a positive skew, with the monthly displaying the most, indicating an increased likelihood of positive returns. This is in contrast to the yearly contracts DEBYc1_return and DEPYc1_return, which have negative skews of -3.47 and -5.08 respectively. This can be interpreted as these contracts being more prone to downside risk compared to the other contracts.

All series exhibit leptokurtic distributions, with the kurtosis significantly higher than that of a normal distribution. DEPYc1_return stands with the highest kurtosis at 108.30, followed by DPc1_return at 90.90. This suggests that DEPYc1_return exhibit more frequent extreme deviations from the mean and an increased likelihood of negative tail events due to its negative skew. Furthermore, we see that the peakload contracts has a higher kurtosis than their baseload counterparts, with an exception of the monthly contracts. Non-normality can also be shown looking at the column for the Jarque Bera (JB) test in Table 6, where all return-series reject the null hypothesis of normality. This rejection highlights the need for models that forecast VaR and ES without assuming normally distributed returns, but instead account for skewed, fat-tailed distributions. Looking at the QQ plots for the different contracts in Figure 8 further confirms our analysis, as they visually illustrate the presence of skewness, high kurtosis, and outliers in the return distributions when compared against the theoretical normal distribution quantiles.

Overall, the observed patterns highlight the varying risk profiles of the contracts. Shorter-term contracts exhibit higher volatility and positive skew, while longer-term contracts, though more stable, carry greater downside risk due to their negative skew and leptokurtic behavior (fat tails). All contracts exhibit characteristics that underscore the importance of accurate distributional assumptions in forecasting models.

7. Empirical Results

This section discusses the result from the empirical study regarding our main research question: *Which univariate joint Var and ES forecasting models are best-performing for different return characteristics*. We divide the section into into two subsection, a subsection discussing model adequacy, followed by a subsection about absolute and relative model performance using the joint Var and ES loss functions. Note that certain CAViaR models (SAV, AS, Adaptive) are excluded from the reported tables due to non-convergence, highlighting computational challenges associated with these specifications.

7.1. Model Adequacy and Misspecificaiton

Tables 8 and 9 summarize the number of backtests rejected at a 5% significance level for VaR and ES forecasts across the 2.5% and 5% quantiles, respectively. The null hypothesis assumes that each model is correctly specified. Consequently, a higher number of rejections indicates a greater degree of model misspecification, where the model fails to align with theoretically sound assumptions.

At the 2.5% quantile, the results reveal distinct patterns in model performance depending on contract durations. Across both base- and peakload contracts, the FHS-EWMA and

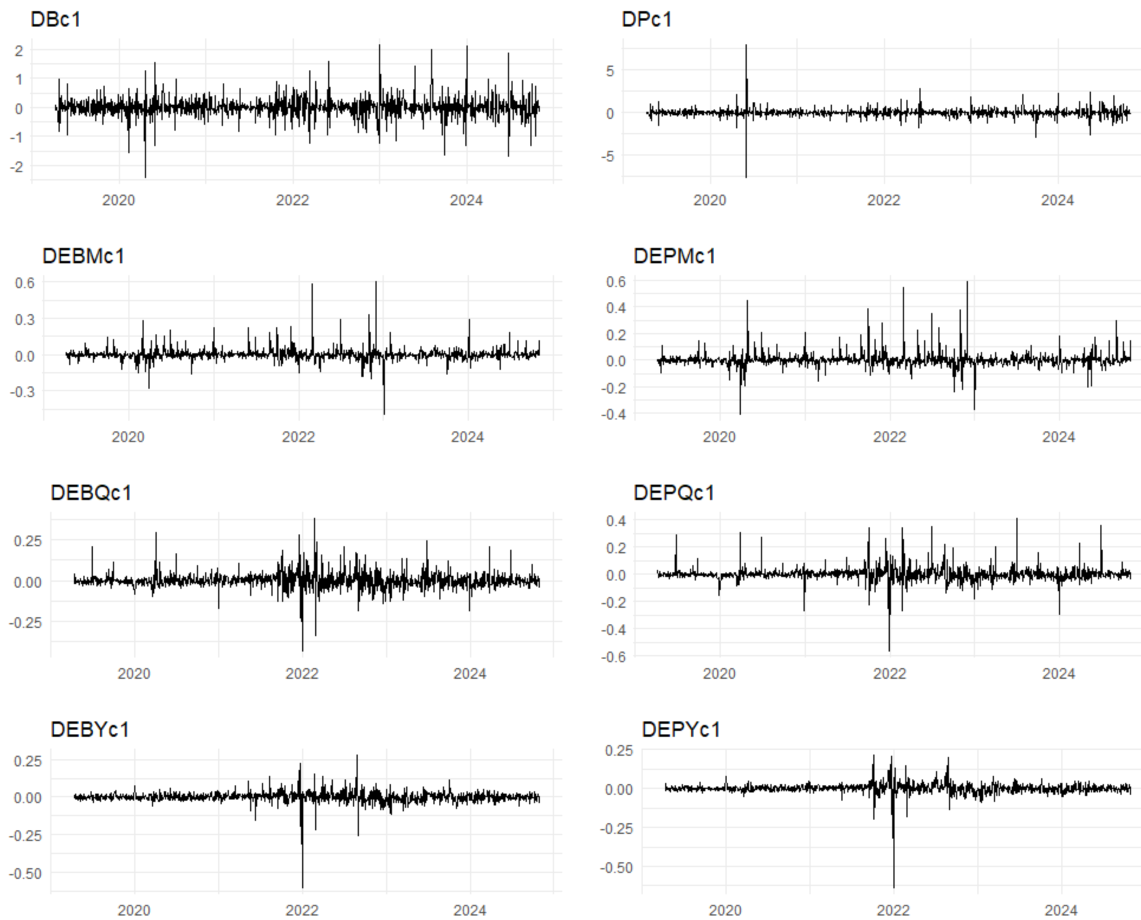

Figure 7: Log returns of the eight different futures.

Table 6
Descriptive statistics for the return series of the different contracts, with the JB column presenting the p-values of the Jarque-Bera test.

Variable	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	JB
DBc1_return	0.00	0.33	0.00	-2.43	2.14	4.57	0.23	9.84	0.00
DPc1_return	0.00	0.51	-0.01	-7.70	7.92	15.61	0.37	90.90	0.00
DEBMc1_return	0.00	0.05	0.00	-0.50	0.60	1.09	2.72	44.01	0.00
DEPMc1_return	0.00	0.05	0.00	-0.41	0.59	1.00	2.98	36.82	0.00
DEBQc1_return	0.00	0.05	0.00	-0.44	0.38	0.82	0.24	13.26	0.00
DEPQc1_return	0.00	0.05	0.00	-0.57	0.41	0.97	0.47	24.63	0.00
DEBYc1_return	0.00	0.04	0.00	-0.60	0.27	0.88	-3.47	63.92	0.00
DEPYc1_return	0.00	0.03	0.00	-0.64	0.21	0.85	-5.08	108.30	0.00

FHS-gjrGARCH-sstd models stand out, as they consistently exhibit fewer rejections across all durations. In contrast, HS and CAViaR-IGARCH-AR perform well only for short duration contracts. For longer durations, the CAViaR-IGARCH-MULT model shows promising characteristics, maintaining relatively few violations. Notably, the standalone GARCH models, regardless of distributional assumptions, are consistently misspecified. On the other hand, some of the EVT-based hybrid models exhibit performance for some longer-duration contracts, but this fails to generalize across all quarter- and year contracts.

Table 7
Annualized standard deviation of the different contracts each year.

Year	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
2019	0.3677	0.4511	0.0393	0.0373	0.0382	0.0452	0.0210	0.0176
2020	0.4440	1.0138	0.0617	0.0695	0.0461	0.0496	0.0222	0.0188
2021	0.3123	0.3830	0.0530	0.0633	0.0748	0.0854	0.0720	0.0737
2022	0.4299	0.5324	0.0987	0.1070	0.0928	0.0887	0.0658	0.0534
2023	0.4962	0.5675	0.0593	0.0386	0.0597	0.0648	0.0407	0.0319
2024	0.5676	0.8179	0.0557	0.0672	0.0523	0.0602	0.0309	0.0277

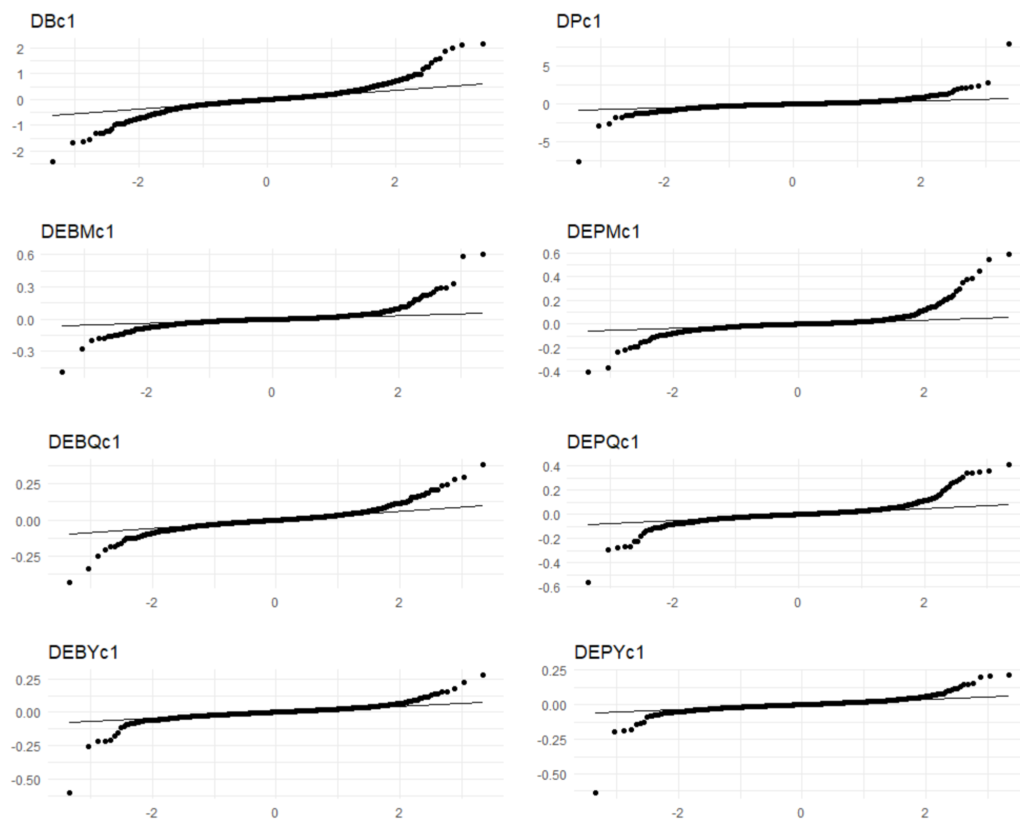

Figure 8: QQ plots of the returns of the baseload and peakload futures with their respective delivery period.

Turning to the 5% quantile results, there are similar patterns, although with a few notable differences. Here, the CAViaR-IGARCH-MULT model not only continues to perform well but also achieves low rejection numbers across all contract durations, marking it as one of the most robust specifications. Additionally, the CAViaR-Adaptive-MULT model demonstrates relatively few violations, particularly for the month-peak, month-base, and quarter-base contracts. This indicates an improved performance of the CAViaR-based models at the 5% quantile compared to their behavior at the 2.5% level. We also see that the EVT has worse results in the 5% quantile at 2.5%, which can be expected due to its focus on extreme quantiles.

7.2. Forecasting performance

7.2.1. Overall performance

The forecasting performance results, summarized in Tables 10 and 11, show the average joint loss function values at the 2.5% and 5% quantiles. The smallest and second smallest losses are outlined, as well as inclusion in the 5% and 10% SSM, marked with * and ** respectively. The results reveal clear patterns in model performance across different specifications and durations.

At the 2.5% quantile, the HS model performs well for short-term contracts like day-base and day-peak, achieving the smallest losses and being part of the 10% SSM. However, its performance worsens for longer durations. Standalone

Table 8
Number of backtests rejected across all contracts at 2.5% quantile.
Formatting: Full outline denote the smallest number of violations in each column, dashed outline denote the second smallest number.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	$\boxed{1}$	$\boxed{0}$	3	2	3	3	3	3
FHS-EWMA	4	2	$\boxed{1}$	2	2	1	$\boxed{0}$	$\boxed{0}$
FHS-sGARCH-norm	7	6	6	4	8	4	5	3
FHS-sGARCH-std	6	6	5	5	7	8	9	9
FHS-sGARCH-sstd	8	7	4	5	5	5	5	9
FHS-gjrGARCH-norm	8	8	8	8	5	6	6	4
FHS-gjrGARCH-std	7	6	5	5	7	5	9	9
FHS-gjrGARCH-sstd	4	1	$\boxed{1}$	$\boxed{1}$	4	1	$\boxed{0}$	1
sGARCH-norm	7	6	6	4	8	4	5	3
sGARCH-std	6	6	5	5	7	8	9	9
sGARCH-sstd	8	7	4	5	5	5	5	9
gjrGARCH-norm	8	8	8	8	5	6	6	4
gjrGARCH-std	7	6	5	5	7	5	9	9
gjrGARCH-sstd	8	6	4	5	5	6	2	9
EVT-sGARCH-norm	8	6	5	6	2	6	3	4
EVT-sGARCH-std	8	5	7	8	3	2	9	9
EVT-sGARCH-sstd	8	4	7	8	3	2	1	1
EVT-gjrGARCH-norm	9	8	5	6	4	$\boxed{0}$	5	4
EVT-gjrGARCH-std	9	9	7	6	1	4	9	9
EVT-gjrGARCH-sstd	9	7	7	7	$\boxed{0}$	6	4	2
GAS-norm	7	3	4	2	5	4	3	3
GAS-std	6	5	3	4	4	4	3	2
GAS-sstd	7	5	4	4	3	2	4	4
CAViaR-IGARCH-AR	3	1	2	$\boxed{1}$	2	4	2	1
CAViaR-IGARCH-MULT	4	2	5	4	1	$\boxed{0}$	3	$\boxed{0}$

GARCH models show overall weak results across all durations. In contrast, the FHS-EWMA model is among the most robust, consistently achieving small losses across all contracts and being included in the 10% SSM.

Another performing model is FHS-gjrGARCH-sstd, delivering small losses across short and long durations, and appearing in the 10% SSM for all contracts. EVT hybrid models in general also show strong performance, with EVT-sGARCH-norm exhibiting among the lowest loss per contract, but they fall short particularly for day contracts.

The GAS models perform well for longer duration contracts, with the variation using the std distribution being included in the 10% SSM for all durations. However, the GAS models do not dominate as broadly as the FHS or the CAViaR models, which perform across all durations.

The CAViaR-IGARCH-MULT model demonstrates competitive performance with low losses across most durations. However, for the year-duration, its performance deviates more significantly from the best-performing model. Despite this, it is consistently included in the 10% SSM across all durations.

At the 5% quantile, similar trends emerge, though the CAViaR models demonstrate a notable improvement overall. The EVT-based hybrid models remain strong at the 5% quantile, particularly EVT-sGARCH-norm, which achieves the smallest loss for yearly durations. However, their performance weakens slightly compared to the 2.5% level.

The loss values for GAS models improved, where GAS-std now exhibits among the lowest loss functions for both quarter peak- and baseload, and yearly baseload.

Table 9
Number of backtests rejected across all contracts at 5% quantile.
Formatting: Full outline denote the smallest number of violations in each column, dashed outline denote the second smallest number.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	0	1	3	5	3	3	3	3
FHS-EWMA	4	2	2	3	1	2	0	0
FHS-sGARCH-norm	3	2	8	7	7	7	4	2
FHS-sGARCH-std	6	7	5	5	8	7	9	9
FHS-sGARCH-sstd	8	7	7	5	7	6	7	9
FHS-gjrGARCH-norm	3	2	8	7	8	8	4	1
FHS-gjrGARCH-std	7	5	5	6	6	8	9	9
FHS-gjrGARCH-sstd	6	2	3	5	1	2	0	0
sGARCH-norm	3	2	8	7	7	7	4	2
sGARCH-std	6	7	5	5	8	7	9	9
sGARCH-sstd	8	7	7	5	7	6	7	9
gjrGARCH-norm	3	2	8	7	8	8	4	1
gjrGARCH-std	7	5	5	6	6	8	9	9
gjrGARCH-sstd	9	6	8	5	5	5	6	9
EVT-sGARCH-norm	8	6	8	7	7	9	2	2
EVT-sGARCH-std	8	3	5	6	8	7	9	9
EVT-sGARCH-sstd	7	1	5	8	8	8	4	5
EVT-gjrGARCH-norm	9	7	7	7	6	7	2	1
EVT-gjrGARCH-std	9	5	5	6	6	7	9	9
EVT-gjrGARCH-sstd	9	5	5	7	6	5	0	4
GAS-norm	4	3	6	6	4	4	5	3
GAS-std	9	5	7	7	3	6	4	1
GAS-sstd	9	5	7	6	4	3	5	3
CAViaR-IGARCH-AR	5	4	2	4	4	3	5	1
CAViaR-IGARCH-MULT	1	1	5	2	2	0	0	0

Tables 16 and17 from Appendix A show loss function for VaR with 2.5% and 5% quantiles, respectively, computed with QL. Here we find several similar results. However a notable difference is the relative performance of standalone GARCH models with both std and sstd, particularly for mid-duration contracts.

Tables 18 and 19 from Appendix A show the number of contracts for which each backtest was rejected.

In summary, FHS-gjrGARCH-sstd, CAViaR-IGARCH-MULT, and EVT-GARCH models deliver the most robust performance across both quantiles and durations. Simpler models, such as HS and standalone GARCH, perform adequately for shorter contracts, but struggle with longer durations.

7.2.2. *Impact of misspecification*

The relationship between model misspecification, as indicated by the number of backtesting rejections, and forecasting performance, measured by the average joint loss, is analyzed using Figure 9. The boxplot groups joint loss values for both quantiles, by the number of hypothesis rejections at the 5% significance level.

A visual inspection suggest that models with fewer backtesting rejections generally exhibit lower and more stable loss values, while models with higher rejections tend to have larger and more variable losses. As backtesting is designed to evaluate whether the model aligns with theoretically desired properties, we should not necessarily expect that fewer rejections should directly correspond to lower forecast losses. Instead, the observed relationship serves as an empirical

Table 10
Average joint loss function results at 2.5% quantile.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Superior Set of Models (SSM), and entries marked with ** are included in the 10% SSM.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	1.066**	1.646**	−0.949**	−0.900**	−0.837*	−0.765**	−0.997	−1.128
FHS-EWMA	1.178**	1.717**	−0.984**	−0.910**	−1.032**	−0.895**	−1.399**	−1.652**
FHS-sGARCH-norm	1.688	2.507	−0.597**	−0.541*	−0.957*	−0.710**	−1.215**	−1.642
FHS-sGARCH-std	1.299**	2.147**	−0.862**	−0.640*	−0.981**	−0.727	NULL	NULL
FHS-sGARCH-sstd	1.525	2.406	−0.879**	−0.222	−0.544	−0.160	−0.865	NULL
FHS-gjrGARCH-norm	1.763	2.654	−0.174**	0.139	−0.856*	−0.513	−1.136	−1.513
FHS-gjrGARCH-std	1.251**	2.043**	−0.839**	−0.630*	−1.005**	−0.832**	NULL	NULL
FHS-gjrGARCH-sstd	1.182**	1.740**	−0.979**	−0.902**	−1.020**	−0.894**	−1.377**	−1.635**
sGARCH-norm	1.688	2.507	−0.597**	−0.541*	−0.957*	−0.710**	−1.215**	−1.642
sGARCH-std	1.299**	2.147**	−0.862**	−0.640*	−0.981**	−0.727	NULL	NULL
sGARCH-sstd	1.525	2.406	−0.879**	−0.222	−0.544	−0.160	−0.865	NULL
gjrGARCH-norm	1.763	2.654	−0.174**	0.139*	−0.856*	−0.508	−1.136	−1.513
gjrGARCH-std	1.251**	2.043**	−0.839**	−0.630*	−1.005**	−0.832**	NULL	NULL
gjrGARCH-sstd	1.440*	2.300**	−0.846**	−0.439	−0.787	−0.710	−0.781	NULL
EVT-sGARCH-norm	2.548	2.609	−1.011**	−0.885**	−1.148**	−0.999**	−1.434**	−1.761**
EVT-sGARCH-std	2.986	2.480**	−0.981**	−0.906**	−1.129**	−0.970**	NULL	NULL
EVT-sGARCH-sstd	2.992	2.381**	−0.950**	−0.820**	−1.115**	−0.948**	−1.186	−1.767**
EVT-gjrGARCH-norm	3.291	3.144	−0.963**	−0.860**	−1.098**	NULL	−1.314**	−1.658**
EVT-gjrGARCH-std	3.386	3.277	−0.918**	−0.889**	−1.143**	−0.955**	NULL	NULL
EVT-gjrGARCH-sstd	3.383	3.421	−0.901**	−0.605**	−1.126**	−0.981**	−1.174	−1.703**
GAS-norm	1.272**	2.076**	−0.968**	−0.866**	−1.036**	−0.868**	−0.968	−1.657**
GAS-std	1.289**	2.020**	−0.853**	−0.742**	−1.139**	−1.033**	−1.403**	−1.695**
GAS-sstd	1.256**	1.955**	−0.691**	−0.615*	−1.044**	−0.998**	−0.799	−1.598
CAViaR-IGARCH-AR	1.124**	1.723**	−0.794**	−0.871**	−0.997**	−0.987**	−1.135**	−1.153
CAViaR-IGARCH-MULT	1.091**	1.670**	−0.913**	−0.839**	−1.015**	−0.965**	−1.068**	−1.626**

insight: models that pass backtests more frequently appear to provide more reliable risk forecasts.

Models like FHS-EWMA, FHS-gjrGARCH-sstd, and CAViaR-IGARCH-MULT, which experience few rejections, consistently achieve smaller joint losses across durations. These models not only align with the characteristics tested in backtesting but also demonstrate accuracy in minimizing joint losses. In contrast, models with many backtesting rejections—such as standalone GARCH variants and poorly performing CAViaR models tend to show larger losses, reflecting their inability to model the joint dynamics of VaR and ES effectively.

Interestingly, EVT-based models like EVT-sGARCH-norm show strong performance. However, they still experience notable backtesting rejections. This indicates that while EVT models deliver accurate joint VaR and ES forecasts, their assumptions may lead to inconsistencies when evaluated against broader backtesting criteria.

In summary, the boxplot highlights an empirical association between backtesting performance and joint loss values. Models with fewer backtesting rejections often yield better forecasts, but this is not guaranteed. Evaluating model adequacy alongside forecasting performance provides a more complete assessment, ensuring both alignment with theoretically desired properties and accuracy in minimizing joint losses.

Table 11
Average joint loss function results at 5% quantile.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Superior Set of Models (SSM), and entries marked with ** are included in the 10% SSM.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	0.845**	1.283**	−1.254**	−1.217**	−1.115	−1.101**	−1.353	−1.485
FHS-EWMA	0.919**	1.354**	−1.277**	−1.198**	−1.240**	−1.204**	−1.600**	−1.843**
FHS-sGARCH-norm	0.932**	1.534**	−1.163**	−1.071*	−1.200	−1.082	−1.478	−1.817
FHS-sGARCH-std	1.114**	1.883	−1.206**	−0.854	−1.160	−0.948	NULL	NULL
FHS-sGARCH-sstd	1.384	2.132	−1.172**	−0.406	−0.706	−0.360	−1.058	NULL
FHS-gjrGARCH-norm	0.951**	1.518**	−1.096**	−0.892*	−1.190	−1.046	−1.469	−1.783
FHS-gjrGARCH-std	1.040	1.683	−1.193	−0.855	−1.177	−1.062	NULL	NULL
FHS-gjrGARCH-sstd	0.935**	1.373**	−1.273**	−1.189**	−1.234**	−1.195**	−1.610**	−1.849**
sGARCH-norm	0.932**	1.534*	−1.163**	−1.071*	−1.200	−1.082*	−1.478	−1.817
sGARCH-std	1.114**	1.883	−1.206**	−0.854	−1.160	−0.948	NULL	NULL
sGARCH-sstd	1.384	2.132	−1.172**	−0.406	−0.706	−0.360	−1.058	NULL
gjrGARCH-norm	0.951**	1.518**	−1.096**	−0.892*	−1.190	−1.044	−1.469	−1.783
gjrGARCH-std	1.040**	1.683**	−1.193**	−0.855	−1.177	−1.062*	NULL	NULL
gjrGARCH-sstd	1.300	2.001	−1.165**	−0.656	−0.954	−0.912	−1.074	NULL
EVT-sGARCH-norm	1.232	1.560	−1.242**	−1.157**	−1.293**	−1.189**	−1.645**	−1.929**
EVT-sGARCH-std	1.444	1.520**	−1.116**	−1.103**	−1.267**	−1.109*	NULL	NULL
EVT-sGARCH-sstd	1.462	1.495**	−1.092**	−1.003	−1.249**	−1.071	−1.394	−1.911**
EVT-gjrGARCH-norm	1.492	1.725	−1.241**	−1.185**	−1.292**	NULL	−1.605**	−1.895**
EVT-gjrGARCH-std	1.522	1.684	−1.059**	−1.056**	−1.290**	−1.100	NULL	NULL
EVT-gjrGARCH-sstd	1.526	1.743	−1.034	−1.169**	−1.268**	−1.078*	−1.490	−1.907**
GAS-norm	0.940**	1.522	−1.172**	−1.096**	−1.221**	−1.116*	−1.360**	−1.825*
GAS-std	0.957**	1.463**	−1.168**	−1.069**	−1.304**	−1.243**	−1.615**	−1.871**
GAS-sstd	0.934**	1.431**	−1.038**	−0.986*	−1.233**	−1.208**	−1.268	−1.804
CAViaR-IGARCH-AR	0.896**	1.319**	−1.193**	−1.144**	−1.188**	−1.218**	−1.468**	−1.720
CAViaR-IGARCH-MULT	0.875**	1.320**	−1.179**	−1.181**	−1.259**	−1.273**	−1.574**	−1.840**

7.2.3. Performance Across Durations

The performance of joint VaR and ES forecasting models varies significantly across contract durations. A notable observation is that the loss function values are substantially higher for day durations compared to longer durations. This highlights the inherent challenges of modeling these contracts, where increased volatility and shorter trading windows before delivery, create difficulties in producing reliable forecasts.

At longer durations, another issue emerges: the GARCH-std fitting process, and often for the GARCH-sstd as well, failed to converge or fit effectively. This fitting issue propagated to hybrid models such as FHS and EVT that rely on GARCH-based components, leading to their failure for longer durations. This underscores not only the technical challenges associated with parameter estimation for longer-term contracts but also the practical complications in implementing these models.

7.2.4. Peak- vs. Baseload Contracts

The comparison between loss values for the peak- and baseload contracts reveals distinct patterns in model performance. Peakload contracts generally exhibit higher losses compared to their baseload counterparts, reflecting the increased volatility and price spikes inherent to peak demand periods. However, at both quantiles, models such as FHS-EWMA, FHS-gjrGARCH-sstd, EVT-sGARCH-norm, and

Table 12
Number of backtests rejected across all contracts at 2.5% quantile for the stressed period.
Formatting: Full outline denote the smallest number of violations in each column, dashed outline denote the second smallest number.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	0	0	4	1	3	5	3	5
FHS-EWMA	1	2	4	1	1	3	3	4
FHS-sGARCH-norm	7	4	3	2	3	4	3	3
FHS-sGARCH-std	4	1	7	6	6	6	9	9
FHS-sGARCH-sstd	5	1	7	7	4	5	3	9
FHS-gjrGARCH-norm	7	7	5	6	2	1	5	2
FHS-gjrGARCH-std	5	3	6	6	4	4	9	9
FHS-gjrGARCH-sstd	0	1	4	1	1	1	1	2
sGARCH-norm	7	4	3	2	3	4	3	3
sGARCH-std	4	1	7	6	6	6	9	9
sGARCH-sstd	5	1	7	7	4	5	3	9
gjrGARCH-norm	7	7	5	6	2	1	5	2
gjrGARCH-std	5	3	6	6	4	4	9	9
gjrGARCH-sstd	6	5	5	5	4	5	3	9
EVT-sGARCH-norm	7	4	7	1	3	6	4	2
EVT-sGARCH-std	7	5	7	5	4	4	9	9
EVT-sGARCH-sstd	7	4	7	4	4	2	7	3
EVT-gjrGARCH-norm	7	4	6	5	3	4	6	5
EVT-gjrGARCH-std	7	8	7	4	3	5	9	9
EVT-gjrGARCH-sstd	7	8	7	7	5	6	2	2
GAS-norm	7	4	1	1	6	2	2	7
GAS-std	4	3	1	0	1	2	1	3
GAS-sstd	4	4	2	2	2	3	3	4
CAViaR-IGARCH-AR	2	0	6	2	1	6	1	1
CAViaR-IGARCH-MULT	3	0	3	0	1	0	4	3

GAS-std demonstrate robust performance for both peak- and baseload contracts, with only a slight decrease in loss.

Interestingly, for year-duration contracts, the forecasts tend to be more accurate for peakload contracts compared to baseload counterparts.

7.3. Sensitivity analysis

7.3.1. *Performance across distributions*

The choice of distributional assumptions plays an important role in determining model performance, with significant variation observed across model classes. Distributions such as norm, std, and sstd are incorporated into different model frameworks, including GARCH, FHS, EVT, and GAS, each exhibiting varying levels of sensitivity to the selected distribution.

For standalone GARCH models, the std and sstd distributions yield lower losses than the norm distribution at the 2.5% quantile, particularly for shorter duration contracts. However, for longer durations or at the 5% quantile, the norm distribution performs the best, suggesting that the complexity introduced by heavy tails is less effective in capturing longer-term dynamics.

Similarly, for FHS models, std and sstd distributions provide superior performance at shorter durations and the 2.5% quantile. However, for longer durations, models employing sstd in combination with gjrGARCH achieve lower losses than those using std or norm. Interestingly, FHS models with standard GARCH and sstd perform worse than their norm-based counterparts, highlighting the interplay between distributional assumptions and GARCH specifications.

Table 13
Number of backtests rejected across all contracts at 5% quantile for the stressed period.
Formatting: Full outline denote the smallest number of violations in each column, dashed outline denote the second smallest number.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	1	0	3	3	3	5	6	4
FHS-EWMA	3	2	2	0	1	7	2	3
FHS-sGARCH-norm	3	1	5	2	5	4	6	4
FHS-sGARCH-std	4	0	7	5	5	4	9	9
FHS-sGARCH-sstd	4	0	8	5	6	5	6	9
FHS-gjrGARCH-norm	1	0	6	5	5	6	7	5
FHS-gjrGARCH-std	3	4	7	7	5	5	9	9
FHS-gjrGARCH-sstd	3	3	3	0	3	6	2	6
sGARCH-norm	3	1	5	2	5	4	6	4
sGARCH-std	4	0	7	5	5	4	9	9
sGARCH-sstd	4	0	8	5	6	5	6	9
gjrGARCH-norm	1	0	6	5	5	6	7	5
gjrGARCH-std	3	4	7	7	5	5	9	9
gjrGARCH-sstd	5	7	7	6	4	5	3	9
EVT-sGARCH-norm	8	3	6	1	4	4	4	4
EVT-sGARCH-std	8	4	4	4	4	4	9	9
EVT-sGARCH-sstd	8	4	4	4	5	4	9	5
EVT-gjrGARCH-norm	8	4	7	4	4	5	7	5
EVT-gjrGARCH-std	8	5	4	3	3	4	9	9
EVT-gjrGARCH-sstd	8	7	4	4	4	5	6	5
GAS-norm	5	1	3	2	3	3	1	3
GAS-std	5	3	0	1	3	3	2	3
GAS-sstd	6	3	3	2	3	4	1	2
CAViaR-IGARCH-AR	2	3	4	3	1	6	1	1
CAViaR-IGARCH-MULT	3	2	5	0	1	0	1	3

7.3.2. Performance across GARCH specifications

For standalone GARCH models, the impact of GARCH specifications is less pronounced compared to hybrid models. While gjrGARCH improves performance slightly, standalone GARCH models still exhibit weaker overall performance across both short and long durations. This suggests that standalone GARCH models, even when using more sophisticated specifications, are less effective in capturing the joint dynamics of VaR and ES compared to hybrid approaches.

In FHS models, the choice of GARCH specification significantly influences performance. FHS-gjrGARCH consistently achieves lower losses than FHS-standard GARCH across durations and quantiles. For example, FHS-gjrGARCH-sstd records one of the smallest losses , reflecting its ability to model asymmetric volatility clustering. By contrast, FHS-standard GARCH with sstd or norm distributions often results in higher losses, particularly for longer durations, indicating the importance of aligning GARCH specifications with the flexibility of the FHS framework.

EVT models benefit less from advanced GARCH specifications compared to FHS. While EVT-gjrGARCH often performs better than EVT-sGARCH, the improvement is less substantial and inconsistent across durations. EVT models inherently focus on extreme quantiles, and their performance appears to be driven more by distributional assumptions than by GARCH specification. This makes the choice of GARCH specification secondary in these models.

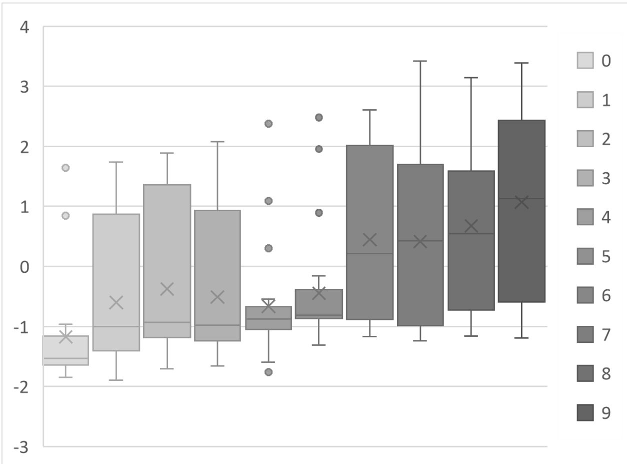

Figure 9: Boxplot of average joint loss values by number of backtest rejections

7.4. Stress Testing and Robustness

Stress testing provides a critical evaluation of how forecasting models perform under extreme market conditions. This analysis is essential for understanding the robustness of models, particularly their ability to generate reliable joint VaR and ES forecasts during periods of heightened volatility and market stress. We identified year 2021 and 2022 as the stressed periods, based on the elevated volatility for longer duration contracts and the fact that prices deviated significantly from the long run mean. A one-year rolling window estimation was employed, yielding one year of stressed joint VaR and ES forecasts to which we apply the same backtests and loss function as previously. The results from backtesting at 2.5% and 5% quantiles are summarized in Tables 12 and 13. Average loss for each of the two quantiles are shown in Tables 14 and 15.

The results show that hybrid models, particularly FHS-gjrGARCH-sstd, and GAS-std, are the most robust, consistently achieving low violation counts and losses across stressed conditions. For instance, FHS-gjrGARCH-sstd records zero violations for DBc1 and DEPMc1 at the 2.5% quantile, while GAS-std achieves minimal violations for contracts like DEBQc1 and DEBYc1. These models also deliver the lowest loss values.

Standalone GARCH models, on the other hand, show significant weaknesses under stress. They frequently exhibit high violation counts, particularly for longer durations, and their loss values are markedly higher, indicating poor robustness.

EVT models demonstrate moderate performance, with EVT-sGARCH-norm achieving competitive losses for longer durations but struggling with shorter contracts. This highlights their sensitivity to calibration and specification.

7.5. Comparative analysis

7.5.1. Model Confidence Set

At the 2.5% quantile, FHS-gjrGARCH-sstd and GAS-std frequently feature in the 10% SSM across most contracts, demonstrating their strong overall performance. For instance, FHS-gjrGARCH-sstd is included in the 10% SSM for both short- and long-duration contracts, such as day-duration and year-duration, reflecting its consistent accuracy in joint VaR and ES forecasting. Similarly, GAS-std is often present in the SSM, particularly for contracts with medium to long durations, highlighting its robustness under varying conditions.

Standalone GARCH models show limited inclusion in the SSM, particularly at longer durations. For example, sGARCH-sstd is excluded from the 10% SSM for most long-duration contracts, underscoring its weaker predictive performance. In contrast, EVT-based models such as EVT-sGARCH-norm occasionally appear in the SSM for longer contracts like DEPYc1, although their performance is less consistent across durations and quantiles.

At the 5% quantile, the composition of the SSM broadens slightly. Models such as FHS-EWMA and CAViaR-IGARCH-MULT gain more frequent inclusion, particularly for shorter durations. For example, CAViaR-IGARCH-MULT is part of the 10% SSM for DEBQc1 and DEBYc1, reflecting its ability to adapt well to the higher quantile's less extreme tails. However, EVT-based models generally lose prominence in the SSM at this quantile, likely due to their focus on extreme quantile behavior, which is less relevant at the 5% level.

8. Conclusion

Value at Risk (VaR) and Expected Shortfall (ES) are widely recognized metrics for assessing potential financial losses, mandated for regulatory compliance in banking and widely applied in risk management. For electricity producers, the transition toward renewable energy and geopolitical conflicts has introduced heightened volatility in the energy sector, making these measures increasingly vital. Additionally, the growing availability of data and advancements in computing power have expanded the potential for more accurate risk forecasting. This further underscores the importance of VaR and ES in evaluating and managing risks associated with energy derivatives such as futures, which act as hedging instruments against initial exposures like electricity price fluctuations.

This study evaluates the forecasting performance of 31 joint VaR and ES models by applying them empirically to German electricity futures spanning April 9, 2019, to October 31, 2024, comprising 1,239 log-return observations. The analysis covers front-day, quarter, month, and year contracts for two underlying products: peak-load and base-load. These contracts exhibit diverse return characteristics and market conditions, including periods of heightened volatility and varying trading durations. Out-of-sample forecasts are evaluated using the AL log score loss function to measure joint VaR and ES accuracy. Additionally, backtests such as

Table 14
Average joint loss function results at 2.5% quantile for the stressed period.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Superior Set of Models (SSM), and entries marked with ** are included in the 10% SSM.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	1.054**	1.398**	−0.579**	−0.426**	−0.588**	−0.662**	−0.688*	−0.940
FHS-EWMA	1.194**	1.515**	−0.447**	−0.317**	−0.621	−0.559	−0.840	−1.135
FHS-sGARCH-norm	2.151**	2.347**	−0.284**	−0.221**	−0.768**	−0.676**	−1.060**	−1.355**
FHS-sGARCH-std	1.488**	1.778**	−0.589**	−0.257**	−0.582	−0.331	NULL	NULL
FHS-sGARCH-sstd	1.558**	1.849**	−0.545**	0.161	−0.011	0.386	0.730	NULL
FHS-gjrGARCH-norm	2.273**	2.270**	0.961**	2.796	−0.715**	−0.630**	−0.801**	−1.049
FHS-gjrGARCH-std	1.539**	1.799**	−0.470**	−0.316**	−0.61	−0.519	NULL	NULL
FHS-gjrGARCH-sstd	1.179**	1.478**	−0.386**	−0.195**	−0.633**	−0.636**	−0.872*	−1.197
sGARCH-norm	2.151**	2.347**	−0.284**	−0.221**	−0.768**	−0.676**	−1.060**	−1.355**
sGARCH-std	1.488**	1.778**	−0.589**	−0.257**	−0.582	−0.331	NULL	NULL
sGARCH-sstd	1.558**	1.849**	−0.545**	0.161	−0.011	0.386	0.730	NULL
gjrGARCH-norm	2.273**	2.270**	0.961**	2.796	−0.715**	−0.630**	−0.801**	−1.049**
gjrGARCH-std	1.539**	1.799**	−0.470**	−0.316**	−0.610	−0.519	NULL	NULL
gjrGARCH-sstd	1.713**	2.027**	−0.559**	−0.149**	−0.253	−0.130	0.955	NULL
EVT-sGARCH-norm	3.584	2.253**	−0.610**	−0.556**	−0.921**	−0.843**	−1.238**	−1.465**
EVT-sGARCH-std	4.661	2.267**	−0.538**	−0.558**	−0.899**	−0.833**	NULL	NULL
EVT-sGARCH-sstd	4.702	2.185**	−0.615**	−0.557**	−0.884**	−0.792**	−0.388	−1.470**
EVT-gjrGARCH-norm	4.507	2.692*	−0.322**	−0.194**	−0.876**	NULL	−0.976**	−1.238**
EVT-gjrGARCH-std	4.676	2.958	−0.551**	−0.465**	−0.888**	−0.841**	NULL	NULL
EVT-gjrGARCH-sstd	4.531	2.971	−0.507**	−0.032**	−0.879**	−0.802**	−0.626	−1.414**
GAS-norm	1.433**	1.716**	−0.715**	−0.617**	−0.807**	−0.743**	−1.020**	−1.059
GAS-std	1.377**	1.582**	−0.675**	−0.383**	−0.839**	−0.753**	−1.119**	−1.330**
GAS-sstd	1.429**	1.648**	−0.485**	−0.417**	−0.826**	−0.741**	−1.132**	−1.319**
CAViaR-IGARCH-AR	1.045**	1.477**	0.089**	−0.177**	−0.539**	−0.577**	−0.886**	−1.186**
CAViaR-IGARCH-MULT	1.032**	1.362**	−0.477**	−0.252**	−0.585**	−0.583**	−0.852**	−1.145**

UC, CC, DQ, and VDT for VaR, as well as ER and ESR tests for ES, are employed. The MCS test is further utilized to assess model ranking. The key findings are summarized as follows:

(1) No single model consistently outperforms others across all contracts in terms of both accuracy and backtesting results. Performance depends on model class, specification, and the assumed distribution, as well as the specific duration and underlying product of the futures contract. Notably, high accuracy does not always equate to adequacy, as demonstrated by the mixed performance of Generalized Autoregressive Score (GAS) and Extreme Value Theory (EVT) hybrid models. However, there is some indication that adequacy may imply high accuracy.

(2) The traditional Historical Simulation (HS) model demonstrates notable underperformance in terms of both accuracy and adequacy for year- and quarter-duration contracts. This underscores the opportunity for agents still relying on HS to enhance their risk evaluation of joint VaR and ES by adopting more advanced modeling approaches.

(3) The Filtered Historical Simulation (FHS) and Conditional Autoregressive VaR (CAViaR) models, FHS-EWMA, FHS-gjrGARCH-sstd, and CAViaR-IGARCH-MULT, demonstrate strong accuracy and are generally well-specified across all contracts. This highlights the effectiveness of both parametric (distribution-based) and semi-parametric approaches for risk forecasting. Notably, FHS models excel at the 2.5% quantile, while CAViaR-IGARCH-MULT performs

Table 15
Average joint loss function results at 5% quantile for the stressed period.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Superior Set of Models (SSM), and entries marked with ** are included in the 10% SSM.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	0.966**	1.133**	−0.838**	−0.823**	−0.806**	−0.910**	−1.040**	−1.284**
FHS-EWMA	1.005**	1.251**	−0.822**	−0.730**	−0.827**	−0.891**	−1.063**	−1.321**
FHS-sGARCH-norm	0.929**	1.217**	−0.828**	−0.727**	−0.907**	−0.908**	−1.168**	−1.439**
FHS-sGARCH-std	1.003**	1.218**	−0.813**	−0.503	−0.765*	−0.570	NULL	NULL
FHS-sGARCH-sstd	1.072**	1.239**	−0.761**	−0.040	−0.224	0.111	0.034**	NULL
FHS-gjrGARCH-norm	0.949**	1.025**	−0.699**	0.404	−0.893**	−0.919**	−1.087**	−1.328**
FHS-gjrGARCH-std	1.029	1.172	−0.776	−0.495	−0.808	−0.769	NULL	NULL
FHS-gjrGARCH-sstd	0.994**	1.237**	−0.833**	−0.753**	−0.809**	−0.868**	−1.051**	−1.299**
sGARCH-norm	0.929**	1.217**	−0.828**	−0.727**	−0.907**	−0.908**	−1.168**	−1.439**
sGARCH-std	1.003**	1.218**	−0.813**	−0.503	−0.765*	−0.570	NULL	NULL
sGARCH-sstd	1.072**	1.239**	−0.761**	−0.040	−0.224	0.111	0.034	NULL
gjrGARCH-norm	0.949**	1.025**	−0.699**	0.404	−0.893**	−0.919**	−1.087**	−1.328**
gjrGARCH-std	1.029**	1.172**	−0.776**	−0.495	−0.808**	−0.769**	NULL	NULL
gjrGARCH-sstd	1.236**	1.371**	−0.790**	−0.339	−0.505	−0.447	−0.081**	NULL
EVT-sGARCH-norm	1.676**	1.336**	−0.880**	−0.814**	−0.978**	−0.956**	−1.279**	−1.528**
EVT-sGARCH-std	2.122	1.271**	−0.821**	−0.821**	−0.945**	−0.863**	NULL	NULL
EVT-sGARCH-sstd	2.141	1.242**	−0.844**	−0.823**	−0.935**	−0.824**	−0.359	−1.451**
EVT-gjrGARCH-norm	2.040	1.385**	−0.809**	−0.696**	−0.957**	NULL	−1.202**	−1.434**
EVT-gjrGARCH-std	2.112	1.483**	−0.776**	−0.646**	−0.941**	−0.858	*NULL***	*NULL***
EVT-gjrGARCH-sstd	2.056*	1.487**	−0.706**	−0.631**	−0.934**	−0.825**	−0.707	−1.497**
GAS-norm	1.079**	1.246**	−0.928**	−0.893**	−0.899**	−0.890**	−1.172**	−1.140
GAS-std	1.101**	1.253**	−0.932**	−0.730**	−0.955**	−0.904**	−1.238**	−1.398**
GAS-sstd	1.109**	1.293**	−0.764**	−0.778**	−0.942**	−0.884**	−1.268**	−1.416**
CAViaR-IGARCH-AR	1.031**	1.322**	−0.478**	−0.528**	−0.779**	−0.864**	−1.023**	−1.435**
CAViaR-IGARCH-MULT	0.990**	1.279**	−0.644**	−0.769**	−0.837**	−0.855**	−1.062**	−1.361**

particularly well at the 5% quantile. Stress testing during the volatile period of 2021–2022 further underscores the robustness of these models, reaffirming their reliability for VaR and ES forecasting in volatile market conditions.

(4) In addition to their slight advantages at specific quantiles, the choice between FHS and CAViaR model specifications balances simplicity and intepretability. FHS-EWMA and FHS-gjrGARCH-sstd are industry-friendly approaches, closely resembling the widely-used HS method and offering open-source implementations. Both FHS-EWMA and CAViaR-IGARCH-MULT avoid residual distribution assumptions, reducing the risk of misspecification. Additionally, the CAViaR models provide a clear understanding of VaR and ES dynamics through their autoregressive framework, with parameters directly tied to the evolution of VaR, enhancing interpretability.

Based on these findings, we suggest that effective risk evaluation for portfolios of diverse electricity futures necessitates either customizing models to align with the characteristics of each contract or utilizing a weighted combination of forecasts, carefully adjusted to reflect the portfolio composition and the dynamics of the underlying products.

Risk assessment in electricity futures calls for further research, particularly in two key areas: enhancing the precision of risk forecasts and evaluating their economic implications in portfolio management.

First, enhancing the precision of risk forecasts requires further exploration of methodological aspects such as rolling window size, reparameterization frequency, and alternative GARCH specifications with different conditional distributions. While these areas have been individually studied, a comprehensive comparative analysis in the context of electricity futures is needed to clarify their strengths and limitations. Comparing traditional methods to multivariate or machine learning models could highlight the trade-offs between model complexity and forecast reliability. Similarly, the inclusion of high-frequency data to capture intraday dynamics or exogenous variables requires systematic testing alongside parametric and hybrid approaches to evaluate their specific contributions and suitability for electricity futures.

Second, understanding the economic implications of risk forecasts is essential, particularly in the context of energy commodities. Unlike financial institutions governed by strict regulatory frameworks such as Basel III, electricity producers and traders often have the discretion to choose risk measures that align with their strategic objectives. This flexibility raises important questions about the consequences of following different risk measures, such as relying solely on VaR versus adopting a ES framework. Research should also evaluate how these risk measures influence the effectiveness of futures contracts in managing initial exposures, such as price movements, for which these instruments were originally intended. Additionally, backtesting frameworks require further exploration, with a focus on how model misspecification, ranking, and the choice of loss functions affect economic outcomes. Such studies could provide practitioners in electricity markets with actionable insights while also informing risk management strategies in other energy sectors facing similar challenges.

While tailored for electricity futures, identifying generalized principles applicable to other energy commodities could uncover universal strategies for risk assessment across markets like natural gas and oil.

In conclusion, without claiming to be exhaustive, this work provides an analysis of joint VaR and ES forecasting models, revealing actionable insights for stakeholders in the electricity market while highlighting key areas for further academic investigation.

Table 16
Average Quantile Loss Function Results at 2.5% Quantile.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Model Confidence Set (MCS), and entries marked with ** are included in the 10% MCS.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	2.545**	3.95**	0.334**	0.333**	0.383**	0.411**	0.32	0.279*
FHS-EWMA	2.777**	4.639	0.322**	0.333**	0.351**	0.398**	0.285**	0.235**
FHS-sGARCH-norm	2.741**	4.485	0.344	0.361	0.31**	0.366**	0.267**	0.205**
FHS-sGARCH-std	2.76**	4.07**	0.304**	0.305**	0.305**	0.35**	NULL	NULL
FHS-sGARCH-sstd	2.727**	4.095**	0.296**	0.307**	0.301**	0.346**	0.315	NULL
FHS-gjrGARCH-norm	2.763**	4.505	0.358	0.372	0.328**	0.378**	0.268	0.213**
FHS-gjrGARCH-std	2.821**	4.251**	0.315**	0.315**	0.313**	0.363**	NULL	NULL
FHS-gjrGARCH-sstd	2.786**	4.614	0.322**	0.333**	0.35**	0.388**	0.285**	0.237**
sGARCH-norm	2.741**	4.485	0.344	0.361	0.31**	0.366**	0.267**	0.205**
sGARCH-std	2.76**	4.07**	0.304**	0.305**	0.305**	0.35**	NULL	NULL
sGARCH-sstd	2.727**	4.095**	0.296**	0.307**	0.301**	0.346**	0.315	NULL
gjrGARCH-norm	2.763**	4.505	0.358	0.372	0.328**	0.379**	0.268	0.213**
gjrGARCH-std	2.821**	4.251**	0.315**	0.315**	0.313**	0.363**	NULL	NULL
gjrGARCH-sstd	2.784**	4.215**	0.304**	0.316**	0.31**	0.358**	0.305	NULL
EVT-sGARCH-norm	3.762	4.932	0.333**	0.352	0.307**	0.346**	0.26**	0.206**
EVT-sGARCH-std	3.831	4.538	0.321**	0.331**	0.311**	0.354**	NULL	NULL
EVT-sGARCH-sstd	3.838	4.516**	0.318**	0.345	0.316**	0.36**	0.338	0.206**
EVT-gjrGARCH-norm	4.038	5.181	0.339	0.358	0.324**	NULL	0.269**	0.215**
EVT-gjrGARCH-std	4.044	5.09	0.322**	0.328**	0.309**	0.36**	NULL	NULL
EVT-gjrGARCH-sstd	4.017	5.136	0.325	0.384	0.313**	0.359**	0.327	0.21**
GAS-norm	2.811**	21.405	0.35	0.362	0.431**	0.431	0.301	8.036
GAS-std	2.704**	4.274**	0.354	0.373	0.326**	0.396**	0.254**	0.25*
GAS-sstd	2.709**	4.212**	0.475	0.46	0.357	0.458	0.298	0.256
CAViaR-Adaptive-AR	518.296	164.621	0.362	0.363	0.373**	0.415**	0.356	0.327
CAViaR-SAV-AR	15.765	33.576	1.014	1.2	1.407	1.803	1.388	0.868
CAViaR-AS-AR	2.825**	4.417**	0.478	0.49	0.52	0.598	0.444	0.404
CAViaR-IGARCH-AR	2.594**	4.004**	0.342	0.325**	0.355**	0.382**	0.298**	0.255**
CAViaR-Adaptive-MULT	889.329	755.321	0.37	0.378	0.391	0.436	0.348	0.31
CAViaR-SAV-MULT	2.611**	4.183**	0.346	0.336	0.373**	1.233	0.99	0.551
CAViaR-AS-MULT	7.168	7.366	0.858	0.878	1.277	1.016	0.91	0.739
CAViaR-IGARCH-MULT	2.603**	4.148**	0.346	0.349	0.365**	0.395**	0.307**	0.23**

A. Additional Material

References

Adamko, P., Spuchľáková, E., Valášková, K., 2015. The history and ideas behind var. Procedia Economics and Finance 24, 18–24. URL: https://www.sciencedirect.com/science/article/pii/S2212567115006073, doi:https://doi.org/10.1016/S2212-5671(15)00607-3. international Conference on Applied Economics (ICOAE) 2015, 2-4 July 2015, Kazan, Russia.

Amaro, R., Pinho, C., 2022. Energy commodities: A study on model selection for estimating value-at-risk. Applied Econometrics 68, 5 – 27. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148287208&doi=10.22394%2f1993-7601-2022-68-5-27&partnerID=40&md5=5a339677f168f21c1912ba7964fa92d7, doi:10.22394/

Table 17
Average Quantile Loss Function Results at 5% Quantile.
Formatting: Full outline denote the smallest loss in each column, dashed outline denote the second smallest loss. Entries marked with * are included in the 5% Model Confidence Set (MCS), and entries marked with ** are included in the 10% MCS.

Model	DBc1	DPc1	DEBMc1	DEPMc1	DEBQc1	DEPQc1	DEBYc1	DEPYc1
HS	4.026**	5.737**	0.483**	0.489**	0.576	0.583	0.448	0.387**
FHS-EWMA	4.275**	6.359	0.475**	0.5*	0.552	0.565	0.43**	0.358**
FHS-sGARCH-norm	4.194**	6.609	0.539	0.57	0.509	0.567	0.408	0.328**
FHS-sGARCH-std	4.199**	5.888**	0.453**	0.464**	0.491**	0.528**	NULL	NULL
FHS-sGARCH-sstd	4.194**	5.913**	0.45**	0.464**	0.492**	0.524**	0.448	NULL
FHS-gjrGARCH-norm	4.179**	6.499	0.545	0.567	0.524	0.576	0.413	0.334**
FHS-gjrGARCH-std	4.237	6.009	0.466	0.478	0.497	0.527	NULL	NULL
FHS-gjrGARCH-sstd	4.289**	6.352	0.475**	0.498*	0.554	0.562	0.428**	0.357**
sGARCH-norm	4.194**	6.609	0.539	0.57	0.509	0.567	0.408	0.328**
sGARCH-std	4.199**	5.888**	0.453**	0.464**	0.491**	0.528**	NULL	NULL
sGARCH-sstd	4.194**	5.913**	0.45**	0.464**	0.492**	0.524**	0.448	NULL
gjrGARCH-norm	4.179**	6.499	0.545	0.567	0.524	0.576	0.413	0.334**
gjrGARCH-std	4.237**	6.009**	0.466**	0.478**	0.497**	0.527**	NULL	NULL
gjrGARCH-sstd	4.246**	6.036**	0.458**	0.47**	0.491**	0.532**	0.444	NULL
EVT-sGARCH-norm	4.53*	6.414	0.481**	0.508	0.496**	0.521**	0.394**	0.324**
EVT-sGARCH-std	4.656	6.205**	0.472**	0.487**	0.505	0.552**	NULL	NULL
EVT-sGARCH-sstd	4.735	6.245**	0.47**	0.525	0.514	0.571	0.562	0.33**
EVT-gjrGARCH-norm	4.688	6.498	0.485**	0.504	0.508**	NULL	0.4**	0.329**
EVT-gjrGARCH-std	4.712	6.203	0.484**	0.5	0.5**	0.563	*NULL***	*NULL***
EVT-gjrGARCH-sstd	4.706	6.238	0.496	0.504	0.508	0.582	0.508	0.326**
GAS-norm	4.328**	34.877	0.555	0.575	0.711	0.669	0.446**	13.466
GAS-std	4.167**	6.034**	0.545	0.572	0.533	0.628	0.399**	0.394**
GAS-sstd	4.158**	5.947**	0.736	0.708	0.579	0.727	0.432	0.393*
CAViaR-Adaptive-AR	16.231	44.686	0.5**	0.526	0.581	0.589	0.462	0.405*
CAViaR-SAV-AR	11.103	25.975	1.081	1.714	1.375	1.29	0.932	1.06
CAViaR-AS-AR	4.308**	6.474	0.839	0.775	0.846	0.821	0.643	0.596
CAViaR-IGARCH-AR	4.112**	5.909**	0.499	0.504	0.549	0.56**	0.43**	0.359**
CAViaR-Adaptive-MULT	157.499	387.019	0.5**	0.524	0.577	0.59	0.449	0.398
CAViaR-SAV-MULT	4.127**	6.197**	0.497**	0.49**	0.676	0.621	0.501	0.501
CAViaR-AS-MULT	7.15	8.767	0.935	0.932	1.209	1.162	1.125	0.775
CAViaR-IGARCH-MULT	4.125**	5.866**	0.504	0.499*	0.543	0.564**	0.44**	0.357**

1993-7601-2022-68-5-27. cited by: 0.

Amaro, R., Pinho, C., Madaleno, M., 2022. Forecasting the value-at-risk of energy commodities: A comparison of models and alternative distribution functions. Applied Econometrics 65, 77–101. doi:10.22394/1993-7601-2022-65-77-101.

Antwi, A., Kyei, K.A., Gill, R.S., 2020. The use of mutual information to improve value-at-risk forecasts for exchange rates. IEEE Access 8, 179881 – 179900. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092790811&doi=10.1109%2fACCESS.2020.3027631&partnerID=40&md5=44618a92ca2e61aedb75aca028dd8185, doi:10.1109/ACCESS.2020.3027631. cited by: 5; All Open Access, Gold Open Access.

Ardia, D., Boudt, K., Catania, L., 2019. Generalized autoregressive score models in r: The gas package. Journal of Statistical Software 88, 1–28.

Table 18
Summary of Rejected Backtests and Average AE Results at 2.5% quantile.
Column Descriptions: Average AE represents the average Actual Exceedance Rate. **UC** (Unconditional Coverage), **CC** (Conditional Coverage), **VDT** (Violation Duration Test), **DQ** (Dynamic Quantile Test), **ER** (Exceedance Residual Test), **CoC** (Conditional Calibration Test), and **AESR** (Auxiliary Expected Shortfall Residual Test), **SESR** (Strict Expected Shortfall Residual Test), **IESR** (Intercept Expected Shortfall Residual Test) denote the number of backtests rejected for each criterion. Lower values indicate better compliance with backtesting criteria. Fully outlined entries denote the smallest number of rejections, while dashed outlines denote the second smallest.

Model	Average AE	UC	CC	VDT	DQ	ER	CoC	AESR	SESR	IESR
HS	1.151	0	5	5	6	0	0	1	1	0
FHS-EWMA	1.057	0	1	1	4	1	1	2	2	0
FHS-sGARCH-norm	0.838	5	4	0	4	7	8	6	5	4
FHS-sGARCH-std	1.111	4	5	0	3	5	6	6	5	3
FHS-sGARCH-sstd	1.172	2	3	1	3	7	7	6	6	4
FHS-gjrGARCH-norm	1.083	3	5	3	6	7	6	8	7	8
FHS-gjrGARCH-std	1.277	3	4	1	3	4	6	6	5	3
FHS-gjrGARCH-sstd	1.053	0	1	1	5	2	1	1	2	0
sGARCH-norm	0.838	5	4	0	4	7	8	6	5	4
sGARCH-std	1.111	4	5	0	3	5	6	6	5	3
sGARCH-sstd	1.172	2	3	1	3	7	7	6	6	4
gjrGARCH-norm	1.083	3	5	3	6	7	6	8	7	8
gjrGARCH-std	1.277	3	4	1	3	4	6	6	5	3
gjrGARCH-sstd	1.297	2	3	1	4	6	6	6	5	3
EVT-sGARCH-norm	2.146	7	7	4	6	5	8	1	1	1
EVT-sGARCH-std	2.217	4	4	0	5	6	6	3	3	2
EVT-sGARCH-sstd	1.949	5	4	0	6	5	6	3	3	2
EVT-gjrGARCH-norm	2.625	7	7	4	7	3	7	2	2	2
EVT-gjrGARCH-std	2.558	4	3	2	3	4	5	4	4	3
EVT-gjrGARCH-sstd	2.519	6	5	2	6	4	6	4	4	3
GAS-norm	0.781	5	6	1	4	6	7	2	2	0
GAS-std	1.020	5	5	1	3	6	7	2	2	0
GAS-sstd	1.026	5	5	1	3	8	7	3	3	0
CAViaR-IGARCH-AR	1.282	2	2	0	5	2	0	2	1	0
CAViaR-IGARCH-MULT	1.302	3	5	1	6	0	1	3	3	0

URL: https://www.jstatsoft.org/index.php/jss/article/view/v088i06, doi:10.18637/jss.v088.i06.

Argyropoulos, C., Panopoulou, E., 2019. Backtesting var and es under the magnifying glass. International Review of Financial Analysis 64, 22–37. URL: https://www.sciencedirect.com/science/article/pii/S1057521918305507, doi:https://doi.org/10.1016/j.irfa.2019.04.005.

Authority, E.B., 2024. Guidelines on stressed value at risk (stressed var). https://www.eba.europa.eu/guidelines-stressed-value-risk-stressed-var?version=Default#activity-versions. Accessed: 2024-12-16.

Badáňová, M., 2015. Forecasting in Futures Markets: Front, Back, and Rolling Contracts. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. URL: https://dspace.cuni.cz/bitstream/handle/20.500.11956/64304/DPTX_2012_2_11230_0_392512_0_138304.pdf?sequence=1&isAllowed=y.

Bayer, S., Dimitriadis, T., 2020. Regression-based expected shortfall backtesting*. Journal of Financial Econometrics 20, 437–471. URL: https://doi.org/10.1093/jjfinec/nbaa013, doi:10.1093/jjfinec/nbaa013, arXiv:https://academic.oup.com/jfec/article-pdf/20/3/437/43974203/nbaa013.pdf.

Bayer, S., Dimitriadis, T., 2023. esback: Expected Shortfall Backtesting. URL: https://CRAN.R-project.org/package=esback, doi:10.32614/CRAN.package.esback. r package version 0.3.1.

Behera, J., Pasayat, A.K., Behera, H., Kumar, P., 2023. Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence 120.

Table 19

Summary of Rejected Backtests and Average AE Results at 5% quantile.

Column Descriptions: Average AE represents the average Actual Exceedance Rate. **UC** (Unconditional Coverage), **CC** (Conditional Coverage), **VDT** (Violation Duration Test), **DQ** (Dynamic Quantile Test), **ER** (Exceedance Residual Test), **CoC** (Conditional Calibration Test), and **AESR** (Auxiliary Expected Shortfall Residual Test), **SESR** (Strict Expected Shortfall Residual Test), **IESR** (Intercept Expected Shortfall Residual Test) denote the number of backtests rejected for each criterion. Lower values indicate better compliance with backtesting criteria. Fully outlined entries denote the smallest number of rejections, while dashed outlines denote the second smallest.

Model	Average AE	UC	CC	VDT	DQ	ER	CoC	AESR	SESR	IESR
HS	1.076	0	6	6	6	0	0	2	1	0
FHS-EWMA	1.056	0	1	0	5	0	2	3	3	0
FHS-sGARCH-norm	0.651	5	5	1	4	0	7	6	8	4
FHS-sGARCH-std	1.103	4	4	0	2	5	6	6	6	5
FHS-sGARCH-sstd	1.075	5	5	1	2	7	7	7	7	6
FHS-gjrGARCH-norm	0.776	3	5	5	5	0	5	6	8	4
FHS-gjrGARCH-std	1.133	4	3	1	4	5	6	5	5	4
FHS-gjrGARCH-sstd	1.062	0	1	0	5	1	4	3	3	2
sGARCH-norm	0.651	5	5	1	4	0	7	6	8	4
sGARCH-std	1.103	4	4	0	2	5	6	6	6	5
sGARCH-sstd	1.075	5	5	1	2	7	7	7	7	6
gjrGARCH-norm	0.776	3	5	5	5	0	5	6	8	4
gjrGARCH-std	1.133	4	3	1	4	5	6	5	5	4
gjrGARCH-sstd	1.157	3	4	2	4	7	6	6	7	5
EVT-sGARCH-norm	1.018	5	5	4	5	5	6	7	7	5
EVT-sGARCH-std	1.055	4	4	0	3	6	5	5	5	5
EVT-sGARCH-sstd	0.913	6	6	0	6	5	7	6	6	4
EVT-gjrGARCH-norm	1.239	3	5	3	6	3	5	5	5	4
EVT-gjrGARCH-std	1.224	5	4	1	3	4	6	5	5	5
EVT-gjrGARCH-sstd	1.203	6	6	1	5	4	7	5	5	4
GAS-norm	0.630	6	6	1	5	6	7	3	2	0
GAS-std	0.834	8	7	1	4	6	7	3	3	1
GAS-sstd	0.804	8	8	1	6	6	8	4	3	1
CAViaR-IGARCH-AR	1.041	0	4	1	5	6	3	3	4	4
CAViaR-IGARCH-MULT	1.052	0	3	1	5	0	0	2	2	0

URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146874760&doi=10.1016%2fj.engappai.2023.105843&partnerID=40&md5=e8423bbf13dd9d74c3338d3b07bdd60a, doi:10.1016/j.engappai.2023.105843. cited by: 46.

Benito, S., Lopez, C., Navarro, M., 2023. Assessing the importance of the choice threshold in quantifying market risk under the pot approach. Risk Management 25, 1. doi:10.1057/s41283-022-00106-w.

Berger, T., Moys, G., 2021. Value-at-risk backtesting: Beyond the empirical failure rate. Expert Systems with Applications 177, 114893. URL: https://www.sciencedirect.com/science/article/pii/S0957417421003341, doi:https://doi.org/10.1016/j.eswa.2021.114893.

BIS, 2006. International convergence of capital measurement and capital standards. URL: https://www.bis.org/publ/bcbs107.htm.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327. URL: https://www.sciencedirect.com/science/article/pii/0304407686900631, doi:https://doi.org/10.1016/0304-4076(86)90063-1.

Buczyński, M., Chlebus, M., 2020. Old-fashioned parametric models are still the best: A comparison of value-at-risk approaches in several volatility states. Journal of Risk Model Validation 14, 1 – 20. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087618722&doi=10.21314%2fJRMV.2020.222&partnerID=40&md5=2b3c89590b4e7108e188fe24288c2ab1, doi:10.21314/JRMV.2020.222. cited by: 5.

Băra, A., Georgescu, I.A., Oprea, S.V., Cristescu, M.P., 2024. Exploring the dynamics of brent crude oil, sp500 and bitcoin prices amid economic instability. IEEE Access 12, 31366 – 31385. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85186957328&doi=10.1109%2fACCESS.2024.3370029&partnerID=40&md5=91e9715a6e0837c1aa3afde437bb06ef, doi:10.1109/ACCESS.2024.3370029. cited by: 2; All Open Access, Gold Open Access.

Candia, C., Herrera, R., 2024. An empirical review of dynamic extreme value models for forecasting value at risk, expected shortfall and expectile. Journal of Empirical Finance 77. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85189012463&doi=10.1016%2fj.jempfin.2024.101488&partnerID=40&md5=8313b8093e6b0f95b47960778c39ab28, doi:10.1016/j.jempfin.2024.101488. cited by: 1.

Cartesan, L., Penzo, J., 2020. FUTURES HEDGING IN THEELECTRICITY MARKET. Master's thesis. Copenhagen Business School.

Chen, Q., Huang, Z., Liang, F., 2024. Forecasting volatility and value-at-risk for cryptocurrency using garch-type models: the role of the probability distribution. Applied Economics Letters 31, 1907 – 1914. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85159028992&doi=10.1080%2f13504851.2023.2208824&partnerID=40&md5=3ab919f20dbb2985b864b4ee55a2e166, doi:10.1080/13504851.2023.2208824. cited by: 1.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics 2, 84–108. URL: https://doi.org/10.1093/jjfinec/nbh004, doi:10.1093/jjfinec/nbh004, arXiv:https://academic.oup.com/jfec/article-pdf/2/1/84/2840715/nbh004.pdf.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862. URL: http://www.jstor.org/stable/2527341.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1279, doi:https://doi.org/10.1002/jae.1279, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.1279.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253–263. URL: https://www.tandfonline.com/doi/abs/10.1080/07350015.1995.10524599, doi:10.1080/07350015.1995.10524599, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07350015.1995.10524599.

Diks, C., Fang, H., 2020. Comparing density forecasts in a risk management context. International Journal of Forecasting 36, 531 – 551. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076225367&doi=10.1016%2fj.ijforecast.2019.07.006&partnerID=40&md5=810dc4d1304858175be084127d0cf57f, doi:10.1016/j.ijforecast.2019.07.006. cited by: 6; All Open Access, Green Open Access.

DNV, 2020. Liquidity and transaction costs. URL: https://www.acer.europa.eu/sites/default/files/documents/en/Electricity/MARKET-CODES/CAPACITY-ALLOCATION-AND-CONGESTION-MANAGEMENT/Documents/200406%20DNV%20GL%20report_final.pdf.

Du, Z., Escanciano, J.C., 2017. Backtesting expected shortfall: Accounting for tail risk. Management Science 63, 940–958. URL: https://doi.org/10.1287/mnsc.2015.2342, doi:10.1287/mnsc.2015.2342, arXiv:https://doi.org/10.1287/mnsc.2015.2342.

Echaust, K., Just, 2020. Value at risk estimation using the garch-evt approach with optimal tail selection. Mathematics 8, 114. doi:10.3390/math8010114.

EEX, . Power derivatives market. URL: https://www.eex.com/en/markets/power.

EEX, 2024a. Eex glossary. URL: https://www.eex.com/en/glossary. accessed: December 5, 2024.

EEX, 2024b. EEX Press Release - EEX Group Annual Volumes 2023: Markets Continue to Grow. URL: https://www.eex.com/en/newsroom/detail?tx_news_pi1%5Baction%5D=detail&tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Bnews%5D=7736&cHash=6f447a9ff798d7bd5f7cf45c262c5b93. accessed: 2024-12-19.

Engle, R.F., Manganelli, S., 2004. Caviar. Journal of Business & Economic Statistics 22, 367–381. URL: https://doi.org/10.1198/073500104000000370, doi:10.1198/073500104000000370, arXiv:https://doi.org/10.1198/073500104000000370.

EU Council, 2024. Impact of Russia's Invasion of Ukraine on the Markets: EU Response. URL: https://www.consilium.europa.eu/en/policies/eu-response-russia-military-aggression-against-ukraine/impact-of-russia-s-invasion-of-ukraine-on-the-markets-eu-response/. archived content, not updated since 15 January 2024.

Fantazzini, D., 2024. Adaptive conformal inference for computing market risk measures: An analysis with four thousand crypto-assets. Journal of Risk and Financial Management 17. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85197113134&doi=10.3390%2fjrfm17060248&partnerID=40&md5=b60e8f2dc8ff8347881923f924f560ae, doi:10.3390/jrfm17060248. cited by: 0; All Open Access, Gold Open Access.

Filippi, C., Guastaroba, G., Speranza, M., 2020. Conditional value-at-risk beyond finance: a survey. International Transactions in Operational Research 27, 1277 – 1319. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073972281&doi=10.1111%2fitor.12726&partnerID=40&md5=f5a73c044d7a2b97dccd4d4cc847ec4d.

Fissler, T., Ziegel, J.F., 2016. Higher order elicitability and Osband's principle. The Annals of Statistics 44, 1680 – 1707. URL: https://doi.org/10.1214/16-AOS1439, doi:10.1214/16-AOS1439.

Fiszeder, P., Małecka, M., Molnár, P., 2024. Robust estimation of the range-based garch model: Forecasting volatility, value at risk and expected shortfall of cryptocurrencies. Economic Modelling 141. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85205135788&doi=10.1016%2fj.econmod.2024.106887&partnerID=40&md5=32fec73ad0d674372f30e17d5cfe2dfd, doi:10.1016/j.econmod.2024.106887. cited by: 0.

Fu, J.Y., Lin, J.G., Hao, H.X., 2023. Volatility analysis for the garch–itô–jumps model based on high-frequency and low-frequency financial data. International Journal of Forecasting 39, 1698 – 1712. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85139026817&doi=10.1016%2fj.ijforecast.2022.08.006&partnerID=40&md5=a0dbad2ed436da6b589c20dbfd488669, doi:10.1016/j.ijforecast.2022.08.006. cited by: 2.

Fülle, M.J., Herwartz, H., 2024. Predicting tail risks by a markov switching mgarch model with varying copula regimes. Journal of Forecasting 43, 2163 – 2186. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85188525161&doi=10.1002%2ffor.3117&partnerID=40&md5=bfa50594757d5ec2431778ef59fc3c20, doi:10.1002/for.3117. all Open Access, Hybrid Gold Open Access.

Galanos, A., 2024. rugarch: Univariate GARCH models. R package version 1.5-3.

Garcia-Jorcano, L., Novales, A., 2021. Volatility specifications versus probability distributions in var forecasting. Journal of Forecasting 40, 189 – 212. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087954135&doi=10.1002%2ffor.2697&partnerID=40&md5=c5f21788cb2a480f595bcf7f823b3e8a, doi:10.1002/for.2697. cited by: 9; All Open Access, Green Open Access.

Gerlach, R., Wang, C., 2020. Semi-parametric dynamic asymmetric laplace models for tail risk forecasting, incorporating realized measures. International Journal of Forecasting 36, 489 – 506. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85080998803&doi=10.1016%2fj.ijforecast.2019.07.003&partnerID=40&md5=e7294b1438f9a25180e9c451cbb85a86, doi:10.1016/j.ijforecast.2019.07.003. cited by: 19; All Open Access, Green Open Access.

Gerlach, R., Wang, C., 2022. Bayesian semi-parametric realized conditional autoregressive expectile models for tail risk forecasting. Journal of Financial Econometrics 20, 105 – 138. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85142263465&doi=10.1093%2fjjfinec%2fnbaa002&partnerID=40&md5=b781a6511892a62d40c8dcc0ed431d1b, doi:10.1093/jjfinec/nbaa002. cited by: 9.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2006.00718.x, doi:https://doi.org/10.1111/j.1468-0262.2006.00718.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2006.00718.x.

GLOSTEN, L.R., JAGANNATHAN, R., 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 48, 1779–1801. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1993.tb05128.x, doi:https://doi.org/10.1111/j.1540-6261.1993.tb05128.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1993.tb05128.x.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378. URL: https://doi.org/10.1198/016214506000001437, doi:10.1198/016214506000001437, arXiv:https://doi.org/10.1198/016214506000001437.

Golodnikov, A., Kuzmenko, V., Uryasev, S., 2019. Cvar regression based on the relation between cvar and mixed-quantile quadrangles. Journal of Risk and Financial Management 12. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85100766464&doi=10.3390%2fjrfm12030107&partnerID=40&md5=c756964ec66fcae48482401196af4e57, doi:10.3390/jrfm12030107. cited by: 8; All Open Access, Gold Open Access, Green Open Access.

González-Rivera, G., Lee, T.H., Mishra, S., 2004. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting 20, 629–645. URL: https://www.sciencedirect.com/science/article/pii/S0169207003001420, doi:https://doi.org/10.1016/j.ijforecast.2003.10.003.

Gupta, S., Sisir, G., 2023. Comparison of value-at-risk forecast using regime-dependent models for the sp 200 energy index. Procedia Computer Science 224, 328–335. URL: https://www.sciencedirect.com/science/article/pii/S1877050923010918, doi:https://doi.org/10.1016/j.procs.2023.09.044. 18th International Conference on Future Networks and Communications / 20th International Conference on Mobile Systems and Pervasive Computing / 13th International Conference on Sustainable Energy Information Technology.

ter Haar, R., 2010. On modelling the electricity futures curve. URL: https://essay.utwente.nl/60867/1/BSc_Rick_ter_Haar.pdf.

Halbleib, R., Pohlmeier, W., 2012. Improving the value at risk forecasts: Theory and evidence from the financial crisis. Journal of Economic Dynamics and Control 36, 1212–1228. URL: https://www.sciencedirect.com/science/article/pii/S0165188912000887, doi:https://doi.org/10.1016/j.jedc.2011.10.005. quantifying and Understanding Dysfunctions in Financial Markets.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business Economic Statistics 23, 365–380.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA5771, doi:https://doi.org/10.3982/ECTA5771.

Happersberger, D., Lohre, H., Nolte, I., 2020. Estimating portfolio risk for tail risk protection strategies. European Financial Management 26, 1107–1146. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/eufm.12256, doi:https://doi.org/10.1111/eufm.12256.

Hill, C., Mccullough, B.D., 2020. On the accuracy of garch estimation in r packages 4, 133–156.

Iqbal, F., Zahid, M., Koutmos, D., 2023. Cryptocurrency trading and downside risk. Risks 11. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85166440341&doi=10.3390%2frisks11070122&partnerID=40&md5=69bca2425e8f4e05389b965010eb0b62, doi:10.3390/risks11070122. cited by: 2; All Open Access, Gold Open Access.

Jongadsayakul, W., 2021. Value at risk estimation of the set50 index: Comparison between stock exchange of thailand and thailand futures exchange. Journal of International Studies 14, 227 – 240. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85105342487&doi=10.14254%2f2071-8330.2021%2f14-1%2f16&partnerID=40&md5=b8d1659419f775dfad8b3c3648a2b5c7, doi:10.14254/2071-8330.2021/14-1/16.

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

Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Economics 27, 791–817. URL: https://www.sciencedirect.com/science/article/pii/S0140988305000228.

Kuang, W., 2022. Oil tail-risk forecasts: from financial crisis to covid-19. Risk Management 24, 420 – 460. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137679432&doi=10.1057%2fs41283-022-00100-2&partnerID=40&md5=fdda0361f779cd6d5836d68e35eeb97d, doi:10.1057/s41283-022-00100-2. cited by: 1; All Open Access, Bronze Open Access.

Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3, 73–84. URL: http://dx.doi.org/10.3905/jod.1995.407942, doi:10.3905/jod.1995.407942.

Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy 293, 116983. URL: https://www.sciencedirect.com/science/article/pii/S0306261921004529, doi:https://doi.org/10.1016/j.apenergy.2021.116983.

Laporta, A.G., Merlo, L., Petrella, L., 2018. Selection of value at risk models for energy commodities. Energy Economics 74, 628–643. URL: https://www.sciencedirect.com/science/article/pii/S0140988318302548, doi:https://doi.org/10.1016/j.eneco.2018.07.009.

Lazar, E., Xue, X., 2020. Forecasting risk measures using intraday data in a generalized autoregressive score framework. International Journal of Forecasting 36, 1057 – 1072. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078808274&doi=10.1016%2fj.ijforecast.2019.10.007&partnerID=40&md5=5dbbab3b75e586658d11a2126ffd107a, doi:10.1016/j.ijforecast.2019.10.007. cited by: 27; All Open Access, Green Open Access, Hybrid Gold Open Access.

Le, T.H., 2024. Forecasting var and es in emerging markets: The role of time-varying higher moments. Journal of Forecasting 43, 402 – 414. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85173488156&doi=10.1002%2ffor.3039&partnerID=40&md5=6146f6c6e688706adb53d965d0b12609, doi:10.1002/for.3039. cited by: 1.

Leccadito, A., Staino, A., Toscano, P., 2024. A novel robust method for estimating the covariance matrix of financial returns with applications to risk management. Financial Innovation 10. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85200244182&doi=10.1186%2fs40854-024-00642-2&partnerID=40&md5=ed30cf655a9f04bb282f0facdcaa3eed, doi:10.1186/s40854-024-00642-2. all Open Access, Gold Open Access.

Lee, D., Kramer, B., 2023. Multifidelity conditional value-at-risk estimation by dimensionally decomposed generalized polynomial chaos-kriging. Reliability Engineering and System Safety 235. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85150026397&doi=10.1016%2fj.ress.2023.109208&partnerID=40&md5=48383fdac2cb259c1d5adffdf3b1520b, doi:10.1016/j.ress.2023.109208. cited by: 1; All Open Access, Bronze Open Access.

Letmathe, S., 2024. quarks: Simple Methods for Calculating and Backtesting Value at Risk and Expected Shortfall. URL: https://CRAN.R-project.org/package=quarks, doi:10.32614/CRAN.package.quarks. r package version 1.1.4.

Liu, W., Semeyutin, A., Lau, C.K.M., Gozgor, G., 2020. Forecasting value-at-risk of cryptocurrencies with riskmetrics type models. Research in International Business and Finance 54. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086856341&doi=10.1016%2fj.ribaf.2020.101259&partnerID=40&md5=1f217fb23f907dd9b52603b366831a0f, doi:10.1016/j.ribaf.2020.101259. cited by: 37; All Open Access, Green Open Access.

Lu, X., Liu, C., Lai, K.K., Cui, H., 2023. Risk measurement in bitcoin market by fusing lstm with the joint-regression-combined forecasting model. Kybernetes 52, 1487 – 1502. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85121657874&doi=10.1108%2fK-07-2021-0620&partnerID=40&md5=f3797cbf3f25c788b3727cee4f09dda1, doi:10.1108/K-07-2021-0620. cited by: 7.

Štefan Lyócsa, Plíhal, T., Výrost, T., 2024. Forecasting day-ahead expected shortfall on the eur/usd exchange rate: The (i)relevance of implied volatility. International Journal of Forecasting 40, 1275–1301. URL: https://www.sciencedirect.com/science/article/pii/S0169207023001115, doi:https://doi.org/10.1016/j.ijforecast.2023.11.003.

Ma, Y., Yan, L., Pan, D., 2024. The power of news data in forecasting tail risk: evidence from china. Empirical Economics 67, 2607 – 2642. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85197697422&doi=10.1007%2fs00181-024-02620-0&partnerID=40&md5=939ceb67fe11e09f9f1192b9ac319a00, doi:10.1007/s00181-024-02620-0. cited by: 0.

Manh, T., 2017. Value-at-Risk estimates. Ph.D. thesis. Aston University.

Marcato, G., Ward, C., 2007. Back from beyond the bid–ask spread: Estimating liquidity in international markets. Real Estate Economics 35, 599–622. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6229.2007.00202.x, doi:https://doi.org/10.1111/j.1540-6229.2007.00202.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6229.2007.00202.x.

Marchese, M., Kyriakou, I., Tamvakis, M., Di Iorio, F., 2020. Forecasting crude oil and refined products volatilities and correlations: New evidence from fractionally integrated multivariate garch models. Energy Economics 88. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084832574&doi=10.1016%2fj.eneco.2020.104757&partnerID=40&md5=603542e32cc8c8bb28b3df414d842a1a, doi:10.1016/j.eneco.2020.104757. cited by: 30; All Open Access, Green Open Access.

May, N.G., Neuhoff, K., Richstein, J., 2019. Europe's power system in transition: What are the evolving roles of future markets and hedging? EconStor Research Reports 204599. ZBW - Leibniz Information Centre for Economics. URL: https://ideas.repec.org/p/zbw/esrepo/204599.html.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. volume 101.

McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7, 271–300. URL: https://www.sciencedirect.com/science/article/pii/S0927539800000128, doi:https://doi.org/10.1016/S0927-5398(00)00012-8. special issue on Risk Management.

Michalsky, J., Perez, R., Seals, R., Ineichen, P., 1994. Degradation of solar concentrator performance in the aftermath of mount pinatubo. Solar Energy 52, 205–213. URL: https://www.sciencedirect.com/science/article/pii/0038092X94900701, doi:https://doi.org/10.1016/0038-092X(94)90070-1.

Molineaux, B., Ineichen, P., 1996. Impact of pinatubo aerosols on the seasonal trends of global, direct and diffuse irradiance in two northern mid-latitude sites. Solar Energy 58, 91–101. URL: https://www.sciencedirect.com/science/article/pii/0038092X96000515, doi:https://doi.org/10.1016/0038-092X(96)00051-5. selected Proceedings of ISES 1995: Solar World Congress. Part I.

Müller, F.M., Righi, M.B., 2024. Comparison of value at risk (var) multivariate forecast models. Computational Economics 63, 75 – 110. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85141673841&doi=10.1007%2fs10614-022-10330-x&partnerID=40&md5=51b788ca958dd1da8afa18c951cef6a4, doi:10.1007/s10614-022-10330-x. cited by: 4; All Open Access, Bronze Open Access, Green Open Access.

Naimoli, A., Gerlach, R., Storti, G., 2022. Improving the accuracy of tail risk forecasting models by combining several realized volatility estimators. Economic Modelling 107, 105701. URL: https://www.sciencedirect.com/science/article/pii/S026499932100290X, doi:https://doi.org/10.1016/j.econmod.2021.105701.

Naimoli, A., Storti, G., 2021. Forecasting volatility and tail risk in electricity markets. Journal of Risk and Financial Management 14. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146136283&doi=10.3390%2fjrfm14070294&partnerID=40&md5=1414a634f728ca0c98c697b2d94e38a8, doi:10.3390/jrfm14070294. cited by: 3; All Open Access, Gold Open Access.

Nolde, N., Ziegel, J.F., 2017. Elicitability and backtesting: Perspectives for banking regulation. The Annals of Applied Statistics 11, 1833 – 1874. URL: https://doi.org/10.1214/17-AOAS1041, doi:10.1214/17-AOAS1041.

Owusu Junior, P., Tiwari, A.K., Tweneboah, G., Asafo-Adjei, E., 2022. Gas and garch based value-at-risk modeling of precious metals. Resources Policy 75, 102456. URL: https://www.sciencedirect.com/science/article/pii/S0301420721004645, doi:https://doi.org/10.1016/j.resourpol.2021.102456.

Owusu Junior, P., Tiwari, A.K., Tweneboah, G., Asafo-Adjei, E., 2022. Gas and garch based value-at-risk modeling of precious metals. Resources Policy 75. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85118845364&doi=10.1016%2fj.resourpol.2021.102456&partnerID=40&md5=3ec4118a015b7a53c7a2255e7c8b3902, doi:10.1016/j.resourpol.2021.102456. cited by: 32.

Peña, J.I., Rodríguez, R., Mayoral, S., 2020. Tail risk of electricity futures. Energy Economics 91, 104886. URL: https://www.sciencedirect.com/science/article/pii/S0140988320302267, doi:https://doi.org/10.1016/j.eneco.2020.104886.

Pietz, M., 2009. Risk premia in the german electricity futures market, in: 2009 3rd International Conference on Energy and Environment (ICEE), pp. 160–170.

Población, J., Serna, G., 2021. Measuring bulk shipping prices risk. Maritime Economics and Logistics 23, 291 – 309. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074032569&doi=10.1057%2fs41278-019-00129-3&partnerID=40&md5=8e60defcfbf69a6d858153f0f83888d6, doi:10.1057/s41278-019-00129-3. cited by: 1.

Pourrezaee, A., Hajizadeh, E., 2024. Forecasting bitcoin volatility and value-at-risk using stacking machine learning models with intraday data. Computational Economics URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85204708720&doi=10.1007%2fs10614-024-10713-2&partnerID=40&md5=e0df5a17a1d0fb58f12fbea367a538d8, doi:10.1007/s10614-024-10713-2. cited by: 0.

Qihui Su, Zhongling Qin, L.P., Qin, G., 2021. Efficiently backtesting conditional value-at-risk and conditional expected shortfall. Journal of the American Statistical Association 116, 2041–2052. URL: https://doi.org/10.1080/01621459.2020.1763804, doi:10.1080/01621459.2020.1763804, arXiv:https://doi.org/10.1080/01621459.2020.1763804.

Rice, G., Wirjanto, T., Zhao, Y., 2020. Forecasting value at risk with intra-day return curves. International Journal of Forecasting 36, 1023 – 1038. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85080058787&doi=10.1016%2fj.ijforecast.2019.10.006&partnerID=40&md5=04ea579a40eb89879cc2ee7cd195eb4e, doi:10.1016/j.ijforecast.2019.10.006. cited by: 10; All Open Access, Green Open Access.

Rowley, J., Slack, F., 2004. Conducting a literature review. Management Research News 27, 31–39. doi:10.1108/01409170410784185.

Saidane, M., 2024. Risk assessment in cryptocurrency portfolios: a composite hidden markov factor analysis framework. Statistics, Optimization and Information Computing 12, 463 – 487. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85186945322&doi=10.19139%2fsoic-2310-5070-1837&partnerID=40&md5=a9c5614a7a6650c40f4e861a8c3ed036, doi:10.19139/soic-2310-5070-1837. cited by: 2; All Open Access, Hybrid Gold Open Access.

Saralees Nadarajah, B.Z., Chan, S., 2014. Estimation methods for expected shortfall. Quantitative Finance 14, 271–291. URL: https://doi.org/10.1080/14697688.2013.816767, doi:10.1080/14697688.2013.816767.

Serrano Bautista, R., Núñez Mora, J.A., 2021. Value-at-risk predictive performance: a comparison between the caviar and garch models for the mila and asean-5 stock markets. Journal of Economics, Finance and Administrative Science 26, 197 – 221. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85119662328&doi=10.1108%2fJEFAS-03-2021-0009&partnerID=40&md5=c9952e46f121f1c2adf0bbbf72d559c9, doi:10.1108/JEFAS-03-2021-0009. cited by: 4; All Open Access, Gold Open Access.

Shayya, R., Sorrosal-Forradellas, M.T., Terceño, A., 2023. Value-at-risk models: a systematic review of the literature. Journal of Risk 25, 1 – 23. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85159065851&doi=10.21314%2fJOR.2022.053&partnerID=40&md5=

f7e65fdaa154c490d32fafd8f24f211f, doi:10.21314/JOR.2022.053.

Siu, Y.W., 2019. Impact of expected shortfall approach on capital requirement under basel. Review of Pacific Basin Financial Markets and Policies 22. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078355524&doi=10.1142%2fS0219091519500255&partnerID=40&md5=d4e8e91a58ad37f03d11beb3daf0b99c, doi:10.1142/S0219091519500255. cited by: 3.

Song, S., Li, H., 2022. Predicting var for china's stock market: A score-driven model based on normal inverse gaussian distribution. International Review of Financial Analysis 82, 102180. URL: https://www.sciencedirect.com/science/article/pii/S1057521922001429, doi:https://doi.org/10.1016/j.irfa.2022.102180.

Song, S., Li, H., 2023. A method for predicting var by aggregating generalized distributions driven by the dynamic conditional score. Quarterly Review of Economics and Finance 88, 203 – 214. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85147569785&doi=10.1016%2fj.qref.2023.01.006&partnerID=40&md5=2808c53a2dea8f91dfb4c991590bdf4d, doi:10.1016/j.qref.2023.01.006. cited by: 1; All Open Access, Green Open Access.

Song, S., Tian, F., Li, H., 2021. An intraday-return-based value-at-risk model driven by dynamic conditional score with censored generalized pareto distribution. Journal of Asian Economics 74. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85104454003&doi=10.1016%2fj.asieco.2021.101314&partnerID=40&md5=51772233fc69ef2e1a188db40e0aa6ec, doi:10.1016/j.asieco.2021.101314. cited by: 6.

Tasche, D., 2002. Expected shortfall and beyond. Journal of Banking & Finance 26, 1519–1533. URL: https://www.sciencedirect.com/science/article/pii/S0378426602002728, doi:https://doi.org/10.1016/S0378-4266(02)00272-8.

Taylor, J.W., 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. Journal of Business & Economic Statistics 37, 121–133. URL: https://doi.org/10.1080/07350015.2017.1281815, doi:10.1080/07350015.2017.1281815, arXiv:https://doi.org/10.1080/07350015.2017.1281815.

Taylor, J.W., 2020. Forecast combinations for value at risk and expected shortfall. International Journal of Forecasting 36, 428–441. URL: https://www.sciencedirect.com/science/article/pii/S0169207019301918, doi:https://doi.org/10.1016/j.ijforecast.2019.05.014.

Taylor, J.W., 2022. Forecasting value at risk and expected shortfall using a model with a dynamic omega ratio. Journal of Banking and Finance 140. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129765596&doi=10.1016%2fj.jbankfin.2022.106519&partnerID=40&md5=3b0a67b97f9dd3813c317bcb6f50d010, doi:10.1016/j.jbankfin.2022.106519. cited by: 10; All Open Access, Green Open Access.

Tong, B., Diao, X., Li, X., 2024. Forecasting vars via hybrid evt with normal and non-normal filters: A comparative analysis from the chinese stock market. Pacific Basin Finance Journal 83. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184470215&doi=10.1016%2fj.pacfin.2024.102271&partnerID=40&md5=c829bec50383d7b5ee8657b81182276d, doi:10.1016/j.pacfin.2024.102271. cited by: 0.

Trabelsi, N., Tiwari, A.K., 2023. Co2 emission allowances risk prediction with gas and garch models. Computational Economics 61, 775 – 805. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123077962&doi=10.1007%2fs10614-021-10231-5&partnerID=40&md5=8b3b4c0a86074662d607fefbc3255f4e, doi:10.1007/s10614-021-10231-5. cited by: 5.

Trucíos, C., Taylor, J.W., 2023. A comparison of methods for forecasting value at risk and expected shortfall of cryptocurrencies. Journal of Forecasting 42, 989 – 1007. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85142645378&doi=10.1002%2ffor.2929&partnerID=40&md5=416f2d8c4c852362a48860ee59d78a98, doi:10.1002/for.2929. cited by: 11.

Tsoukala, A., Tsiotas, G., 2021. Assessing green bond risk: an empirical investigation. Green Finance 3, 222 – 252. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134533689&doi=10.3934%2fGF.2021012&partnerID=40&md5=61ee56c3bf63e30b192855d83250bfa8, doi:10.3934/GF.2021012. cited by: 13; All Open Access, Gold Open Access.

Tsukahara, H., 2024. Estimation and backtesting of risk measures with emphasis on distortion risk measures. Japanese Journal of Statistics and Data Science URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85198047716&doi=10.1007%2fs42081-024-00264-z&partnerID=40&md5=9eb8e1f78716bb1af008c83a3aae58c5, doi:10.1007/s42081-024-00264-z. cited by: 0.

Wang, C., Gerlach, R., Chen, Q., 2023. A semi-parametric conditional autoregressive joint value-at-risk and expected shortfall modeling framework incorporating realized measures. Quantitative Finance 23, 309 – 334. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148087310&doi=10.1080%2f14697688.2022.2157322&partnerID=40&md5=bf4d547438c651b9bd1ba00de7352653, doi:10.1080/14697688.2022.2157322. cited by: 2.

Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting 30, 1030–1081. URL: https://www.sciencedirect.com/science/article/pii/S0169207014001083, doi:https://doi.org/10.1016/j.ijforecast.2014.08.008.

Westgaard, S., Arhus, G.H., Frydenberg, M., Frydenberg, S., 2019. Value-at-risk in the european energy market: A comparison of parametric, historical simulation and quantile regression value-at-risk. The Journal of Risk Model Validation 13, TBA.

Westgaard, S., Paraschiv, F., Ekern, L.L., Naustadal, I., Roland, M., 2018. Forecasting price distributions in the german electricity market. Routledge .

Wu, Z., Cai, G., 2024. Can intraday data improve the joint estimation and prediction of risk measures? evidence from a variety of realized measures. Journal of Forecasting 43, 1956 – 1974. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85187160570&doi=10.1002%2ffor.3111&partnerID=40&md5=9e58884d33b9113d5898ed5bf9433839, doi:10.1002/for.3111. cited by: 0.

Zaevski, T.S., Nedeltchev, D.C., 2023. From basel iii to basel iv and beyond: Expected shortfall and expectile risk measures. International Review of Financial Analysis 87, 102645. URL: https://www.sciencedirect.com/science/article/pii/S1057521923001618, doi:https://doi.org/10.1016/j.irfa.2023.102645.

Zhao, L.T., Zheng, Z.Y., Fu, Y., Liu, Z.X., Li, M.F., 2020. Google index-driven oil price value-at-risk forecasting: A decomposition ensemble approach. IEEE Access 8, 183351 – 183366. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102785630&doi=10.1109%2fACCESS.2020.3028124&partnerID=40&md5=9b6c35894f84c9bf73a2abce1e3ee85c, doi:10.1109/ACCESS.2020.3028124. cited by: 4; All Open Access, Gold Open Access.