NTNU
Norwegian University of
Science and Technology

Forecasting the EPEX spot price distribution using fundamental variables

Lina Lassesen Ekern
Ingrid Naustdal
Malene Roald

Industrial Economics and Technology Management
Submission date: June 2017
Supervisor: Sjur Westgaard, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Problem description

Electricity price forecasts are important inputs to energy companies' decision-making. Value-at-Risk (VaR) is the market standard for risk management, yet VaR forecasting of electricity prices remains under-researched. Our work contributes to this literature and analyses the performance of fundamental VaR forecasting approaches.

Forecasting VaR amounts to predicting the electricity price distribution. Quantile regression (QR) has several promising properties in this respect. We employ the traditional QR model, as well as novel extensions that are not yet explored for electricity prices. We benchmark these against common VaR models in literature. An important step in this analysis is to select variables with high predictive ability. We perform our study on the EPEX spot price, as the German market is the main reference for power trading in Europe. Moreover, we are not aware of previous studies that apply QR for forecasting in this market.

The purpose of our work is to identify modelling approaches that yield accurate and robust VaR forecasts, and thus, provide important insight to market operators.

Preface

This work is our Master's thesis in Financial Engineering, a specialisation within Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU).

We would like to thank our supervisor, Professor Sjur Westgaard, at the Department of Industrial Economics and Technology Management. We are thankful for the inspiration, help and excellent guidance he has provided throughout the process. We also express special gratitude to Florentina Paraschiv, Associate Professor of Financial Economics at NTNU Business School, for supplying the data set and offering valuable contributions.

Trondheim, June 9th, 2017

Lina L. Ekern | Ingrid Naustdal | Malene Roald

Abstract

The increased focus on risk management in electricity markets underpins the importance of distribution forecasting. We contribute to this literature by providing empirical evidence for the use of state-of-the-art fundamental distribution forecasting approaches. More specifically, we forecast Value-at-Risk (VaR) for the hourly German spot prices. We focus on quantile regression (QR) approaches, because of their simplicity, possibility to include fundamentals, and promising ability to capture the complex features of electricity prices.

In addition to traditional QR, we employ sophisticated extensions of the model that are not previously applied to electricity prices. Exponentially weighted QR (EWQR) and exponentially weighted double kernel QR (EWDKQR) intend to capture swift distribution changes. We benchmark these models against common VaR approaches in literature: Skewed student-t GARCH, asymmetric slope CAViaR, and symmetric absolute value CAViaR.

We argue that the greatest advantage of QR is that it models the quantiles separately. In contrast to existing works, we take advantage of this through the variable selection. We propose using a separate set of variables for each quantile and trading period, and our results support that this improves forecasting accuracy. Moreover, our findings highlight the importance of variable selection, and show that it in many cases is as important as the choice of model.

The empirical study shows that EWQR is the best model overall. It consistently exhibits good performance across trading periods, and performs particularly well in the outer tail quantiles. This suggests that EWQR is able to capture the changing market conditions in Germany. The CAViaR models are the best performing benchmarks, but their performance is inconsistent. The GARCH model captures clustering of exceedances the best. However, it performs poor overall. Our results indicate that both EWQR and EWDKQR suffer from overfitting. This is evident in the mid-region of the distribution where traditional QR outperforms the more complex models.

Based on the encouraging results in this thesis, we recommend further studies to investigate EWQR. Its ease of implementation, transparency and low computational complexity increases the probability of industry adoption.

Sammendrag

Vi ser et økt fokus på risikostyring i elektrisitetsmarkeder. Dette gjør det viktig å predikere sannsynlighetsfordelingen for fremtidige priser. Vi bidrar til litteraturen ved å fremlegge empiriske resultater for bruk av fundamentale metoder for å estimere prisfordelinger. Mer presist beregner vi risikomålet *Value-at-Risk (VaR)* for tyske spottpriser på timesbasis. Vi fokuserer på kvantilregresjonsmetoder (QR), fordi de er enkle, kan brukes med fundamentale variabler og har vist seg å være lovende for å fange de komplekse egenskapene til elektrisitetspriser.

I tillegg til tradisjonell QR, bruker vi mer avanserte versjoner av QR som ikke tidligere er anvendt på elektrisitetspriser. Eksponentielt vektet QR (EWQR) og eksponentielt vektet dobbel kernel QR (EWDKQR) er metoder ment for å fange opp raske endringer i fordelinger. Vi sammenligner disse modellene med vanlige referanser i litteraturen: Skewed student-t GARCH, asymmetric slope CAViaR og symmetric absolute value CAViaR.

Den største fordelen ved QR er at kvantilene modelleres separat. I motsetning til eksisterende studier, utnytter vi dette gjennom variabelseleksjonen. Vi foreslår å bruke separate forklaringsvariabler for hver kvantil og time. Resultatene våre støtter at denne typen seleksjon forbereder kvaliteten på estimatene. De viser også at en god variabelseleksjon i mange tilfeller er like viktig som valget av modell.

EWQR er den beste modellen samlet sett. Den leverer gode estimater for alle timer og gjør det spesielt bra i de ytre kvantilene. Dette indikerer at EWQR klarer å fange opp utviklingen av det tyske markedet over tid. CAViaR-modellene gjør det best blant referansemodellene, men kvaliteten på estimatene varierer. GARCH-modellen fanger opphopning av overskridelser best, men gjør det dårlig samlet sett. Resultatene våre viser at overtilpasning er et problem for både EWQR og EWDKQR. Dette ser vi i den midtre delen av fordelingen, hvor tradisjonell QR utkonkurrerer de mer komplekse modellene.

Basert på de lovende resultatene for EWQR, anbefaler vi at denne modellen utforskes videre i fremtidige studier. EWQR kan enkelt implementeres og tolkes, og krever begrenset regnekraft. Dette øker sannsynligheten for at modellen tas i bruk i industrien.

Table of Contents

List of Tables

List of Figures

Acronyms

CAViaR	Conditional autoregressive Value-at-Risk
CC	Conditional coverage
DQ	Dynamic conditional quantile
EEX	European Energy Exchange
EPEX	European Power Exchange
EWDKQR	Exponentially weighted double kernel quantile regression
EWQR	Exponentially weighted quantile regression
GARCH	General autoregressive conditional heteroscedacity
QR	Quantile regression
SIC	Schwarz Information Criterion
UC	Unconditional coverage
VaR	Value-at-Risk

Chapter 1

Introduction

Electricity price forecasts are important inputs to energy companies' decision-making. For day-to-day market operations, accurate forecasts of short-term prices are crucial. These serve as aids for producers, retailers, and speculators who seek to determine their optimal short-term strategies for production, consumption, hedging and trading (Bunn et al., 2016). In this thesis, we focus on *distribution forecasting models* and investigate their performance for the EPEX spot price.

Electricity is non-storable by nature, and a stable power system requires a constant demand-supply balance. This makes electricity a unique commodity with price dynamics not observed in any other market (Chan and Gray, 2006). Prices are highly volatile, feature high levels of skewness and kurtosis, and display significant seasonality as well as volatility clustering. Thus, forecasting in electricity markets is arguably more challenging than in traditional financial markets.

Uncontrolled exposure to market price risk can have devastating consequences for market participants (Deng and Oren, 2006). This has led to an increased focus on risk management in power markets. As stakeholders require explicit control of the risk of both high and low extreme prices, point forecasts are inadequate in many cases (Paraschiv et al., 2016). Distribution forecasts, on the other hand, provide a more comprehensive picture. According to Bunn et al. (2016), forecasting the tails of price distributions is often more crucial than central expectations. And Chatfield (2000) argues that distribution forecasts are important for planning different strategies for a range of possible outcomes.

Value-at-Risk (VaR) is a particular type of distribution forecast, and can be regarded as a market standard for risk measurement. VaR is popularly used to evaluate down-side risk in financial markets, and is favoured because it is effective and easy to interpret (Senera et al., 2012). α-VaR is defined as the threshold loss value, such that the probability that a loss over a certain time horizon exceeds this value is $\alpha\%$. When applied to electricity prices, it gives the value that the price will stay below with probability $\alpha\%$. This is equivalent to the quantiles of the distribution. Although VaR commonly refers to the tail quantiles, we use these terms interchangeably.

Despite the importance of risk management in power markets, Weron (2014) finds that distribution forecasting is "barely touched upon" in electricity price forecasting literature. This is supported by Bunn et al. (2016), who argue that for electricity markets, VaR forecasting remains under-researched. Maciejowska et al. (2016) state that the lack of such research is likely due to the increased complexity of the problem compared to point forecasting. The sparseness in current literature, combined with the importance of the problem, is our motivation for considering VaR forecasting of electricity spot prices.

There is significant evidence in literature that electricity prices adapt to fundamental variables (e.g. Karakatsani and Bunn (2008), Weron (2014) and Paraschiv et al. (2014)). In a comprehensive review of electricity forecasting literature, Weron (2014) finds that the majority of models include fundamentals. While the use of such variables has been successful in traditional point forecasting of electricity prices, literature on fundamental *distribution* forecasting is sparse. However, we find examples of fundamental *modelling* of electricity price distributions, e.g. in Paraschiv et al. (2016) and Hagfors et al. (2016a,b). These studies do not consider forecasting, but focus on in-sample fit and revealing price drivers in different parts of the distribution. A main finding from these papers is that the impact of fundamentals varies significantly across the price distribution and trading periods.

Quantile regression (QR) approaches estimate each quantile with a distinct regression. Moreover, they are simple, insensitive to outliers, avoid distributional assumptions and facilitate the use of fundamentals. Thus, they promise several attractive features for capturing the complex properties of electricity prices. QR has already received some attention in electricity price forecasting, e.g. in Bunn et al. (2016) and Lundby and Uppheim (2011). We argue that the greatest advantage of QR is that it models quantiles separately. In contrast to the aforementioned works, we take advantage of this through the variable selection. Since we know from literature that the effects of fundamentals vary, we propose using a separate set of variables for each quantile and trading period.

By using knowledge of market conditions, we form a set of fundamental factors and perform a variable selection for each trading period and quantile. We use the selected variables to forecast VaR with traditional QR, as well as with exponentially weighted QR (EWQR) and exponentially weighted double kernel QR (EWDKQR). The latter two models are more sophisticated extensions of QR, and not yet explored for electricity prices. These models were proposed by Taylor (2008b), with the goal of capturing swift distribution changes. This is an attractive property for EPEX, since the input mix and market conditions change over time. Both EWQR and EWDKQR show promising results in other domains. Thus, we investigate if they offer improvements for electricity price forecasting. We compare the predictive performance of the QR type models to some of the most common benchmarks in literature; parametric GARCH models and CAViaR.

The subject of our analysis is the hourly spot price in Germany, the "Phelix" spot traded on EPEX [1]. We study the German market, as it is the main reference for electricity price trading in Europe. Moreover, we are not aware of previous studies that apply QR for

[1] The term spot price refers to the hourly day-ahead price, following the European convention.

forecasting in this market.

The overall goal of our work is threefold. Firstly, we want to identify appropriate fundamental variables for selected hours and quantiles of the price distribution. Secondly, we assess the gain of using more complex QR models compared to traditional QR. Thirdly, we benchmark the QR models against common VaR approaches in literature. From this, we hope to show that QR is an appropriate choice for capturing the unique features of electricity prices.

The thesis is structured as follows: In Chapter 2 we review relevant literature on fundamental electricity price modelling and VaR forecasting. Next, we describe the German power market and price formation process in Chapter 3. In Chapter 4 we present and analyse the data set. This is followed by Chapter 5, where we give a detailed explanation of the models and how we implement and evaluate them. We present and discuss the empirical results in Chapter 6, and consider the practical implications of our work in Chapter 7. Chapter 8 concludes this thesis, before we recommend further work in Chapter 9.

Chapter 2

Literature review

As motivated in Chapter 1, we aim to forecast the VaR of electricity spot prices using fundamental variables. We therefore position ourselves between the following groups of literature: i) Fundamental electricity price modelling, and ii) VaR forecasting. This thesis contributes to filling the missing link between these literature streams for the German market.

In this chapter, we review research belonging to both of these literature groups. Our focus is on electricity price forecasting, but we also supplement with research from other domains.

2.1 Fundamental electricity price modelling

Fundamental models try to capture price dynamics by modelling the impact of exogenous factors on the electricity price (Weron, 2014). The main motivation for using such models is that characteristic electricity price patterns are results of adaption to fundamentals (Paraschiv et al., 2014). Note that by the term fundamentals we refer to exogenous variables, and that we use these terms interchangeably. In literature, fundamental models are also referred to as structural models. This is in contrast to reduced-form models, which solely rely on intrinsic properties (Weron, 2014).

Chen and Bunn (2010) argue that prices are functions of different drivers in specific trading periods. Bunn et al. (2016) find evidence that while fundamentals have substantial impact on UK spot prices, this impact varies across quantiles and over time. This is backed by Karakatsani and Bunn (2008), who argue that models accounting for the time-varying effects of fundamental drivers are the most effective and useful in practice.

The findings for UK spot prices are also confirmed for the German market. Hagfors et al. (2016a) find that the effect of fundamentals varies substantially across both trading periods and the price distribution. Paraschiv et al. (2016) come to similar conclusions. They emphasise the importance of using fundamentals, and find variables for renewable power particularly influential. Paraschiv et al. also stress that electricity prices' dependence upon fundamental factors is dynamic. This is because evolving factors, like technology, market structure and participant conduct, affect the underlying price formation. They argue that the dynamicity of the price distribution can be captured by modelling the dependence of fundamentals that evolve over time.

Gonzales et al. (2012) find improved accuracy by including fundamentals when forecasting UK spot prices. Moreover, they observe that the variable coefficients in their models evolve remarkably over time. Thus, they argue that dynamic specifications are necessary, and that forecasting models should be re-estimated day by day. They suggest constant monitoring of market conditions in order to select the appropriate model specification and fundamental drivers.

Maciejowska and Weron (2016) find that inclusion of fundamentals generally improves the forecasting performance of UK baseload prices. However, they emphasise that variable selection is crucial. For example, they observe that including gas prices increases forecasting performance, whereas variables related to system-wide demand and CO_2 prices worsen price predictions. The authors conclude that there is no general answer as to which fundamentals are the best, and that the optimal selection depends on both forecasting horizon and trading period.

Hagfors et al. (2016b) show that including fundamental variables is useful because it enables scenario analysis. They create scenarios by changing the values of the fundamentals and evaluating the corresponding price changes. Market participants can use these in risk-management, by planning for a range of price scenarios given different input ranges for the fundamental variables.

Weron (2014) points to several challenges regarding the inclusion of fundamentals in forecasting models. The first is data availability, as data is usually not public. Moreover, Weron finds that it is an open question how to select a minimum set of the most effective input variables. He argues that it is unlikely that one universal set can be found. Despite these challenges, Weron finds that the majority of models in electricity price forecasting literature include fundamentals.

2.2 VaR forecasting

The definition of VaR does not specify how to calculate it. Consequently, many different forecasting methods are proposed in literature (Senera et al., 2012). The best approach in each case depends both on what is being modelled, and on which information the forecast should give.

VaR forecasting is complicated by the fact that most financial assets, e.g. electricity prices, exhibit non-standard statistical properties. According to Hartz et al. (2006), there are particularly two aspects an adequate VaR model must be able to capture: i) Time-varying volatility and volatility clustering, and ii) excess kurtosis relative to the normal distribution.

2.2.1 Classification of VaR forecasting models

Kuester et al. (2006) provide a comprehensive review of VaR prediction strategies. They argue that the approaches for obtaining VaR forecasts can be classified into four main categories:

- *Historical simulation* is a simple approach that computes empirical quantiles based on past data.
- *Extreme value theory* (EVT) models the tails of the distribution separately.
- *Fully parametric models* assume a price/return distribution.
- *Quantile regression* directly models specific quantiles, rather than the whole distribution.

We find examples of models belonging to all four classes in literature on electricity price VaR forecasting. *Historical simulation* is used to a limited extent since it is not able to capture time-varying volatility, and thus, has poor performance in practice (Kuester et al., 2006). Filtered historical simulation (FHS) attempts to overcome this issue by prefiltering the data with location-scale models, such as ARMA and GARCH (see below). Gurrola-Perez and Murphy (2015) evaluate FHS models for energy markets. However, we are not aware of recent applications to electricity prices.

At present, research on *EVT* for estimating VaR in energy markets is sparse. However, examples are found in Bystrom (2005), Chan and Gray (2006) and Florentina and Hadzi-Mishev (2016), who all report that the results are encouraging.

Fully parametric models are often based on models of volatility dynamics. Examples include the RiskMetrics model widely applied in banking (J.P.Morgan/Reuters, 1996), and Generalised Autoregressive Conditional Heteroscedacity (GARCH) models. GARCH models were first introduced by Bollerslev (1986), and build upon the ARCH model of Engle (1982). The idea is to let conditional variance change over time as a function of past variance and error terms. The models are specifically designed to capture volatility clustering, one of the key characteristics of electricity prices. For this reason, they are common benchmarks in electricity price forecasting literature. However, Alexander (2008b) finds that GARCH models are used to a limited degree in practice, because they are hard to calibrate.

The last class from Kuester et al.'s classification is *quantile regression* (QR). This class consists of all approaches that model the specific quantiles of a distribution. This includes the traditional QR model by Koenker and Bassett Jr. (1978), as well as the Conditional Autoregressive Value-at-Risk (CAViaR) models by Engle and Manganelli (2004). In this thesis, we refer to CAViaR as a separate group of models.

The QR model by Koenker and Bassett Jr. (1978) has the advantage of reducing sensitivity to large outlying observations, and is solved as a linear program. The model can easily be extended to include explanatory variables. Since each quantile is modelled separately, it is possible to capture varying effects of factors across the distribution. Moreover, QR has the appeal of avoiding distributional assumptions. This makes it appropriate for modelling time series with complex statistical properties, for instance electricity prices.

CAViaR approaches model quantiles directly, similar to the approach of QR. However, these models specify the evolution of quantiles as autoregressive processes. In the original paper, Engle and Manganelli (2004) consider stock market returns. They argue that since the volatilities of returns cluster over time, their distribution is autocorrelated. Consequently, the quantiles must exhibit similar behaviour. The same argument may be used for electricity prices. This makes CAViaR a popular benchmark in electricity forecasting literature (Bunn et al., 2016).

In addition to the aforementioned categories, there are several hybrid approaches and advanced statistical techniques for VaR forecasting. Kuester et al. (2005) name long memory, Markov-switching and stochastic volatility models as other commonly used techniques. Furthermore, computational intelligence receives increasing attention in all types of forecasting literature. Maciel et al. (2017) apply a fuzzy modelling approach for VaR estimation of the S&P500, and report that it achieves higher performance than alternative econometric models. Mostafa et al. (2017) provide a review of neural network applications for modelling VaR. Examples of computational intelligence techniques applied to electricity price forecasting are found in Keles et al. (2016), who use a neural network approach, and in Neupane et al. (2017), who use an ensemble model. However, none of these forecast VaR or price distributions. We consider computational intelligence models to be interesting subjects for future studies. However, we limit our scope to traditional statistical approaches in this work, since these models currently are more accepted by industry practitioners.

Among the approaches from Kuester et al.'s classification, we find QR models to be particularly promising for forecasting electricity prices. We focus on QR models rather than the alternative approaches, because of their simplicity, possibility to include fundamentals, and promising ability to capture the complex features of electricity prices. Due to the wide use in literature, we use GARCH- and CAViaR models as benchmarks. In the next section, we review literature on using QR-, GARCH- and CAViaR models for forecasting. Moreover, we consider extensions of the QR model. For further description of the models we review in this section, please refer to Kuester et al. (2006) and Alexander (2008d).

2.2.2 Performance of VaR forecasting models in literature

Bunn et al. (2016) forecast the VaR of UK spot prices. They apply three QR models; one with price lags only, one including fundamentals, and one with fundamentals as well as a volatility variable. The QR models are benchmarked against CAViaR models, and GARCH models with both a gaussian- and a skewed student-t distribution. Bunn et al. find that the QR model including fundamentals and volatility outperforms all other models. QR is also favoured due to its ease of implementation. Furthermore, the authors conclude that gaussian GARCH models are *"seriously flawed"* for electricity forecasting. This is unsurprising, as the kurtosis and skewness of the data is lost under the normality assumption. However, they observe significant improvement with a skewed student-t distribution.

The work of Bunn et al. (2016) has several similarities to that of Lundby and Uppheim (2011), who forecast the VaR of Nord Pool spot prices. Lundby and Uppheim implement

QR, EWQR (see Section 2.2.3), and different CAViaR specifications. They also extend one of the CAViaR models to include explanatory variables. GARCH models with gaussian- and skewed student-t distributions serve as benchmarks. The results show that the extended CAViaR model outperforms the rest, and that CAViaR models generally perform well. However, due to the simplicity and relatively good performance of QR, the authors argue that market participants may favour it over the more sophisticated extended CAViaR model.

Garcia et al. (2005) use two GARCH models to forecast spot prices in the Spanish and Californian market; one with price as the only variable, and one including demand. They benchmark these against an ARIMA model. They find that GARCH with price only outperforms ARIMA when volatility and price spikes are present. Moreover, adding demand as explanatory variable further improves the forecasting performance. This version of GARCH consistently outperforms ARIMA. Florentina and Hadzi-Mishev (2016) use a combination of GARCH and EVT to investigate the tails of the German electricity price change distribution. They find that the model delivers relatively precise quantile estimates, but that the quality of the estimates is sensitive to the threshold selected for the tail.

QR models are also applied in domains other than electricity prices. Bremnes (2006) uses a QR model to forecast wind power production, and compares the model to a gaussian model and the Nadaraya–Watson (NW) estimator. He concludes that QR outperforms the gaussian model, but that the differences between QR and the NW estimator are minor. Chen and Chen (2002) apply a QR model to forecast VaR for the returns of the Nikkei 225 index. They find that it outperforms the conventional variance-covariance approach.

2.2.3 Extensions of the QR model

Taylor (2008b) introduces exponentially weighted quantile regression (EWQR). The extension is motivated by the trade-off between including too few observations and getting large sampling errors, and including too many and getting a model that reacts slowly to changes in the true distribution. EWQR attempts to resolve this by placing exponentially decaying weights on the observations, which gives greater emphasis to newer observations. This could be useful for modelling the EPEX spot price, since we observe significant changes in input mix and market structure over time. Taylor uses EWQR to estimate VaR for stock returns and finds that it outperforms GARCH- and CAViaR models. In Taylor (2007), he uses an intercept-only EWQR model to forecast daily supermarket sales. He reports that the empirical results are encouraging, with improvements over traditional QR.

Gelper et al. (2010) argue that despite the simplicity of exponential smoothing methods, they are still competitive with more complicated forecasting models. However, De Livera et al. (2011) warns that in modelling complex time-series, the exponential smoothing models, used by e.g. Taylor, suffer from over-parameterisation.

To the best of our knowledge, EWQR has received little attention in electricity price forecasting literature. The only example we find is in the aforementioned study by Lundby and Uppheim (2011). They report poor results and find that an intercept-only EWQR model is outperformed by both traditional QR and the CAViaR methods.

It is challenging to estimate extreme quantiles due to the sparseness of observations in the tails. This is Taylor's (2008b) motivation for extending the EWQR model further, to exponentially weighted double kernel quantile regression (EWDKQR). The EWDKQR method is based on the paper by Jones and Yu (1998), who argue that double-kernel methods are useful for calculating quantiles. In empirical studies, Taylor finds that EWDKQR performs worse than EWQR in terms of hit percentage. However, the dynamic properties of the quantiles are better explained by the EWDKQR model. To the best of our knowledge, this model has not previously been applied to electricity prices.

2.3 Key findings

We conclude the literature review by summarising the key findings and discussing similarities with our work.

In Section 2.1, we reviewed literature on modelling electricity prices with fundamental variables. We find evidence in literature that different fundamentals affect specific trading periods and quantiles. Moreover, current research suggests that the dependence upon fundamentals varies over time and that forecasting models should be dynamic. While the use of fundamentals is encouraged for improving forecasting performance, selecting the optimal input variables is challenging.

Section 2.2 reviewed different types of VaR forecasts. VaR forecasting approaches can be classified into i) historical simulation, ii) EVT, iii) fully parametric models, and iv) quantile regression. The latter two approaches are the most common for electricity prices. Within these classes, GARCH and CAViaR models are widely applied. Thus, they are appropriate benchmarks. We argue that QR models are promising for forecasting electricity prices, because they model each quantile separately. This enables them to capture the varying effect of fundamentals across the distribution.

We examined research on the predictive performance of QR, GARCH and CAViaR models in Section 2.2.2. QR exhibits good predictive accuracy, and generally outperforms GARCH and traditional CAViaR models for electricity price forecasting. However, QR may be inferior to more sophisticated CAViaR models. Accurate forecasting of extreme quantiles is challenging due to the sparseness of data in the tails. This is a particular problem when the distribution changes rapidly, as is the case for EPEX. EWQR and EWDKQR are extensions of QR intended to account for this, as discussed in Section 2.2.3.

The studies of Bunn et al. (2016) and Lundby and Uppheim (2011) are most similar to what we do in this thesis. However, we analyse the German market, while the aforementioned papers consider the UK- and Nord Pool markets. We also apply more sophisticated QR models, and perform a variable selection across quantiles. We are not aware of any previous studies that consider the predictive ability of fundamentals across the distribution for electricity spot prices. We argue that these distinctions make our work a valuable contribution to current literature.

Chapter 3

Market description

In this chapter, we describe the German electricity market to provide an understanding of price drivers and the price formation process. This serves as motivation for our choice of fundamental variables.

The European Power Exchange (EPEX) is the main trading platform for electricity prices in Europe. It offers trading, clearing and settlement in both the day-ahead- and intraday markets. The day-ahead, hourly prices in Germany are traded on EPEX and are referred to as ”Phelix”. The day-ahead market is the primary market for power trading. Here, buyers and sellers make hourly contracts for delivery of power the following day. This happens through a daily auction at 12.00pm, where the market clearing price is determined by matching demand and supply. The intraday market supplements the day-ahead market and helps secure necessary demand-supply balance.

Seasonal fluctuations, substantial volatility clustering, large spikes and increasing occurrences of negative prices characterise the German electricity price (Reisch and Micklitz, 2006; Bunn et al., 2016; Hagfors et al., 2016c). In the following sections, we describe the market mechanisms that give rise to this complexity.

3.1 Energy mix

Table 3.1 shows the development of the energy mix in Germany from 2010 to 2016. It illustrates that power production in Germany mainly relies on fossil fuel power, particularly coal with 40.3% of the total production in 2016. Moreover, there is a large share of intermittent renewable energy in the form of wind and solar power.

The increase in renewable energies and reduction in nuclear power are the most notable developments during the period. The latter is due to the German government’s decision to phase out nuclear energy within 2022. Regulatory changes are also the key driving force for the growth in renewables, as several subsidies have been introduced to incentivise expansion of renewable production (Federal Minsitry for Economic Affairs and Energy, 2017).

Table 3.1: Electricity production in Germany by source (%)

Source	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Coal	41.6	42.9	44.1	45.2	43.8	42.1	40.3
Nuclear	22.2	17.6	15.8	15.3	15.5	14.2	13.1
Natural gas	14.1	14.1	12.2	10.6	9.7	9.6	12.4
Oil	1.4	1.2	1.2	1.1	0.9	1.0	0.9
Renewable energies:	16.5	20.1	22.6	23.7	25.8	29.0	29.0
Wind	6.0	8.0	8.1	8.1	9.1	12.3	11.9
Solar	1.9	3.2	4.2	4.9	5.7	6.0	5.9
Biomass	4.6	5.2	6.1	6.3	7.7	6.9	7.0
Hydro power	3.3	2.9	3.5	3.6	3.1	2.9	3.2
Waste to energy	0.7	0.8	0.8	0.8	1.0	0.9	0.9
Other	4.2	4.1	4.1	4.1	4.3	4.1	4.2

Data from AG Energibalanzen e.V. (2017) and Clean Energy Wire (2017)

3.2 Price formation fundamentals

As seen in Chapter 2.1, there is significant evidence in literature that electricity prices adapt to fundamentals. To explain how fundamental variables influence the price formation, we follow the approach of Paraschiv et al. (2014) and categorise the variables into *demand*- and *supply* side factors. In addition, we discuss the effect of *endogenous variables*. By this, we refer to the intrinsic properties of the electricity price, for instance price lags and volatility.

3.2.1 Demand side factors

Since electricity is a flow, it is produced and consumed continuously (Bunn et al., 2016). The non-storable nature of electricity entails that a constant balance between supply and demand is necessary to ensure power system stability (Kaminski, 2013). Hence, hourly price variations are largely due to fluctuations in demand.

Demand is a function of temperature, seasonality and consumer patterns, which give rise to the periodic nature of electricity prices (Mirza and Bergland, 2012). As few options are available to consumers in response to price changes, demand is highly inelastic in the short term (Mirza and Bergland, 2012). Hagfors et al. (2016c) find that positive price spikes are closely related to high demand. This is because producers with market power may offer and create market prices substantially above marginal costs in times of scarcity and high demand (Bunn et al., 2016).

3.2.2 Supply side factors

The *merit order curve* plays a vital role in the electricity price formation process (Paraschiv et al., 2014). This is the sorted marginal cost curve of electricity production, starting with the least expensive technologies to the left. Generally, the plants with the lowest marginal costs are the first to be brought online to meet demand. Thus, we can use the merit order curve to determine the price setting technology, i.e. the production technology located at the intersection between supply and demand. The German merit order curve increases through concave, flat and convex regions (Karakatsani and Bunn, 2008).

During periods of low demand, base load power plants, such as nuclear and coal, usually serve as price setting technologies (Erni, 2012). These plants are inflexible, due to high ramp-up costs. Contrary, in times of high demand, prices are set by expensive peak load plants, like gas and oil. These facilities have high flexibility, high marginal costs, and give rise to the convex shape of the merit order curve. With the lowest marginal cost, renewable energy sources are at the bottom of the merit order curve. Increased supply of renewable energy shifts the curve to the right, and thus lowers power prices[1].

Fuel prices and CO_2 emission allowances

Coal is the largest source of electricity in Germany. Hence, coal is a generating technology in the mid-region of the supply function where demand tends to be most of the time.

Gas is characterised by high operational flexibility and short ramp-up times (Sensfu et al., 2008). In 2016, it supplied about 12% of the electricity in Germany. These power plants are price setting during peak hours when demand for electricity is high. Along with coal, gas is also a price setting technology in the mid-region of the supply function (Paraschiv et al., 2016).

With a share of only 0.9% in 2016, oil is rarely used directly for electricity production in Germany. Thus, it has a relatively small impact on the merit order curve. Nevertheless, Paraschiv et al. (2014) and Erni (2012) argue that oil prices influence the electricity prices in Germany, because of the significant impact on the transportation costs of imported coal.

CO_2-producing companies are obliged to buy emission allowances. Since coal-fired power plants and some gas-fired power plants are CO_2 intensive, the price of CO_2 allowances influence their marginal cost (Paraschiv et al., 2014). Thus, the prices of CO_2 allowances affect the spot prices in general. During periods of high prices for emission allowances, a phenomenon called *fuel switch* may occur. This is a change in the merit order curve, where the marginal production costs of more efficient gas-fired power plants become less than those of CO_2 intense coal-fired power plants (Erni, 2012).

Renewables

Among the renewable energy sources in Table 3.1, wind and solar energy have attracted the most attention in Germany over the past years (Erni, 2012; Paraschiv et al., 2014). In 2016, they contributed to 18% of the total production in Germany. The supply of wind and solar

[1] Note that it reduces the *wholesale prices*. Prices for the final consumers may increase because they must pay the feed-in tariffs for the promotion of renewable energy, see Paraschiv et al. (2014)

energy is determined by meteorological conditions and features seasonal patterns (Erni, 2012; Grothe and Schnieders, 2011). A notable observation is that wind infeed tends to be higher in the early morning and the afternoon hours (Erni, 2012; Paraschiv et al., 2014).

Due to intermittency, renewable energy sources pose significant challenges for modern energy markets (EPEX Spot, 2017). Hours with increased supply of renewable energy cause difficulties for inflexible facilities that should run continuously. This is because the inflexible base load facilities have shutdown and start-up costs, forcing them to accept negative marginal returns in order to generate continuously. This has a lowering effect on electricity prices. Hagfors et al. (2016c) find that negative prices largely are caused by high wind production at times when demand is low. Thus, negative price spikes occur mainly at night.

Reserve margin and other supply side factors

Reserve margin is a commonly considered supply side factor in literature. It is defined as available supply minus demand. Bunn et al. (2016) argue that spot prices are sensitive to demand shocks and plant outages, and that expectations of spot prices involve consideration of the reserve margin.

From Table 3.1, we see that the German energy mix also consists of nuclear and other renewable energy sources, such as biomass, hydro power and waste to energy. Nuclear power plants are must-run facilities that never go off-line and have low marginal costs. Thus, they have little impact on electricity prices (Erni, 2012). Moreover, as previously mentioned, Germany is phasing out nuclear energy. The impact of other renewable energy sources is explored to a limited extent in literature. This is because the respective data is unavailable or incomplete.

3.2.3 Endogenous factors

The majority of the reviewed forecasting models include lagged values of spot prices to capture learning effects. Naturally, lagged prices reflect the current price level. Moreover, they influence market agents' price expectations and risk aversion.

Bunn et al. (2016) find that the price signal from the previous day often has a positive effect on the next day's price. This is because market agents tend to reinforce previously successful offers in the market. As a result, high prices are followed by high prices, and positive spikes cluster. Bunn et al. (2016) argue that this effect is particularly strong at higher quantiles where the market becomes less competitive and gaming more plausible. Although lagged prices are commonly found in forecasting models, Paraschiv et al. (2016) claim that autocorrelation of prices often is better reflected in the adaption process to market fundamentals.

Volatility and volatility lags are other commonly considered endogenous factors in literature. Volatility is an indicator of historic instability and risk. Paraschiv et al. (2014) argue that the coefficient of price volatility can be interpreted as a risk compensation in the market. Bunn et al. (2016) find that in times of low prices, an increase in volatility drives prices even lower, and conversely for high prices.

Chapter 4

Data analysis

In this chapter, we describe and analyse our data set. The selection of fundamental variables is based on the factors we found to be important price drivers in Chapter 3. However, we omit some of the renewable energy sources, for instance biomass, due to data availability. Note that we use expected power plant availability (PPA) as a proxy for reserve margin.

We use data from EPEX observed between 01.01.2010 and 31.08.2016. The main reason for this choice is the Equalisation Mechanism Ordinance, which came into force January 1st 2010. This act induced a significant increase in the use of renewable energy and caused large changes in the EPEX input mix (Paraschiv et al., 2014). Moreover, some of the data, like solar and power plant availability, are incomplete or not available for earlier time periods.

The spot price data has hourly resolution, which means that we have 58 440 price observations. However, since each hour is a separate trading period, we treat the price data as 24 independent time series with 2435 data points each. Our data set consists of the fundamental variables shown in Table 4.1. Please refer to Appendix A for a more detailed description of the variables and the data sources.

Variable	**Daily**	**Hourly**
Phelix spot price		X
Coal price	X	
Gas price	X	
Oil price	X	
CO_2 allowance price EU	X	
Expected wind infeed		X
Expected solar (PV) infeed		X
Expected power plant availability (PPA)	X	
Expected demand		X

Table 4.1: Data granularity

In addition to the exogenous variables above, we include seven lags of the price and volatility as variables. We compute the volatility for each hour separately using a skewed-t GARCH(1,1) model. GARCH is a well-known and frequently used model for estimating volatility, as well as being relatively accurate and simple to implement.

4.1 Descriptive statistics

Figure 4.1 shows how the volatility and average prices vary throughout the day. We see that there are large variations in both prices and volatility. The price has a peak in hour 8 in the morning, and reaches its super-peak in hour 19 in the evening. The off-peak, i.e. the period with the lowest prices, occurs in hour 3 at night. We also observe that the extrema of the price graph coincide with periods of high volatility.

Based on the findings in Figure 4.1, we select hours 3, 8 and 19 as the periods we aim to model. We consider these hours most interesting, as they represent the highest and the lowest prices. Hour 8 and 19 are among the hours with highest volatility, while hour 3 is the most volatile hour at night and has a large number of negative price occurrences. Moreover, this selection allows us to analyse intraday variations of the effects of fundamental factors.

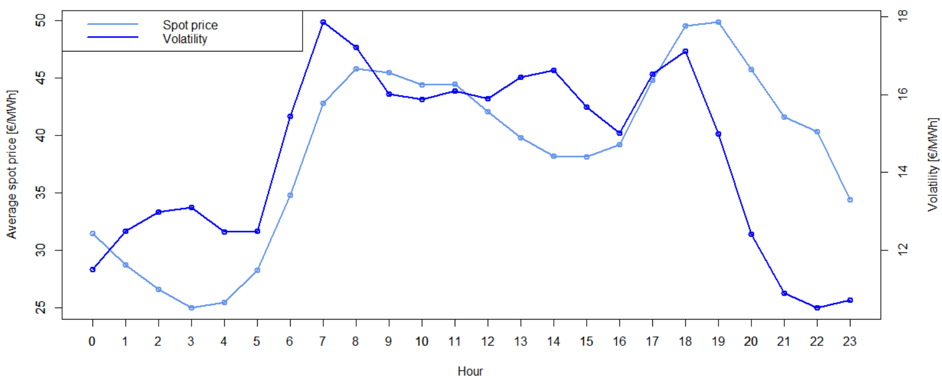

Figure 4.1: Average hourly prices and volatility

Figure 4.2 shows the development of spot prices in the selected hours. We see occurrences of negative price spikes in hour 3, and positive spikes in hours 8 and 19. Moreover, there seems to be volatility clustering in all hours. The period between 2012 and 2013 appears to be particularly volatile, while prices in 2016 are more stable. The seasonality of spot prices is particularly evident in hour 19, where we find price dips in the summer months. We observe some seasonality in hour 8 as well, while prices in hour 3 seem to be less affected by the time of the year. This is unsurprising, as demand at night tends to be relatively stable throughout the year.

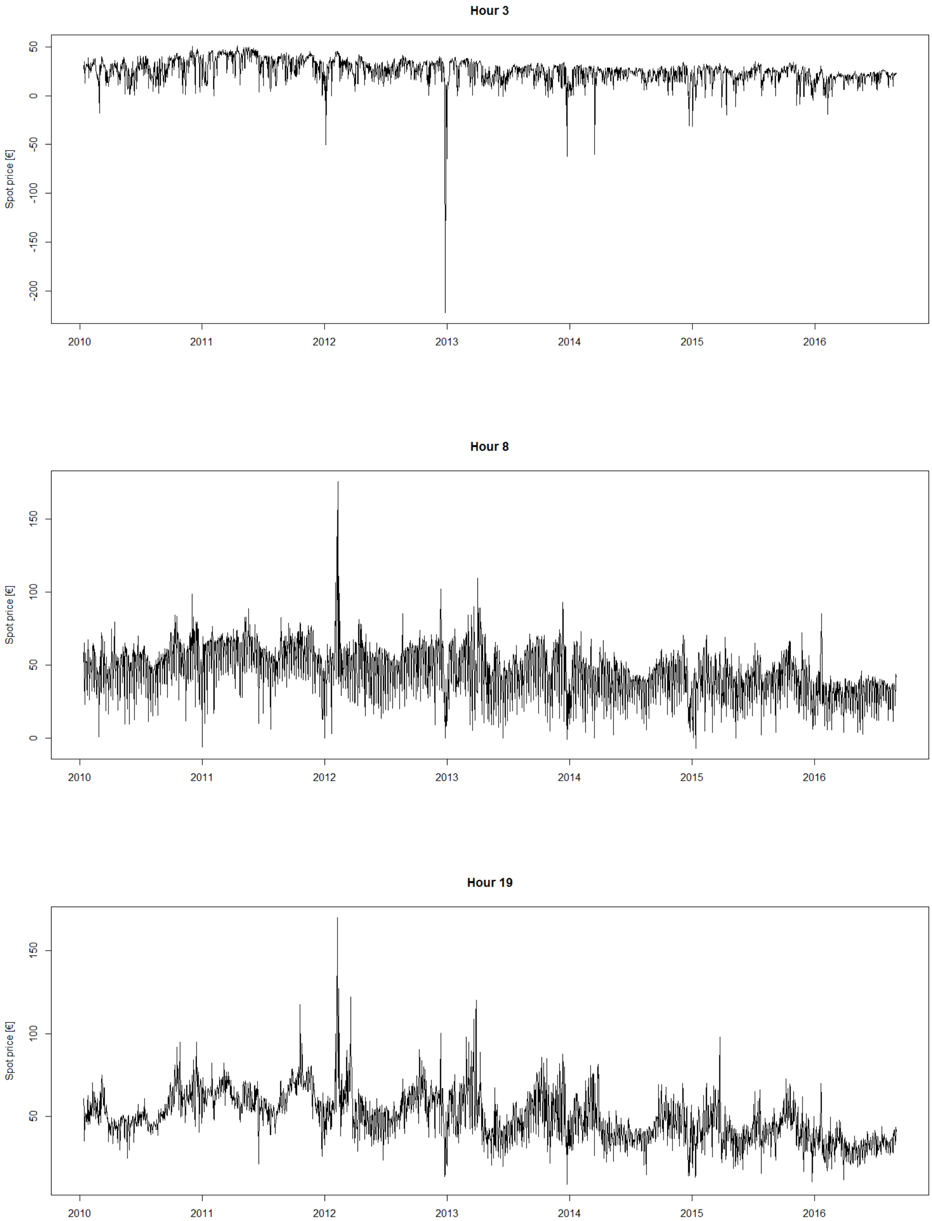

Figure 4.2: Development in hourly spot prices from 2010 to 2016

The descriptive statistics in Table 4.2 reflect these findings. The decreasing mean and median values show that prices overall have fallen from 2010 to 2016. Moreover, we find that the maximum prices have declined. This indicates a lower amplitude of the extreme price spikes in the most recent years. This is also evident from the decreasing standard deviation. From reaching its maximum in 2012, the volatility is declining over time for all hours. Note that we do not have data for the entire 2016, which may explain why we find significantly less volatility this year.

Table 4.2: Descriptive statistics of spot prices, measured in Euros. The tables show the evolution of the price characteristics from 2010 to 2016 for each hour, as well as a total calculated across all years

Hour 3	**Median**	**Mean**	**Min**	**Max**	**Std.div.**	**Skewness**	**Kurtosis**
2010	29.83	27.63	-18.10	50.15	10.60	-0.77	3.47
2011	38.58	34.85	-0.10	51.08	10.71	-1.10	3.47
2012	30.08	26.21	-221.94	45.20	20.99	-8.11	86.69
2013	25.90	23.29	-62.03	39.67	10.78	-2.03	13.98
2014	23.98	21.14	-60.26	34.46	9.00	-3.09	23.16
2015	24.02	21.29	-31.41	34.92	9.21	-1.82	7.73
2016	20.10	18.48	-19.30	30.01	6.27	-1.99	8.95
Total	25.67	25.00	-221.90	51.08	13.10	-5.37	84.24

Hour 8	**Median**	**Mean**	**Min**	**Max**	**Std.div.**	**Skewness**	**Kurtosis**
2010	51.55	50.07	1.06	98.71	14.66	-0.50	3.83
2011	60.63	57.44	-5.95	88.78	13.83	-1.18	5.10
2012	53.24	51.38	-0.09	175.55	19.35	1.15	10.20
2013	46.61	46.71	-0.98	109.36	18.14	-0.10	3.02
2014	41.03	39.65	0.05	72.94	13.81	-0.35	2.82
2015	40.46	38.85	-6.86	71.92	13.88	-0.43	3.05
2016	34.10	31.17	2.59	85.05	10.53	0.03	5.86
Total	46.37	45.76	-6.86	175.60	17.21	0.19	2.02

Hour 19	**Median**	**Mean**	**Min**	**Max**	**Std.div.**	**Skewness**	**Kurtosis**
2010	50.85	53.52	24.76	95.00	10.78	0.92	4.15
2011	62.53	62.37	21.49	117.49	9.79	0.20	6.20
2012	55.00	56.39	13.70	169.90	15.85	1.87	12.79
2013	49.55	51.29	9.28	120.16	15.67	0.85	4.45
2014	42.44	44.20	14.34	81.51	11.56	0.67	3.73
2015	42.11	42.53	10.55	98.05	11.39	0.44	4.40
2016	33.15	33.04	11.79	70.03	6.99	0.75	6.74
Total	48.95	49.85	9.28	169.90	14.99	0.74	2.65

We find that the distribution in hour 3 is highly leptokurtic, with a negative skew. This indicates a fat lower tail and high probability of observing low prices. Hour 8 is slightly skewed, but the skew varies between being negative and positive. We find excess kurtosis in hour 8 for most years. This suggests a greater probability of observing extreme values relative to the normal distribution. Hour 19 has positive skew and excess kurtosis. This implies that the probability of observing prices in the upper tail is greater than in the lower tail. For all hours, we find that 2012 was a particularly volatile year with high kurtosis. This is also visible in in Figure 4.2, where we observe particularly many price spikes in 2012. Overall, we observe swift changes in the price distribution of all hours.

Table 4.3 displays the results from testing the spot prices for stationarity and normality.

The Augmented Dickey Fuller (ADF) test reveals that prices are stationary at a 1% significance level. As mentioned in Chapter 3, electricity prices are known to exhibit mean reversion, and hence, this result is as expected. We perform the test using seven lags, and test for both unit root with drift and unit root with drift and trend. The stationarity of prices entails that we can forecast prices directly without differentiating. Moreover, the Jarque-Bera test rejects normal distribution of prices for all three time series. This is unsurprising, based on the skewness and kurtosis values in Table 4.2.

Table 4.3: Tests of stationarity- and normality of spot prices

	Hour 3	Hour 8	Hour 19
ADF	-10.11	-6.08	-5.45
Jaque-Bera	731610	425	939

Results from the Augmented Dickey Fuller (ADF) test for stationarity with seven lags and unit root with drift, and Jaque-Bera test for normality. The critical values at the 1% significance level are -3.43 and 9.21 for ADF and Jaque-Bera, respectively.

To test for autocorrelation of prices, we use a Ljung-Box test with seven lags. It confirms that autocorrelation is present in all hours, as show in Table 4.4. The autocorrelation is strongest for lag 1 and lag 7, i.e. the prices one and seven days earlier. Hour 19 has the highest degree of autocorrelation overall.

Table 4.4: Autocorrelation of prices

Hour	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	LB(1)	LB(7)
3	0,649	0,431	0,333	0,311	0,318	0,348	0,323	1025	2778
8	0,528	0,243	0,200	0,174	0,185	0,444	0,714	676	2795
19	0,738	0,614	0,529	0,519	0,551	0,629	0,675	1323	6391

The table shows the correlation between the price and previous price lags, as well as the test statistics for a Ljung-Box (LB) test with one and seven lags. The critical value at 1% significance for the Ljung-Box test is 18.48

High correlation with lag 7 indicates that there is a weekday effect in the market. This effect is particularly strong for hours 8 and 19. Interestingly, hour 8 has the lowest autocorrelation of prices overall, but the highest weekday effect. We expect this to be due to the differences in electricity demand on a weekday morning compared to weekends.

Table 4.5 displays descriptive statistics of the fundamental variables. For the variables with hourly resolution, i.e. wind, solar and demand, we find distributional differences across the trading periods. As expected, solar power has close to zero mean in hour 3. Moreover, we see that there is substantially more solar infeed in hour 8 than in hour 19. Due to the significant impact on prices, it is particularly interesting to analyse the differences in demand. Hours 8 and 19 have approximately the same mean demand, whereas the mean demand in hour 3 is considerably lower. However, we note that demand in hour 8 is substantially more volatile. This could be due to the weekday effect we identified in the autocorrelation analysis.

Table 4.5: Descriptive statistics of fundamental variables

Hour 3	Median	Mean	Min	Max	Std.div.	Skewness	Kurtosis
Wind	4491	6067	286	37322	5265	1.9	7.4
Solar	0.0	0.1	0.0	255.0	5.2	49.3	2433
Demand	31078	31219	19127	45071	3821	0.2	3.1
Coal	60.4	64.2	37.6	99.0	14.1	0.3	2.2
Gas	22.1	21.4	11.0	39.5	4.7	-0.3	2.7
Oil	45.3	40.4	15.0	56.7	10.1	-0.6	2.1
Co2	7.2	8.5	2.7	16.8	3.8	0.8	2.3
PPA	55531	55323	40016	64169	4863	-0.2	2.1

Hour 8	Median	Mean	Min	Max	Std.div.	Skewness	Kurtosis
Wind	4075	5875	229	35663	5399	1.8	6.9
Solar	2087	3011	0.0	11665	2849	0.8	2.6
Demand	48673	45193	22783	62594	7800	-0.8	2.3
Coal	60.4	64.2	37.6	99.0	14.1	0.3	2.2
Gas	22.1	21.4	11.0	39.5	4.7	-0.3	2.7
Oil	45.3	40.4	15.0	56.7	10.1	-0.6	2.1
Co2	7.2	8.5	2.7	16.8	3.8	0.8	2.3
PPA	55531	55323	40016	64169	4863	-0.2	2.1

Hour 19	Median	Mean	Min	Max	Std.div.	Skewness	Kurtosis
Wind	4473	6101	270	33522	5225	1.7	6.3
Solar	74.0	736	0.0	4730	1047	1.3	3.5
Demand	45947	45496	30768	60966	5840	-0.3	2.4
Coal	60.4	64.2	37.6	99.0	14.1	0.3	2.2
Gas	22.1	21.4	11.0	39.5	4.7	-0.3	2.7
Oil	45.3	40.4	15.0	56.7	10.1	-0.6	2.1
Co2	7.2	8.5	2.7	16.8	3.8	0.8	2.3
PPA	55531	55323	40016	64169	4863	-0.2	2.1

Note that coal, gas, oil, CO_2 and PPA has a daily data granularity, and therefore show the same numbers for all hours.

Table 4.6 shows the correlation between the spot prices and the explanatory variables. We see that hour 3 has the highest absolute correlation with wind. This correlation is negative, implying that high values of wind tend to coincide with low prices. Hour 8 and 19 are closest correlated with demand and coal, respectively.

Table 4.6: Correlation between spot prices and fundamental variables

Hour	Wind	Solar	Demand	Coal	Gas	Oil	Co2	PPA	GARCH
3	-0,571	-0,003	0,264	0,370	0,151	0,222	0,316	-0,074	-0,249
8	-0,378	-0,224	0,699	0,441	0,300	0,321	0,308	0,132	0,211
19	-0,394	-0,368	0,538	0,553	0,425	0,427	0,336	0,182	0,346

We observe multicollinearity in the data set, i.e. high correlation between the explanatory variables. The correlations are shown in Tables B.3 - B.4 in Appendix B. We find particularly high correlations between the fossil fuels. This is as expected, as the fossils are substitutes in many industries. The multicollinearity entails that models including several fossils should not be used for analysis of the individual coefficients. This is because the presence of multicollinearity means that small changes to the input data can lead to large changes in the model, and even changes in signs of the parameter estimates.

4.1.1 Log-transformation of variables

Logarithmic transformations of both prices and independent variables are commonly applied in electricity price forecasting literature (Maciejowska and Weron (2016) and Bunn et al. (2016)). This is to limit the influence of price spikes and to variance stabilise the data.

Negative prices occur in the German electricity market. Thus, we cannot transform the prices directly. To overcome this problem, we shift each price series by the amount that makes its most negative entry equal to one. This approach has the advantage of not distorting the quantiles. We also observe values below one in the solar data, since there is no sun at night. We resolve this by setting all values below one to one, before transforming the data. As there is negligible solar infeed in hour 3, and the mean values of solar are 3011 in hour 8 and 736 in hour 19, this small adjustment is unlikely to affect the results (see Table 4.5).

Chapter 5

Method

In this chapter, we describe how we implement and test the QR- and benchmark models reviewed in Chapter 2. We start with a detailed explanation of the implementation in Section 5.1. In Section 5.2, we present the model evaluation criteria. Finally, in Section 5.3, we describe our variable selection approach.

We divide the data set described in Chapter 4 into two subsets; an in-sample set and an out-of-sample set. The data is split at 31 August 2014. This leaves 30% of the data for out-of-sample testing to evaluate the predictive performance. We use the in-sample set to fit the model parameter, and the entire data set to perform the variable selection.

5.1 Model implementation

We implement three different QR models; traditional QR, EWQR, and EWDKQR. Additionally, we test some of the most common benchmark models in literature; GARCH(1,1) with skewed student-t distribution, symmetric absolute value CAViaR and asymmetric slope CAViaR.

From the literature review in Chapter 2, we know that electricity price models should account for dynamicity. We do this by applying a rolling window formulation[1], which works as follows: If the window size is 365, observations [1, 365] are used for forecasting the VaR of observation 366. Next, we re-estimate the model with observations $[2, 366]$ and forecast the VaR of observation 367, and so on. We test window sizes of 250, 365, 548, 730 and 913 days.

[1] Alternatively, we could have used an expanding window. However, we believe that a rolling window is a better choice for swift adaption to changing market conditions, since it disregards observations far back in time. Preliminary testing supported this hypothesis. Thus, we make this limitation of scope.

5.1.1 Linear quantile regression

We start with the original QR model by Koenker and Bassett Jr. (1978). This is given by

$$Q_\theta(lnP_{i,t+1}) = \beta^\theta_{i,0} + \sum_{n=1}^{N} \beta^\theta_{i,n} X_{n,t}. \tag{5.1}$$

Here, $\theta \in \{1\%, 5\%, 10\%, 25\%, 50\%, 75\%, 90\%, 95\%, 99\% \}$ denotes the quantile, $i \in [0, 23]$ is the hour, and n indexes the set of explanatory variables which has N elements.

Further, we define $\mathbf{X}_{i,t}$ as the vector of explanatory variables at time t, and $\boldsymbol{\beta}^\theta_i$ as the vector of regression coefficients. The quantiles are found by solving the linear minimisation problem given by

$$\min_{\beta^\theta_i} \sum_{t=1}^{T} \big(lnP_{i,t} - \mathbf{X}_{i,t}\boldsymbol{\beta}^\theta_i\big)\big(\theta - I(lnP_{i,t} \leq \mathbf{X}_{i,t}\boldsymbol{\beta}^\theta_i)\big), \tag{5.2}$$

where $I(\cdot)$ represents the indicator function, returning 0 or 1. We solve the minimisation using the "quantreg" package in R.

5.1.2 Exponentially weighted quantile regression

By adding a weighting parameter λ to Equation 5.2, we get the EWQR model by Taylor (2008b). λ decays exponentially, amounting to simple exponential smoothing of the cumulative distribution function. Thus, the EWQR minimisation has the form

$$\min_{\beta^\theta_i} \sum_{t=1}^{T} \lambda^{T-t}\big(lnP_{i,t} - \mathbf{X}_{i,t}\boldsymbol{\beta}^\theta_i\big)\big(\theta - I(lnP_{i,t} \leq +\mathbf{X}_{i,t}\boldsymbol{\beta}^\theta_i)\big). \tag{5.3}$$

As before, the vector of explanatory variables is given by $\mathbf{X}_{i,t}$, and $\boldsymbol{\beta}^\theta_i$ is the vector of regression coefficients. T is the time of the last observation in the window. The θ-quantiles are given by Equation 5.1. Again, we solve the minimisation using R's "quantreg" package.

The value of λ determines how fast the weights decay. If the distribution changes rapidly, a relatively low value is needed to ensure that the model adapts swiftly. However, larger values may be necessary in the extreme quantiles to give significant weight to a higher number of observations. We follow Taylor's approach to optimise the λ-values. This is done by using a rolling window to produce one step-ahead quantile forecasts for the observations in the in-sample set, and selecting the λ that yields the minimum QR sum. This is the summation in the standard form of QR in Equation 5.2.

Since λ depends on all parts of the model specification, we perform this optimisation for all combinations of hours, quantiles, explanatory variables and window sizes. We test a window of λ-values between 0.9 and 1, with a step size of 0.001.

5.1.3 Exponentially weighted double kernel quantile regression

We expand the EWQR model further to EWDKQR, following the approach of Taylor (2008b). In this model, we replace the observations $lnP_{i,t}$ from Equation 5.3 with a kernel function K_{h_2}.

In this context, the term *kernel* refers to a non-negative function with zero mean that integrates to one. Kernel density estimation (KDE) is perhaps the most important application of these functions. This is a non-parametric approach to estimating the probability density function of a random variable and an alternative to using histograms. The latter is similar to how traditional QR and EWQR works. In EWQR, a rapidly decaying weighting parameter is analogous to using a small number of observations to construct the histogram. When few observations are available, kernels often improve the density estimates from histograms. Thus, Taylor argues that introducing kernels may allow faster decay of the EWQR weighting parameter, and consequently, better adaption to swift distribution changes.

The minimisation problem is formulated as

$$\min_{\beta_i^\theta} \sum_{t=1}^{T} K_{h_1}(x - x_t)\Big(\int_{-\infty}^{\infty} (y - \mathbf{x}_t'\beta_i^\theta)(\theta - I(y < \mathbf{x}_t'\boldsymbol{\beta_i^\theta}))K_{h_2}(y - y_t)dy\Big). \tag{5.4}$$

In this equation, $y = lnPrice_T$ and $y_t = lnPrice_t$. K_{h_1} and K_{h_2} are kernel functions in the x- and y-dimension, respectively. K_{h_1} is selected to be exponentially weighted with $K_{h_1}(x - x_t) = \lambda^{T-t}$ and K_{h_2} to be gaussian. With these kernels defined, Taylor shows that the minimisation becomes

$$\min_{\beta_i^\theta} \sum_{t=1}^{T} \lambda^{T-t}(\theta(lnP_{i,t} - Q_t) + (Q_t - lnP_{i,t})\Theta((Q_t - lnP_{i,t})/h_2) + h_2\phi((Q_t - lnP_{i,t})h_2)). \tag{5.5}$$

Here, Θ is the standard gaussian cdf and ϕ is the gaussian pdf. $Q_t^\theta = \mathbf{x}_t'\boldsymbol{\beta_i^\theta}$, as specified in Equation 5.1. Further, h_2 is the bandwidth of kernel K_{h2}. It determines the degree of smoothing in the kernel density estimate. Taylor explains that h_2 and λ have a negative relationship. Lower values of λ entail faster exponential decay and that less historical information is captured. Thus, there is a need for more kernel smoothing, which we achieve through larger values of h_2. Note that EWDKQR with h_2 set to zero corresponds to EWQR.

We estimate the optimal bandwidth using the same approach as for the weighting parameter in EWQR. We test values for h_2 between 0 and 0.045, with a step size of 0.005. Due to the computational complexity of this procedure, the λ values are kept fixed during the optimisation. We use the same λ estimates as in EWQR.

To perform the minimisation, we use the ”nlm” nonlinear optimisation solver in R .

5.1.4 Fully parametric GARCH models

In Section 2.2.2, we found GARCH with a skewed student-t distribution to perform significantly better than gaussian GARCH. This is the reason why we only implement skewed student-t GARCH.

We follow the approach of Bunn et al. (2016), and assume a model where the conditional mean, μ_t, is a linear function of exogenous variables, and the conditional volatility follows a GARCH(1,1) process. This gives

$$lnP_t = \mathbf{X}_{i,t}\boldsymbol{\beta}_i + \sigma_t z_t \tag{5.6}$$

where $\mathbf{X}_{i,t}$ is as defined in Section 5.1.1, and σ_t is given by

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1\sigma_{t-1}^2 + \alpha_2\varepsilon_{t-1}^2}. \tag{5.7}$$

Further, z_t is an i.i.d random variable, assumed to have a skewed student-t distribution. Thus, we get the one-step forecast of the θ-quantile of lnP_t, conditional on the information up to time t, from

$$Q_\theta(lnP_{t+1}) = \mathbf{X}_{i,t}\boldsymbol{\beta}_i + \hat{\sigma}_t Q_\theta(z_t). \tag{5.8}$$

$Q_\theta(z_t)$ is the θ-quantile of z_t. We can compute this because of the distributional assumption. The skew- and shape parameters of the distribution are re-estimated in each window.

We estimate this model in two steps. First, we estimate the β-parameters in Equation 5.6 using OLS. Then, we fit a GARCH(1,1) model to the residuals from the regression using the "fGarch" package in R.

Note that with this approach, the regression coefficients $\boldsymbol{\beta}_i$ are not specific to the quantiles. Instead, the explanatory variables model the mean of the distribution. This is a limitation, because it means that the variables have the same effect across the distribution.

5.1.5 Conditional autoregressive Value-at-Risk models

The CAViaR models by Engle and Manganelli (2004) specify the evolution of a quantile over time as an autoregressive process. They derive expressions for four different CAViaR processes; symmetric absolute value, asymmetric slope, adaptive, and indirect GARCH(1,1). We use the first two as benchmarks, as these outperformed the others in preliminary empirical testing. $Q_t(\theta)$ is the θ-quantile at time t, the residual term y_t is the return at time t, and α_i are parameters to be decided. $I(\cdot)$ represents the indicator function, returning values 0 or 1.

Symmetric absolute value CAViaR is given by

$$Q_t(\theta) = \alpha_1 + \alpha_2 Q_{t-1}(\theta) + \alpha_3|y_{t-1}|, \tag{5.9}$$

and asymmetric slope CAViaR is defined as

$$Q_t(\theta) = \alpha_1 + \alpha_2 Q_{t-1}(\theta) + \alpha_3 max(y_{t-1}, 0) - \alpha_4 min(y_{t-1}, 0). \tag{5.10}$$

Both models are mean-reverting in the sense that the coefficient of the lagged quantile estimate, Q_{t-1}, can take any value. However, they differ in how they respond to past returns. While the first responds symmetrically, the second may respond differently to positive and negative past returns.

Following the approach of Jeon and Taylor (2013), we estimate the parameters of the CAViaR models by solving the minimisation problem

$$\min_{\alpha}\Big[\sum_{t|y_t>Q_t}\theta \mid y_t - Q_t(\theta)\mid \sum_{t|y_t<Q_t}(1-\theta)\mid y_t - Q_t(\theta)\mid\Big]. \tag{5.11}$$

This is equivalent to minimising the objective function in the quantile regression model of Koenker and Bassett Jr. (1978).

As standard CAViaR specifications do not include exogenous variables, we follow the approach of Bunn et al. (2016). They formulate a linear regression model for the mean of the distribution on the form

$$lnP_t = \mathbf{X}_{i,t}\boldsymbol{\beta}_i + \varepsilon_t, \tag{5.12}$$

where the beta coefficients are found using OLS. A CAViaR model can then be fitted on the residuals. This means that we replace the return terms in in Equations 5.9 and 5.10 with the error term from Equation 5.12, ε_t. Thus, we define the quantiles of the price distribution as

$$Q_\theta(lnP_{t+1}) = \mathbf{X}_{i,t}\boldsymbol{\beta}_i + Q_\theta(\varepsilon_t). \tag{5.13}$$

This approach has the same limitation as the GARCH models in that the beta-coefficients are not specific to the quantiles.

5.2 Out-of-sample performance analysis

To test the predictive performance of the models, we use Kupiec's unconditional coverage (UC) test (1995), Christoffersen's conditional coverage (CC) test (1998), and the dynamic conditional quantile (DQ) tests by Engle and Manganelli (2004).

The simplest way to test a quantile model is to find the percentage of observations falling below the estimated quantile. Ideally, this percentage should be close to θ. The *UC test* of Kupiec (1995) is based on this idea. In our context, we let $\{I_t\}_{t=1}^{T}$ be a sequence of i.i.d Bernoulli variables, i.e.

$$I_t = \begin{cases} 1, & \text{if } Y_t < Q_t(\theta) \\ 0, & \text{if } Y_t \geqslant Q_t(\theta) \end{cases}, \tag{5.14}$$

where Y_t is the observed price and Q_t is the predicted quantile at time *t*.

The UC test is a test of the null hypothesis that the indicator sequence has an exceedance percentage sufficiently close to the quantile. The test statistic is a likelihood ratio statistic given by

$$LR_{uc} = -2\log\left(\frac{\pi_{exp}^{n_1}(1-\pi_{exp})^{n_0}}{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}\right) \overset{asy}{\sim} X_1^2. \tag{5.15}$$

Here, n_1 and n_0 are the number of exceedances and non-exceedances, respectively. π_{exp} is the expected proportion of exceedances, and $\pi_{obs} = n_1/(n_0 + n_1)$ is the observed proportion of exceedances. Clustering of violations can invalidate a VaR model, indicating that it is not sufficiently responsive to changing market conditions (Alexander, 2008d). A weakness of the UC test is therefore that it only counts the number of hits, ignoring clustering of exceedances.

The *CC test* extends the UC test by also specifying that the hit sequence should be independent over time. This is a combined test for both coverage and independence. The test statistic is given by

$$LR_{cc} = -2\log\left(\frac{\pi_{exp}^{n_1}(1-\pi_{exp})^{n_0}}{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}\right) \overset{asy}{\sim} X_2^2, \tag{5.16}$$

where n_{ij} is defined as the number of times an observation with indicator value i is followed by indicator value j. For example, n_{01} is the number of times a non-exceedance is followed by an exceedance. Further, we define $\pi_{01} = n_{01}/(n_{00} + n_{01})$ and $\pi_{11} = n_{11}/(n_{11} + n_{10})$. As before, n_1 and n_0 represent the number of exceedances and non-exceedances, respectively. The CC test only considers autocorrelation of order one in the hit sequence, i.e. whether an exceedance today affects the probability of an exceedance tomorrow. This means that the test has limited power to detect general patterns of clustering.

Some VaR models are misspecified because they do not utilise all the information available in the market. Since the UC- and CC tests only use information on past quantile violations, they do not have the power to detect such misspecifications (Engle and Manganelli, 2004). Arguing that the conditional probability of an exceedance also depends on the quantile estimate itself, Engle and Manganelli (2004) introduce the dynamic conditional quantile (DQ) tests. These tests are based on a linear regression model that links the exceedances to a set of explanatory variables. The first test, DQ1, is formulated as

$$I_t = \beta_0 + \sum_{k=1}^{7}\beta_k I_{t-k} + \epsilon_t, \tag{5.17}$$

where I_t is the hit sequence defined by Equation (5.14). We choose to test seven lags based on our findings in the data analysis in Chapter 4. The second test, DQ2, is an extension of DQ1, where the quantile estimate itself, Q_t, is included. This test is formulated as

$$I_t = \beta_0 + \sum_{k=1}^{7}\beta_k I_{t-k} + \beta_8 Q_t + \epsilon_t. \tag{5.18}$$

We use a standard F-test to test the null hypothesis that all coefficients in Equation (5.17) are equal to zero. Here, H_0: $\beta_i = 0$, where $i \in \{1..7\}$, is tested against the alternative hypothesis that at least one of the coefficients is different from zero. This provides insight into clustering. Further, we use a t-test to test the null hypothesis that the individual coefficient, β_8, in Equation (5.18) is equal to zero, H_0: $\beta_8 = 0$. In other words, given that there may be clustering, we test whether the quantile exceedances are linked to the scale of the quantile forecasts.

5.3 Variable selection

Variable selection is a crucial step in building a good prediction model (e.g. Diebold (2015)). For distribution forecasting this is a complex process, and the standard goodness-of-fit tests are not sufficient. In Chapter 2, we saw evidence that fundamental variables affect specific hours and quantiles differently. It is therefore necessary to perform variable selections for all combinations of hours and quantiles, to take full advantage of modelling each quantile separately.

In Chapter 3, we found that there are significant changes in the EPEX energy mix from 2010 to 2016. Thus, we argue that it is of greater interest to perform variable selection using the entire data set, rather than just the in-sample set. This is to catch the latest distributional changes. However, we recognise that this results in some optimism in the overall performance of our models.

We emphasise that we are searching for a model with high *predictive power*, rather than *explanatory* or *descriptive power*. Explanatory and descriptive models aim to test causal hypotheses and describe how a set of factors X affect a factor Y. Predictive models, on the other hand, anticipate future observations. The choice between these objectives often results in different models (Shmueli, 2010). To achieve high predictive power, variable selection should be based on the quality of the *association* between predictors and responses, rather than causal relationships. In fact, prediction accuracy is often improved by removing statistically significant variables. This is because removing variables reduces estimation variance, a gain that may outweigh the bias introduced by omitting them (Shmueli, 2010).

Diebold (2015) argues that overfitting is one of the greatest pitfalls when building prediction models. Thus, variable selection methods must reward *parsimony* and validation procedures should use *out-of-sample* metrics. Commonly used criteria include the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC). These both measure in-sample fit, and introduce a penalty term for the number of parameters in the model. For distribution forecasting models, out-of-sample validation tests include the aforementioned UC-, CC- and DQ tests.

Variable selection for the distribution quantiles

We apply SIC to select variables for the quantiles, as it is commonly used in this context (Schwarz et al., 1978; Koenker et al., 1994; Lee et al., 2014) [2]. We use a version adapted to QR by Behl et al. (2014), given by

$$SIC_{QR,\theta,i} = ln\left(\sum_{t=1}^{T}\left(lnP_{i,t} - Q_i^{\theta}\right)\left(\theta - I(lnP_{i,t} \leq Q_i^{\theta})\right)\right) + \frac{K \times ln(T)}{2}, \quad (5.19)$$

where K is the number of variables (including the intercept) and T is the number of observations. Note that the first term is the objective function value of QR, as in Equation 5.2. Since we use a rolling window, we let this be the sum of objective function values for all windows.

[2] We could have applied more advanced methods, like lasso, adaptive lasso and elastic net. As variable selection procedures are not our main focus, we leave this for future work.

We perform the variable selection using a traditional QR model, with a rolling window of 2 years [3]. For each hour and quantile, we choose the combination of variables that yields the lowest SIC score while passing the out-of-sample tests. This means that we let parsimony and in-sample performance decide in cases where more than one model pass the out-of-sample tests. We perform two variable selections; one where the models must pass the UC test, and one where they are restricted to pass both the UC- and CC tests.

With 24 variables to choose from, a brute force approach testing all variable combinations amounts to performing $2^{24}–1$ tests. To reduce the number of tests, we therefore define classes of variables. Within each class, we create subsets of variables and restrict the tests to include at most one subset from each class. For further description of the combinations we test, please refer to Appendix C.

Note that we determine the variable combinations without considering multicollinearity, since it does not affect the models' predictive performance (Shmueli, 2010). However, it means the information value of the individual coefficients is limited, as discussed in Chapter 4.

Variable selection for the distribution mean

As previously discussed, the explanatory variables in the GARCH and CAViaR methods model the mean, not the specific quantiles. Consequently, the optimal variables for QR may be inappropriate choices for these models. Thus, we also perform a variable selection to find the variables that best model the mean of the price distribution.

We follow the recommendation of Diebold (2015), and apply SIC rather than AIC. We use the standard version (Schwarz et al., 1978) given by

$$SIC = -2lnL + KlnT. \tag{5.20}$$

Here, lnL is the log-likelihood, K is the number of variables (including the intercept) and T is the number of observations. Since we use a rolling window, the overall log-likelihood of the model is given by the sum of log-likelihoods for all windows.

We proceed by fitting a linear regression model for each of the variable combinations, with a window size of 2 years. We choose the set of variables that yields the lowest SIC, and emphasise that this is a simplified approach. Finding the optimal model for the distribution mean is beyond the scope of our work. Still, we consider it interesting to analyse the performance of the GARCH and CAViaR model with variables chosen this way.

[3]The choice of window size is not trivial. Our choice is based on the window size used by Bunn et al. (2016). We address the issue further in the model evaluation.

Chapter 6

Results and discussion

In Chapter 1, we identified the three main goals of this thesis. In this chapter, we present and evaluate empirical results to fulfil these goals.

Our first goal is to identify appropriate fundamental variables for selected hours and quantiles. We do this in Section 6.1. In Section 6.2, we address the second and third goals; we evaluate the predictive performance of our models, and assess the gain of using more complex QR models. Moreover, we benchmark these models against common VaR models in literature. In Section 6.3, we analyse the impact of the variable selection on the models' predictive performance. Finally, we summarise our main findings in Section 6.4.

6.1 Variable selection

Tables 6.1 and 6.2 summarise the results of the variable selection. Table 6.1 shows the variables selected using the SIC and the UC test. Table 6.2 displays the results when the models are also restricted to pass the CC test. From here on, we refer to these variable combinations as VarGroup1 and VarGroup2, respectively.

Overall, we find that $Coal$ and $P.lag1$ have the highest predictive power. This is unsurprising, since we know that there is a strong autocorrelation of prices and that coal is by far the largest source of energy production in Germany.

Further, the results demonstrate that the variables with the highest predictive ability vary for different hours and parts of the price distribution. This is particularly evident in Table 6.1. Here, we find that $Wind$ is the best predictor for extremely low prices in hour 3, and that $Demand$ should be used for predicting the highest peaks in hours 8 and 19. These findings are well aligned with the market characteristics described in Chapter 3. In times of scarcity and high demand, producers can set asking prices well above marginal costs, and thereby contribute to price jumps. Contrary, negative price spikes that occur at night are to a large extent caused by high wind production at times when demand is low.

Table 6.1: Results of variable selection based on Schwarz Information Criterion (SIC) and the unconditional coverage (UC) test, referred to as VarGroup1

Quantile	Hour 3	Hour 8	Hour 19
Q0.01	Wind	V.lag1	Coal
Q0.05	P.lag1	P.lag1	Coal
Q0.10	P.lag1	Coal	Coal, Demand
Q0.25	P.lag1	Coal, Demand	Coal
Q0.50	Coal, Demand	Coal, Gas	Coal, Gas
Q0.75	Coal, Gas	Coal, Gas	Coal
Q0.90	Coal, V.lag1	Coal	Coal
Q0.95	Coal	Coal	P.lag1
Q0.99	Coal	Demand	Demand

P.lag1 is the first lag of the spot price, V.lag1 is the first lag of volatility.

We observe that $Coal$ and Gas are good predictors for the quantiles in the middle of the price distribution for all hours. This also coincides with the discussion in Chapter 3. Here, we found that coal and gas are price-setting technologies in the mid-region of the supply function. Another interesting observation is that that the majority of the selected sets include exogenous variables.

Table 6.2: Results of variable selection based on Schwarz Information Criterion (SIC), the unconditional coverage (UC)- and conditional coverage (CC) test, referred to as VarGroup2

Quantile	Hour 3	Hour 8	Hour 19
Q0.01	Demand, PPA	V.lag1	P.lag1, Avg.P.lag2-7, Coal
Q0.05	P.lag1, Avg.P.lag2-7	P.lag1	P.lag1, Avg.P.lag2-7, Coal
Q0.10	P.lag1, Avg.P.lag2-7	P.lag1, Avg.P.lag2-7	P.lag1, Avg.P.lag2-7, Coal
Q0.25	P.lag1, Avg.P.lag2-7, V.lag1	Coal, Demand*	P.lag1, Avg.P.lag2-7, Coal
Q0.50	Coal, Demand*	Coal, Gas*	P.lag1, P.lag2, P.lag3, P.lag4, P.lag5, P.lag6, P.lag7, Coal, Co2
Q0.75	Coal, Gas*	Coal, Gas*	P.lag1, P.lag2, P.lag3, P.lag4, P.lag5, P.lag6, P.lag7, Coal, Co2, PPA
Q0.90	P.lag1, P.lag2, P.lag3, P.lag4, P.lag5, P.lag6, P.lag7, Coal, Gas, Demand	Coal*	P.lag1, Avg.P.lag2-7, V.lag1, PPA
Q0.95	P.lag1, Avg.P.lag2-7, V.lag1, Avg.V.lag2-7, Coal, Oil, Demand	Coal*	V.lag1
Q0.99	P.lag1, Avg.P.lag2-7, Coal, Demand	V.lag1, PPA	V.lag1

Combinations marked with * do not pass the CC test. These are selected in scenarios where no variable combinations pass the CC test. P.lagX is the Xth lag of the spot price, V.lag1 is the first lag of volatility, Avg.P.lag2-7 is the average of price lags 2 to 7.

There are significant differences between VarGroup1 and VarGroup2. Firstly, from Table 6.2 we see that more variables are necessary to pass the CC test. As previously mentioned, this test is stricter than UC because it also considers clustering of exceedances. Secondly, price- and volatility lags are used to a greater extent in VarGroup2, indicating that they help capturing the dynamicity of a distribution. Moreover, the high autocorrelation of hours 3 and 19 explains the wide use of price lags (see Chapter 4). Thirdly, we note that also the *exogenous* variables differ between VarGroup1 and VarGroup2.

In VarGroup2, extremely low prices in hour 3 can be predicted using $Demand$ and PPA. This is probably because low prices occur when demand is low and the reserve margins are high. The peaks in hours 8 and 19 are best captured by $V.lag1$ and PPA. This supports the findings in the market description; when prices are high, high volatility drives prices even higher. We emphasise that our model is not designed to reveal causal relationships. However, these results indicate that there is a strong association between certain fundamental variables and extreme events.

We perform the variable selection using the simplest model, traditional QR. Models that are more sophisticated may be able to capture dynamic properties with fewer variables. Thus, we use both VarGroup1 and VarGroup2 in the model evaluation.

For predicting the mean, we find [$P.lag1$, $Avg.P.lag2-7$, $Coal$, $Co2$, $Demand$, $Solar$, $Wind$] to be the best variable combination for all three hours. The inclusion of price lags is unsurprising, as Table 4.4 shows high autocorrelation of prices. From Table 4.6, we see that the included variables have high correlations with the spot prices, in particular coal and demand for hours 8 and 19, and wind for hour 3.

6.2 Model evaluation

In this section, we assess the predictive performance of the models presented in Section 5.1. As evaluation criteria, we use the UC-,CC- and DQ tests from Section 5.2. Tables 6.3, 6.4, and 6.5 show the detailed results for hours 3, 8, and 19, respectively. Here, we only present each model's best run, meaning that we select one window size and one variable combination per hour. For results beyond this, please refer to Appendix D.

It is difficult to draw general conclusions as to which model is the "best" based on our results. The model with the highest predictive performance varies across both the distribution and the trading periods. Moreover, the four evaluation criteria favour different models in many cases. Thus, both application and preferences must be accounted for when selecting the most appropriate model. Unless otherwise noted, we refer to the total number of rejected tests when we rate one model as better than another. We place particular focus on the UC test, as we argue that there is limited value in a model with independent exceedances if the hit percentage is wrong.

We start by presenting high-level results and overall performance in Section 6.2.1. Next, we go more in-depth into the performance in each hour in Section 6.2.2, and across the distribution in Section 6.2.3.

Table 6.3: Predictive performance in hour 3

Model	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
QR	0.01	1.09E-02	8.01E-01	0.00E+00	1.86E-01	2.61E-02
	0.05	4.92E-02	9.25E-01	3.07E-01	1.32E-01	1.65E-01
	0.10	8.48E-02	1.61E-01	2.73E-01	6.11E-01	6.07E-01
	0.25	2.42E-01	6.22E-01	1.08E-01	7.81E-03	1.21E-01
	0.50	5.62E-01	7.51E-04	0.00E+00	5.54E-31	3.38E-02
	0.75	7.44E-01	7.17E-01	0.00E+00	2.46E-38	2.16E-01
	0.90	8.88E-01	2.81E-01	3.61E-02	5.28E-04	3.75E-01
	0.95	9.51E-01	9.25E-01	6.71E-01	1.56E-04	9.64E-01
	0.99	9.84E-01	1.11E-01	0.00E+00	1.90E-04	2.83E-02
# Rejections			1	5	6	3
EWQR	0.01	1.37E-02	3.44E-01	0.00E+00	4.95E-01	9.23E-02
	0.05	5.06E-02	9.39E-01	3.52E-01	1.85E-01	2.70E-01
	0.10	9.44E-02	6.10E-01	7.24E-01	7.26E-01	7.97E-01
	0.25	2.46E-01	8.14E-01	8.27E-01	1.41E-02	1.50E-01
	0.50	4.77E-01	2.22E-01	0.00E+00	1.14E-23	8.24E-03
	0.75	6.87E-01	1.17E-04	0.00E+00	5.96E-27	7.97E-03
	0.90	8.82E-01	1.21E-01	4.32E-02	2.70E-02	1.22E-01
	0.95	9.51E-01	9.25E-01	6.71E-01	1.56E-04	9.64E-01
	0.99	9.82E-01	5.66E-02	7.74E-02	1.60E-01	4.90E-01
# Rejections			1	4	5	2
EWDKQR	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	7.49E-02
	0.05	3.97E-02	1.84E-01	3.13E-01	8.41E-01	8.72E-01
	0.10	8.62E-02	2.03E-01	4.30E-01	4.58E-01	4.73E-01
	0.25	2.37E-01	4.02E-01	0.00E+00	3.98E-43	2.79E-10
	0.50	4.90E-01	5.79E-01	4.45E-02	3.44E-02	3.83E-08
	0.75	7.35E-01	3.40E-01	0.00E+00	6.60E-85	1.48E-26
	0.90	9.53E-01	9.73E-08	1.53E-07	1.84E-01	4.63E-18
	0.95	9.53E-01	6.62E-01	4.51E-11	2.12E-25	2.17E-12
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	2.29E-07
# Rejections			1	7	4	6

Model	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
GARCH	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	1.80E-01
	0.05	6.29E-02	1.22E-01	3.03E-01	3.54E-01	5.66E-01
	0.10	1.07E-01	5.50E-01	5.09E-01	6.24E-01	9.45E-01
	0.25	2.38E-01	4.52E-01	7.26E-04	1.49E-05	4.28E-02
	0.50	8.81E-01	0.00E+00	0.00E+00	7.93E-13	1.46E-02
	0.75	6.13E-01	3.33E-16	0.00E+00	3.60E-16	2.03E-06
	0.90	8.84E-01	1.52E-01	1.18E-01	3.11E-01	2.29E-09
	0.95	9.49E-01	9.39E-01	7.62E-01	6.04E-02	1.37E-06
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	2.07E-02
# Rejections			2	5	3	6
Asym. slope CAViaR	0.01	1.37E-02	3.44E-01	1.74E-01	9.98E-01	3.13E-01
	0.05	5.75E-02	3.66E-01	3.86E-01	8.26E-01	2.71E-01
	0.10	9.44E-02	6.10E-01	1.32E-02	1.24E-02	2.59E-01
	0.25	2.28E-01	1.74E-01	2.20E-04	7.02E-05	9.86E-02
	0.50	5.27E-01	1.49E-01	7.05E-09	6.33E-14	1.72E-07
	0.75	7.55E-01	7.48E-01	9.26E-04	2.97E-11	3.32E-06
	0.90	9.21E-01	5.43E-02	5.29E-05	5.49E-05	3.99E-09
	0.95	9.67E-01	2.34E-02	1.79E-03	1.30E-03	5.79E-01
	0.99	9.86E-01	3.44E-01	0.00E+00	1.21E-05	2.64E-01
# Rejections			1	7	7	3
Sym. abs. value CAViaR	0.01	1.09E-02	8.01E-01	1.89E-01	2.98E-03	2.26E-01
	0.05	5.61E-02	4.58E-01	1.77E-01	1.26E-01	1.20E-01
	0.10	1.11E-01	3.38E-01	4.76E-03	3.11E-02	9.41E-01
	0.25	1.44E-01	1.82E-12	0.00E+00	1.09E-18	2.30E-02
	0.50	5.40E-01	2.90E-02	0.00E+00	2.47E-39	1.59E-04
	0.75	6.80E-01	2.05E-05	0.00E+00	6.18E-43	4.10E-04
	0.90	8.96E-01	7.22E-01	4.45E-01	5.17E-01	4.05E-09
	0.95	9.52E-01	7.91E-01	8.11E-01	9.57E-01	2.12E-10
	0.99	9.88E-01	5.44E-01	0.00E+00	9.97E-01	3.06E-11
# Rejections			3	5	5	6

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration. Asymmetric slope CAViaR uses VarGroup1, while the rest use VarGroup2. The models use the following window sizes: QR 730, EWQR 730, EWDKQR 365, GARCH 365, Asym. slope CAViaR 730, and Sym. abs. value CAViaR 365.

Table 6.4: Predictive performance in hour 8

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
QR	0.01	9.58E-03	9.08E-01	0.00E+00	9.99E-01	2.73E-01
	0.05	3.15E-02	1.39E-02	4.60E-02	6.06E-01	9.98E-01
	0.10	7.52E-02	2.01E-02	6.10E-02	1.96E-02	6.43E-01
	0.25	2.54E-01	7.82E-01	6.66E-16	9.75E-20	2.52E-02
	0.50	4.90E-01	5.79E-01	1.37E-14	6.91E-49	1.66E-01
	0.75	7.15E-01	3.35E-02	3.26E-13	6.64E-23	5.32E-02
	0.90	9.18E-01	9.66E-02	6.16E-13	1.30E-28	1.41E-04
	0.95	9.49E-01	9.39E-01	2.33E-15	1.11E-42	1.86E-05
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	4.56E-01
# Rejections			3	8	6	3
EWQR	0.01	1.23E-02	5.44E-01	0.00E+00	9.97E-01	3.30E-01
	0.05	4.79E-02	7.91E-01	9.35E-01	6.77E-01	6.50E-01
	0.10	9.85E-02	8.92E-01	3.08E-01	9.48E-04	1.23E-01
	0.25	2.60E-01	5.38E-01	8.48E-10	4.87E-10	2.86E-01
	0.50	4.66E-01	6.98E-02	5.07E-12	6.05E-52	7.33E-03
	0.75	7.10E-01	1.40E-02	3.72E-10	2.90E-15	1.28E-03
	0.90	9.12E-01	2.53E-01	2.63E-12	7.48E-21	4.84E-01
	0.95	9.44E-01	4.58E-01	7.71E-11	7.32E-24	1.44E-01
	0.99	9.86E-01	3.44E-01	0.00E+00	7.70E-02	1.34E-04
# Rejections			1	7	6	3
EWDKQR	0.01	1.50E-02	2.02E-01	1.37E-02	9.01E-08	4.42E-01
	0.05	5.20E-02	8.07E-01	7.50E-01	2.35E-04	9.10E-01
	0.10	9.99E-02	9.90E-01	1.65E-06	3.50E-07	8.85E-01
	0.25	2.50E-01	9.83E-01	1.40E-02	1.11E-59	1.47E-01
	0.50	5.06E-01	7.39E-01	0.00E+00	4.43E-29	4.71E-01
	0.75	7.47E-01	8.48E-01	6.50E-07	5.43E-08	5.04E-08
	0.90	9.06E-01	6.10E-01	2.55E-15	3.80E-28	3.67E-01
	0.95	9.53E-01	6.62E-01	3.49E-06	2.82E-17	4.76E-11
	0.99	9.88E-01	5.44E-01	2.05E-01	7.89E-04	9.62E-01
# Rejections			0	7	9	2

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
GARCH	0.01	1.37E-02	3.44E-01	0.00E+00	5.08E-01	1.01E-03
	0.05	3.15E-02	1.39E-02	4.60E-02	2.53E-02	1.52E-03
	0.10	7.66E-02	2.85E-02	8.16E-03	1.21E-01	1.25E-01
	0.25	1.86E-01	3.72E-05	4.86E-07	1.08E-02	8.81E-02
	0.50	6.05E-01	1.35E-08	1.66E-11	2.79E-06	2.22E-08
	0.75	7.28E-01	1.69E-01	6.20E-06	2.20E-06	2.17E-11
	0.90	8.77E-01	4.35E-02	1.10E-02	1.03E-01	4.04E-12
	0.95	9.44E-01	4.58E-01	4.21E-01	1.56E-01	5.50E-10
	0.99	9.89E-01	8.01E-01	0.00E+00	1.38E-09	4.36E-05
# Rejections			5	8	5	7
Asym. slope CAViaR	0.01	3.83E-02	4.43E-09	2.36E-08	4.04E-09	5.66E-03
	0.05	8.62E-02	4.33E-05	1.47E-05	2.59E-12	1.39E-01
	0.10	1.38E-01	1.07E-03	5.48E-04	5.51E-25	1.59E-02
	0.25	2.75E-01	1.23E-01	4.23E-07	1.25E-05	3.79E-02
	0.50	4.47E-01	4.36E-03	3.50E-06	2.12E-57	8.85E-16
	0.75	6.51E-01	2.72E-09	1.23E-12	2.60E-29	7.47E-08
	0.90	8.73E-01	1.80E-02	1.56E-09	3.33E-20	1.67E-03
	0.95	9.36E-01	8.88E-02	2.36E-06	6.59E-12	7.45E-04
	0.99	9.90E-01	9.08E-01	0.00E+00	6.20E-02	5.99E-01
# Rejections			6	9	8	7
Sym. abs. value CAViaR	0.01	1.78E-02	5.66E-02	0.00E+00	7.79E-01	5.74E-02
	0.05	6.29E-02	1.22E-01	6.77E-02	3.68E-10	2.44E-01
	0.10	9.85E-02	8.92E-01	1.64E-01	4.60E-13	2.52E-02
	0.25	2.60E-01	5.38E-01	2.50E-08	1.24E-09	1.74E-01
	0.50	4.40E-01	1.27E-03	0.00E+00	1.33E-65	6.42E-04
	0.75	6.59E-01	4.56E-08	0.00E+00	1.11E-33	2.50E-04
	0.90	9.02E-01	8.92E-01	4.12E-09	8.98E-21	5.26E-03
	0.95	9.38E-01	1.66E-01	9.57E-06	1.52E-12	5.68E-03
	0.99	9.90E-01	9.08E-01	0.00E+00	6.20E-02	7.16E-01
# Rejections			2	7	7	5

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration. All of the QR- and CAViaR models are run with VarGroup2, and GARCH with mean variables. The models use the following window sizes: QR 730, EWQR 730, EWDKQR 250, GARCH 913, Asym. slope CAViaR 548, and Sym. abs. value CAViaR 548.

Table 6.5: Predictive performance in hour 19

Model	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
QR	0.01	1.23E-02	5.44E-01	0.00E+00	3.24E-02	9.21E-03
	0.05	5.20E-02	8.07E-01	9.70E-01	2.80E-04	5.39E-02
	0.10	9.58E-02	7.01E-01	9.22E-01	4.30E-04	2.00E-01
	0.25	2.48E-01	8.81E-01	5.95E-01	1.06E-08	5.55E-01
	0.50	4.99E-01	9.70E-01	4.95E-01	6.61E-01	1.34E-01
	0.75	7.55E-01	7.48E-01	1.31E-01	5.58E-01	3.70E-04
	0.90	9.19E-01	7.30E-02	1.50E-03	1.33E-02	2.40E-01
	0.95	9.85E-01	4.18E-07	7.84E-11	5.81E-26	8.27E-01
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	9.31E-01
# Rejections			1	4	6	2
EWQR	0.01	1.64E-02	1.11E-01	0.00E+00	3.63E-01	2.58E-02
	0.05	6.57E-02	6.32E-02	5.85E-02	5.40E-04	5.44E-03
	0.10	1.16E-01	1.52E-01	2.80E-01	1.91E-02	9.33E-03
	0.25	2.56E-01	7.17E-01	4.70E-01	4.32E-07	4.05E-01
	0.50	5.03E-01	8.53E-01	7.07E-01	6.33E-01	7.39E-02
	0.75	7.36E-01	3.84E-01	1.18E-01	2.19E-01	2.38E-02
	0.90	9.02E-01	8.92E-01	7.78E-02	6.55E-02	6.79E-02
	0.95	9.59E-01	2.52E-01	2.97E-10	5.45E-24	2.77E-01
	0.99	9.95E-01	1.78E-01	1.81E-02	2.76E-08	8.19E-01
# Rejections			0	3	5	4
EWDKQR	0.01	1.09E-02	8.01E-01	1.89E-01	1.70E-01	1.25E-01
	0.05	4.79E-02	7.91E-01	0.00E+00	8.60E-02	3.97E-01
	0.10	9.71E-02	7.95E-01	9.66E-01	4.56E-03	7.12E-02
	0.25	2.56E-01	7.17E-01	1.55E-15	2.13E-18	5.74E-01
	0.50	5.05E-01	7.96E-01	1.32E-01	2.24E-10	2.71E-02
	0.75	7.47E-01	8.48E-01	0.00E+00	3.66E-41	5.41E-01
	0.90	9.15E-01	1.61E-01	9.68E-02	5.92E-02	4.90E-02
	0.95	9.70E-01	7.85E-03	3.54E-04	1.63E-04	5.97E-01
	0.99	9.93E-01	3.62E-01	4.86E-02	1.25E-04	7.33E-01
# Rejections			1	5	6	2

Model	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
GARCH	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	2.84E-02
	0.05	4.51E-02	5.40E-01	7.51E-01	8.82E-02	4.33E-01
	0.10	8.62E-02	2.03E-01	3.46E-01	4.85E-01	3.29E-01
	0.25	2.13E-01	2.01E-02	8.09E-04	3.14E-02	4.67E-04
	0.50	4.87E-01	4.82E-01	5.84E-02	1.10E-02	7.33E-05
	0.75	6.83E-01	4.19E-05	2.65E-09	9.73E-05	3.65E-08
	0.90	8.80E-01	7.41E-02	6.92E-02	2.52E-01	2.26E-05
	0.95	9.45E-01	5.64E-01	7.28E-01	8.35E-01	2.39E-04
	0.99	9.95E-01	1.78E-01	0.00E+00	2.97E-08	6.07E-01
# Rejections			2	4	4	6
Asym. slope CAViaR	0.01	2.33E-02	2.12E-03	0.00E+00	8.08E-01	2.41E-04
	0.05	5.61E-02	4.58E-01	6.73E-02	1.55E-02	4.35E-02
	0.10	1.01E-01	9.12E-01	2.29E-01	2.75E-02	6.87E-01
	0.25	2.68E-01	2.61E-01	4.36E-01	6.51E-08	6.19E-01
	0.50	4.92E-01	6.84E-01	8.61E-01	8.03E-01	5.39E-02
	0.75	7.28E-01	1.69E-01	7.67E-02	1.87E-01	6.64E-05
	0.90	9.15E-01	1.61E-01	9.68E-02	4.24E-01	1.42E-01
	0.95	9.48E-01	8.07E-01	2.89E-08	9.87E-15	6.05E-01
	0.99	9.85E-01	2.02E-01	5.59E-04	6.51E-12	9.58E-01
# Rejections			1	3	5	3
Sym. abs. value CAViaR	0.01	1.37E-02	3.44E-01	0.00E+00	1.34E-01	6.44E-01
	0.05	5.06E-02	9.39E-01	7.22E-01	2.96E-01	6.95E-01
	0.10	1.05E-01	6.33E-01	9.03E-02	2.60E-04	4.84E-01
	0.25	2.76E-01	1.04E-01	2.00E-01	2.00E-06	2.06E-01
	0.50	5.06E-01	7.39E-01	8.29E-01	7.46E-01	2.21E-02
	0.75	7.35E-01	3.40E-01	3.04E-01	4.87E-01	3.42E-05
	0.90	9.21E-01	5.43E-02	1.25E-01	3.55E-01	5.33E-02
	0.95	9.59E-01	2.52E-01	2.97E-10	1.31E-24	7.20E-01
	0.99	9.92E-01	6.15E-01	1.82E-03	8.77E-18	2.31E-01
# Rejections			0	3	4	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration. All models are run with VarGroup2. The models use the following window sizes: QR 548, EWQR 365, EWDKQR 365, GARCH 730, Asym. slope CAViaR 365, and Sym. abs. value CAViaR 365.

6.2.1 Overall performance

Table 6.6: Total number of test rejections per model

	UC (27)	CC (27)	DQ1 (27)	DQ2 (27)	Total (108)
EWQR	2	14	16	9	41
QR	5	17	18	8	48
Sym. abs CAViaR	5	15	16	13	49
EWDKQR	2	19	19	10	50
GARCH	9	17	12	19	57
Asym. slope CAViaR	8	21	18	12	59

The table displays the total number of test rejections per model at the 5% significance level. The numbers in parentheses give the maximum number of rejections. A high number of rejections indicates poor calibration. UC is the unconditional coverage test, CC is the conditional coverage test, and DQ1 and DQ2 are the two dynamic conditional quantile tests, as described in Section 5.2.

In Table 6.6 we display the total number of test rejections per model. Based on this, we rate EWQR as the best model overall; it outperforms both the other QR type models and the benchmarks in terms of test rejections.

Another important observation from Table 6.6, is that clustering of exceedances is challenging to capture for all models. This is evident from the scores on the CC- and DQ1 tests. Interestingly, our results suggest that the GARCH model is the best in this regard. It has the fewest DQ1 rejections, and the majority of the CC rejections are due to incorrect hit percentage. However, the GARCH model is inferior to all other models, as measured by the UC and DQ2 tests. Thus, we rate this model as weak overall, despite its promising ability to capture clustering.

The DQ2 scores indicate that the QR type models are less prone to correlations between the quantile forecasts and the exceedances, compared to the benchmarks.

Aggregating the test scores as we do here hides a lot of important insight. To perform a thorough model evaluation, we break down the analysis into performance in each hour and parts of the distribution in the next sections.

6.2.2 Predictive performance across hours

In this section, we compare the models' performance across trading periods and consider how the characteristics of each hour affect the results. The results for each hour are summarised in Table 6.7. We find that our models perform far better in hours 3 and 19 than in hour 8, where the results are generally not satisfactory. For more detailed results, please refer to Tables 6.3, 6.4, and 6.5.

EWQR is the best model overall in hour 3, but traditional QR and GARCH also exhibit relatively good performance. Interestingly, asymmetric slope is the best CAViaR model for hour 3, unlike in hours 8 and 19. From Table 6.7, we see that asymmetric slope marginally outperforms symmetric absolute value CAViaR in terms of total rejections. However, by looking at Table 6.3, we find that asymmetric slope provides a far better score on the important UC test. As discussed in Chapter 4, the price distribution in hour

Table 6.7: Total number of test rejections per hour

Rating	Model	Rejections
1	EWQR	12
2	QR	15
3	Student-t GARCH	16
4	EWDKQR	18
5	Asym. slope CAViaR	18
6	Sym. abs CAViaR	19

Hour 3 (36)

Rating	Model	Rejections
1	EWQR	17
2	EWDKQR	18
3	QR	20
4	Sym. abs CAViaR	21
5	Student-t GARCH	25
6	Asym. slope CAViaR	30

Hour 8 (36)

Rating	Model	Rejections
1	Sym. abs CAViaR	9
2	Asym. slope CAViaR	11
3	EWQR	12
4	QR	13
5	EWDKQR	14
6	Student-t GARCH	16

Hour 19 (36)

The table displays the total number of test rejections per model at the 5% significance level. The numbers in parentheses give the maximum number of rejections. A high number of rejections indicates poor calibration.

3 is significantly more skewed than the other trading periods. This explains why it is advantageous to use a CAViaR model with asymmetric responses to previous error terms.

Our results are notably weak in hour 8 compared to the other hours, particularly when it comes to clustering. EWQR is the model with the fewest rejections in total, but it also exhibits poor performance on the CC- and DQ1 tests. However, we note that all QR type models outperform the benchmarks.

Among the investigated periods, hour 8 has the highest volatility and the strongest weekday effect, as seen in Chapter 4. We believe that both of these characteristics contribute to the poor results. The DQ1 test catches up to seven lags of clustering. Thus, it is natural to believe that failure of capturing the weekday effect is partly responsible for the weak DQ1 results.

Several models perform well in hour 19. The CAViaR models marginally outperform EWQR, while all models exhibit relatively good accuracy. Note that the CAViaR models exhibit far better performance in hour 19 than in any other trading period. In Chapter 4, we found hour 19 to have the strongest autocorrelation. This could explain why the CAViaR models perform particularly well in this hour.

6.2.3 Predictive performance across the distribution

In the previous section, we revealed substantial differences in model performance across the trading periods. In this section, we assess performance across the distribution. For risk management purposes, it is particularly important to consider accuracy in the tails. Thus, we divide the distribution into three parts: i) The lower tail with quantiles 0.01%, 0.05% and 0.10%, ii) the mid-region with quantiles 0.25%, 0.50% and 0.75%, and iii) the upper tail with quantiles 0.90%, 0.95% and 0.99%.

Table 6.8: Total number of test rejections in sections of the distribution

Rating	Model	Rejections
1	EWDKQR	8
2	EWQR	9
3	Sym. abs CAViaR	9
4	Student-t GARCH	11
5	QR	12
6	Asym. slope CAViaR	17

Lower tail (36)

Rating	Model	Rejections
1	QR	17
2	EWQR	19
3	Asym. slope CAViaR	20
4	EWDKQR	22
5	Sym. abs CAViaR	25
6	Student-t GARCH	31

Mid-region (36)

Rating	Model	Rejections
1	EWQR	13
2	Student-t GARCH	15
3	Sym. abs CAViaR	15
4	QR	19
5	EWDKQR	20
6	Asym. slope CAViaR	22

Upper tail (36)

The table displays the total number of test rejections per model at the 5% significance level. The numbers in parentheses give the maximum number of rejections. A high number of rejections indicates poor calibration.

Table 6.8 displays the total number of test rejections for parts i), ii) and iii). Overall, we find that the models perform better in the tails than in the mid-region. Our results differ between the parts of the distribution. While traditional QR outperforms all other models in the mid-region, it is among the bottom half in both tails. EWQR shows strong performance overall, but scores particularly well in the tails. EWDKQR is inconsistent; it is the top-performer for the lower tail, but is the second worst in the upper tail.

Sparseness of observations in the tails is particularly problematic when the distribution changes rapidly. If this is the case, previous tail observations may stem from an underlying distribution with different characteristics. This means that high performance in the tails indicates ability to capture dynamicity. Thus, the findings in Table 6.8 suggest that EWQR is the best at accounting for the changing market conditions in EPEX. Moreover, EWDKQR exhibits the same ability for the lower tail.

By selecting appropriate values for bandwidth and lambda, EWDKQR can be simplified to both EWQR and QR. This means that whenever EWQR or EWDKQR perform worse than QR, it is due to overfitting. Thus, our results suggest that both EWQR and EWDKQR suffer from this problem. However, in the tails where data is sparse, the gain of increased sophistication outweighs the adverse effect of overfitting. Hence, both EWQR and EWDKQR outperform traditional QR in these parts of the distribution. EWDKQR is extra prone to overfitting, as it introduces an additional parameter. We believe that this is the reason why EWDKQR is inferior to EWQR.

6.2.4 Other findings

We find an inverse relationship between the complexity of the QR models and the selected window sizes. EWDKQR always performs the best with a window that is smaller, or the same size, as EWQR. The same relation holds for EWQR and QR. Thus, adding exponential smoothing or double kernels reduces the number of observations needed to produce accurate VaR forecasts. This finding suggests that the more complex models are appropriate in scenarios where data is sparse, or when the data generating process changes rapidly.

6.3 Impact of variable selection on model performance

In this section, we analyse how our choice of variables affects the models' performance. Moreover, we present results to highlight the importance of variable selection. Table 6.9 displays test results that offer valuable insights in this regard. To limit the number of parameters, we restrict our analysis to QR models with window size 730.

Table 6.9: Total number of test rejections for various variable combinations

Variables	Model	UC (27)	CC (27)	DQ1 (27)	DQ2 (27)	Total (108)
VarGroup2	QR	7	17	16	8	48
VarGroup1	EWQR	3	19	21	11	54
Mean variables	QR	14	19	15	9	57
VarGroup1	QR	7	24	21	9	61
All variables	QR	20	21	15	15	71
Intercept only	QR	18	27	22	8	75

The table displays the total number of test rejections per model at the 5% significance level. The numbers in parentheses give the maximum number of rejections. A high number of rejections indicates poor calibration. UC is the unconditional coverage test, CC is the conditional coverage test, and DQ1 and DQ2 are the two dynamic conditional quantile tests, as described in Section 5.2. All models have window size 730 days.

It is evident from Table 6.9 that the intercept-only model is seriously flawed, and performs the worst overall. By including regressors, the quantiles become dynamic because they evolve as the regressors evolve. Thus, it is unsurprising that the intercept-only model is inferior to all other models when it comes to clustering of exceedances.

Selecting a model with *all* the available variables offers significant improvement in the dynamicity tests, i.e. CC and DQ1. However, this model still provides incorrect coverage in the majority of tests. This illustrates a previously emphasised point: Using too many explanatory variables leads to overfitting and inflation of estimation variance.

Table 6.9 also shows that the QR model with VarGroup2 outperforms the QR model with VarGroup1, but that both models have the same number of UC rejections. Recall that we selected VarGroup1 solely based on the UC test, whereas we also considered the CC test for VarGroup2. Thus, this result is as expected. However, it illustrates that the choice of variables plays a vital role in capturing dynamicity.

Furthermore, we find that the QR model with mean variables is inferior to the QR model with VarGroup2. This supports our hypothesis that using different variables across the distribution improves predictive performance. Note that the model with mean variables outperforms the QR model with VarGroup1. However, VarGroup1 performs significantly better on the UC test, which was its selection criterion. Thus, we argue that this result does not undermine our hypothesis. Instead, it illustrates an unsurprising, yet important result; *you make what you measure*. We find that the variable groups perform best on the tests we used in the selection procedure. This implies that if end-users require a model with high performance on certain tests, these tests should be included in the variable selection.

Our results indicate that variable selection in many cases is more crucial than the model formulation. This means that it might be wise to put more effort into identifying appropriate variables, rather than using a more complex model. To illustrate this point, we have included an EWQR model with VarGroup1 in Table 6.9. We find that if you have a traditional QR model with VarGroup1, you will experience a greater improvement by switching to VarGroup2, than by using a more complex EWQR formulation.

Based on the results we present in this section, it is unsurprising that we find VarGroup2 as the best variable combination for all the QR type models. It is more interesting to note that VarGroup2 is also the best set of explanatory variables for the majority of the GARCH and CAViaR models (see Tables 6.3, 6.4, and 6.5). Although the variables in GARCH and CAViaR model the mean, our results suggest that it is advantageous to filter observations with variables chosen for the specific quantiles.

Finally, we note that several variable combinations that pass both the UC- and CC tests in the variable selection, fail when tested on the out-of-sample set. Recall that we use data from 2010 to 2016 for the variable selection, but perform the out-of-sample analysis with data from 2014 to 2016. Thus, this finding indicates that the relation between the fundamentals and the spot prices has changed over this time period. This illustrates that there is a need to investigate procedures that enable continuous re-estimation of variable selection.

6.4 Summary of main findings

The results of the variable selection reflect our findings from the market description in Chapter 3. The first price lag and the coal price are the best variables overall. However, we find that the variables with the highest predictive ability vary across the price distribution and between the trading periods.

We rate EWQR as the best model overall. This model has the fewest test rejections in total, and shows particularly good performance in the tails. This indicates that it is able to account for the changing market dynamics in Germany. We argue that the CAViaR models are the best performing benchmarks, while noting that their performance is inconsistent across trading periods. Capturing clustering is challenging for all models, but the GARCH model outperforms the others in this regard.

Both EWQR and EWDKQR suffer from overfitting, and this it is particularly evident for EWDKQR. This is because EWDKQR requires estimation of an additional parameter.

The overfitting is visible in the mid-region of the distribution, where traditional QR outperforms the more complex models. However, in the tails where data is sparse, the gain of increased sophistication outweighs the adverse effect of overfitting.

Our results support that it is advantageous to tailor input variables to specific quantiles. We also find that selecting appropriate variables is important for capturing dynamicity. Moreover, our results reveal that variable selection in many cases is as crucial as the model formulation.

Chapter 7

Practical implications

In this section, we consider the practical implications of our work. An important step in this analysis is to compare our results to existing literature. Moreover, we explain how market operators can use the insight from our work to create value.

It is not straightforward to compare our findings to previous work. In Chapter 6, we found that the models' performance varies across trading periods. Consequently, it is unsurprising that results from studies on other markets in many cases differ from ours.

The study that is most similar to this work, Bunn et al. (2016), come to conclusions that are much alike ours. They find that a traditional QR model outperforms the skewed student-t GARCH and the CAViaR models. Moreover, they also report that the GARCH model suffers from exceedances being correlated with the quantile estimate itself, but that it shows little clustering. In contrast to our results, they draw the same conclusion for the CAViaR models. We note that this work does not consider the mid-region of the distribution, thus, we cannot compare performance across the distribution.

We argue that the work by Bunn et al. largely supports our findings. However, they do not consider what is arguably the most interesting aspect of our work, namely EWQR and EWDKQR. To the best of our knowledge, Lundby and Uppheim's analysis of Nordpool prices (2011) is the only other application of EWQR to electricity prices. However, they find that the model performs poorly, and that the CAViaR models are superior. A possible explanation is that they do not include fundamentals in EWQR, and use a window size of only 250 observations. From our experience, both of these decisions degrade performance.

Our results support the findings by Taylor (2008b). In his original paper, EWQR shows encouraging results. Taylor finds that EWQR generally outperforms traditional QR for forecasting of stock returns. Moreover, he comes to conclusions similar to ours regarding the performance of EWDKQR. This model is largely inferior to EWQR, but is more competitive in the extreme quantiles. The overfitting we observe for EWQR and EWDKQR may be examples of the over-parameterisation that De Livera et al. (2011) warns of.

We have no knowledge of previous research that considers the predictive ability of fundamentals across the distribution for electricity spot prices. However, our results are further evidence of the differences that exist across the distribution and trading periods. Moreover, the variables we select are well aligned with the current literature on the price formation process, e.g. Paraschiv et al. (2014) and Hagfors et al. (2016a). Our results underpin the importance of variable selection, as highlighted by e.g. Weron (2014), Diebold (2015) and Shmueli (2010). This implies that significant effort should be devoted to this stage of the modelling process.

For electricity price forecasting in practice, it is also important to consider the computational complexity of the models. Time is a limited resource in short-term forecasting. Thus, models with shorter run-times are more likely to be used in situations with little time from information becomes available until the forecast is needed. This criterion favours QR and EWQR. Our results suggest that these models provide the best performance, as well as being the least computationally demanding. EWQR requires slightly more computation than QR, as the weighting parameter must be estimated. However, our results suggest that it also improves performance. The EWDKQR and CAViaR models are particularly challenging to fit, but we do generally not observe a gain from the increased complexity. Moreover, the simplicity and transparency of the QR type models increases the probability of adoption by industry professionals.

The approaches we explore in this thesis are of great value to market operators. Both producers and consumers of electricity can take advantage of sudden price changes by forecasting VaR. We illustrate this with an example: A producer who forecasts high upside risk, should offer a high volume to gain from the potential price spikes. Contrary, in times of high downside risk, producers should restrict their production and offer lower volumes, as the probability of a price lower than the cost of production increases. The opposite applies for electricity consumers. Thus, VaR forecasts can be used for determining optimal bidding and consumption strategies, both for profit maximisation and risk reduction.

Chapter 8

Conclusion

In this thesis, we forecast VaR for the EPEX spot price. Despite the importance of these forecasts to market participants, VaR forecasting of electricity prices remains under-researched. We focus on QR approaches because of their simplicity, possibility to include fundamentals, and promising ability to capture the complex features of electricity prices.

Our work has a threefold goal. Firstly, we identify variables with high predictive power for selected quantiles and trading periods. Secondly, we assess the gain of using sophisticated extensions of QR by investigating the performance of EWQR and EWDKQR. Thirdly, we benchmark the QR type models against common VaR models in literature: Skewed student-t GARCH, asymmetric slope CAViaR, and symmetric absolute value CAViaR.

The German market is the subject of our study, as it is the main reference for power trading in Europe. In the data analysis, we found hours 3, 8 and 19 as the most interesting to model. These hours represent the off-peak, the morning peak and the evening super-peak.

The greatest advantage of QR models is that they model each quantile separately. In contrast to existing studies, we take advantage of this through the variable selection. We propose using a separate set of variables for each quantile and hour. This is motivated by evidence in literature that the impact of fundamentals differs across the distribution and between trading periods. Our variable selection shows that the variables with highest predictive ability do in fact vary. Moreover, the variables we select reflect current research on the price formation process. Our findings highlight the importance of variable selection, and show that it in many cases is as important as the choice of model. Thus, we argue that market operators should devote significant attention to this procedure.

It is difficult to draw general conclusions as to which model is the best based on our results, and both application and preferences must be accounted for when selecting the most appropriate model. However, we rate EWQR as the best model overall based on the total number of test rejections. EWQR is among the top-performers for all trading periods, and performs particularly well in the outer quantiles. The latter is a crucial property for risk management purposes.

Our results indicate that both EWQR and EWDKQR suffer from overfitting. This is evident in the mid-region of the distribution, where these models are outperformed by traditional QR. Moreover, we believe overfitting to be the reason why EWDKQR is inferior to EWQR.

The CAViaR models are the best performing benchmarks, but their performance is inconsistent across the hours. We find these models to perform particularly well in periods with high autocorrelation. While asymmetric slope CAViaR performs well for highly skewed periods, it is outperformed by symmetric absolute value CAViaR for periods with less skew.

Clustering of exceedances is challenging to capture for all models, but the GARCH model shows the best performance. However, GARCH is inferior to the other models, as measured by the important UC test. Thus, we rate this model as weak overall, despite its promising ability to capture clustering.

Accounting for computational complexity and transparency favours QR and EWQR. These models are simple to implement and interpret, as well as being less computationally demanding than the benchmarks. Moreover, they generally outperform the models with higher complexity.

In summary, we have contributed to electricity price forecasting literature by providing important empirical evidence for the use of fundamental VaR forecasting approaches. We have demonstrated that EWQR exhibits particularly promising ability to capture electricity price dynamics. Moreover, we presented results to underpin the importance of variable selection, and identified fundamentals with predictive ability across the distribution. These insights can be applied by market participants who seek to determine optimal bidding, production and consumption strategies.

Chapter 9

Further work

In this chapter, we briefly discuss recommendations for further work. This includes both improvements of methods and possible extensions of the research.

We suggest to conduct similar studies on other electricity markets and trading periods, to see if our findings are confirmed and to provide further empirical evidence. Moreover, additional benchmark models can be added to the study. We find it particularly interesting to measure the performance of QR type models against EVT and computational intelligence approaches, like ANNs.

Our results indicate that the predictive performance of variables changes over time. Thus, we suggest to devote attention to techniques for re-estimation of variable selection for QR, e.g. adaptive lasso methods. We also recommend to further extend the set of fundamental variables from which the variable selection is performed.

The models we analyse in this thesis generally struggle to capture clustering. Our findings indicate that both the model formulation and the variable selection are important in this respect. Thus, these components should be considered when attempting to resolve the clustering issue. Our suggestions include adding more variables to capture the weekday effect, and to use the DQ tests as variable selection criteria.

Although we find the results for EWQR encouraging, both EWQR and EWDKQR show signs of overfitting. Thus, we argue that more effort should be made to improve the estimation procedures of the lambda- and bandwidth parameters. We recommend to investigate approaches where these parameters are re-estimated in each window.

Bibliography

AG Energibalanzen e.V., 2017. Stromerzeugung nach Energieträgern (Strommix) von 1990 bis 2016 (in TWh) Deutschland insgesamt.
URL www.ag-energiebilanzen.de/28-0-Zusatzinformationen

Alexander, C., 2008b. Market Risk Analysis, Volume II, Practical Financial Econometrics. John Wiley.

Alexander, C., 2008d. Market Risk Analysis, Volume IV, Value-at-Risk Models. John Wiley.

Behl, P., Claeskens, G., Dette, H., 2014. Focused model selection in quantile regression. Statistica Sinica 24, 601–624.

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

Bremnes, J., 2006. A comparison of a few statistical models for making quantile wind power forecasts. Wind Energy 2006 9, 3–11.

Bunn, D., Andresen, A., Chen, D., Westgaard, S., 2016. Analysis and forecasting of electricity price risks with quantile factor models. The Energy Journal 37 (2), 169–190.

Bystrom, H. N., 2005. Extreme value theory and extremely large electricity price changes. International Review of Economics and Finance 14 (1), 41 – 55.

Chan, K. F., Gray, P., 2006. Using extreme value theory to measure value at risk for daily electricity spot prices. International Journal of Forecasting 22 (2), 283 – 300.

Chatfield, C., 2000. Time-Series Forecasting. Chapman & Hall/CRC.

Chen, D., Bunn, D., 2010. Analysis of the nonlinear response of electricity prices to fundamental and strategic factors. IEEE Transactions on Power Systems 25 (2), 595–606.

Chen, M.-Y., Chen, J.-E., 2002. Application of quantile regression to estimation of value at risk. Review of Financial Risk Management 1 (2), 15.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Clean Energy Wire, 2017. Germany's energy consumption and power mix in charts. URL https://www.cleanenergywire.org/search/site/factsheets

De Livera, A. M., Hyndman, R. J., Snyder, R. D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association 106 (496), 1513–1527.

Deng, S.-J., Oren, S. S., 2006. Electricity derivatives and risk management. Energy 31 (6), 940–953.

Diebold, F., 2015. Forecasting. Department of Economics, University of Pennsylvania.

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

Engle, R., Manganelli, S., 2004. Caviar conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics 22 (4), 367–381.

EPEX Spot, 2017. Renewable energy: Increasingly important role in Europe. URL https://www.epexspot.com/en/renewables

Erni, D., 2012. Day-ahead electricity spot prices - fundamental modelling and the role of expected wind electricity infeed at the european energy exchange. Ph.d dissertation, University of St. Gallen.

Federal Minsitry for Economic Affairs and Energy, 2017. For a future of green energy. URL www.bmwi.de/Redaktion/EN/Dossier/renewable-energy

Florentina, P., Hadzi-Mishev, R., 2016. Extreme value theory for heavy tails in electricity prices. Journal of Energy Markets 9 (2).

Frauendorfer, K., Paraschiv, F., Schüle, M., 2016. Cross-border effects of the German electricity market fundamentals on the Swiss electricity prices.

Garcia, R. C., Contreras, J., Van Akkeren, M., Garcia, J. B. C., 2005. A garch forecasting model to predict day-ahead electricity prices. IEEE Transactions on power systems 20 (2), 867–874.

Gelper, S., Fried, R., Croux, C., 2010. Robust forecasting with exponential and holt–winters smoothing. Journal of Forecasting 29, 285–300.

Gonzales, V., Contreras, J., Bunn, D., 2012. Forecasting power prices using a hybrid fundamental-econometric model. IEEE Transactions on power systems 27 (1), 363–372.

Grothe, O., Schnieders, J., 2011. Spatial dependence in wind and optimal wind power allocation: A copula-based analysis. Energy Policy 39, 4742–4754.

Gurrola-Perez, P., Murphy, D., 2015. Filtered historical simulation value-at-risk models and their competitors. Bank of England (525).

Hagfors, L., Paraschiv, F., Molnar, P., Westgaard, S., 2016a. Using quantile regression to analyze the effect of renewables on eex price formation. Renewable Energy and Environmental Sustainability 1 (32).

Hagfors, L. I., Bunn, D., Kristoffersen, E., Staver, T. T., Westgaard, S., 2016b. Modeling the {UK} electricity price distributions using quantile regression. Energy 102, 231 – 243.

Hagfors, L. I., Kamperud, H. H., Paraschiv, F., Prokopczuk, M., Sator, A., Westgaard, S., 2016c. Prediction of extreme price occurrences in the german day-ahead electricity market. Quantitative Finance 16 (2), 1929–1948.

Hartz, C., Mittnik, S., Paolella, M., 2006. Accurate value at risk forecasting based on the normal garch model. Computational Statistics and Data Analysis 51 (4), 2295 – 2312.

Jeon, J., Taylor, J., 2013. Using caviar models with implied volatility for value at risk estimation. Journal of Forecasting 32, 62–74.

Jones, M., Yu, K., 1998. Local linear quantile regression. Journal of the American Statistical Association 93, 228–237.

J.P.Morgan/Reuters, 1996. RiskMetrics TM —Technical Document. Tech. rep., Morgan Guaranty Trust Company, Reuters Ltd.

Kaminski, V., 2013. Energy markets. Risk Books.

Karakatsani, N. V., Bunn, D. W., 2008. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. International Journal of Forecasting 24, 764–785.

Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Applied Energy 162, 218–230.

Koenker, R., Bassett Jr., G., 1978. Regression quantiles. The Econometric Society 46 (1), 33–50.

Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81, 673–680.

Kuester, K., Mittnik, S., Paoella, M., 2006. Value at risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 9 (2).

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 2, 173–184.

Lee, E. R., Noh, H., Park, B. U., 2014. Model selection via bayesian information criterion for quantile regression models. Journal of the American Statistical Association 109 (505), 216–229.

Lundby, M., Uppheim, K., 2011. Fundamental risk analysis and var forecasts of the nord pool system price. Master's thesis, Norwegian University of Science and Technology (NTNU).

Maciejowska, K., Weron, R., 2016. Short- and mid-term forecasting of baseload electricity prices in the u.k.: The impact of intra-day price relationships and market fundamentals. IEEE Transactions on power systems 31 (2), 957–965.

Maciejowska, K., et al., 2016. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. International Journal of Forecasting 37 (3), 957–965.

Maciel, L., Ballini, R., Gomide, F., 2017. Evolving Possibilistic Fuzzy Modeling and Application in Value at Risk Estimation. pp. 119–139.

Mirza, F., Bergland, O., 2012. Transmission congestion and market power: the case of the norwegian electricity market. The Journal of Energy Markets 5 (2), 59–88.

Mostafa, F., Dillon, T., Chang, E., 2017. Computational intelligence applications to option pricing, volatility forecasting and value at risk. Studies in Computational intelligence 697, 137–147.

Neupane, B., Woon, W. L., Aung, Z., 2017. Ensemble prediction model with expert selection for electricity price forecasting. Energies 10 (1).

Paraschiv, F., Bunn, D., Westgaard, S., 2016. Estimation and applications of fully parametric multifactor quantile regression with dynamic coefficients. University of St. Gallen, School of Finance Research.

Paraschiv, F., Erni, D., Pietsch, R., 2014. The impact of renewable energies on eex dayahead electricity prices. Energy Policy 73, 196–210.

Reisch, L., Micklitz, H., 2006. Consumers and deregulation of the electricity market in germany. J Consum Policy 29, 399–415.

Schwarz, G., et al., 1978. Estimating the dimension of a model. The annals of statistics 6 (2), 461–464.

Senera, E., Baronyana, S., Mengütürkb, L., 2012. Ranking the predictive performances of value-at-risk estimation methods. International Journal of Forecasting 28 (4), 849–873.

Sensfu, F., Ragwitz, M., Genoese, M., 2008. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Energy Policy 36, 3086–3094.

Shmueli, G., 2010. To explain or to predict? Statistical Science 25, 289–310.

Taylor, J. W., 2007. Forecasting daily supermarket sales using exponentially weighted quantile regression. European Journal of Operational Research 178 (1), 154–167.

Taylor, J. W., 2008a. Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics, 231–252.

Taylor, J. W., 2008b. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. Journal of Financial Econometrics 6 (3), 382–406.

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.

Appendix A

Variable description

Variable	Units	Description	Data Source
Spot price	EUR/MWh	Market clearing price	European Energy Exchange: http://www.eex.com
Coal price	EUR/12,000 t	Latest available price (daily auctioned) of the front-month ARA futures contract before the electricity price auction takes place	European Energy Exchange: http://www.eex.com
Gas price	EUR/MWh	Latest price of the NCG Day Ahead Natural Gas Spot Price on the day before the electricity price auction takes place	Bloomberg, ticker: EGTHDAHD Index
Oil price	EUR/bbl	Latest price of the active ICE Brent Crude futures contract on the day before the electricity price auction takes place	Bloomberg, ticker: CO1 Comdty
CO2 price	EUR 0.01/EUA 1,000 t CO2	Latest available price of the EEX Carbon Index (Carbix), daily auctioned at 10.30 am	European Energy Exchange: http://www.eex.com
Expected wind and PV infeed	MWh	Sum of expected infeed of wind electricity into the grid, published by German transmission systems operators in the late afternoon following the electricity price auction	Transmission system operators: http://www.50Hertz.com, http://www.amprion.de, http://www.transnetbw.de, http://www.tennettso.de
Expected PPA	MWh	Forecast of expected power plant availability production (voluntary publication) on the delivery day (daily granularity), published at 10:00 am	European Energy Exchange & transmission energy operators: ftp://infoproducts.eex.com
Expected demand	MWh	Sum of the total vertical system load and actual wind infeed for the same hour on the last relevant delivery day	Transmission system operators: http://www.50Hertz.com, http://www.amprion.de, http://www.transnetbw.de, http://www.tennettso.de

Figure A.1: Overview of fundamental variables, as in Frauendorfer et al. (2016)

Appendix B

Additional statistics

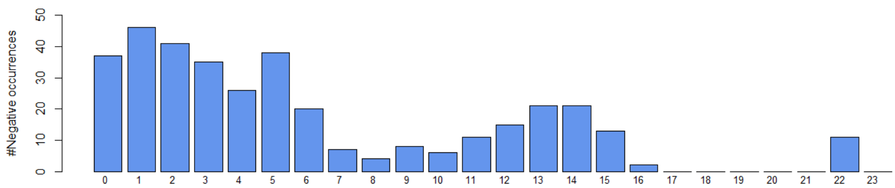

Figure B.1: Negative spot price occurrences per hour

Table B.1: Correlation between the spot price and volatility lags per hour

Hour	Vol_lag1	Vol_lag2	Vol_lag3	Vol_lag4	Vol_lag5	Vol_lag6	Vol_lag7	Vol_AVlag
Spot hour 3	-0,409	-0,192	-0,105	-0,105	-0,132	-0,144	-0,087	-0,179
Spot hour 8	-0,088	-0,035	-0,039	-0,075	-0,103	-0,202	-0,277	-0,190
Spot hour 19	-0,009	-0,041	-0,071	-0,078	-0,079	-0,081	-0,087	-0,104

Table B.2: Above: Correlation between fundamentals for hour 3

Table B.3: Below: Correlation between fundamentals for hour 8

Variables	Wind	Solar	Demand	Coal	Gas	Oil	Co2	PPA	Vol_lag1	Vol_lag2	Vol_lag3	Vol_lag4	Vol_lag5	Vol_lag6	Vol_lag7	Vol_AVlag
Wind	1,000	0,029	0,223	-0,232	-0,120	-0,273	-0,162	0,238	0,271	0,176	0,122	0,113	0,116	0,113	0,089	0,193
Solar	0,029	1,000	0,007	-0,010	0,006	-0,020	-0,010	0,023	0,006	0,010	0,004	-0,009	-0,009	-0,010	-0,007	-0,004
Demand	0,223	0,007	1,000	0,192	0,064	0,004	0,244	0,552	-0,035	-0,004	-0,010	-0,017	-0,049	-0,091	-0,106	-0,050
Coal	-0,232	-0,010	0,192	1,000	0,456	0,638	0,586	-0,165	0,064	0,064	0,064	0,063	0,062	0,061	0,060	0,151
Gas	-0,120	0,006	0,064	0,456	1,000	0,745	-0,190	0,077	0,106	0,108	0,111	0,113	0,111	0,110	0,112	0,189
Oil	-0,273	-0,020	0,004	0,638	0,745	1,000	-0,027	-0,167	0,057	0,057	0,058	0,059	0,059	0,060	0,061	0,114
Co2	-0,162	-0,010	0,244	0,586	-0,190	-0,027	1,000	-0,077	-0,013	-0,014	-0,016	-0,017	-0,017	-0,018	-0,020	0,010
PPA	0,238	0,023	0,552	-0,165	0,077	-0,167	-0,077	1,000	0,094	0,102	0,098	0,088	0,067	0,053	0,083	0,112
Vol_lag1	0,271	0,006	-0,035	0,064	0,106	0,057	-0,013	0,094	1,000	0,691	0,396	0,228	0,185	0,200	0,213	0,497
Vol_lag2	0,176	0,010	-0,004	0,064	0,108	0,057	-0,014	0,102	0,691	1,000	0,691	0,396	0,228	0,185	0,200	0,598
Vol_lag3	0,122	0,004	-0,010	0,064	0,111	0,058	-0,016	0,098	0,396	0,691	1,000	0,691	0,396	0,228	0,185	0,677
Vol_lag4	0,113	-0,009	-0,017	0,063	0,113	0,059	-0,017	0,088	0,228	0,396	0,691	1,000	0,691	0,396	0,228	0,729
Vol_lag5	0,116	-0,009	-0,049	0,062	0,111	0,059	-0,017	0,067	0,185	0,228	0,396	0,691	1,000	0,691	0,396	0,756
Vol_lag6	0,113	-0,010	-0,091	0,061	0,110	0,060	-0,018	0,053	0,200	0,185	0,228	0,396	0,691	1,000	0,691	0,742
Vol_lag7	0,089	-0,007	-0,106	0,060	0,112	0,061	-0,020	0,083	0,213	0,200	0,185	0,228	0,396	0,691	1,000	0,663
Vol_AVlag	0,193	-0,004	-0,050	0,151	0,189	0,114	0,010	0,112	0,497	0,598	0,677	0,729	0,756	0,742	0,663	1,000

Variables	Wind	Solar	Demand	Coal	Gas	Oil	Co2	PPA	Vol_lag1	Vol_lag2	Vol_lag3	Vol_lag4	Vol_lag5	Vol_lag6	Vol_lag7	Vol_AVlag
Wind	1,000	-0,248	0,060	-0,202	-0,092	-0,245	-0,147	0,259	0,165	0,152	0,116	0,135	0,135	0,129	0,125	0,217
Solar	-0,248	1,000	-0,081	-0,306	-0,107	-0,036	-0,365	-0,565	0,037	0,044	0,048	0,050	0,055	0,053	0,058	0,063
Demand	0,060	-0,081	1,000	0,084	0,049	0,035	0,077	0,305	-0,033	0,051	0,025	0,000	-0,021	-0,153	-0,238	-0,075
Coal	-0,202	-0,306	0,084	1,000	0,456	0,638	0,586	-0,165	-0,154	-0,155	-0,157	-0,157	-0,159	-0,160	-0,160	-0,267
Gas	-0,092	-0,107	0,049	0,456	1,000	0,745	-0,190	0,077	0,154	0,141	0,142	0,141	0,131	0,125	0,123	0,135
Oil	-0,245	-0,036	0,035	0,638	0,745	1,000	-0,027	-0,167	0,029	0,028	0,027	0,027	0,026	0,026	0,027	-0,024
Co2	-0,147	-0,365	0,077	0,586	-0,190	-0,027	1,000	-0,077	-0,306	-0,306	-0,306	-0,306	-0,305	-0,303	-0,303	-0,427
PPA	0,259	-0,565	0,305	-0,165	0,077	-0,167	-0,077	1,000	0,131	0,163	0,159	0,151	0,141	0,102	0,088	0,215
Vol_lag1	0,165	0,037	-0,033	-0,154	0,154	0,029	-0,306	0,131	1,000	0,501	0,367	0,363	0,334	0,258	0,349	0,642
Vol_lag2	0,152	0,044	0,051	-0,155	0,141	0,028	-0,306	0,163	0,501	1,000	0,501	0,367	0,363	0,333	0,257	0,674
Vol_lag3	0,116	0,048	0,025	-0,157	0,142	0,027	-0,306	0,159	0,367	0,501	1,000	0,501	0,367	0,363	0,333	0,701
Vol_lag4	0,135	0,050	0,000	-0,157	0,141	0,027	-0,306	0,151	0,363	0,367	0,501	1,000	0,501	0,367	0,362	0,708
Vol_lag5	0,135	0,055	-0,021	-0,159	0,131	0,026	-0,305	0,141	0,334	0,363	0,367	0,501	1,000	0,501	0,367	0,704
Vol_lag6	0,129	0,053	-0,153	-0,160	0,125	0,026	-0,303	0,102	0,258	0,333	0,363	0,367	0,501	1,000	0,501	0,683
Vol_lag7	0,125	0,058	-0,238	-0,160	0,123	0,027	-0,303	0,088	0,349	0,257	0,333	0,362	0,367	0,501	1,000	0,651
Vol_AVlag	0,217	0,063	-0,075	-0,267	0,135	-0,024	-0,427	0,215	0,642	0,674	0,701	0,708	0,704	0,683	0,651	1,000

iv

Table B.4: Correlation between fundamentals for hour 19

Variables	Wind	Solar	Demand	Coal	Gas	Oil	Co2	PPA	Vol_lag1	Vol_lag2	Vol_lag3	Vol_lag4	Vol_lag5	Vol_lag6	Vol_lag7	Vol_AVlag
Wind	1,000	-0,211	0,167	-0,198	-0,084	-0,235	-0,143	0,247	0,166	0,153	0,148	0,143	0,139	0,136	0,135	0,170
Solar	-0,211	1,000	-0,384	-0,209	-0,113	-0,010	-0,271	-0,627	0,081	0,090	0,096	0,101	0,105	0,107	0,110	0,130
Demand	0,167	-0,384	1,000	0,179	0,119	0,064	0,169	0,545	0,025	-0,001	-0,038	-0,054	-0,057	-0,053	-0,020	-0,044
Coal	-0,198	-0,209	0,179	1,000	0,456	0,638	0,586	-0,165	-0,171	-0,171	-0,170	-0,168	-0,166	-0,164	-0,164	-0,217
Gas	-0,084	-0,113	0,119	0,456	1,000	0,745	-0,190	0,077	0,099	0,090	0,088	0,080	0,077	0,074	0,069	0,077
Oil	-0,235	-0,010	0,064	0,638	0,745	1,000	-0,027	-0,167	-0,041	-0,040	-0,038	-0,036	-0,034	-0,033	-0,031	-0,056
Co2	-0,143	-0,271	0,169	0,586	-0,190	-0,027	1,000	-0,077	-0,309	-0,311	-0,311	-0,312	-0,313	-0,312	-0,313	-0,377
PPA	0,247	-0,627	0,545	-0,165	0,077	-0,167	-0,077	1,000	0,099	0,087	0,073	0,068	0,068	0,071	0,083	0,083
Vol_lag1	0,166	0,081	0,025	-0,171	0,099	-0,041	-0,309	0,099	1,000	0,881	0,766	0,673	0,619	0,599	0,612	0,818
Vol_lag2	0,153	0,090	-0,001	-0,171	0,090	-0,040	-0,311	0,087	0,881	1,000	0,881	0,766	0,673	0,619	0,600	0,864
Vol_lag3	0,148	0,096	-0,038	-0,170	0,088	-0,038	-0,311	0,073	0,766	0,881	1,000	0,881	0,766	0,674	0,620	0,897
Vol_lag4	0,143	0,101	-0,054	-0,168	0,080	-0,036	-0,312	0,068	0,673	0,766	0,881	1,000	0,881	0,767	0,674	0,913
Vol_lag5	0,139	0,105	-0,057	-0,166	0,077	-0,034	-0,313	0,068	0,619	0,673	0,766	0,881	1,000	0,881	0,767	0,911
Vol_lag6	0,136	0,107	-0,053	-0,164	0,074	-0,033	-0,312	0,071	0,599	0,619	0,674	0,767	0,881	1,000	0,881	0,888
Vol_lag7	0,135	0,110	-0,020	-0,164	0,069	-0,031	-0,313	0,083	0,612	0,600	0,620	0,674	0,767	0,881	1,000	0,846
Vol_AVlag	0,170	0,130	-0,044	-0,217	0,077	-0,056	-0,377	0,083	0,818	0,864	0,897	0,913	0,911	0,888	0,846	1,000

Appendix C

Variable selection

C.1 Tested variable combinations

We define the variable groups in Table C.1 to simplify the variable selection. We test all combinations consisting of one or zero subsets from each variable group. The groups and subsets are specified based on results from preliminary testing. For example, the preliminary tests revealed that Coal has significantly higher predictive power than Co2, Gas and Oil. Thus, we restrict the tests with fossil variables to include Coal and at most one other fossil variable.

Table C.1: Definition of variable groups

Price lags
P.lag1
P.lag1, P.lag2, P.lag3, P.lag4, P.lag5, P.lag6, P.lag.7
P.lag1, Avg.P.lag2-7
Volatility lags
V.lag1
V.lag1, V.lag2, V.lag3, V.lag4, V.lag5, V.lag6, V.lag.7
V.lag1, Avg.V.lag2-7
Fossils
Coal
Coal, Co2
Coal, Gas
Coal, Oil
Renewables
Wind
Solar
Wind, Solar
Other
PPA
Demand
PPA, Demand

Appendix D

Additional result tables

In this chapter, we provide additional result tables to support the discussion in Chapter 6. We include all three hours; 3, 8 and 19. For traditional QR we show results for both VarGroup1 and VarGroup2, to illustrate the distinction between the two sets. For the rest of the models, we only attach result tables for VarGroup2, as these are the most successful runs.

- Traditional QR results is shown in Tables D.1 - D.3. Window sizes 365, 548, 730 and 913.
- EWQR results is shown in Tables D.4 and D.5. Window sizes 365, 548, 730 and 913.
- EWDKQR results is shown in Tables D.6 and D.7. Window sizes 250, 365, 548 and 730.
- GARCH results is shown in Tables D.8 and D.9. Window sizes 365, 548, 730 and 913.
- Asymmetric slope CAViaR results are shown in Tables D.10 and D.11. Window sizes 365, 548, 730 and 913.
- Symmetric absolute value CAViaR results are shown in Tables D.12 and D.13. Window sizes 365, 548, 730 and 913.

Table D.1: Traditional QR results for hour 3

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	1.22E-02
	0.05	4.51E-02	5.40E-01	4.27E-01	5.56E-01	5.03E-01
	0.10	8.76E-02	2.53E-01	4.99E-01	6.13E-01	9.90E-01
VarGroup1,	0.25	2.39E-01	5.06E-01	5.51E-01	6.64E-03	6.85E-01
w = 365	0.50	5.42E-01	2.40E-02	0.00E+00	1.73E-36	1.23E-01
	0.75	6.81E-01	2.94E-05	0.00E+00	1.10E-29	7.88E-03
	0.90	9.21E-01	5.43E-02	8.44E-12	2.48E-19	9.96E-01
	0.95	9.45E-01	5.64E-01	5.80E-10	5.30E-21	1.06E-01
	0.99	9.82E-01	5.66E-02	2.26E-05	2.72E-15	7.20E-04
# Rejections			2	6	6	3
	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	9.38E-01
	0.05	4.92E-02	9.25E-01	6.71E-01	5.80E-01	6.50E-01
	0.10	9.30E-02	5.25E-01	8.08E-01	6.47E-01	5.30E-01
VarGroup1,	0.25	2.27E-01	1.48E-01	2.27E-01	4.36E-03	7.45E-01
w = 548	0.50	5.61E-01	9.81E-04	0.00E+00	2.01E-32	1.16E-01
	0.75	6.92E-01	4.21E-04	0.00E+00	1.65E-32	2.39E-03
	0.90	9.23E-01	2.85E-02	8.60E-13	1.53E-21	9.55E-01
	0.95	9.47E-01	6.81E-01	1.39E-11	3.03E-24	5.82E-02
	0.99	9.81E-01	2.72E-02	2.39E-05	4.23E-13	2.53E-06
# Rejections			4	6	6	2
	0.01	5.47E-03	1.78E-01	0.00E+00	1.00E+00	8.61E-02
	0.05	4.92E-02	9.25E-01	6.71E-01	3.28E-01	9.22E-01
	0.10	8.48E-02	1.61E-01	2.73E-01	2.30E-01	4.09E-01
VarGroup1,	0.25	2.28E-01	1.74E-01	6.98E-02	1.06E-04	6.54E-01
w = 730	0.50	5.62E-01	7.51E-04	0.00E+00	5.54E-31	3.38E-02
	0.75	7.44E-01	7.17E-01	0.00E+00	2.46E-38	2.16E-01
	0.90	9.08E-01	4.46E-01	5.88E-15	4.44E-26	1.71E-01
	0.95	9.40E-01	2.20E-01	1.45E-13	4.56E-30	5.41E-04
	0.99	9.81E-01	2.72E-02	6.79E-07	6.69E-19	4.88E-07
			2	6	6	3
	0.01	2.74E-03	1.93E-02	0.00E+00	1.00E+00	1.96E-03
	0.05	4.92E-02	9.25E-01	9.81E-01	7.37E-02	3.02E-01
	0.10	8.48E-02	1.61E-01	3.53E-01	2.89E-01	3.90E-01
VarGroup1,	0.25	2.38E-01	4.52E-01	7.32E-02	6.19E-05	7.72E-01
w = 913	0.50	5.53E-01	4.36E-03	0.00E+00	3.12E-32	1.49E-02
	0.75	7.46E-01	7.82E-01	0.00E+00	1.01E-33	8.68E-01
	0.90	8.97E-01	8.15E-01	0.00E+00	1.14E-35	3.55E-01
	0.95	9.32E-01	3.02E-02	0.00E+00	6.90E-39	7.59E-03
	0.99	9.82E-01	5.66E-02	5.13E-07	1.03E-22	1.52E-05
# Rejections			3	6	6	4

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.23E-02	5.44E-01	0.00E+00	9.97E-01	3.13E-03
	0.05	4.79E-02	7.91E-01	9.35E-01	3.99E-01	9.94E-01
	0.10	9.03E-02	3.74E-01	6.08E-01	5.23E-01	6.73E-01
VarGroup2,	0.25	2.46E-01	8.14E-01	6.12E-01	3.74E-02	7.91E-02
w = 365	0.50	5.42E-01	2.40E-02	0.00E+00	1.73E-36	1.23E-01
	0.75	6.81E-01	2.94E-05	0.00E+00	1.10E-29	7.88E-03
	0.90	8.60E-01	7.15E-04	1.74E-03	3.70E-01	4.46E-02
	0.95	9.64E-01	5.96E-02	1.69E-01	7.27E-02	3.88E-01
	0.99	9.77E-01	2.12E-03	6.30E-03	6.76E-01	2.56E-01
# Rejections			4	5	3	3
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	1.82E-02
	0.05	4.92E-02	9.25E-01	6.71E-01	3.73E-01	2.32E-01
	0.10	8.89E-02	3.10E-01	5.58E-01	5.23E-01	8.12E-01
VarGroup2,	0.25	2.44E-01	6.84E-01	1.38E-01	1.86E-03	1.58E-01
w = 548	0.50	5.61E-01	9.81E-04	0.00E+00	2.01E-32	1.16E-01
	0.75	6.92E-01	4.21E-04	0.00E+00	1.65E-32	2.39E-03
	0.90	8.62E-01	1.07E-03	5.48E-04	1.36E-03	2.68E-02
	0.95	9.56E-01	4.31E-01	3.39E-01	6.69E-02	3.59E-01
	0.99	9.82E-01	5.66E-02	7.74E-02	1.51E-03	1.78E-02
# Rejections			3	4	5	4
	0.01	1.09E-02	8.01E-01	0.00E+00	1.86E-01	2.61E-02
	0.05	4.92E-02	9.25E-01	3.07E-01	1.32E-01	1.65E-01
	0.10	8.48E-02	1.61E-01	2.73E-01	6.11E-01	6.07E-01
VarGroup2,	0.25	2.42E-01	6.22E-01	1.08E-01	7.81E-03	1.21E-01
w = 730	0.50	5.62E-01	7.51E-04	0.00E+00	5.54E-31	3.38E-02
	0.75	7.44E-01	7.17E-01	0.00E+00	2.46E-38	2.16E-01
	0.90	8.88E-01	2.81E-01	3.61E-02	5.28E-04	3.75E-01
	0.95	9.51E-01	9.25E-01	6.71E-01	1.56E-04	9.64E-01
	0.99	9.84E-01	1.11E-01	0.00E+00	1.90E-04	2.83E-02
# Rejections			1	5	6	3
	0.01	9.58E-03	9.08E-01	0.00E+00	1.76E-04	7.67E-04
	0.05	5.34E-02	6.81E-01	7.52E-01	1.19E-01	9.80E-01
	0.10	8.76E-02	2.53E-01	9.09E-02	3.75E-01	5.12E-01
VarGroup2,	0.25	2.61E-01	4.83E-01	3.35E-02	5.31E-05	2.63E-01
w = 913	0.50	5.53E-01	4.36E-03	0.00E+00	3.12E-32	1.49E-02
	0.75	7.46E-01	7.82E-01	0.00E+00	1.01E-33	8.68E-01
	0.90	8.80E-01	7.41E-02	5.08E-03	5.34E-03	8.48E-01
	0.95	9.52E-01	7.91E-01	6.00E-01	1.04E-04	6.72E-01
	0.99	9.82E-01	5.66E-02	6.19E-04	3.34E-08	4.36E-02
# Rejections			1	6	7	3

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.2: Traditional QR results for hour 8

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.05E-02	1.23E-02	0.00E+00	9.21E-01	3.31E-02
	0.05	4.65E-02	6.62E-01	5.17E-01	5.56E-03	7.90E-01
	0.10	1.09E-01	4.01E-01	3.96E-05	3.93E-37	5.60E-03
VarGroup1,	0.25	2.74E-01	1.45E-01	1.14E-12	2.54E-15	1.46E-01
w = 365	0.50	4.51E-01	8.59E-03	2.04E-13	5.91E-59	3.78E-03
	0.75	6.89E-01	2.25E-04	0.00E+00	4.85E-30	9.96E-04
	0.90	9.26E-01	1.40E-02	3.33E-16	7.96E-29	4.18E-02
	0.95	9.62E-01	1.31E-01	1.62E-08	2.13E-23	1.28E-02
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	5.77E-01
# Rejections			4	8	7	6
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	1.31E-02
	0.05	3.69E-02	8.98E-02	2.37E-01	2.48E-01	6.47E-01
	0.10	1.07E-01	5.50E-01	7.38E-05	2.43E-32	9.97E-03
VarGroup1,	0.25	2.46E-01	8.14E-01	7.87E-14	1.39E-16	9.31E-02
w = 548	0.50	4.64E-01	4.99E-02	1.55E-15	3.44E-60	1.07E-02
	0.75	6.85E-01	8.36E-05	0.00E+00	3.46E-31	1.31E-03
	0.90	9.29E-01	6.35E-03	3.33E-16	1.72E-32	7.87E-03
	0.95	9.62E-01	1.31E-01	6.89E-10	2.69E-27	2.88E-03
	0.99	9.90E-01	9.08E-01	1.48E-01	3.35E-15	3.81E-01
# Rejections			3	7	7	6
	0.01	9.58E-03	9.08E-01	0.00E+00	9.99E-01	2.73E-01
	0.05	3.15E-02	1.39E-02	4.60E-02	6.06E-01	9.98E-01
	0.10	1.07E-01	5.50E-01	2.72E-04	1.12E-29	4.75E-02
VarGroup1,	0.25	2.54E-01	7.82E-01	6.66E-16	9.75E-20	2.52E-02
w = 730	0.50	4.90E-01	5.79E-01	1.37E-14	6.91E-49	1.66E-01
	0.75	7.15E-01	3.35E-02	3.26E-13	6.64E-23	5.32E-02
	0.90	9.18E-01	9.66E-02	6.16E-13	1.30E-28	1.41E-04
	0.95	9.49E-01	9.39E-01	2.33E-15	1.11E-42	1.86E-05
	0.99	9.96E-01	6.92E-02	0.00E+00	6.07E-18	4.78E-01
			2	9	7	4
	0.01	1.23E-02	5.44E-01	1.10E-02	1.29E-08	8.04E-02
	0.05	5.47E-02	5.64E-01	8.96E-05	2.18E-09	6.54E-01
	0.10	1.07E-01	5.50E-01	3.15E-08	3.56E-11	3.15E-02
VarGroup1,	0.25	2.67E-01	2.99E-01	1.22E-15	4.85E-30	8.10E-01
w = 913	0.50	4.49E-01	5.50E-03	3.33E-16	1.33E-24	1.00E+00
	0.75	7.58E-01	6.22E-01	0.00E+00	1.98E-59	5.83E-01
	0.90	9.25E-01	2.01E-02	1.88E-14	7.16E-27	2.26E-01
	0.95	9.86E-01	1.11E-07	2.28E-07	8.31E-02	2.38E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	3.26E-01
# Rejections			4	9	7	1

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.05E-02	1.23E-02	0.00E+00	9.21E-01	3.31E-02
	0.05	4.65E-02	6.62E-01	5.17E-01	5.56E-03	7.90E-01
	0.10	9.17E-02	4.46E-01	3.64E-01	2.28E-06	5.77E-01
VarGroup2,	0.25	2.74E-01	1.45E-01	1.14E-12	2.54E-15	1.46E-01
w = 365	0.50	4.51E-01	8.59E-03	2.04E-13	5.91E-59	3.78E-03
	0.75	6.89E-01	2.25E-04	0.00E+00	4.85E-30	9.96E-04
	0.90	9.26E-01	1.40E-02	3.33E-16	7.96E-29	4.18E-02
	0.95	9.62E-01	1.31E-01	1.62E-08	2.13E-23	1.28E-02
	0.99	9.90E-01	9.08E-01	1.48E-01	1.19E-04	3.84E-01
# Rejections			4	6	8	5
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	1.31E-02
	0.05	3.69E-02	8.98E-02	2.37E-01	2.48E-01	6.47E-01
	0.10	8.76E-02	2.53E-01	4.28E-01	2.71E-04	6.34E-01
VarGroup2,	0.25	2.46E-01	8.14E-01	7.87E-14	1.39E-16	9.31E-02
w = 548	0.50	4.64E-01	4.99E-02	1.55E-15	3.44E-60	1.07E-02
	0.75	6.85E-01	8.36E-05	0.00E+00	3.46E-31	1.31E-03
	0.90	9.29E-01	6.35E-03	3.33E-16	1.72E-32	7.87E-03
	0.95	9.62E-01	1.31E-01	6.89E-10	2.69E-27	2.88E-03
	0.99	9.92E-01	6.15E-01	0.00E+00	6.51E-03	4.17E-01
# Rejections			3	7	7	5
	0.01	9.58E-03	9.08E-01	0.00E+00	9.99E-01	2.73E-01
	0.05	3.15E-02	1.39E-02	4.60E-02	6.06E-01	9.98E-01
	0.10	7.52E-02	2.01E-02	6.10E-02	1.96E-02	6.43E-01
VarGroup2,	0.25	2.54E-01	7.82E-01	6.66E-16	9.75E-20	2.52E-02
w = 730	0.50	4.90E-01	5.79E-01	1.37E-14	6.91E-49	1.66E-01
	0.75	7.15E-01	3.35E-02	3.26E-13	6.64E-23	5.32E-02
	0.90	9.18E-01	9.66E-02	6.16E-13	1.30E-28	1.41E-04
	0.95	9.49E-01	9.39E-01	2.33E-15	1.11E-42	1.86E-05
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	4.56E-01
# Rejections			3	8	6	3
	0.01	1.37E-02	3.44E-01	0.00E+00	1.34E-01	3.97E-02
	0.05	4.79E-02	7.91E-01	0.00E+00	1.32E-03	7.01E-01
	0.10	9.99E-02	9.90E-01	9.92E-01	3.54E-02	3.43E-01
VarGroup2,	0.25	2.52E-01	9.15E-01	5.46E-01	2.19E-09	5.76E-01
w = 913	0.50	5.10E-01	5.79E-01	4.69E-01	4.72E-01	1.21E-01
	0.75	7.78E-01	7.25E-02	1.36E-02	1.26E-01	3.40E-03
	0.90	9.29E-01	6.35E-03	2.49E-03	1.30E-01	9.43E-01
	0.95	9.93E-01	2.39E-11	1.50E-11	1.25E-04	7.26E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	2.09E-01
# Rejections			3	6	4	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.3: Traditional QR results for hour 19

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.37E-02	3.44E-01	1.92E-01	8.89E-07	3.17E-07
	0.05	8.07E-02	4.44E-04	7.86E-09	4.63E-10	8.50E-03
	0.10	1.22E-01	5.71E-02	4.76E-11	2.73E-14	1.73E-02
VarGroup1,	0.25	2.79E-01	7.30E-02	0.00E+00	4.51E-35	4.19E-02
w = 365	0.50	4.45E-01	2.71E-03	0.00E+00	4.72E-35	3.14E-02
	0.75	7.66E-01	3.12E-01	0.00E+00	7.09E-55	2.36E-01
	0.90	9.17E-01	1.26E-01	1.11E-16	8.23E-32	2.99E-01
	0.95	9.75E-01	5.06E-04	4.79E-04	2.88E-02	2.99E-01
	0.99	9.93E-01	3.62E-01	4.86E-02	1.48E-13	4.27E-01
# Rejections			3	8	9	5
	0.01	1.37E-02	3.44E-01	1.32E-02	2.13E-08	1.23E-03
	0.05	5.34E-02	6.81E-01	4.27E-04	4.26E-06	1.96E-01
	0.10	1.09E-01	4.01E-01	4.75E-07	8.43E-09	1.17E-01
VarGroup1,	0.25	2.63E-01	4.32E-01	3.19E-14	2.34E-27	3.31E-01
w = 548	0.50	4.47E-01	4.36E-03	0.00E+00	6.67E-37	1.50E-01
	0.75	7.66E-01	3.12E-01	0.00E+00	2.43E-60	3.03E-01
	0.90	9.23E-01	2.85E-02	3.33E-16	1.00E-29	1.38E-01
	0.95	9.81E-01	1.34E-05	4.10E-05	3.52E-01	6.00E-01
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	2.53E-01
# Rejections			3	9	7	1
	0.01	1.23E-02	5.44E-01	1.10E-02	8.80E-11	4.47E-03
	0.05	4.92E-02	9.25E-01	5.27E-03	7.51E-05	3.17E-01
	0.10	1.04E-01	7.22E-01	3.16E-07	2.65E-09	1.63E-02
VarGroup1,	0.25	2.72E-01	1.69E-01	8.99E-15	2.05E-28	4.11E-01
w = 730	0.50	4.66E-01	6.98E-02	0.00E+00	6.11E-32	7.30E-01
	0.75	7.59E-01	5.63E-01	0.00E+00	3.27E-63	9.80E-02
	0.90	9.22E-01	3.97E-02	1.11E-15	4.55E-28	5.99E-02
	0.95	9.86E-01	1.11E-07	2.28E-07	8.89E-03	9.76E-01
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	7.25E-01
			3	9	8	2
	0.01	1.23E-02	5.44E-01	1.10E-02	1.29E-08	8.04E-02
	0.05	5.47E-02	5.64E-01	8.96E-05	2.18E-09	6.54E-01
	0.10	1.07E-01	5.50E-01	3.15E-08	3.56E-11	3.15E-02
VarGroup1,	0.25	2.67E-01	2.99E-01	1.22E-15	4.85E-30	8.10E-01
w = 913	0.50	4.49E-01	5.50E-03	3.33E-16	1.33E-24	1.00E+00
	0.75	7.58E-01	6.22E-01	0.00E+00	1.98E-59	5.83E-01
	0.90	9.25E-01	2.01E-02	1.88E-14	7.16E-27	2.26E-01
	0.95	9.86E-01	1.11E-07	2.28E-07	8.31E-02	2.38E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	3.26E-01
# Rejections			4	9	7	1

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.64E-02	1.11E-01	0.00E+00	2.99E-03	2.50E-02
	0.05	6.70E-02	4.41E-02	3.91E-02	4.48E-03	1.51E-01
	0.10	1.03E-01	8.15E-01	5.05E-01	2.77E-03	4.69E-02
VarGroup2,	0.25	2.59E-01	5.95E-01	4.43E-01	3.45E-09	4.70E-01
w = 365	0.50	5.05E-01	7.96E-01	2.28E-01	1.08E-01	3.62E-02
	0.75	7.40E-01	5.38E-01	2.21E-01	3.06E-01	3.38E-02
	0.90	9.19E-01	7.30E-02	1.32E-02	2.55E-02	1.15E-01
	0.95	9.78E-01	9.33E-05	1.81E-08	9.01E-16	1.43E-01
	0.99	9.96E-01	6.92E-02	4.38E-03	5.21E-18	7.73E-01
# Rejections			2	5	7	4
	0.01	1.23E-02	5.44E-01	0.00E+00	3.24E-02	9.21E-03
	0.05	5.20E-02	8.07E-01	9.70E-01	2.80E-04	5.39E-02
	0.10	9.58E-02	7.01E-01	9.22E-01	4.30E-04	2.00E-01
VarGroup2,	0.25	2.48E-01	8.81E-01	5.95E-01	1.06E-08	5.55E-01
w = 548	0.50	4.99E-01	9.70E-01	4.95E-01	6.61E-01	1.34E-01
	0.75	7.55E-01	7.48E-01	1.31E-01	5.58E-01	3.70E-04
	0.90	9.19E-01	7.30E-02	1.50E-03	1.33E-02	2.40E-01
	0.95	9.85E-01	4.18E-07	7.84E-11	5.81E-26	8.27E-01
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	9.31E-01
# Rejections			1	4	6	2
	0.01	1.37E-02	3.44E-01	0.00E+00	1.34E-01	1.23E-02
	0.05	5.61E-02	4.58E-01	4.55E-01	3.53E-04	1.15E-01
	0.10	9.71E-02	7.95E-01	9.66E-01	2.91E-03	1.99E-01
VarGroup2,	0.25	2.49E-01	9.49E-01	6.61E-01	1.17E-09	6.59E-01
w = 730	0.50	5.01E-01	9.70E-01	7.60E-01	7.56E-01	1.05E-01
	0.75	7.61E-01	5.06E-01	3.83E-01	7.55E-01	3.12E-03
	0.90	9.30E-01	4.16E-03	3.48E-03	4.86E-01	6.01E-01
	0.95	9.92E-01	1.74E-10	1.54E-10	6.61E-03	4.42E-01
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	9.53E-01
# Rejections			3	4	4	2
	0.01	1.37E-02	3.44E-01	0.00E+00	1.34E-01	3.97E-02
	0.05	4.79E-02	7.91E-01	0.00E+00	1.32E-03	7.01E-01
	0.10	9.99E-02	9.90E-01	9.92E-01	3.54E-02	3.43E-01
VarGroup2,	0.25	2.52E-01	9.15E-01	5.46E-01	2.19E-09	5.76E-01
w = 913	0.50	5.10E-01	5.79E-01	4.69E-01	4.72E-01	1.21E-01
	0.75	7.78E-01	7.25E-02	1.36E-02	1.26E-01	3.40E-03
	0.90	9.29E-01	6.35E-03	2.49E-03	1.30E-01	9.43E-01
	0.95	9.93E-01	2.39E-11	1.50E-11	1.25E-04	7.26E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	2.09E-01
# Rejections			3	6	4	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

xii

Table D.4: EWQR results for window sizes 365 and 548

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.37E-02	3.44E-01	1.92E-01	4.97E-01	3.58E-05
	0.05	4.79E-02	7.91E-01	9.35E-01	3.99E-01	9.94E-01
	0.10	9.71E-02	7.95E-01	8.73E-01	5.53E-01	4.15E-01
Hour 3	0.25	2.49E-01	9.49E-01	9.32E-01	8.71E-03	6.80E-02
w = 365	0.50	4.75E-01	1.71E-01	0.00E+00	2.58E-24	7.71E-03
	0.75	6.87E-01	1.17E-04	0.00E+00	2.13E-25	3.31E-03
	0.90	8.52E-01	5.15E-05	1.42E-04	2.16E-01	6.52E-03
	0.95	9.63E-01	8.98E-02	0.00E+00	6.20E-01	6.99E-01
	0.99	9.77E-01	2.12E-03	6.30E-03	6.76E-01	2.56E-01
# Rejections			3	5	3	4
	0.01	2.19E-02	5.25E-03	0.00E+00	8.33E-01	3.31E-02
	0.05	4.79E-02	7.91E-01	6.00E-01	9.37E-03	7.08E-01
	0.10	9.30E-02	5.25E-01	7.84E-01	1.62E-05	8.69E-01
Hour 8	0.25	2.60E-01	5.38E-01	2.14E-11	7.21E-11	3.33E-01
w = 365	0.50	4.72E-01	1.29E-01	2.34E-12	3.52E-54	2.91E-03
	0.75	7.05E-01	5.31E-03	1.10E-10	3.11E-17	1.49E-03
	0.90	9.12E-01	2.53E-01	2.63E-12	7.48E-21	4.63E-01
	0.95	9.41E-01	2.86E-01	7.27E-13	2.47E-28	8.62E-02
	0.99	9.78E-01	5.25E-03	2.71E-03	1.45E-02	2.07E-03
# Rejections			3	7	8	4
	0.01	1.64E-02	1.11E-01	0.00E+00	3.63E-01	2.58E-02
	0.05	6.57E-02	6.32E-02	5.85E-02	5.40E-04	5.44E-03
	0.10	1.16E-01	1.52E-01	2.80E-01	1.91E-02	9.33E-03
Hour 19	0.25	2.56E-01	7.17E-01	4.70E-01	4.32E-07	4.05E-01
w = 365	0.50	5.03E-01	8.53E-01	7.07E-01	6.33E-01	7.39E-02
	0.75	7.36E-01	3.84E-01	1.18E-01	2.19E-01	2.38E-02
	0.90	9.02E-01	8.92E-01	7.78E-02	6.55E-02	6.79E-02
	0.95	9.59E-01	2.52E-01	2.97E-10	5.45E-24	2.77E-01
	0.99	9.95E-01	1.78E-01	1.81E-02	2.76E-08	8.19E-01
			0	3	5	4

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.23E-02	5.44E-01	0.00E+00	3.33E-01	7.11E-02
	0.05	4.92E-02	9.25E-01	6.71E-01	3.73E-01	2.32E-01
	0.10	9.17E-02	4.46E-01	6.52E-01	6.05E-01	9.71E-01
Hour 3	0.25	2.42E-01	6.22E-01	6.08E-01	1.43E-02	1.48E-01
w = 548	0.50	4.77E-01	2.22E-01	0.00E+00	1.14E-23	8.20E-03
	0.75	6.87E-01	1.17E-04	0.00E+00	5.96E-27	7.58E-03
	0.90	8.62E-01	1.07E-03	5.48E-04	1.36E-03	2.68E-02
	0.95	9.56E-01	4.31E-01	3.39E-01	6.69E-02	3.59E-01
	0.99	9.82E-01	5.66E-02	7.74E-02	1.51E-03	1.78E-02
# Rejections			2	4	5	4
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	1.31E-02
	0.05	4.92E-02	9.25E-01	9.81E-01	6.87E-01	6.27E-01
	0.10	9.30E-02	5.25E-01	6.37E-01	2.83E-04	7.79E-01
Hour 8	0.25	2.60E-01	5.38E-01	5.87E-12	5.47E-12	3.54E-01
w = 548	0.50	4.69E-01	9.59E-02	2.03E-12	6.18E-52	7.58E-03
	0.75	7.10E-01	1.40E-02	3.72E-10	1.04E-14	9.43E-04
	0.90	9.12E-01	2.53E-01	2.63E-12	7.48E-21	4.78E-01
	0.95	9.41E-01	2.86E-01	1.35E-11	1.11E-24	5.71E-02
	0.99	9.82E-01	5.66E-02	0.00E+00	2.27E-03	4.33E-04
# Rejections			1	7	7	4
	0.01	1.50E-02	2.02E-01	0.00E+00	2.22E-01	3.22E-02
	0.05	6.57E-02	6.32E-02	5.85E-02	2.02E-03	2.73E-02
	0.10	1.05E-01	6.33E-01	8.07E-01	9.47E-02	3.81E-03
Hour 19	0.25	2.50E-01	9.83E-01	7.22E-01	2.30E-07	5.47E-01
w = 548	0.50	5.01E-01	9.70E-01	3.34E-01	6.10E-01	7.22E-02
	0.75	7.39E-01	4.83E-01	8.91E-02	2.57E-01	1.99E-03
	0.90	9.17E-01	1.26E-01	1.22E-02	1.57E-01	1.12E-01
	0.95	9.60E-01	1.84E-01	2.25E-09	4.22E-21	1.92E-01
	0.99	9.95E-01	1.78E-01	1.81E-02	2.76E-08	7.78E-01
# Rejections			0	4	4	4

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.5: EWQR results for window sizes 730 and 913

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.37E-02	3.44E-01	0.00E+00	4.95E-01	9.23E-02
	0.05	5.06E-02	9.39E-01	3.52E-01	1.85E-01	2.70E-01
	0.10	9.44E-02	6.10E-01	7.24E-01	7.26E-01	7.97E-01
Hour 3	0.25	2.46E-01	8.14E-01	8.27E-01	1.41E-02	1.50E-01
w = 730	0.50	4.77E-01	2.22E-01	0.00E+00	1.14E-23	8.24E-03
	0.75	6.87E-01	1.17E-04	0.00E+00	5.96E-27	7.97E-03
	0.90	8.82E-01	1.21E-01	4.32E-02	2.70E-02	1.22E-01
	0.95	9.51E-01	9.25E-01	6.71E-01	1.56E-04	9.64E-01
	0.99	9.82E-01	5.66E-02	7.74E-02	1.60E-01	4.90E-01
# Rejections			1	4	5	2
	0.01	1.23E-02	5.44E-01	0.00E+00	9.97E-01	3.30E-01
	0.05	4.79E-02	7.91E-01	9.35E-01	6.77E-01	6.50E-01
	0.10	9.85E-02	8.92E-01	3.08E-01	9.48E-04	1.23E-01
Hour 8	0.25	2.60E-01	5.38E-01	8.48E-10	4.87E-10	2.86E-01
w = 730	0.50	4.66E-01	6.98E-02	5.07E-12	6.05E-52	7.33E-03
	0.75	7.10E-01	1.40E-02	3.72E-10	2.90E-15	1.28E-03
	0.90	9.12E-01	2.53E-01	2.63E-12	7.48E-21	4.84E-01
	0.95	9.44E-01	4.58E-01	7.71E-11	7.32E-24	1.44E-01
	0.99	9.86E-01	3.44E-01	0.00E+00	7.70E-02	1.34E-04
# Rejections			1	7	6	3
	0.01	1.09E-02	8.01E-01	0.00E+00	1.88E-01	2.12E-02
	0.05	5.75E-02	3.66E-01	6.37E-01	1.32E-04	6.41E-02
	0.10	1.09E-01	4.01E-01	5.54E-01	6.63E-02	5.00E-04
Hour 19	0.25	2.48E-01	8.81E-01	7.11E-01	1.79E-07	8.25E-01
w = 730	0.50	4.95E-01	7.96E-01	7.75E-01	6.44E-01	1.21E-01
	0.75	7.46E-01	7.82E-01	3.28E-01	8.23E-01	2.25E-03
	0.90	9.17E-01	1.26E-01	1.22E-02	1.21E-01	2.39E-01
	0.95	9.60E-01	1.84E-01	2.25E-09	4.22E-21	1.73E-01
	0.99	9.95E-01	1.78E-01	1.81E-02	2.76E-08	9.38E-01
			0	4	4	3

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.37E-02	3.44E-01	0.00E+00	4.95E-01	1.21E-01
	0.05	5.34E-02	6.81E-01	7.52E-01	1.19E-01	9.80E-01
	0.10	9.30E-02	5.25E-01	7.84E-01	7.84E-01	9.90E-01
Hour 3	0.25	2.46E-01	8.14E-01	7.25E-01	1.75E-02	2.47E-01
w = 913	0.50	4.79E-01	2.51E-01	0.00E+00	1.82E-24	8.87E-03
	0.75	6.88E-01	1.63E-04	0.00E+00	9.25E-28	7.37E-03
	0.90	8.77E-01	4.35E-02	5.37E-03	6.67E-02	2.00E-01
	0.95	9.47E-01	6.81E-01	4.14E-01	1.02E-04	9.08E-01
	0.99	9.79E-01	1.23E-02	2.59E-02	4.66E-01	8.71E-01
# Rejections			3	5	4	2
	0.01	1.37E-02	3.44E-01	0.00E+00	9.93E-01	1.10E-01
	0.05	3.97E-02	1.84E-01	4.10E-01	3.98E-01	4.08E-01
	0.10	9.71E-02	7.95E-01	4.48E-01	1.36E-04	3.34E-01
Hour 8	0.25	2.61E-01	4.83E-01	3.42E-12	2.41E-12	3.53E-01
w = 913	0.50	4.69E-01	9.59E-02	2.03E-12	6.18E-52	7.45E-03
	0.75	7.10E-01	1.40E-02	3.72E-10	2.90E-15	1.30E-03
	0.90	9.12E-01	2.53E-01	2.63E-12	7.48E-21	4.84E-01
	0.95	9.44E-01	4.58E-01	7.71E-11	7.32E-24	1.43E-01
	0.99	9.82E-01	5.66E-02	0.00E+00	2.27E-03	8.47E-04
# Rejections			1	7	7	3
	0.01	1.50E-02	2.02E-01	0.00E+00	2.60E-01	2.24E-02
	0.05	4.79E-02	7.91E-01	0.00E+00	1.32E-03	7.01E-01
	0.10	1.07E-01	5.50E-01	5.36E-01	1.38E-01	2.09E-03
Hour 19	0.25	2.52E-01	9.15E-01	6.62E-01	2.53E-10	9.47E-01
w = 913	0.50	5.10E-01	5.79E-01	7.55E-01	8.38E-01	2.19E-01
	0.75	7.35E-01	3.40E-01	2.38E-01	9.14E-01	7.74E-03
	0.90	9.21E-01	5.43E-02	8.21E-02	5.59E-01	5.27E-01
	0.95	9.60E-01	1.84E-01	2.25E-09	4.22E-21	1.65E-01
	0.99	9.95E-01	1.78E-01	1.81E-02	2.76E-08	9.31E-01
# Rejections			0	4	4	3

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

xiv

Table D.6: EWDKQR results for window sizes 250 and 365

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.23E-02	5.44E-01	0.00E+00	3.30E-01	1.77E-04
	0.05	6.02E-02	2.20E-01	0.00E+00	6.30E-36	4.53E-06
	0.10	9.85E-02	8.92E-01	9.18E-14	1.69E-25	2.58E-06
Hour 3	0.25	2.57E-01	6.55E-01	0.00E+00	3.31E-52	1.46E-11
w = 250	0.50	5.03E-01	8.53E-01	0.00E+00	1.42E-111	4.62E-24
	0.75	7.14E-01	2.72E-02	0.00E+00	3.53E-97	3.11E-25
	0.90	8.78E-01	5.71E-02	0.00E+00	3.00E-53	2.55E-14
	0.95	9.36E-01	8.88E-02	4.55E-15	5.15E-35	1.32E-11
	0.99	9.82E-01	5.66E-02	6.19E-04	1.81E-16	3.24E-05
# Rejections			1	9	8	9
	0.01	1.50E-02	2.02E-01	1.37E-02	9.01E-08	4.42E-01
	0.05	5.20E-02	8.07E-01	7.50E-01	2.35E-04	9.10E-01
	0.10	9.99E-02	9.90E-01	1.65E-06	3.50E-07	8.85E-01
Hour 8	0.25	2.50E-01	9.83E-01	1.40E-02	1.11E-59	1.47E-01
w = 250	0.50	5.06E-01	7.39E-01	0.00E+00	4.43E-29	4.71E-01
	0.75	7.47E-01	8.48E-01	6.50E-07	5.43E-08	5.04E-08
	0.90	9.06E-01	6.10E-01	2.55E-15	3.80E-28	3.67E-01
	0.95	9.53E-01	6.62E-01	3.49E-06	2.82E-17	4.76E-11
	0.99	9.88E-01	5.44E-01	2.05E-01	7.89E-04	9.62E-01
# Rejections			0	7	9	2
	0.01	1.23E-02	5.44E-01	2.05E-01	4.70E-02	1.13E-02
	0.05	5.47E-02	5.64E-01	6.03E-04	4.48E-08	8.68E-05
	0.10	1.11E-01	3.38E-01	2.23E-12	2.36E-17	2.62E-02
Hour 19	0.25	2.50E-01	9.83E-01	1.11E-16	7.09E-21	5.43E-01
w = 250	0.50	4.99E-01	9.70E-01	0.00E+00	2.10E-40	4.16E-01
	0.75	7.29E-01	1.97E-01	0.00E+00	4.88E-41	2.38E-01
	0.90	8.99E-01	9.12E-01	3.33E-16	1.40E-28	7.20E-01
	0.95	9.59E-01	2.52E-01	1.95E-03	2.59E-05	6.15E-01
	0.99	9.89E-01	8.01E-01	1.89E-01	3.03E-03	4.23E-01
			0	7	9	3

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	7.49E-02
	0.05	3.97E-02	1.84E-01	3.13E-01	8.41E-01	8.72E-01
	0.10	8.62E-02	2.03E-01	4.30E-01	4.58E-01	4.73E-01
Hour 3	0.25	2.37E-01	4.02E-01	0.00E+00	3.98E-43	2.79E-10
w = 365	0.50	4.90E-01	5.79E-01	4.45E-02	3.44E-02	3.83E-08
	0.75	7.35E-01	3.40E-01	0.00E+00	6.60E-85	1.48E-26
	0.90	9.53E-01	9.73E-08	1.53E-07	1.84E-01	4.63E-18
	0.95	9.53E-01	6.62E-01	4.51E-11	2.12E-25	2.17E-12
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	2.29E-07
# Rejections			1	7	4	6
	0.01	2.05E-02	1.23E-02	4.10E-04	6.08E-07	7.34E-02
	0.05	4.79E-02	7.91E-01	2.57E-01	1.59E-03	9.06E-01
	0.10	1.16E-01	1.52E-01	3.53E-06	3.64E-07	4.34E-01
Hour 8	0.25	2.48E-01	8.81E-01	1.31E-02	1.07E-58	2.10E-01
w = 365	0.50	5.16E-01	3.95E-01	0.00E+00	1.38E-29	3.74E-01
	0.75	7.54E-01	8.14E-01	1.83E-06	9.58E-08	5.89E-08
	0.90	9.10E-01	3.74E-01	1.11E-15	6.47E-28	4.47E-01
	0.95	9.56E-01	4.31E-01	7.60E-07	2.27E-18	1.47E-10
	0.99	9.89E-01	8.01E-01	1.89E-01	9.61E-03	4.12E-01
# Rejections			1	7	9	2
	0.01	1.09E-02	8.01E-01	1.89E-01	1.70E-01	1.25E-01
	0.05	4.79E-02	7.91E-01	0.00E+00	8.60E-02	3.97E-01
	0.10	9.71E-02	7.95E-01	9.66E-01	4.56E-03	7.12E-02
Hour 19	0.25	2.56E-01	7.17E-01	1.55E-15	2.13E-18	5.74E-01
w = 365	0.50	5.05E-01	7.96E-01	1.32E-01	2.24E-10	2.71E-02
	0.75	7.47E-01	8.48E-01	0.00E+00	3.66E-41	5.41E-01
	0.90	9.15E-01	1.61E-01	9.68E-02	5.92E-02	4.90E-02
	0.95	9.70E-01	7.85E-03	3.54E-04	1.63E-04	5.97E-01
	0.99	9.93E-01	3.62E-01	4.86E-02	1.25E-04	7.33E-01
			1	5	6	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.7: EWDKQR results for window sizes 548 and 730

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	5.47E-03	1.78E-01	0.00E+00	1.00E+00	1.01E-01
	0.05	2.74E-02	2.20E-03	2.64E-03	3.52E-01	9.30E-01
	0.10	1.05E-01	6.33E-01	8.31E-13	2.94E-24	1.92E-06
Hour 3	0.25	2.01E-01	1.76E-03	4.34E-03	2.78E-01	4.63E-02
w = 548	0.50	5.05E-01	7.96E-01	0.00E+00	1.47E-108	3.56E-23
	0.75	8.33E-01	5.90E-08	8.27E-14	1.68E-09	2.70E-18
	0.90	9.07E-01	5.25E-01	0.00E+00	2.64E-39	1.00E-16
	0.95	9.73E-01	2.20E-03	2.64E-03	3.72E-01	3.23E-17
	0.99	9.96E-01	6.92E-02	0.00E+00	3.10E-15	4.80E-03
# Rejections			0	0	0	0
	0.01	1.92E-02	2.72E-02	5.33E-04	1.17E-08	3.32E-01
	0.05	4.79E-02	7.91E-01	2.57E-01	1.47E-04	9.46E-01
	0.10	1.15E-01	1.88E-01	5.71E-07	5.97E-08	2.42E-01
Hour 8	0.25	2.48E-01	8.81E-01	1.31E-02	1.07E-58	2.07E-01
w = 548	0.50	5.21E-01	2.51E-01	0.00E+00	1.51E-30	3.95E-01
	0.75	7.65E-01	3.55E-01	0.00E+00	7.26E-30	3.28E-01
	0.90	9.12E-01	2.53E-01	1.44E-15	2.42E-27	4.04E-01
	0.95	9.55E-01	5.40E-01	1.20E-07	7.58E-19	1.82E-10
	0.99	9.92E-01	6.15E-01	0.00E+00	6.60E-03	6.74E-01
# Rejections			0	0	0	0
	0.01	1.23E-02	5.44E-01	2.05E-01	3.34E-01	4.24E-02
	0.05	4.92E-02	9.25E-01	0.00E+00	2.00E-03	2.28E-01
	0.10	9.58E-02	7.01E-01	8.86E-01	2.79E-02	9.76E-02
Hour 19	0.25	2.60E-01	5.38E-01	1.22E-15	5.06E-18	5.67E-01
w = 548	0.50	5.06E-01	7.39E-01	1.29E-01	6.11E-10	3.59E-02
	0.75	7.51E-01	9.49E-01	0.00E+00	1.28E-38	4.68E-01
	0.90	9.15E-01	1.61E-01	9.68E-02	5.92E-02	5.06E-02
	0.95	9.71E-01	4.26E-03	1.42E-04	4.41E-05	4.07E-01
	0.99	9.89E-01	8.01E-01	0.00E+00	4.57E-11	5.26E-07
# Rejections			1	5	7	3

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	6.95E-02
	0.05	5.34E-02	6.81E-01	4.19E-09	3.51E-17	1.55E-05
	0.10	9.71E-02	7.95E-01	3.64E-13	1.23E-22	3.49E-06
Hour 3	0.25	2.49E-01	9.49E-01	0.00E+00	1.33E-49	6.51E-11
w = 730	0.50	4.88E-01	5.29E-01	2.96E-02	5.48E-02	2.48E-08
	0.75	7.46E-01	7.82E-01	0.00E+00	3.53E-87	9.52E-27
	0.90	9.33E-01	1.68E-03	3.87E-04	3.04E-03	3.82E-17
	0.95	9.55E-01	5.40E-01	1.48E-11	1.39E-26	2.57E-12
	0.99	9.96E-01	6.92E-02	0.00E+00	3.10E-15	4.57E-03
# Rejections			0	0	0	0
	0.01	1.64E-02	1.11E-01	1.23E-02	1.47E-03	1.02E-01
	0.05	4.79E-02	7.91E-01	6.00E-01	3.12E-03	6.45E-01
	0.10	1.16E-01	1.52E-01	3.53E-06	1.03E-06	1.69E-01
Hour 8	0.25	3.05E-01	7.72E-04	0.00E+00	1.50E-24	9.34E-01
w = 730	0.50	5.24E-01	1.95E-01	0.00E+00	1.83E-33	3.11E-01
	0.75	7.65E-01	3.55E-01	0.00E+00	7.26E-30	3.25E-01
	0.90	9.12E-01	2.53E-01	1.44E-15	2.42E-27	4.04E-01
	0.95	9.55E-01	5.40E-01	3.55E-10	1.03E-28	1.65E-01
	0.99	9.93E-01	3.62E-01	0.00E+00	1.26E-04	8.85E-01
# Rejections			0	0	0	0
	0.01	1.50E-02	2.02E-01	1.58E-01	2.56E-01	7.92E-03
	0.05	4.79E-02	7.91E-01	0.00E+00	2.08E-02	3.10E-01
	0.10	1.14E-01	2.31E-01	7.73E-08	1.28E-09	1.62E-01
Hour 19	0.25	2.59E-01	5.95E-01	4.44E-16	1.08E-18	6.66E-01
w = 730	0.50	5.08E-01	6.84E-01	0.00E+00	2.05E-37	6.84E-01
	0.75	7.54E-01	8.14E-01	0.00E+00	8.05E-53	1.18E-01
	0.90	9.11E-01	3.10E-01	2.32E-01	3.23E-02	8.14E-01
	0.95	9.73E-01	2.20E-03	4.70E-04	4.51E-04	4.74E-01
	0.99	9.93E-01	3.62E-01	4.86E-02	1.25E-04	7.16E-01
# Rejections			1	7	8	1

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

XVI

Table D.8: GARCH results for window sizes 365 and 548

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	6.84E-03	3.62E-01	0.00E+00	1.00E+00	1.80E-01
	0.05	6.29E-02	1.22E-01	3.03E-01	3.54E-01	5.66E-01
	0.10	1.07E-01	5.50E-01	5.09E-01	6.24E-01	9.45E-01
Hour 3	0.25	2.38E-01	4.52E-01	7.26E-04	1.49E-05	4.28E-02
w = 365	0.50	8.81E-01	0.00E+00	0.00E+00	7.93E-13	1.46E-02
	0.75	6.13E-01	3.33E-16	0.00E+00	3.60E-16	2.03E-06
	0.90	8.84E-01	1.52E-01	1.18E-01	3.11E-01	2.29E-09
	0.95	9.49E-01	9.39E-01	7.62E-01	6.04E-02	1.37E-06
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	2.07E-02
# Rejections			2	5	3	6
	0.01	5.20E-02	6.66E-16	0.00E+00	1.05E-222	1.76E-04
	0.05	6.29E-02	1.22E-01	0.00E+00	2.68E-299	6.36E-04
	0.10	1.07E-01	5.50E-01	1.96E-01	2.38E-01	1.58E-02
Hour 8	0.25	7.25E-02	0.00E+00	0.00E+00	0.00E+00	2.17E-03
w = 365	0.50	5.55E-01	2.71E-03	0.00E+00	9.17E-41	7.76E-03
	0.75	9.58E-02	0.00E+00	0.00E+00	2.78E-283	2.30E-03
	0.90	1.12E-01	0.00E+00	0.00E+00	8.98E-256	2.99E-05
	0.95	4.10E-03	0.00E+00	0.00E+00	6.07E-18	6.64E-02
	0.99	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
# Rejections			7	8	8	8
	0.01	9.58E-03	9.08E-01	0.00E+00	9.99E-01	2.51E-01
	0.05	4.65E-02	6.62E-01	0.00E+00	1.28E-01	3.98E-01
	0.10	1.09E-01	4.01E-01	5.02E-01	2.37E-01	7.92E-01
Hour 19	0.25	2.54E-01	7.82E-01	1.20E-01	1.43E-10	7.69E-01
w = 365	0.50	5.95E-01	2.51E-07	5.96E-07	5.91E-01	1.09E-01
	0.75	7.36E-01	3.84E-01	5.16E-01	2.29E-01	1.13E-02
	0.90	9.18E-01	9.66E-02	2.21E-01	9.63E-01	7.53E-01
	0.95	9.85E-01	4.18E-07	9.81E-07	2.37E-04	8.82E-01
	0.99	1.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
			3	5	3	2

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	5.47E-03	1.78E-01	0.00E+00	1.00E+00	9.47E-02
	0.05	6.16E-02	1.66E-01	3.79E-01	3.65E-01	4.91E-01
	0.10	1.08E-01	4.72E-01	6.66E-01	8.23E-01	5.78E-01
Hour 3	0.25	2.46E-01	8.14E-01	4.36E-03	6.59E-05	9.56E-02
w = 548	0.50	9.04E-01	0.00E+00	0.00E+00	1.70E-13	3.17E-01
	0.75	6.28E-01	2.97E-13	0.00E+00	1.23E-24	1.56E-07
	0.90	9.08E-01	4.46E-01	2.33E-01	3.17E-02	6.31E-06
	0.95	9.49E-01	9.39E-01	0.00E+00	2.07E-04	6.93E-04
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	1.63E-01
# Rejections			3	6	5	3
	0.01	3.06E-01	0.00E+00	0.00E+00	1.33E-87	6.40E-08
	0.05	2.74E-03	1.44E-14	0.00E+00	1.00E+00	5.99E-03
	0.10	1.18E-01	1.21E-01	2.43E-01	5.84E-17	2.50E-01
Hour 8	0.25	2.34E-01	3.12E-01	4.29E-09	3.61E-10	1.16E-01
w = 548	0.50	7.72E-01	0.00E+00	0.00E+00	4.17E-27	1.78E-04
	0.75	6.70E-01	1.40E-06	0.00E+00	2.95E-35	4.25E-04
	0.90	1.46E-01	0.00E+00	0.00E+00	2.05E-224	1.91E-04
	0.95	1.52E-01	0.00E+00	0.00E+00	1.46E-207	1.70E-04
	0.99	9.92E-01	6.15E-01	0.00E+00	8.87E-03	5.30E-01
# Rejections			6	8	8	6
	0.01	1.23E-02	5.44E-01	0.00E+00	9.97E-01	3.04E-01
	0.05	4.79E-02	7.91E-01	0.00E+00	6.93E-02	1.44E-01
	0.10	9.71E-02	7.95E-01	9.66E-01	2.21E-01	9.66E-01
Hour 19	0.25	2.45E-01	7.48E-01	1.89E-01	1.21E-10	4.29E-01
w = 548	0.50	5.95E-01	2.51E-07	1.04E-06	7.55E-01	3.21E-01
	0.75	7.66E-01	3.12E-01	3.60E-01	4.58E-01	2.89E-01
	0.90	9.26E-01	1.40E-02	4.25E-02	5.15E-01	3.31E-01
	0.95	9.85E-01	4.18E-07	9.81E-07	2.37E-04	8.82E-01
	0.99	1.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
# Rejections			4	6	3	1

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.9: GARCH results for window sizes 730 and 913

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	6.68E-02
	0.05	2.87E-02	4.26E-03	1.50E-02	2.69E-01	2.97E-01
	0.10	6.16E-02	2.12E-04	9.18E-04	9.80E-01	8.33E-01
Hour 3	0.25	1.59E-01	2.11E-09	3.20E-09	4.91E-01	2.41E-04
w = 730	0.50	5.85E-01	3.57E-06	2.97E-07	8.88E-04	2.87E-14
	0.75	6.66E-01	4.05E-07	4.44E-07	3.38E-03	5.77E-13
	0.90	8.45E-01	4.34E-06	2.58E-05	1.50E-01	2.40E-10
	0.95	9.32E-01	3.02E-02	6.58E-02	4.62E-01	7.11E-09
	0.99	9.95E-01	1.78E-01	0.00E+00	1.00E+00	1.08E-01
# Rejections			7	8	2	5
	0.01	1.37E-02	3.44E-01	0.00E+00	5.08E-01	2.09E-06
	0.05	3.15E-02	1.39E-02	4.60E-02	9.17E-01	1.93E-05
	0.10	7.39E-02	1.40E-02	2.67E-04	1.60E-03	4.00E-02
Hour 8	0.25	1.86E-01	3.72E-05	9.23E-08	1.25E-03	2.10E-01
w = 730	0.50	5.73E-01	7.35E-05	1.46E-06	8.50E-04	1.52E-08
	0.75	7.06E-01	6.82E-03	1.49E-05	7.40E-04	7.29E-15
	0.90	8.62E-01	1.07E-03	2.92E-05	1.22E-02	1.20E-13
	0.95	9.32E-01	3.02E-02	6.58E-02	2.79E-02	3.54E-11
	0.99	9.95E-01	1.78E-01	0.00E+00	1.00E+00	2.58E-03
# Rejections			7	8	6	8
	0.01	1.09E-02	8.01E-01	0.00E+00	9.99E-01	2.84E-02
	0.05	4.51E-02	5.40E-01	7.51E-01	8.82E-02	4.33E-01
	0.10	8.62E-02	2.03E-01	3.46E-01	4.85E-01	3.29E-01
Hour 19	0.25	2.13E-01	2.01E-02	8.09E-04	3.14E-02	4.67E-04
w = 730	0.50	4.87E-01	4.82E-01	5.84E-02	1.10E-02	7.33E-05
	0.75	6.83E-01	4.19E-05	2.65E-09	9.73E-05	3.65E-08
	0.90	8.80E-01	7.41E-02	6.92E-02	2.52E-01	2.26E-05
	0.95	9.45E-01	5.64E-01	7.28E-01	8.35E-01	2.39E-04
	0.99	9.95E-01	1.78E-01	0.00E+00	2.97E-08	6.07E-01
# Rejections			2	4	4	6

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	4.10E-03	6.92E-02	0.00E+00	1.00E+00	4.85E-01
	0.05	4.51E-02	5.40E-01	7.60E-01	1.20E-01	7.71E-01
	0.10	9.44E-02	6.10E-01	5.19E-01	8.17E-01	9.58E-01
Hour 3	0.25	2.59E-01	5.95E-01	1.38E-01	1.66E-04	1.26E-02
w = 913	0.50	9.17E-01	0.00E+00	0.00E+00	2.76E-13	2.01E-04
	0.75	7.63E-01	4.02E-01	0.00E+00	3.98E-33	1.56E-03
	0.90	9.51E-01	5.02E-07	2.63E-06	2.66E-02	2.35E-03
	0.95	9.58E-01	3.34E-01	6.00E-01	4.64E-01	3.02E-06
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	1.28E-01
# Rejections			3	5	4	5
	0.01	1.49E-01	0.00E+00	0.00E+00	1.45E-66	2.94E-07
	0.05	1.37E-03	5.55E-16	0.00E+00	1.00E+00	1.08E-03
	0.10	1.26E-01	2.45E-02	5.83E-02	2.30E-19	1.52E-01
Hour 8	0.25	2.30E-01	2.03E-01	1.09E-10	9.27E-11	5.59E-02
w = 913	0.50	8.67E-01	0.00E+00	0.00E+00	1.10E-21	8.08E-01
	0.75	7.84E-01	3.17E-02	9.44E-14	1.72E-22	8.37E-01
	0.90	3.28E-02	0.00E+00	0.00E+00	2.35E-30	6.70E-05
	0.95	3.56E-02	0.00E+00	0.00E+00	6.91E-45	6.35E-05
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	9.25E-01
# Rejections			8	8	7	4
	0.01	1.23E-02	5.44E-01	0.00E+00	9.97E-01	4.67E-01
	0.05	5.06E-02	9.39E-01	7.62E-01	1.83E-01	1.82E-01
	0.10	1.01E-01	9.12E-01	9.74E-01	2.06E-01	7.29E-01
Hour 19	0.25	2.41E-01	5.63E-01	1.96E-01	7.47E-12	4.05E-01
w = 913	0.50	5.70E-01	1.36E-04	3.44E-04	8.87E-01	2.28E-01
	0.75	7.67E-01	2.72E-01	5.45E-01	9.33E-01	3.35E-01
	0.90	9.34E-01	1.03E-03	2.69E-03	4.67E-01	3.23E-01
	0.95	7.31E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	0.99	7.31E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
# Rejections			4	5	3	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

xviii

Table D.10: Asymmetric slope CAViaR results for window sizes 365 and 548

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.05E-02	1.23E-02	2.59E-02	8.42E-01	2.38E-02
	0.05	5.34E-02	6.81E-01	1.67E-01	1.97E-01	6.00E-02
	0.10	9.99E-02	9.90E-01	8.71E-02	4.57E-01	6.14E-01
Hour 3	0.25	2.41E-01	5.63E-01	8.68E-04	3.82E-05	1.89E-01
w = 365	0.50	5.13E-01	4.82E-01	2.47E-06	4.52E-14	1.59E-03
	0.75	6.74E-01	4.59E-06	1.07E-13	4.15E-16	1.68E-06
	0.90	8.78E-01	5.71E-02	1.57E-01	9.36E-01	1.50E-09
	0.95	9.40E-01	2.20E-01	4.27E-01	9.15E-01	3.21E-08
	0.99	9.79E-01	1.23E-02	0.00E+00	8.38E-01	2.43E-08
# Rejections			3	5	3	6
	0.01	5.47E-02	0.00E+00	2.22E-16	6.09E-14	2.92E-02
	0.05	9.03E-02	6.36E-06	4.38E-06	1.25E-10	9.00E-02
	0.10	1.42E-01	3.10E-04	9.40E-05	2.74E-20	5.43E-02
Hour 8	0.25	2.56E-01	7.17E-01	2.77E-07	1.73E-06	1.46E-01
w = 365	0.50	4.36E-01	5.72E-04	4.74E-07	4.09E-58	1.91E-14
	0.75	6.37E-01	1.56E-11	1.33E-15	2.81E-38	7.78E-08
	0.90	8.69E-01	6.78E-03	4.20E-09	2.02E-17	2.94E-04
	0.95	9.33E-01	4.41E-02	5.49E-04	4.88E-07	1.36E-03
	0.99	9.86E-01	3.44E-01	0.00E+00	2.53E-05	2.74E-02
# Rejections			7	9	9	6
	0.01	2.33E-02	2.12E-03	0.00E+00	8.08E-01	2.41E-04
	0.05	5.61E-02	4.58E-01	6.73E-02	1.55E-02	4.35E-02
	0.10	1.01E-01	9.12E-01	2.29E-01	2.75E-02	6.87E-01
Hour 19	0.25	2.68E-01	2.61E-01	4.36E-01	6.51E-08	6.19E-01
w = 365	0.50	4.92E-01	6.84E-01	8.61E-01	8.03E-01	5.39E-02
	0.75	7.28E-01	1.69E-01	7.67E-02	1.87E-01	6.64E-05
	0.90	9.15E-01	1.61E-01	9.68E-02	4.24E-01	1.42E-01
	0.95	9.48E-01	8.07E-01	2.89E-08	9.87E-15	6.05E-01
	0.99	9.85E-01	2.02E-01	5.59E-04	6.51E-12	9.58E-01
			1	3	5	3

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.05E-02	1.23E-02	4.58E-03	1.17E-01	9.76E-03
	0.05	5.47E-02	5.64E-01	4.03E-01	7.96E-01	2.06E-02
	0.10	9.85E-02	8.92E-01	1.50E-01	2.87E-01	6.60E-01
Hour 3	0.25	2.37E-01	4.02E-01	4.55E-04	6.89E-05	7.19E-02
w = 548	0.50	5.25E-01	1.71E-01	4.71E-09	3.03E-15	4.74E-03
	0.75	6.89E-01	2.25E-04	3.93E-09	2.71E-11	5.06E-07
	0.90	9.11E-01	3.10E-01	2.30E-01	6.96E-01	3.05E-07
	0.95	9.25E-01	3.46E-03	5.21E-03	7.99E-03	1.07E-07
	0.99	9.88E-01	5.44E-01	0.00E+00	9.97E-01	1.55E-03
# Rejections			3	6	4	7
	0.01	3.83E-02	4.43E-09	2.36E-08	4.04E-09	5.66E-03
	0.05	8.62E-02	4.33E-05	1.47E-05	2.59E-12	1.39E-01
	0.10	1.38E-01	1.07E-03	5.48E-04	5.51E-25	1.59E-02
Hour 8	0.25	2.75E-01	1.23E-01	4.23E-07	1.25E-05	3.79E-02
w = 548	0.50	4.47E-01	4.36E-03	3.50E-06	2.12E-57	8.85E-16
	0.75	6.51E-01	2.72E-09	1.23E-12	2.60E-29	7.47E-08
	0.90	8.73E-01	1.80E-02	1.56E-09	3.33E-20	1.67E-03
	0.95	9.36E-01	8.88E-02	2.36E-06	6.59E-12	7.45E-04
	0.99	9.90E-01	9.08E-01	0.00E+00	6.20E-02	5.99E-01
# Rejections			6	9	8	7
	0.01	1.78E-02	5.66E-02	0.00E+00	9.66E-01	6.51E-03
	0.05	5.34E-02	6.81E-01	4.14E-01	4.66E-02	1.05E-01
	0.10	1.07E-01	5.50E-01	2.08E-02	3.85E-04	3.04E-01
Hour 19	0.25	2.65E-01	3.40E-01	1.80E-01	2.18E-05	6.56E-01
w = 548	0.50	4.92E-01	6.84E-01	6.63E-01	9.81E-01	9.76E-02
	0.75	7.35E-01	3.40E-01	1.80E-01	4.84E-01	3.68E-02
	0.90	9.29E-01	6.35E-03	1.18E-02	8.10E-01	1.58E-01
	0.95	9.45E-01	5.64E-01	8.60E-09	5.85E-14	5.09E-01
	0.99	9.89E-01	8.01E-01	7.59E-03	2.82E-09	4.50E-02
# Rejections			1	5	5	3

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.11: Asymmetric slope CAViaR results for window sizes 730 and 913

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.46E-02	8.10E-04	8.88E-06	1.30E-05	5.89E-03
	0.05	5.20E-02	8.07E-01	3.89E-01	5.93E-01	3.20E-02
	0.10	9.58E-02	7.01E-01	1.84E-02	7.70E-02	6.24E-01
Hour 3	0.25	2.46E-01	8.14E-01	8.04E-03	2.97E-04	1.98E-03
w = 730	0.50	5.27E-01	1.49E-01	7.05E-09	6.33E-14	1.72E-07
	0.75	7.55E-01	7.48E-01	9.26E-04	2.97E-11	3.32E-06
	0.90	9.36E-01	6.24E-04	1.53E-03	6.09E-01	1.76E-05
	0.95	9.38E-01	1.66E-01	1.79E-03	8.55E-04	1.32E-10
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	1.34E-02
# Rejections			2	8	5	8
	0.01	3.28E-02	9.29E-07	5.77E-06	1.26E-09	1.34E-02
	0.05	7.52E-02	3.46E-03	2.50E-03	2.94E-10	3.71E-01
	0.10	1.40E-01	7.15E-04	7.70E-04	9.01E-22	3.88E-02
Hour 8	0.25	2.85E-01	3.35E-02	2.18E-07	1.35E-05	8.25E-03
w = 730	0.50	4.60E-01	2.90E-02	2.24E-05	8.74E-58	1.84E-13
	0.75	6.88E-01	1.63E-04	8.31E-07	3.51E-23	3.73E-05
	0.90	8.78E-01	5.71E-02	3.96E-08	9.09E-18	1.19E-04
	0.95	9.34E-01	6.32E-02	1.96E-05	9.46E-12	3.09E-05
	0.99	9.92E-01	6.15E-01	0.00E+00	8.87E-03	1.31E-01
# Rejections			6	9	9	7
	0.01	1.50E-02	2.02E-01	0.00E+00	9.88E-01	6.25E-01
	0.05	5.20E-02	8.07E-01	7.50E-01	3.49E-02	5.15E-01
	0.10	1.07E-01	5.50E-01	8.06E-03	8.53E-06	5.70E-01
Hour 19	0.25	2.60E-01	5.38E-01	2.21E-01	4.63E-07	9.34E-01
w = 730	0.50	4.90E-01	5.79E-01	4.68E-01	9.63E-01	1.14E-01
	0.75	7.51E-01	9.49E-01	4.30E-01	7.05E-01	1.27E-02
	0.90	9.38E-01	2.12E-04	7.97E-04	8.65E-01	2.73E-01
	0.95	9.55E-01	5.40E-01	7.15E-09	1.70E-17	2.58E-01
	0.99	9.90E-01	9.08E-01	0.00E+00	9.99E-01	5.04E-01
# Rejections			1	5	4	1

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.46E-02	8.10E-04	5.16E-07	2.15E-10	7.59E-04
	0.05	6.02E-02	2.20E-01	1.80E-01	2.71E-01	6.10E-01
	0.10	1.07E-01	5.50E-01	2.85E-03	1.61E-02	7.93E-01
Hour 3	0.25	2.53E-01	8.48E-01	3.83E-03	3.69E-08	1.60E-03
w = 913	0.50	5.43E-01	1.97E-02	1.99E-08	2.37E-13	1.37E-09
	0.75	7.98E-01	2.38E-03	3.61E-08	3.17E-12	2.90E-05
	0.90	9.47E-01	4.76E-06	2.84E-05	7.39E-01	2.69E-07
	0.95	9.41E-01	2.86E-01	3.74E-01	8.63E-02	6.39E-20
	0.99	9.95E-01	1.78E-01	0.00E+00	1.00E+00	5.88E-02
# Rejections			4	7	5	6
	0.01	3.01E-02	1.06E-05	5.66E-05	3.50E-05	9.42E-03
	0.05	7.39E-02	5.54E-03	7.01E-03	3.78E-12	2.55E-01
	0.10	1.40E-01	7.15E-04	7.70E-04	2.34E-18	6.03E-02
Hour 8	0.25	2.83E-01	4.11E-02	1.45E-07	5.86E-06	2.75E-03
w = 913	0.50	4.58E-01	2.40E-02	1.89E-04	1.75E-52	4.24E-12
	0.75	6.89E-01	2.25E-04	1.38E-06	2.87E-20	2.48E-04
	0.90	8.80E-01	7.41E-02	9.94E-10	4.47E-23	7.38E-04
	0.95	9.30E-01	2.03E-02	1.53E-08	5.90E-16	1.42E-05
	0.99	9.96E-01	6.92E-02	0.00E+00	1.00E+00	9.26E-01
# Rejections			7	9	8	6
	0.01	1.37E-02	3.44E-01	0.00E+00	9.93E-01	4.83E-02
	0.05	5.06E-02	9.39E-01	7.22E-01	5.80E-02	4.20E-01
	0.10	1.04E-01	7.22E-01	1.36E-02	7.41E-06	7.86E-01
Hour 19	0.25	2.52E-01	9.15E-01	3.33E-01	1.26E-07	7.94E-01
w = 913	0.50	4.92E-01	6.84E-01	8.61E-01	9.99E-01	1.13E-01
	0.75	7.59E-01	5.63E-01	5.56E-01	7.83E-01	1.52E-01
	0.90	9.40E-01	1.20E-04	2.34E-04	2.34E-01	2.40E-01
	0.95	9.56E-01	4.31E-01	1.20E-10	8.30E-22	4.61E-02
	0.99	9.90E-01	9.08E-01	0.00E+00	9.99E-01	7.33E-01
# Rejections			1	5	3	2

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.12: Symmetric absolute value CAViaR results for window sizes 365 and 548

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	1.09E-02	8.01E-01	1.89E-01	2.98E-03	2.26E-01
	0.05	5.61E-02	4.58E-01	1.77E-01	1.26E-01	1.20E-01
	0.10	1.11E-01	3.38E-01	4.76E-03	3.11E-02	9.41E-01
Hour 3	0.25	1.44E-01	1.82E-12	0.00E+00	1.09E-18	2.30E-02
w = 365	0.50	5.40E-01	2.90E-02	0.00E+00	2.47E-39	1.59E-04
	0.75	6.80E-01	2.05E-05	0.00E+00	6.18E-43	4.10E-04
	0.90	8.96E-01	7.22E-01	4.45E-01	5.17E-01	4.05E-09
	0.95	9.52E-01	7.91E-01	8.11E-01	9.57E-01	2.12E-10
	0.99	9.88E-01	5.44E-01	0.00E+00	9.97E-01	3.06E-11
# Rejections			3	5	5	6
	0.01	2.46E-02	8.10E-04	0.00E+00	1.05E-02	5.20E-03
	0.05	7.25E-02	8.70E-03	4.01E-03	7.45E-07	1.86E-02
	0.10	9.85E-02	8.92E-01	1.25E-02	6.08E-19	8.55E-01
Hour 8	0.25	2.71E-01	1.97E-01	1.27E-08	2.53E-11	1.95E-01
w = 365	0.50	4.32E-01	2.45E-04	3.33E-16	6.86E-66	8.40E-04
	0.75	6.83E-01	4.19E-05	1.11E-16	1.08E-25	2.69E-03
	0.90	8.96E-01	7.22E-01	1.44E-09	3.75E-19	4.87E-04
	0.95	9.30E-01	2.03E-02	5.20E-06	1.11E-14	2.37E-03
	0.99	9.85E-01	2.02E-01	1.58E-01	1.85E-05	4.33E-02
# Rejections			5	8	9	7
	0.01	1.37E-02	3.44E-01	0.00E+00	1.34E-01	6.44E-01
	0.05	5.06E-02	9.39E-01	7.22E-01	2.96E-01	6.95E-01
	0.10	1.05E-01	6.33E-01	9.03E-02	2.60E-04	4.84E-01
Hour 19	0.25	2.76E-01	1.04E-01	2.00E-01	2.00E-06	2.06E-01
w = 365	0.50	5.06E-01	7.39E-01	8.29E-01	7.46E-01	2.21E-02
	0.75	7.35E-01	3.40E-01	3.04E-01	4.87E-01	3.42E-05
	0.90	9.21E-01	5.43E-02	1.25E-01	3.55E-01	5.33E-02
	0.95	9.59E-01	2.52E-01	2.97E-10	1.31E-24	7.20E-01
	0.99	9.92E-01	6.15E-01	1.82E-03	8.77E-18	2.31E-01
			0	3	4	2

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.05E-02	1.23E-02	4.10E-04	1.40E-04	3.45E-03
	0.05	5.88E-02	2.86E-01	1.78E-01	1.98E-01	1.21E-01
	0.10	9.71E-02	7.95E-01	3.00E-03	9.14E-03	5.65E-01
Hour 3	0.25	2.45E-01	7.48E-01	7.79E-02	3.54E-04	3.00E-02
w = 548	0.50	5.91E-01	8.10E-07	0.00E+00	4.32E-35	4.00E-03
	0.75	7.11E-01	1.76E-02	0.00E+00	9.71E-43	7.56E-05
	0.90	9.06E-01	6.10E-01	2.17E-01	7.01E-02	1.58E-08
	0.95	9.36E-01	8.88E-02	1.25E-01	6.13E-01	3.41E-10
	0.99	9.89E-01	8.01E-01	0.00E+00	9.99E-01	2.19E-07
# Rejections			3	5	5	7
	0.01	1.78E-02	5.66E-02	0.00E+00	7.79E-01	5.74E-02
	0.05	6.29E-02	1.22E-01	6.77E-02	3.68E-10	2.44E-01
	0.10	9.85E-02	8.92E-01	1.64E-01	4.60E-13	2.52E-02
Hour 8	0.25	2.60E-01	5.38E-01	2.50E-08	1.24E-09	1.74E-01
w = 548	0.50	4.40E-01	1.27E-03	0.00E+00	1.33E-65	6.42E-04
	0.75	6.59E-01	4.56E-08	0.00E+00	1.11E-33	2.50E-04
	0.90	9.02E-01	8.92E-01	4.12E-09	8.98E-21	5.26E-03
	0.95	9.38E-01	1.66E-01	9.57E-06	1.52E-12	5.68E-03
	0.99	9.90E-01	9.08E-01	0.00E+00	6.20E-02	7.16E-01
# Rejections			2	7	7	5
	0.01	1.23E-02	5.44E-01	0.00E+00	4.88E-02	8.30E-01
	0.05	4.79E-02	7.91E-01	9.35E-01	3.25E-02	2.99E-01
	0.10	9.44E-02	6.10E-01	8.30E-02	2.52E-02	9.28E-01
Hour 19	0.25	2.53E-01	8.48E-01	9.89E-02	1.21E-05	5.44E-01
w = 548	0.50	4.98E-01	9.12E-01	7.14E-01	9.65E-01	1.09E-01
	0.75	7.47E-01	8.48E-01	5.99E-01	6.28E-01	1.47E-02
	0.90	9.27E-01	9.51E-03	3.45E-02	3.89E-01	2.52E-01
	0.95	9.64E-01	5.96E-02	1.49E-09	1.73E-21	7.95E-01
	0.99	9.92E-01	6.15E-01	9.57E-02	6.11E-14	4.99E-01
# Rejections			1	3	6	1

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

Table D.13: Symmetric absolute value CAViaR results for window sizes 730 and 913

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	3.15E-02	3.20E-06	2.97E-09	1.64E-07	3.49E-04
	0.05	6.29E-02	1.22E-01	6.77E-02	3.70E-01	8.70E-02
	0.10	9.85E-02	8.92E-01	1.33E-03	2.45E-03	8.03E-01
Hour 3	0.25	2.54E-01	7.82E-01	7.14E-02	4.86E-04	4.01E-02
w = 730	0.50	6.14E-01	5.50E-10	0.00E+00	6.70E-31	4.37E-05
	0.75	8.28E-01	4.47E-07	0.00E+00	3.35E-30	1.58E-01
	0.90	9.32E-01	2.67E-03	1.04E-02	7.96E-01	1.47E-07
	0.95	9.51E-01	9.25E-01	1.02E-01	6.39E-02	3.14E-11
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	6.92E-05
# Rejections			4	6	5	6
	0.01	2.46E-02	8.10E-04	0.00E+00	3.35E-04	5.00E-08
	0.05	6.02E-02	2.20E-01	7.50E-02	1.40E-09	7.59E-01
	0.10	1.04E-01	7.22E-01	2.90E-01	2.97E-16	7.03E-01
Hour 8	0.25	2.64E-01	3.84E-01	4.27E-12	1.34E-13	2.95E-02
w = 730	0.50	4.79E-01	2.51E-01	2.22E-16	3.35E-57	2.99E-02
	0.75	7.31E-01	2.27E-01	9.77E-12	1.05E-19	3.57E-02
	0.90	9.02E-01	8.92E-01	8.55E-12	2.44E-26	9.66E-03
	0.95	9.36E-01	8.88E-02	1.61E-05	1.35E-11	9.03E-03
	0.99	9.93E-01	3.62E-01	0.00E+00	2.36E-04	8.28E-01
# Rejections			1	7	9	6
	0.01	1.64E-02	1.11E-01	0.00E+00	2.34E-01	6.40E-01
	0.05	5.61E-02	4.58E-01	6.73E-02	2.48E-02	6.98E-01
	0.10	1.08E-01	4.72E-01	1.14E-01	1.58E-02	8.56E-01
Hour 19	0.25	2.64E-01	3.84E-01	1.64E-01	9.64E-06	9.05E-01
w = 730	0.50	4.94E-01	7.39E-01	7.59E-01	9.80E-01	2.10E-01
	0.75	7.52E-01	8.81E-01	4.79E-01	7.67E-01	2.64E-02
	0.90	9.37E-01	3.68E-04	1.42E-03	9.17E-01	5.76E-01
	0.95	9.73E-01	2.20E-03	4.07E-06	2.56E-12	1.58E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	6.17E-01
# Rejections			3	4	4	1

	Quantile	Violations	P_UC	P_CC	P_DQ1	P_DQ2
	0.01	2.19E-02	5.25E-03	1.81E-05	4.57E-10	1.18E-03
	0.05	6.29E-02	1.22E-01	1.46E-01	1.80E-01	8.25E-01
	0.10	1.04E-01	7.22E-01	4.83E-04	1.09E-03	6.86E-02
Hour 3	0.25	2.72E-01	1.69E-01	8.46E-03	1.71E-05	3.18E-03
w = 913	0.50	6.79E-01	0.00E+00	0.00E+00	1.57E-35	1.48E-08
	0.75	8.89E-01	0.00E+00	0.00E+00	3.64E-40	3.95E-01
	0.90	9.48E-01	2.31E-06	1.03E-05	2.77E-01	1.19E-06
	0.95	9.48E-01	8.07E-01	9.70E-01	3.73E-01	8.06E-21
	0.99	9.93E-01	3.62E-01	0.00E+00	1.00E+00	5.26E-03
# Rejections			4	7	5	6
	0.01	2.05E-02	1.23E-02	0.00E+00	5.06E-04	6.74E-02
	0.05	5.75E-02	3.66E-01	7.23E-02	1.39E-07	6.63E-01
	0.10	1.09E-01	4.01E-01	6.33E-01	3.15E-22	5.63E-01
Hour 8	0.25	2.56E-01	7.17E-01	3.33E-16	9.17E-18	2.67E-02
w = 913	0.50	4.73E-01	1.49E-01	2.49E-12	1.17E-52	1.65E-01
	0.75	7.51E-01	9.49E-01	1.05E-14	1.36E-25	2.47E-01
	0.90	8.88E-01	2.81E-01	2.27E-11	1.48E-27	5.66E-03
	0.95	9.33E-01	4.41E-02	2.93E-06	2.20E-11	4.72E-02
	0.99	9.97E-01	1.93E-02	0.00E+00	1.00E+00	7.93E-01
# Rejections			3	7	8	3
	0.01	1.23E-02	5.44E-01	0.00E+00	4.88E-02	5.38E-01
	0.05	5.47E-02	5.64E-01	4.25E-01	1.33E-01	5.34E-01
	0.10	9.85E-02	8.92E-01	7.78E-02	7.59E-03	9.06E-01
Hour 19	0.25	2.49E-01	9.49E-01	4.92E-02	1.72E-06	8.82E-01
w = 913	0.50	4.91E-01	6.31E-01	8.89E-01	9.92E-01	1.90E-01
	0.75	7.69E-01	2.36E-01	3.60E-01	6.30E-01	1.11E-01
	0.90	9.34E-01	1.03E-03	4.07E-03	8.53E-01	4.03E-01
	0.95	9.77E-01	2.24E-04	8.21E-08	4.50E-18	2.10E-01
	0.99	9.99E-01	3.19E-03	0.00E+00	1.00E+00	6.16E-01
# Rejections			3	5	4	0

The table displays the p-values of the unconditional coverage test (UC), the conditional coverage test (CC), and the two dynamic conditional quantile tests (DQ1 and DQ2), as described in Section 5.2. P-values highlighted in red are significant at the 5% level, which implies poor model calibration.

xxii