Applying real-time GARCH approaches to forecasting short-term electricity prices and volatility

Nicky Sonnemans
Student number: 506125

Supervisor: dr. R. Lange
Second assessor: dr. O. Kleen

4 July 2021

Abstract

In this thesis we construct interval and point forecasts while incorporating for conditional heteroskedasticity with the aim of predicting electricity prices and their volatility. Due to the increase in renewable energy sources and their corresponding weather dependence, short-lived unanticipated price spikes occur more frequently. To our knowledge, current electricity price forecasting models only incorporate for conditional heteroskedasticity by at most using an asymmetric GARCH component. This thesis focuses on incorporating current information in the volatility process, using the recent developed Real-time GARCH model. We have found that Real-time GARCH constructed more accurate interval forecasts than other GARCH processes. However, the models incorporating for conditional heteroskedasticity are outperformed by the base model, which doesn't include a volatility process, when evaluating the point forecasts. It may be that due to the price shocks only lasting a couple of hours, even the more recent Real-time GARCH model cannot account for this information and is too slow in processing these signals.

Keywords: Electricity Price Point Forecasting, RT-GARCH, GARCH, Electricity Price Interval Forecasting

Contents

1 Introduction

Since the worldwide structural reforms and market liberalization that started in the early 1990s, the introduction of competitive markets in energy have been reshaping the landscape of the conventionally monopolistic and government-controlled energy sectors. Currently, the economic law of supply and demand determines the price in marketplaces where electricity can be traded in either spot or forward contracts. However electricity as a commodity possesses certain unique features not shared by other commodities. Electricity is economically non-storable and the power system stability requires a constant balance between consumption and production ((EPEX, 2021a)). Simultaneously, energy consumption is affected by weather, the intensity of business and time (e.g. weekdays vs weekend). Furthermore, the increase in Renewable Energy Sources (RES) significantly boosts the impact of weather on energy production. These distinct features result in price dynamics not seen in any other market, with seasonality at hourly to monthly levels, as well as abrupt, short-lived, and generally unanticipated price spikes.

As the state of Texas showed in 2021, in the largest forced power outage in U.S. history, it is of importance to hedge not only against volume risk but also against extreme price movements. In the midst of this energy crisis, the largest power generation and transmission cooperative in Texas, Brazos, filed for bankruptcy , see (Hill et al., 2021), as the costs of over-/under contracting and subsequently selling/buying power in the balancing market are typically so high that this often results in huge financial losses or bankruptcy. To prevent power portfolio managers from this phenomenon, price and volatility forecasts have become of particular interest. Prediction of volatile spot electricity prices with a reasonable level of accuracy is very valuable for the operational planning and trading of energy firms, as well as for the demand response and energy consumers in the wholesale electricity market.

A variety of differing models for electricity price forecasting (EPF) have been attempted, with differing degrees of success. Interested readers can find an extensive literature study on the forecasting of electricity prices in the research of (Weron, 2014). This thesis will focus on models incorporating for conditional heteroskedasticity. Empirically, electricity spot prices, like financial time series, exhibit various sorts of nonlinear dynamics, the most important of which is the series strong dependence on its past values (Weron, 2014). Another nonlinear dynamic is due to the time-varying conditional volatility, characterized by the clustering of large shocks. Aforementioned models try to capture this heteroskedasticity and dependence on past values. In their research (Liu & Shi, 2013) applied various ARMA models with GARCH processes, namely ARMA-GARCH models, along with their modified forms, to model and forecast hourly electricity prices. They find that ARMA-GARCH-M (in-mean) models, which model a conditional mean, are an effective tool for modelling and forecasting the mean and volatility of electricity prices, whilst using multiple statistical measures. Furthermore their results and other literature indicate the existence of an inverse leverage effect, the effect that electricity price volatility tends to arise more with positive shocks than negative shocks. However, this effect is not captured in their symmetric models but is captured in their highly competitive ARMA-GJR-GARCH-M model.

Because electricity prices can be subject to unanticipated extreme price movements, it is of essence that not only the tails of the distribution are correctly modelled but the model is faster in its adjustment to a new unconditional level of volatility. Most statistical models used for EPF follow the structure of (Engle, 1982) in modelling volatility of prices/returns by only using past information. (Smetanina, 2017) proposes a new model, which retains the structure of GARCH models, but models the volatility process in the spirit of stochastic volatility (SV) models by using a combination of past and current information. Her model, Real-time GARCH (RT-GARCH), is able to deliver better volatility forecasts compared with standard GARCH models, improve the empirical fit (especially within the tails of the distribution) and adjust faster to new unconditional levels of volatility. The RT-GARCH model, rewritten as in (Lange, 2021), can be defined as:

$$r_t = \mu + \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. Student's } t, \tag{1}$$
$$g_t = \omega + \alpha(r_{t-1} - \mu)^2 + \beta h_{t-1}, \qquad \text{(known at time } t-1\text{)}, \tag{2}$$
$$h_t = g_t + \phi\varepsilon_t^2, \tag{3}$$

where $\omega, \alpha, \phi \geq 0$ and drive the processes g_t, h_t. This modified model has the symmetric GARCH model as a special case when $\phi = 0$.

The new term introduced within the RT-GARCH model can be interpreted as a variation in the information set and as an additional shape parameter of the density of returns/prices, which defines the peakness and/or thickness of the tails of the distribution. This model, like asymmetric GARCH, is able to encompass leverage by extending the model as follows:

$$h_t = g_t + (\phi_1 \mathbb{1}_{\varepsilon_t \leq 0} + \phi_2 \mathbb{1}_{\varepsilon_t > 0})\varepsilon_t^2, \tag{4}$$

where $\phi_1, \phi_2 \geq 0$. The model can be further extended to differentiate between effects of positive and negative returns in driving volatility. The extension of RT-GARCH with leverage and feedback effects is given by:

$$h_t = g_t + (\phi_1 \mathbb{1}_{\varepsilon_t \leq 0} + \phi_2 \mathbb{1}_{\varepsilon_t > 0})\varepsilon_t^2, \tag{5}$$
$$g_t = \omega + (\alpha_1 \mathbb{1}_{r_{t-1} \leq \mu} + \alpha_2 \mathbb{1}_{r_{t-1} > \mu})(r_{t-1} - \mu)^2 + \beta h_{t-1}, \tag{6}$$

where $\alpha_1, \alpha_2 \geq 0$. Notice that this extension has the assymetric GARCH model as its special case when $\phi_1 = \phi_2 = 0$.

In addition to this, extensions on the RT-GARCH model have already been proposed. (Ding, 2021) allows for conditonal heteroskedasticity in the volatility process, improving upon the Real-time GARCH model's in-sample nowcasts and out-of-sample forecasts. This extension is defined as follows:

$$g_t = \omega + \alpha(r_{t-1} - \mu)^2 + \beta h_{t-1}, \tag{7}$$
$$h_t = g_t + (\phi_1 + \phi_2 \cdot g_t)\varepsilon_t^2, \tag{8}$$

Just as in the spirit of (Smetanina, 2017), this baseline model can be extended to encompass leverage and feedback effects as well.

Statistical models used in EPF mostly follow the original idea of (Engle, 1982) and at most encompasses for the inverse-leverage effect using an asymmetric time-varying volatility. Moreover, both (Smetanina, 2017) and (Ding, 2021) assessed their models on financial time series only. As such, it is of interest to evaluate and compare aforementioned models in the context of electricity prices. Specifically, these models will be compared in modelling the volatility and mean of electricity prices. The aim of this thesis is to demonstrate our efforts in this aspect.

Section 2 provides a short literature study. Next, Section 3 describes the market, data and a transformation. Subsequently section 4 and 5 describes, respectively, the models used and the estimation, forecasting and evaluation methodology. Finally, we discuss the results and conclude with the limitations of the research and suggestions for the future.

2 Literature

In their literature study, (Weron, 2014) found that the modelling and forecasting of imbalance prices and intraday prices has not been studied as thoroughly as their day-ahead counterparts. (Andrade et al., 2017) carried out probabilistic price forecasting of intraday electricity prices, but focused on the Spanish and Portugese market (Iberian market). The Iberian market has a lower bound on electricity prices of 0, whereas other European markets may have negative electricity prices. More recently, (Uniejewski et al., 2019) performed short-term intra-day price forecasting of a specific price index (ID_3)[1] for hourly products in the EPEX German Intraday Continous market. Their research showed that using LASSO to help with variable selection, they could create well-performing, parsimonious ARX-type models that would be comparable to the best LASSO-model whilst having on average 3.5 times fewer variables. However, (Narajewski & Ziel, 2020) suggests that there is no more information to extract from transactions data for hourly products, despite the most recent price, when evaluated on the same data. They strongly encourage and also empirically show that it is beneficial to use asinh's transformation on electricity price data, which incorporates for negative values of price data.

As auto-regressive (AR) models are used extensively in EPF, it might also be of interest to capture and model the conditional heteroskedasticity as often present in financial time series. In his research (Bollerslev, 1986) introduced GARCH to allow for past conditional variances in the current conditional variance equation. In their research (Liu & Shi, 2013) combined ARMA models with GARCH models and conducted research regarding ARMA-GARCH models on electricity prices, their results show that ARMA-GARCH-M models are indeed an effective tool for both modeling and forecasting the mean and volatility of short-term intraday electricity prices. They also observed that

[1]The ID3 index is the weighted average price of all continuous trades executed within the last 3 trading hours of a contract. This index focuses on the most liquid timeframe of a continuous contract trading session, see (EPEX, 2021b)

hourly prices exhibited daily, weekly and monthly periodicities along with nonlinear and asymmetric time-varying volatility.

However, (Politis, 2007) points out that standard GARCH models make inefficient use of all available information, in particular the current return. The implication of this inefficient use of information is that GARCH models are unsuitable for scenarios involving unanticipated price spikes, i.e. as when volatility quickly changes to a new level, see (Hansen et al., 2012). In her research (Smetanina, 2017) utilizes all available information in GARCH models and demonstrates that this RT-GARCH model can account for rapid changes in volatility and outperform standard GARCH models in terms of short-term and long-term forecasts. However, comparing RT-GARCH to the Realized GARCH in (Hansen et al., 2012) , the empirical results show that RT-GARCH is slower to adjust to the new level of volatility than Realized GARCH. Extending the work of Smetanina, (Ding, 2021) has allowed for conditional heteroskedasticity in the volatility process. Their research shows a better in-sample fit and out-of-sample forecasts compared with the original RT-GARCH model. Despite comparing adjustment to new volatility levels with Realized GARCH, (Smetanina, 2017) and (Ding, 2021) fail to compare error metrics with this model.

3 Data

The data consists of hourly observations from 2014-01-01 to 2019-12-31 of day-ahead settlement prices and intraday transactions. This section aims to explain the market structure of electricity, give an insight in the data and finally proposes a transformation to normalize the data.

Data	Unit	Description	Source
EPEX day-ahead	EUR/MWh	The price resulting from the EPEX and Nordpool coupled day-ahead auction for the North-west European region.	ENTSOE-E-Transparancy Platform
Intraday transactions	EUR/MWh	Prices from all intraday transactions for all hourly instruments.	EPEX

Table 1: Overview of the data with corresponding sources

3.1 Market description

The electricity market is split into two parts: the spot market, where electrical energy is exchanged for immediate physical delivery, and the futures market. The delivery in the future market is delayed and therefore does not usually require physical delivery. The futures market is typically used for risk management. In general, European spot markets are split into two categories: day-ahead and intraday sessions , see (EPEX, 2021a). We will use data available from the German electricity market from 2014-01-01 to 2019-12-31. This data will include hourly intraday products as well as day-ahead prices from German market.

The German Intraday Continous market opens every day at 15:00 for hourly products and the gate closure is 5 minutes before delivery. The intraday continuous market is preceded by the Day-

Ahead Auction, which takes place daily at 12:00 (EPEX, 2021c). See Figure 1 for a visualization on the German electricity market. In the spirit of (Narajewski & Ziel, 2020), we do not include transactions any later than 30 minutes before delivery on the intraday market.

Notice that unlike financial markets, negative prices can and do occur on electricity markets. Negative prices are a price signal on the power wholesale market that occurs when a high inflexible power generation meets low demand. Inflexible power sources cannot be shut down and restarted in a quick and cost-efficient manner (EPEX, 2021a).

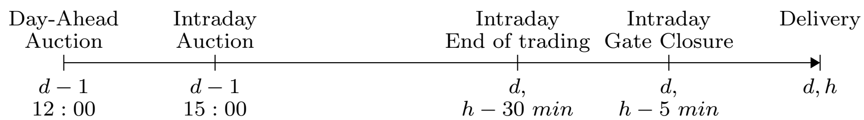

Figure 1: German Electricity Market trading routine. Variables d and h correspond to day of delivery and hour of delivery respectively.

The day-ahead (DA) market receives offers from the consumption (demand) and production (supply) sides until gate closure. Each offer is supplemented by both price and volume in energy (i.e., MW) for a specific hour the following day. When the session closes, an equilibrium point is computed between the supply and demand curves, the price and volume is set accordingly.

Additionally, due to the increasing number of renewable energy sources (RES) and its weather dependence power portfolio managers are forced to balance their production frequently to offset any operational planning errors made in the DA market. Intraday (ID) electricity markets address this problem by letting market participants trade energy continuously until 30 and 5 minutes before delivery begins in respectively the whole market and within control zones. As soon as a buy- and sell-order match, the trade is executed.

3.2 Data description

The distribution and statistics of the German day-ahead settlement prices obtained from ENTSOE-E-Transparancy Platform can be found in the Appendix. The data is from 2014-01-01 00:00 to 2019-12-31 23:00, with a total of 52583 observations (a leap year is included). Furthermore, we find the day-ahead series to be stationary when testing for stationarity using the Augmented Dickey-Fuller (ADF) test.

The intraday transactions from all German energy instruments traded between 2014-01-01 00:00 to 2019-12-31 23:00 are obtained from EPEX. Due to the observations being transactions, we must compute an ID price based on this. However, the ID market does not have a specific definition of an ID price. One cannot consider the latest transaction's price as the product's current price as this would be misleading. Volatility of the prices is highly dependent of the volume traded. (EPEX, 2015) introduced multiple price measures to provide market participants with an accurate price index. One of these price indexes, ID_3 provides a price signal that takes into account price movements in the very short-term before delivery, and hence reveals possible movements shortly before end of trading. Following the definition in (Narajewski & Ziel, 2020), ID_3 is defined as:

$$\text{ID}_3^{d,h} = \frac{1}{\sum_{k \in T_3^{d,h} \cap T^{d,h}} V_k^{d,h}} \sum_{k \in T_3^{d,h} \cap T^{d,h}} V_k^{d,h} P_k^{d,h}, \tag{9}$$

where $T_3^{d,h} = [b(d,h) - 3, b(d,h) - 0.5)$ and b(d,h) be a start of the delivery for hour h on day d. Thus $T_3^{d,h}$ is a timeframe between 3 hours and 30 minutes before delivery. $T^{d,h}$ is a set of transactions regarding hour h on day d, $P_k^{d,h}$ and $V_k^{d,h}$ are respectively the price and volume of k'th transaction within $T_3^{d,h} \cap T^{d,h}$. (Narajewski & Ziel, 2020) also defined the general price value of a product at a particular time during the trading period, ${}_x\text{ID}_y$ is defined as:

$${}_x\text{ID}_y^{d,h} = \frac{1}{\sum_{k \in T_{x,y}^{d,h} \cap T^{d,h}} V_k^{d,h}} \sum_{k \in T_{x,y}^{d,h} \cap T^{d,h}} V_k^{d,h} P_k^{d,h}, \tag{10}$$

where $T_{x,y}^{d,h} = [b(d,h) - x - y, b(d,h) - x), x \geq 0, y > 0$. In the case of no trades, we set the ${}_x\text{ID}_y^{d,h}$ price to the price of the last transaction that occured before time frame $T_{x,y}^{d,h}$. If no trades are present, we use ID-Auction and DA-Auction values. Following this logic, the trades from the hourly instruments are used to compute both ${}_{0.5}\text{ID}_{0.25}$, the first 15 minutes of trading, and $\text{ID}_3 = {}_{0.5}\text{ID}_{2.5}$. Furthermore, we find these price series to be stationary using ADF. Additionally, the distribution and statistics of the ID_3 prices can be found in Figure 2 and the scatterplot can be found in the Appendix.

Moreover, the autocorrelations of ID_3 can be found in Figure 3. As can be seen from this figure, autocorrelation peaks every 12 lags with high peaks occurring every 24 lags. Additionally, the autocorrelation after exactly 168 lags has a high peak as well. These observations all suggest there might exist seasonality within the data.

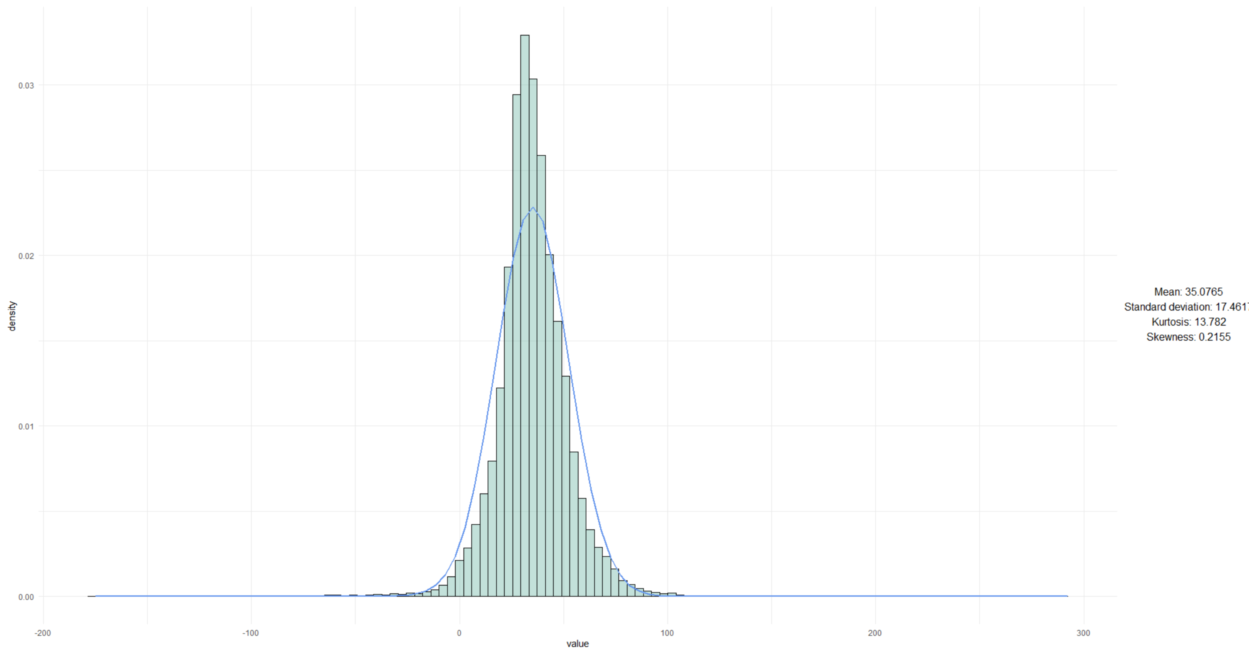

Figure 2: Distribution of ID_3 prices from period 2014-01-01 to 2019-12-31. The blue line depicts the normal distribution scaled with the first and second moment of the intraday prices.

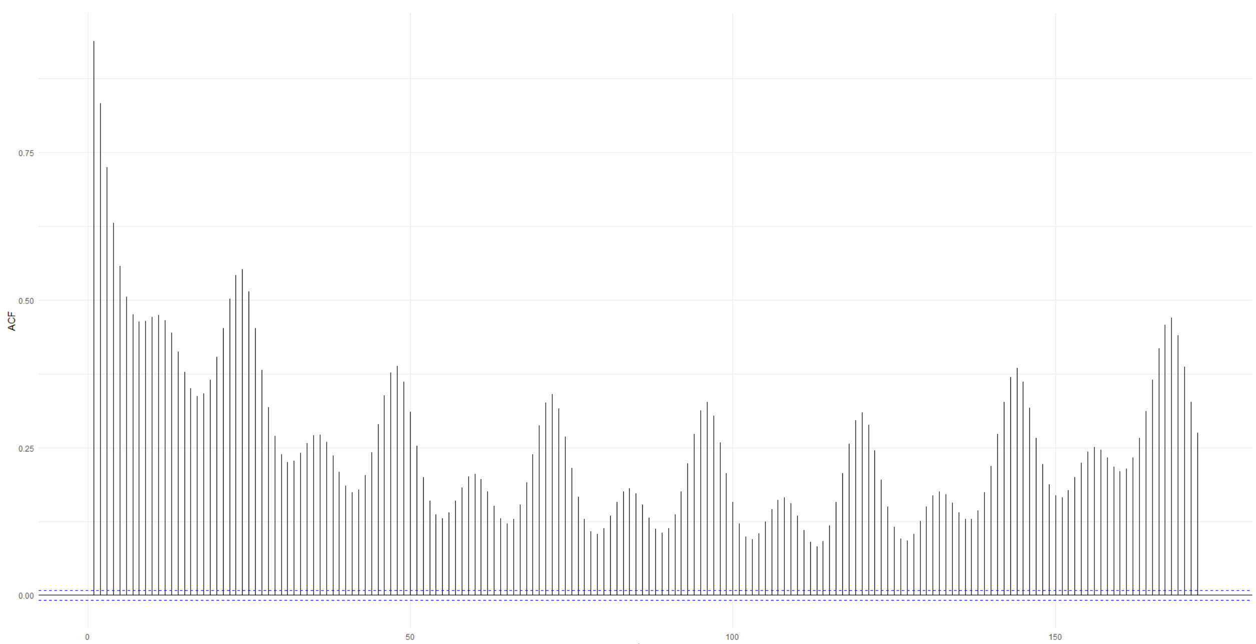

Figure 3: Autocorrelation of the ID_3 prices up to lag 172. The data is from 2014-01-01 to 2019-12-31.

3.3 Data transformation

(Uniejewski et al., 2019) showed that usage of variance stabilizing methods have resulted in higher quality forecasts regarding electricity prices. Following (Narajewski & Ziel, 2020), median normalization and asinh transformation will be applied to the data. The median normalization of price P_t is defined as:

$$p_t = p_{d,h} = \frac{1}{\text{MAD}(\mathbf{P}_{d,h})/z_{0.75}}(P_{d,s} - \widetilde{\mathbf{P}}_{d,h}), \tag{11}$$

where $\widetilde{\mathbf{P}}_{d,h}$ is the 365 days rolling median of $P_{d,h}$, $\text{MAD}(\mathbf{P}_{d,h})$ is the median absolute deviation around the rolling median and $z_{0.75}$ is .75 quantile of the normal distribution. The rolling medians are first calibrated on a given sample, in this case 365 days.

As most European electricity markets allow for negative prices, logarithmic transformation fails. The area hyperbolic sine (asinh) has a logarithmic tail behaviour for both positive and negative values, solving the negative price issue. This is defined as:

$$asinh(p_t) = \ln(p_t + \sqrt{p_t^2 + 1}). \tag{12}$$

As this formula is non-linear, backwards transformation is not as obvious. (Narajewski & Ziel, 2020) provides a mathematically correct way of backwards transformation building on existing literature. The backwards transformation can hence be defined in two ways, an exact and approximate backwards transformation:

$$\hat{P}_t^A = \hat{P}_{d,h}^A = sinh(\hat{p}_{d,h}) \cdot \text{MAD}(\mathbf{P}_{d,h})/z_{0.75} + \widetilde{\mathbf{P}}_{d,h}, \tag{13}$$

$$\hat{P}_t^E = \hat{P}_{d,h}^E = \frac{1}{D}\sum_{j=-D+d}^{d-1}\left[sinh(\hat{p}_{d,h} + \varepsilon_{j,h}) \cdot \text{MAD}(\mathbf{P}_{d,h})/z_{0.75}\right] + \widetilde{\mathbf{P}}_{d,h}, \tag{14}$$

where $\hat{p}_{d,h}$ the forecasted transformed price at day d on hour h (or simply t), $\varepsilon_{j,h}$ the in-sample residuals of the model used to forecast $\hat{p}_{d,h}$ and $\hat{P}_t$ the backwards transformed forecasted price and finally $\hat{P}_t^A$ and $\hat{P}_t^E$ respectively the approximate and exact price. For extra info and the derivations of these formulae please see (Narajewski & Ziel, 2020). Additionally, the distributions of the transformed prices can be found in the Appendix.

4 Models

This section considers different methods of capturing the variability in electricity prices. The models for time-varying volatility all assume some basic properties, which are defined as:

$$\text{asinh}(p_t) = \mu_t + \sqrt{h_t}\varepsilon_t,\ 1 \leq t \leq T, \tag{15}$$

where T denotes the sample size, the ε_t's are i.i.d. shocks with $E[\varepsilon_t] = 0$, $V[\varepsilon_t] = 1$ and $E[\varepsilon_t^4] < \infty$. The existence of the fourth moment is needed for covariance stationarity (as well as a valid forecast), see (Smetanina, 2017). Additionally, for these shocks we assume either $\varepsilon_t \sim N(0,1)$ or $\varepsilon_t \sim t(\nu)$ where $\nu > 4$. This assumption of distribution depends on the model used. Furthermore, μ_t represents the conditional mean and $\text{asinh}(p_t)$ the transformed price.

The final models are summarized in Table 2. A few benchmark models for modeling conditional heteroskedasticity are considered in this thesis to improve upon: the standard GARCH model as in (Bollerslev, 1986), the GJR-GARCH model as in (Glosten et al., 1993) and the Realized GARCH model introduced in (Hansen et al., 2012). As the methodology for the Realized GARCH model is beyond the scope of this paper, we kindly refer to (Hansen et al., 2012). However, we will define the proxy volatility that is used to model the volatility process in Realized GARCH. We compute this proxy volatility as:

$$\sigma^2_{t,proxy} = (\text{asinh}(p_t) - \text{asinh}(p_{t-1}))^2, \tag{16}$$

where p_t the transformed ID_3 price as seen in Equation 12. Notice, that because of the log transformation done in Equation 12, taking the difference of $\text{asinh}(p_t)$ to obtain a proxy for volatility can be justified as this would result in something similar to the log returns known from financial literature.

To model the conditional mean, all models use the same regressors as in the base model. This base model uses the lagged transformed values of ${}_{3.25}\text{ID}_{0.25}$, these lags were based upon the ACF in Figure 14. Furthermore, to capture the seasonality in the conditional mean, calendar variables decomposed in cosine and sine parts[2] are also introduced. Additionally, this base model also contains the transformed DA-auction and ${}_{0.5}\text{ID}_{0.25}$ prices.

Model	Description	Extension(s)
Base	Base model with regressors being the transformed ${}_{3.25}\text{ID}_{0.25}L^p$ for p = 3, 4 , 18, 24, 48, 72, 168. The transformed ${}_{0.5}\text{ID}_{0.25}$ and DA prices; and respectively hour, month and weekday as calendar variables C_H, C_M, C_{WD} decomposed in cosine and sine parts.	-
GARCH(1,1)	Base model to model conditional mean and GARCH(1,1) to capture volatility.	Skewed student-t
GJR-GARCH(1,1)	Base model to model conditional mean and GJR-GARCH(1,1) to capture volatility.	Skewed student-t
Realized GARCH(1,1)	Base model to model conditional mean and Realized GARCH(1,1) to capture volatility.	Skewed student-t
Real-time GARCH(1,1)	Base model to model conditional mean and Real-time GARCH(1,1) to capture volatility.	Leverage and Feedback

Table 2: Overview of the models

4.1 Real-time GARCH

The Real-time GARCH model (RT-GARCH) proposed by (Smetanina, 2017), introduces a new term that can be interpreted as an additional shape parameter which defines the peakness/thickness of the tails of the distribution. The most general model encompassing leverage and feedback effects,

[2] For example, $C_{H,cos} = \cos(2\pi \cdot C_H/T)$ where T = 24.

rewritten in (Lange, 2021), is defined as:

$$y_t = \mu_t + \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. Student's } t, \tag{17}$$

$$\mu_t = \gamma X_t, \qquad \text{Conditional mean,} \tag{18}$$

$$g_t = \omega + (\alpha_1 \mathbb{1}_{y_{t-1} \leq \mu_{t-1}} + \alpha_2 \mathbb{1}_{y_{t-1} > \mu_{t-1}})(y_{t-1} - \mu_{t-1})^2 + \beta h_{t-1}, \qquad \text{(known at time } t-1), \tag{19}$$

$$h_t = g_t + (\phi_1 \mathbb{1}_{\varepsilon_t \leq 0} + \phi_2 \mathbb{1}_{\varepsilon_t > 0})\varepsilon_t^2, \tag{20}$$

where $y_t = \text{asinh}(p_t)$ the transformed ID_3 price, μ_t the conditional mean, X_t the regressors contained in the base model and $\omega, \alpha_1, \alpha_2 \phi_1, \phi_2 \geq 0$ the non-negative parameters driving the volatility. Notice that, as discussed in Section 1, this model nests asymmetric GARCH when $\phi_1 = \phi_2 = 0$. Additionally, when $\alpha_1 = \alpha_2 = \alpha$ the symmetric GARCH model is retrieved.

5 Estimation, forecasting and evaluation

This section summarizes the methods of estimation and forecasting of the different models. Additionally, this section also discusses the methods of evaluation of both the price point predictions and the probabilistic interval predictions.

5.1 Estimation

First, let us define some generalized estimation method using quasi maximum likelihood (QMLE) such that the methods of estimating the benchmark models are clear. As the benchmark models incorporate both normally distributed errors as well as student-t distributed errors, we will define the generalized estimation methods as such:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{ML} &= argmax_{\boldsymbol{\theta}} \sum_{t=1}^{T} \log(f(y_t|I_{t-1};\boldsymbol{\theta})) = argmax_{\boldsymbol{\theta}} \sum_{t=1}^{T} \Big[-\log(\sigma_{t|t-1}) + \log p\Big(\frac{y_t - \mu_t}{\sigma_{t|t-1}}\Big)\Big] \\ &= argmax_{\boldsymbol{\theta}} \sum_{t=1}^{T} \Big[-\log(\sigma_{t|t-1}) + \log p\big(\varepsilon_t\big)\Big],\end{aligned} \tag{21}$$

where $y_t = \text{asinh}(p_t)$ the transformed price, $\sigma_{t|t-1}$ the standard deviation as estimated by the different models and $p(\cdot)$ either the p.d.f. of the student-t distribution or the normal distribution.

The estimation for RT-GARCH generalizes to the following:

$$\hat{\boldsymbol{\theta}}_{ML} = argmax_{\boldsymbol{\theta}} \sum_{t=1}^{T} \Big[\log(\frac{\sqrt{h_t}}{h_t + F(y_t)\varepsilon_t^2}) + \log p\big(\varepsilon_t\big)\Big], \tag{22}$$

where $F(y_t) = \phi_1 \mathbb{1}_{\varepsilon_t \leq 0} + \phi_2 \mathbb{1}_{\varepsilon_t > 0}$. Due to this, the expression in 22 reverts to 21, with student-t distributed errors, when $\phi_1 = \phi_2 = 0$ as $F(y_t) = 0$.

5.2 Forecasting

To obtain the conditional mean forecasts, we simply use the estimated parameter(s) $\hat{\gamma}$ as obtained using QMLE and define the conditional mean forecasts as:

$$\hat{\mu_{t+k}} = \hat{\gamma} X_{t+k}, \tag{23}$$

where X_{t+k} the regressors known at time t for time $t+k$.

In addition to the conditional mean forecasts, we compute the volatility predictions. As RT-GARCH nests both asymmetric GARCH and symmetric GARCH, we derive only the k step-ahead forecasts for the RT-GARCH model. Let us define the constant terms h and ω as follows:

$$h := \frac{\omega + \phi + \alpha_1(\phi_1 K_\varepsilon - \phi)/2 + \alpha_2(\phi_2 K_\varepsilon - \phi)/2}{1 - \alpha - \beta}, \tag{24}$$

$$\omega := h(1 - \alpha - \beta) - \phi - \alpha_1(\phi_1 K_\varepsilon - \phi)/2 - \alpha_2(\phi_2 K_\varepsilon - \phi)/2, \tag{25}$$

where $K_\varepsilon = 3 + 6/(\nu - 4)$, ν the estimated shape of the student-t distribution and $\phi = \phi_1 + \phi_2$. The dynamic equation for h_{t+k} is then defined as follows:

$$h_{t+k} = h + (\alpha + \beta)^{k-1}(g_{t+1} + \phi - h), \qquad k \geq 1. \tag{26}$$

Notice that this dynamic equation nests the symmetric and asymmetric GARCH models under the same aforementioned restrictions. Additionally, we set $\sqrt{h_{t+k}} = \hat{\sigma}_{t+k}$ for evaluation purposes.

5.3 Point prediction evaluation

To evaluate the price point forecasts, the six robust loss functions proposed by (Hansen & Lunde, 2005) are evaluated:

$$\text{RMSE}_1 = \sqrt{T^{-1}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}, \qquad \text{RMSE}_2 = \sqrt{T^{-1}\sum_{t=1}^{T}(y_t^2 - \hat{y}_t^2)^2},$$

$$\text{QLIKE} = T^{-1}\sum_{t=1}^{T}(\ln(\hat{y}_t) + y_t^2/\hat{y}_t^2)^2, \qquad R^2\text{LOG} = T^{-1}\sum_{t=1}^{T}[\ln(y_t^2/\hat{y}_t^2)]^2,$$

$$\text{MAE}_1 = T^{-1}\sum_{t=1}^{T}|y_t - \hat{y}_t|, \qquad \text{MAE}_2 = T^{-1}\sum_{t=1}^{T}|y_t^2 - \hat{y}_t^2|.$$

The criteria RMSE_2 and $R^2\text{LOG}$ can be interpreted as similar to the R^2 of the Mincer- Zarnowitz regressions, whereas QLIKE corresponds to the loss implied by a Gaussian likelihood. As outliers occur frequently in electricity prices, it is interesting to look at robust error measures. Both MAE_1 and MAE_2 are used for this purpose.

The price point forecasts will be evaluated using RMSE_1, RMSE_2, MAE_1 and MAE_2. These forecasts are obtained as $\widehat{\text{asinh}(p_{t+k})}$, the k'th step-ahead forecast, and need to be backwards transformed using either the exact or approximate method discussed in Section 3.3 to evaluate them against the observed prices.

In contrast, the backwards transformed forecasted standard deviations are trickier to compute and can be defined as follows:

$$\hat{\sigma}_{t+k,M} = \hat{P}^M_{t+k} - bwt_M(\widehat{\text{asinh}(p}_{t+k}) - \hat{\sigma}_{t+k}), \tag{27}$$

where $M = \{\text{E}, \text{A}\}$ either the exact or approximate transformation, $\hat{\sigma}_{t+k,M}$ the eventual backwards transformed forecasted standard deviation used for evaluation, $\hat{P}^M_{t+k}$ the backwards transformed forecasted price, $\hat{p}_{t+k}$ the forecasted price, σ_{t+k} the forecasted standard deviation and $bwt_M(.)$ the backwards transformation function.

Additionally, a proxy for the volatility has to be obtained to evaluate the volatility forecasts. To obtain this proxy, we remove the noise/ conditional mean from the price series. Hence, the residuals from the base model estimated on the full sample will be used. This is defined as:

$$\begin{aligned} \varepsilon_t^2 &= (\text{asinh}(p_t) - X_t\beta)^2, \\ \sigma_t &= \hat{\varepsilon}_{t,M} - \hat{\varepsilon}_{t-1,M}, \end{aligned} \tag{28}$$

where $\text{asinh}(p_t)$ the transformed ID_3 price, X a matrix with the regressors contained in the base model, $\hat{\varepsilon}_{t,M}$ the backwards transformed ε_t using the exact/ approximate transformation. This transformation is done using the full-in-sample residuals in case of the exact transformation. Finally, the volatility forecasts are evaluated using all loss functions.

5.4 Probabilistic interval evaluation

To further evaluate the volatility forecasts, the methodology for evaluating interval forecasts as introduced by (Christoffersen, 1998) is used. To evaluate these interval forecasts, we first define a function that indicates the violations of a probabilistic interval prediction:

$$I_{t+k} = \begin{cases} 0 & \text{if } P_{t+k} \in (L_{t+k|t}(q), U_{t+k|t}(q)) \\ 1 & \text{otherwise,} \end{cases} \tag{29}$$

where $L_{t+k|t}(q) = \hat{P}^M_{t+k} - Q(1-\frac{q}{2}) \cdot \hat{\sigma}_{t+k,M}$ the lower bound for the interval, $U_{t+k|t}(q) = \hat{P}^M_{t+k} + Q(1-\frac{q}{2}) \cdot \hat{\sigma}_{t+k,M}$ the upper bound for the interval and $Q(1-\frac{q}{2})$ the quantile function of either the normal distribution or the student-t distribution, this depends on the assumption of errors within the model used. In the case the assumption on student-t distributed errors is used, the shape of the student-t distribution estimated within the model is used as the degrees of freedom in the quantile function.

To test whether interval estimates have correct unconditional coverage, we test whether the fraction of violations is equal to the nominal coverage probability q. We define this test as follows:

$$\begin{aligned} \text{H}_0 &: \mathcal{L}(q; I_{T,I_{T-1},\dots,I_1}) = (1-q)^{T_0} q^{T_1}, \\ \text{H}_1 &: \mathcal{L}(\pi; I_{T,I_{T-1},\dots,I_1}) = (1-\pi)^{T_0} \pi^{T_1}, \\ LR_{uc} &= -2 \log \frac{\mathcal{L}(q; I_{T,I_{T-1},\dots,I_1})}{\mathcal{L}(\pi; I_{T,I_{T-1},\dots,I_1})} \sim \chi^2(1), \end{aligned} \tag{30}$$

where the maximum likelihood (ML) estimate of π is $\hat{\pi} = \frac{T_1}{T_0+T_1}$, T_0 being the number of non-violations and T_1 the amount of violations.

Additionally, we test whether the interval estimates are independent. The violations should be spread out over the sample and not cluster together. We define this test as follows:

$$\begin{aligned}
\text{H}_0 &: \mathcal{L}(\pi_2; I_{T,I_{T-1},\ldots,I_1}) = (1-\pi_2)^{T_{00}+T_{10}} \pi_2^{T_{01}+T_{11}}, \\
\text{H}_1 &: \mathcal{L}(\Pi_1; I_{T,I_{T-1},\ldots,I_1}) = (1-\pi_{01})^{T_{00}} (\pi_{01})^{T_{01}} (1-\pi_{11})^{T_{10}} (\pi_{11})^{T_{11}}, \\
LR_{ind} &= -2 \log \frac{\mathcal{L}(\hat{\pi}_2; I_{T,I_{T-1},\ldots,I_1})}{\mathcal{L}(\hat{\Pi}_1; I_{T,I_{T-1},\ldots,I_1})} \sim \chi^2(1),
\end{aligned} \tag{31}$$

where T_{ij} is the number of observations such that $I_{t+k} = j$ and $I_{t+k-1} = i$, the ML estimate of π_2 is $\hat{\pi}_2 = \frac{T_{01}+T_{11}}{T}$ and the ML estimate of Π_1 is defined as:

$$\hat{\Pi}_1 = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} = \begin{pmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{pmatrix}.$$

6 Results

This section will discuss both the price forecasting and the volatility forecasting results of the aforementioned models. The models are estimated on a rolling-window of 365 days, refitting every 7 days. The models will then use an out-of-sample series to forecast transformed prices and transformed standard deviations for 3 hours ahead for the next 7 days. These results are then backwards transformed as discussed in Section 3.3.

6.1 Price point forecasts

Table 3 contains the error measures for the price point forecasts. We highlight three main results based on this table. First, based on these criteriums, we can see the benchmark base model clearly outperforms all models accounting for conditional heteroskedasticity except for the Realized GARCH(1,1) with skewed student-t distributed errors. This extension on the Realized GARCH model has a lower RMSE_1 and RMSE_2 (by a small margin) compared to the base model. However, when comparing the outlier-robust error measures MAE_1 and MAE_2 the base model outperforms yet again.

Second, Real-time GARCH (including extensions) fails to outperform Realized GARCH and its extension. Moreover, Real-time GARCH does not even seem to outperform a standard GARCH process with skewed student-t distributed errors.

Third, comparing the error measures from the exact and approximate price transformations, one could conclude that incorporating the residuals in the price transformation definitely helped improve the forecasting accuracy (albeit a small improvement). However, notice that this improvement does not necessarily hold for the MAE measures, but rather the $RMSE$ measures.

	Exact				Approximate			
	$RMSE_1$	$RMSE_2$	MAE_1	MAE_2	$RMSE_1$	$RMSE_2$	MAE_1	MAE_2
Benchmark Models with normal distributed errors								
Base	6.005	911.023	3.713	285.469	6.086	926.056	3.707	284.904
GARCH(1,1)	6.180	944.319	3.808	294.277	6.296	967.954	3.808	294.631
GJR-GARCH(1,1)	6.180	944.504	3.808	294.285	6.296	968.197	3.808	294.651
Realized GARCH(1,1)	6.029	914.140	3.751	289.967	6.138	938.028	3.746	290.159
Benchmark Models with skewed student-t distributed errors								
GARCH(1,1) - sstd	6.102	934.178	3.772	291.611	6.205	957.783	3.760	291.100
GJR-GARCH(1,1) - sstd	6.100	934.180	3.772	291.575	6.203	958.271	3.759	291.064
Realized GARCH(1,1) - sstd	6.003	909.791	3.735	288.772	6.097	932.109	3.719	287.832
Real-time GARCH								
Real-time GARCH(1,1)	6.103	934.247	3.772	291.523	6.210	957.929	3.763	291.277
Real-time GARCH(1,1) - Leverage	6.108	937.009	3.772	291.681	6.214	960.858	3.762	291.340
Real-time GARCH(1,1) - Feedback	6.103	934.095	3.771	291.577	6.209	957.640	3.762	291.206

Table 3: Out-of-sample price point forecasts error evaluations (in EUR/MWh) for time period 2014-01-01 to 2019-12-31. The forecasts are generated for 3 hours ahead and the models are refitted every 7 days.

Figure 4 show the aforementioned error measures divided by hour. Based on this figure, we deduct and highlight 3 findings. First, it can be seen that all models follow similar patterns regarding error measures. Especially during night hours the errors seem to be the lowest, reaching a minimum of around 2.9 and 4.4 for respectively MAE_1 and $RMSE_1$ on 04:00. However, a maximum is reached on hour 08:00 for MAE_1 and for MAE_1. As MAE_1 is more outlier-robust, different peaks are formed compared to $RMSE_1$.

Secondly, in MAE_1, the high demand hours can be clearly extracted from the graph. These are respectively, based on value, 08:00-09:00 being the morning rush hour; 19:00-20:00 being the evening high demand hour and finally 12:00-13:00 being demand during the lunch hour. Furthermore, the highest peak of $RMSE_1$ is formed on 12:00-13:00. This can be most likely attributed to more outliers in this hour in the form of extreme weather changes, outages or unexpected peak demands.

Third, during the peaks in Figure 4 the differences between the models become visible. Especially Real-time GARCH - Feedback and GJR-GARCH(1,1) fail to follow the same pattern as the other models and often have higher error measures. Moreover, the two right-most graphs clearly show the decrease in upwards bias whilst using the exact price transformation versus the approximate price transformation.

Lastly, Figure 5 shows the error measures, MAE_1 and $RMSE_1$, over time. From this figure a seasonal pattern can be observed. Additionally, testing for the seasonality proves that seasonality indeed does exist. This seasonality takes on the form of small error measures during the spring and summer contrasted by high error measures during the fall and winter. This seasonality can most likely be attributed to weather.

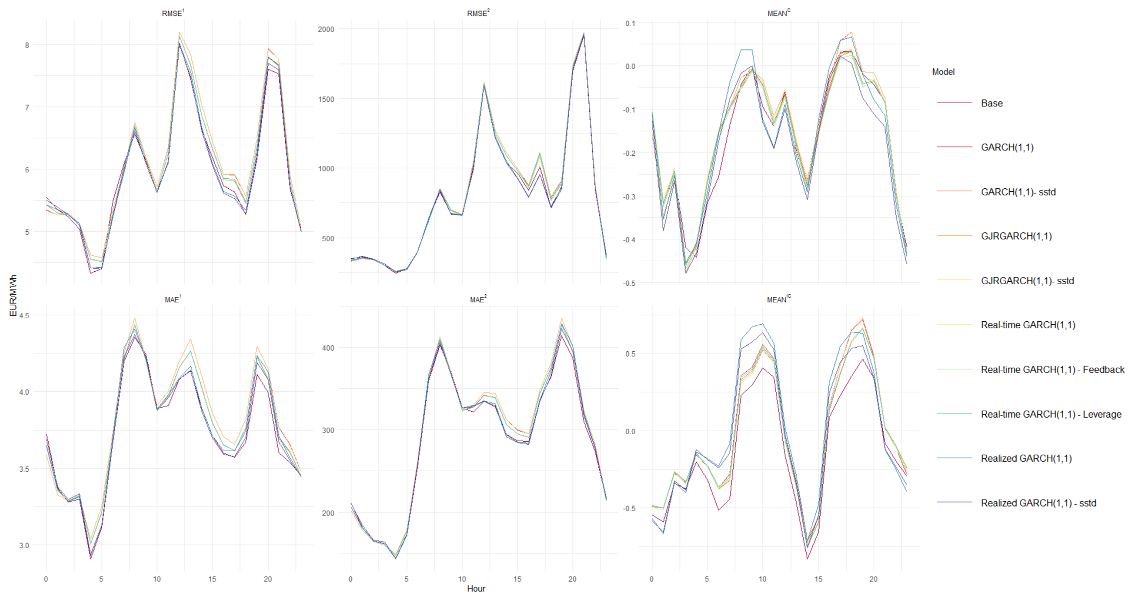

Figure 4: Out-of-sample price point forecast error evaluations (in EUR/MWh) grouped by hour. These errors are evaluated using the exact price transformation, however the mean of the residuals for both exact (north-east) and approximate (south-east) price transformations are shown for comparison.

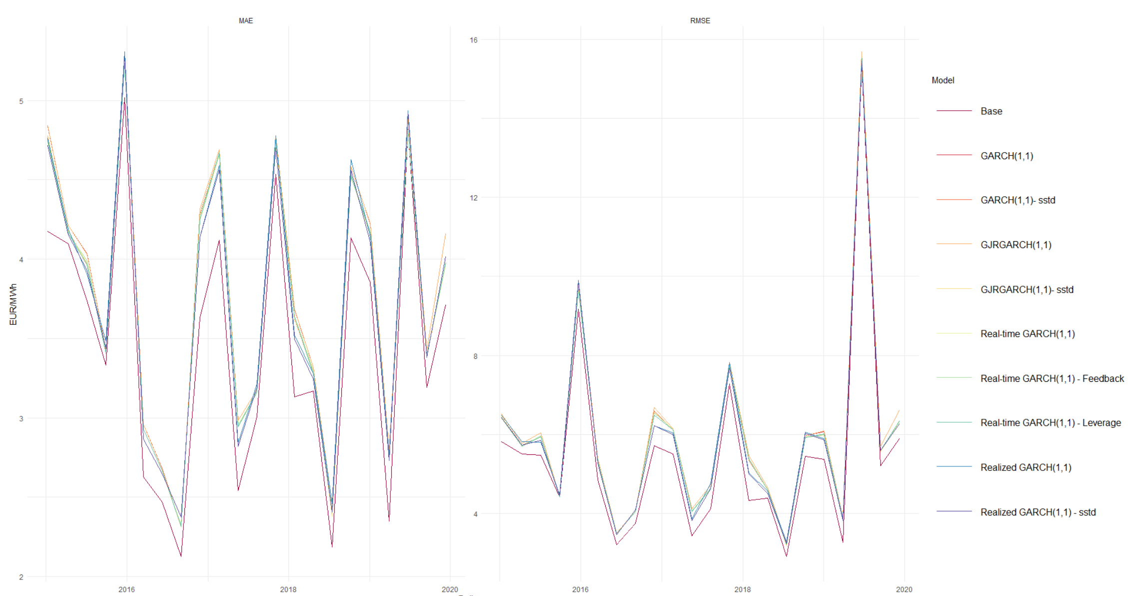

Figure 5: 1-month rolling mean out-of-sample price point forecast error evaluations over time (in EUR/MWh) for time period 2014-01-01 to 2019-12-31. These errors are evaluated using the exact price transformation.

6.2 Volatility forecasts

The averaged coefficients of all parameters are shown in Table 4. Below, we highlight four main results based on this table. First, in contrast to GARCH-like models fitted on financial data, the coefficients for α, α_1 and α_2 are high whereas the coefficient for β is low. However, whilst evaluating the auto correlation from the squared residuals, this event seems justified as the auto correlation disappears fast and with about the same term as α, α_1 and α_2 in all the models. This effect can be intuitively explained as conditional heteroskedasticity arises when there are sudden big spikes in the prices that persist over some period, e.g. sudden big spikes in electricity prices arise due to extreme weather, outages or unexpected demand. However, unlike with financial data where these events (and its effects) can last for months, the aforementioned events often only last for a few hours in electricity markets.

	Normal			Skewed student-t			Real-time GARCH(1,1)		
	GARCH(1,1)	GJR-GARCH(1,1)	Realized GARCH(1,1)	GARCH(1,1)	GJR-GARCH(1,1)	Realized GARCH(1,1)	Regular	Leverage	Leverage and Feedback Feedback
					Volatility				
α	0.539	0.558	0.087	0.617	0.625	0.089	0.482	0.484	-
α_1	-	-	-	-	-	-	-	-	0.446
α_2	-	-	-	-	-	-	-	-	0.488
β	0.142	0.141	0.691	0.113	0.111	0.686	0.145	0.123	0.122
γ	-	**-0.05**	-	-	**-0.026**	-	-	-	-
η_1	-	-	0.032	-	-	0.035	-	-	-
η_2	-	-	0.186	-	-	0.164	-	-	-
ϕ	-	-	-	-	-	-	**0.001**	-	-
ϕ_1	-	-	-	-	-	-	-	0.001	**0**
ϕ_2	-	-	-	-	-	-	-	0.001	**0.001**
					Mean				
ω	0.041	0.041	-0.496	0.04	0.041	-0.49	0.029	0.03	0.03
					Shape				
ν	-	-	-	9.627	9.374	6.455	12.333	10.545	9.946
sk	-	-	-	1.029	1.023	1.017	-	-	-
					Goodness of fit				
LL	2075.07	2071.933	22281.654	2032.548	2028.467	**22119.982**	2032.945	2036.338	2033.804
AIC	-8.973	-6.973	**-10.019**	-6.917	-4.913	-8.003	-4.914	-2.916	-0.915
BIC	**12.26**1	21.339	25.371	21.395	30.477	34.464	30.476	39.552	48.631

Table 4: Coefficients of the parameters and goodness of fit measures of all models averaged over all estimation periods (sample size of 365 days, refitted every 7 days) from 2014-01-01 to 2019-12-31. The winners of goodness of fit measures and parameters of interest are shown in bold.

Second, we find a confirmation of the existence of inverse leverage in electricity prices with a small inverse leverage effect contained in the GJR-GARCH(1,1) models.

Third, the real time shape parameters ϕ, ϕ_1 and ϕ_2 are extremely small compared to the results from financial data in the work of (Smetanina, 2017). This implies that either a small parameter is sufficient enough to capture the 'real-time' effect, or the parameter is insignificant. As the auto regressive parameters (e.g. α, β) differ from the GARCH model, and using the results shown later in this section we can argue the former is true.

Finally, comparing the goodness of fit measures amongst models, we find GARCH with normal distributed errors to have the lowest BIC value. This can be attributed to the fact that GARCH only has 3 parameters and hence is the most parsimonious. Furthermore, we find Realized GARCH with normal distributed errors to have the lowest AIC among all models, whilst also having a log-likelihood much larger than the other models.

In Table 5 we can see the error measures of the 3-hour ahead volatility forecasts compared to the proxy obtained for volatility. First, we find that the models that incorporate for current information, Realized GARCH and Real-time GARCH, outperform models that only rely on past information. Realized GARCH and Real-time GARCH - Feedback improve drastically upon the worst-performing model, GJR-GARCH (1,1) - sstd, for both the exact and approximate transformation. For example, this improvement in RMSE_1 can be as much as 47.1% and 45.6% for respectively Realized GARCH and Real-time GARCH - Feedback upon the worst-performing model using the exact price transformation.

Secondly, comparing the real-time GARCH model and its extensions, we find the Feedback extension provides the overall best results for forecasting the volatility. However, Realized GARCH still seems to consistently beat all extensions of the Real-time GARCH model by at least some 2.5% and 6.2% for respectively RMSE_1 and MAE_1. Moreover, when comparing the general results of all models for an exact and approximate price transformation, we find that in the case of forecasting volatility the approximate transformation outperforms the exact transformation.

To better understand the origin of the errors, the error evaluations are grouped by hour (similar to Figure 4) in Figure 6. As can be deducted from the figure, the error measures from both the Realized GARCH model and Real-time GARCH models lie a significant amount below those of models incorporating only past returns. Additionally, the peaks in MAE_1 get smoothed out in these models, with only a peak remaining at 08:00-09:00. This holds especially true for the Realized GARCH model. However, this out performance of the Realized GARCH model can be partly explained by the fact that a proxy for the volatility, see Equation 16, is directly fed into the model (unlike with the other models which do not contain such a volatility proxy). Additionally, the peaks in both MAE_1 and RMSE_1 tell us the same story as was found in Figure 4: the hours 08:00-09:00, 19:00-20:00 and 12:00-13:00 are the most unpredictable hours during the day.

Using these volatility forecasts, violation statistics can be computed as explained in Section 5.4. The full tables with all relevant statistics can be found in the Appendix. Three main findings can be brought to attention here. First, all models reject the null hypothesis of independence across all confidence levels, implying that $\pi_{01} \neq \pi_{11}$ and hence the violations are clustered together. The

	Exact						Approximate					
	$RMSE_1$	$RMSE_2$	QLIKE	R^2LOG	MAE_1	MAE_2	$RMSE_1$	$RMSE_2$	QLIKE	R^2LOG	MAE_1	MAE_2
Benchmark Models with normal distributed errors												
GARCH(1,1)	6.619	443.001	7.762	34.287	4.495	72.117	6.253	415.328	8.649	32.586	4.223	65.469
GJR-GARCH(1,1)	6.532	398.493	7.815	34.028	4.444	70.480	6.169	371.090	8.723	32.340	4.176	63.947
Realized GARCH(1,1)	4.088	231.011	102.642	-	2.662	30.524	4.008	230.693	129.363	-	2.571	29.230
Benchmark Models with skewed student-t distributed errors												
GARCH(1,1)- sstd	7.654	603.282	7.499	36.703	4.993	91.288	7.215	553.909	8.284	34.911	4.690	82.517
GJR-GARCH(1,1)- sstd	7.663	563.533	7.534	36.595	4.978	91.301	7.212	512.154	8.327	34.805	4.674	82.345
Realized GARCH(1,1) - sstd	4.104	231.238	98.565	-	2.671	30.760	4.022	230.859	124.814	-	2.578	29.427
Real-time GARCH												
Real-time GARCH(1,1)	4.332	232.387	46.947	-	2.977	34.618	4.208	231.331	54.376	-	2.847	32.752
Real-time GARCH(1,1) - Leverage	4.272	232.029	48.395	-	2.918	33.773	4.157	231.093	56.384	-	2.795	32.027
Real-time GARCH(1,1) - Feedback	4.205	231.525	59.774	-	2.858	32.810	4.098	230.736	69.363	-	2.741	31.194

Table 5: Out-of-sample volatility forecasts error evaluations (in EUR/MWh) using the proxy volatility as the objective. These criteria are computed over time period 2014-01-01 to 2019-12-31. The volatility forecasts are generated for 3 hours ahead and the models are refitted every 7 days. A value of "-" means it wasn't numerically possible to compute a result.

Figure 6: Out-of-sample volatility forecast error evaluations (in EUR/MWh) grouped by hour. These errors are evaluated using the exact price transformation, however the mean of the residuals for both exact (north-east) and approximate (south-east) price transformations are shown for comparison.

explanation here is that the models do not have sufficient time to change to a new level of conditional volatility before the volatility drops down again, and thus not capturing the large price spikes. This is due to the price spikes often only lasting for a couple hours (see the large $\alpha, \alpha_1, \alpha_2$ in Table 4 and the autocorrelation of the squared residuals in the Appendix).

In contrast, the null hypothesis for unconditional coverage is accepted for a few models across some confidence levels. In particular, RT-GARCH - Feedback using the approximate transformation with a confidence level of 66.66% and RT-GARCH - Feedback using the exact transformation with a confidence level of 90%. However, due to the amount of data (44114 out-of-sample observations) the test might be too strict. Observing the $\hat{\pi}$ instead, we notice that these values are not too different from q across all confidence levels for especially the RT-GARCH - Feedback and RT-GARCH - Leverage models. Additionally, when comparing the difference between $\hat{\pi}$ and q for all models and across all confidence levels we exclusively find the RT-GARCH - Feedback and RT-GARCH - Leverage models to have the lowest difference. Hence, we can conclude that these aforementioned models capture the volatility more accurately than the other models with the goal of constructing probabilistic interval forecasts. However, compared to GARCH and GJR-GARCH, that often overestimate the interval bounds and hence have significantly lower $\hat{\pi}$'s, RT-GARCH (which computes more precise intervals) is more subject to extreme outliers. The effect of outliers on the size of the maximum violation can be seen in Figure 8 in which the less conservative models, RT-GARCH and Realized GARCH, have a higher violation size due to not being conservative in their interval forecasts. Notice however, when comparing the average of all violation sizes across all models, we find these models are not significantly different from each other. More importantly, comparing the average intervals of all models, we find RT - GARCH (but also Realized GARCH) to have significantly smaller intervals. Thus, we can conclude that RT - GARCH and its extensions predict more accurate and precise intervals than its GARCH counterparts.

Finally, in contrast to the conservative estimates of GARCH and GJR-GARCH, the Realized GARCH models actually underestimates the interval bounds. This phenomenon actually seems to be very time dependent, as visualized in Figure 7. Between 2016-2017 and in the end of 2019, the Realized GARCH model consistently underestimates the interval bounds. This is likely due to the variance targeting within the Realized GARCH model as its error measures were competitive as seen in Table 5 but the model doesn't hold up when evaluating its interval forecasts. Since the Realized GARCH model uses a proxy volatility (which is very similar to the volatility computed for evaluation), it has lower error measures than other models but fails to capture the 'real' volatility process which can be deduced from the interval forecasts. Using this observation, we can conclude that even though Realized GARCH constructs even smaller intervals, its intervals often do not include the observed price and can not outperform the predicted intervals from RT - GARCH.

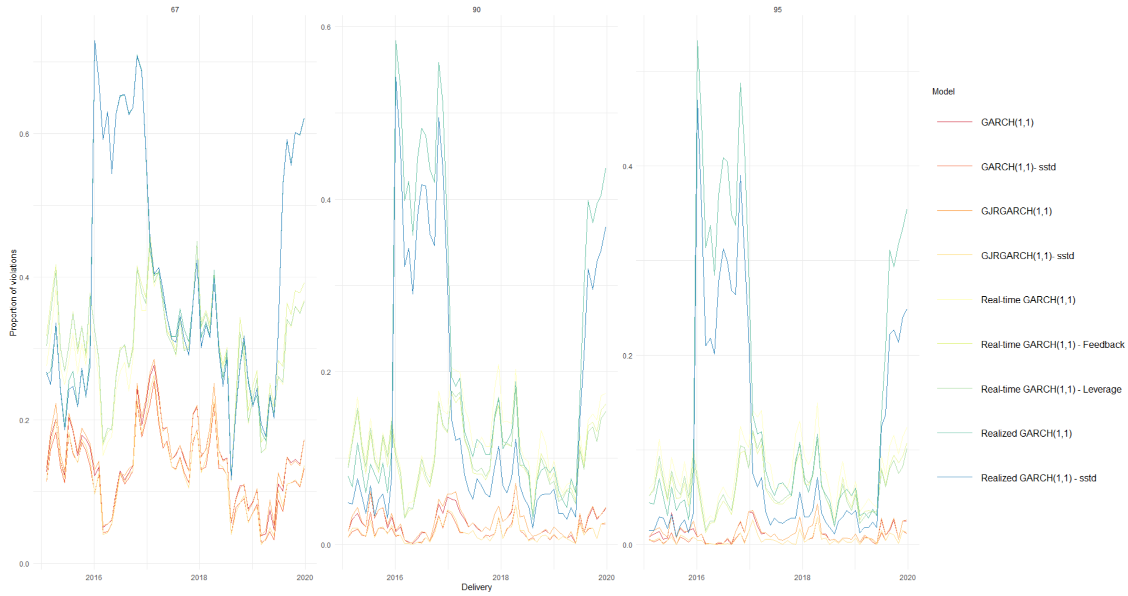

Figure 7: 1-month rolling mean of violation statistic I_{t+k} as in Equation 29 using the exact price transformation. This is visualized for time period 2014-01-01 to 2019-12-31.

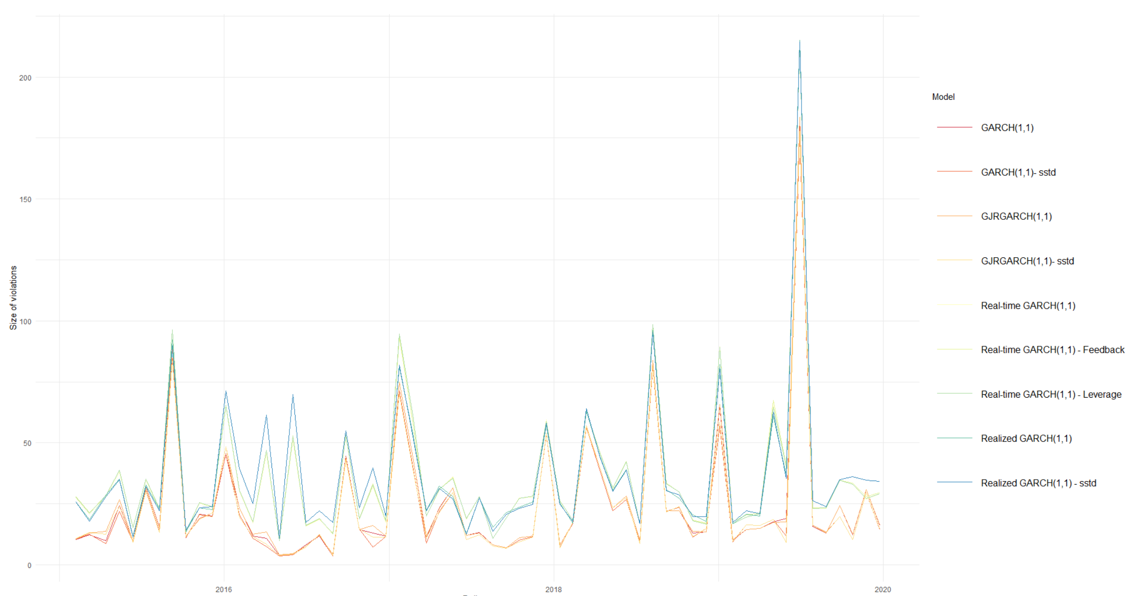

Figure 8: 1-month rolling mean of maximum violation size using the exact price transformation with a confidence of 66.67%. This can be mathematically expressed as $y_{k,t} = \max_{i \in \{t-k+1, t-k, \ldots, t\}} I_t \cdot |\hat{\sigma}_{i,M} - \hat{\varepsilon}_{i,M}|$ where $\hat{\sigma_{i,M}}$ the backwards transformed forecasted volatility, $\hat{\varepsilon_{i,M}}$ the residuals from the backwards transformed forecasted price and $k = 30$. These statistics are visualized for time period 2014-01-01 to 2019-12-31.

7 Conclusion

To conclude, RT - GARCH and especially its Feedback extension can provide more accurate and narrow forecasted intervals for ID_3 prices for 3 hours ahead than its GARCH counterparts can. However, accounting for conditional heteroskedasticity in forecasting the ID_3 prices neither decreases, nor significantly increases the error measures compared to the base model, unlike what was hypothesized. Additionally, we can confirm the finding of (Narajewski & Ziel, 2020) of an increase in electricity price forecasting performance using their exact backwards transformation over a larger time period (2014-2019). However, this increase in performance does not necessarily hold for computing volatility forecasts.

Whilst forecasting 3 hours ahead, RT- GARCH and its extensions have significantly smaller intervals than the other models, while also keeping correct unconditional coverage. Due to volatility clustering and short price spikes, tests for independence are rejected for all models. In contrast to what was hypothesized, accounting for conditional heteroskedasticity does not seem to (positively) impact the error measures and hence it can be concluded that the models are either too slow in incorporating the conditional heteroskedasticity or the price spikes are too short to be of extra information within the hourly granularity. Additionally, Realized GARCH has better error measures for forecasting volatility however this is most likely due to the fact that the proxy volatility which the error measures are evaluated against and the proxy volatility fed into the Realized GARCH model is computed in the same manner. When comparing the interval forecasts of the Realized GARCH model, we observe that it actually gets outperformed by RT - GARCH and its extensions. We conclude that whilst RT - GARCH appears most helpful when constructing interval forecasts compared to other GARCH approaches, accounting for conditional heteroskedasticity does not seem to improve the price point forecasting accuracy of the base model. Additionally, we show that the exact (correct) asinh's backward transformation from (Narajewski & Ziel, 2020) give significantly better forecasts over a longer period of time (2014-2019) than the time period chosen in their paper (2014-2015).

As suggested by the literature section, conditional heteroskedasticity models could provide better fits to the data if the extreme outliers had been removed. Especially considering the size of the maximum violations, this could have definitely improved some models. Furthermore, as the very recent price index ${}_{0.5}ID_{0.25}$ (3 hours before delivery) is used to model ID_3, it becomes very hard to improve upon this model. We agree with the weak-form efficiency suggestion made by (Narajewski & Ziel, 2020) and price combinations provide very little, to no additional information at all. Thus, we suspect that incorporating other information like weather and power forecasts, both in the conditional mean as in the volatility process, will increase the model's performance. Moreover, we suggest to use regime-switching models to capture different price regimes and use a multivariate approach in which the hours are separated.

References

Andrade, J. R., Filipe, J., Reis, M., & Bessa, R. J. (2017). Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model. *Sustainability*, *9*(11).

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

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841–862.

Ding, Y. (2021). Augmented Real-Time GARCH: A joint model for returns, volatility and volatility of volatility. *SSRN Electronic Journal*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

EPEX. (2015). New ID3 Price index on continuous German intraday market [Accessed: 2021-05-14]. https://www.epexspot.com/en/news/new-id3-price-index-continuous-german-intraday-market

EPEX. (2021a). Basics of the power market [Accessed: 2021-05-14]. https://www.epexspot.com/en/basicspowermarket#day-ahead-and-intraday-the-backbone-of-the-european-spot-market

EPEX. (2021b). Indices [Accessed: 2021-06-19]. https://www.epexspot.com/en/indices

EPEX. (2021c). Trading Products [Accessed: 2021-05-14]. https://www.epexspot.com/en/tradingproducts#intraday-trading

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.

Hill, J., Gismatullin, E., & Morison, R. (2021). Texas power firm hit with 2.1 billion bill files for bankruptcy. https://www.bloomberg.com/news/articles/2021-03-01/a-texas-power-firm-files-for-bankruptcy-after-historic-outages

Lange, R. (2021). A short explanation of RT-GARCH (with proofs).

Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166.

Narajewski, M., & Ziel, F. (2020). Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, *19*, 100107.

Politis, D. (2007). Model-free versus model-based volatility prediction. *Journal of Financial Econometrics*, *5*(3), 358–389.

Smetanina, E. (2017). Real-Time GARCH. *Journal of Financial Econometrics*, *15*(4), 561–601.

Uniejewski, B., Marcjasz, G., & Weron, R. (2019). Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *International Journal of Forecasting*, *35*(4), 1533–1547.

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*(4), 1030–1081.

Appendix

Figures

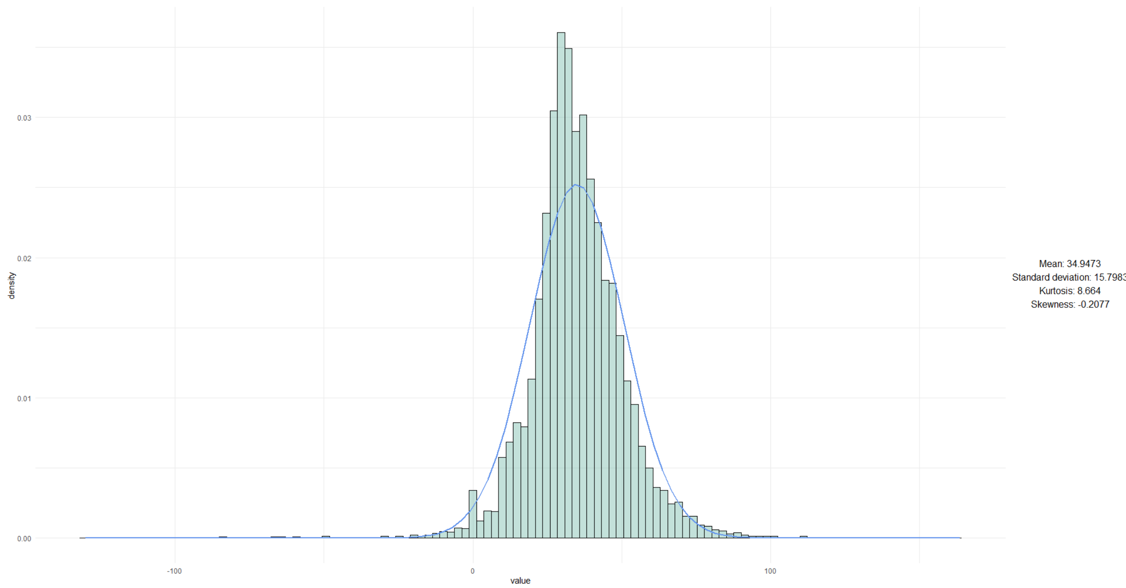

Figure 9: Distribution of day-ahead settlement prices from 2014-01-01 to 2019-12-31. The blue line depicts the normal distribution scaled with the first and second moment of the day-ahead settlement prices.

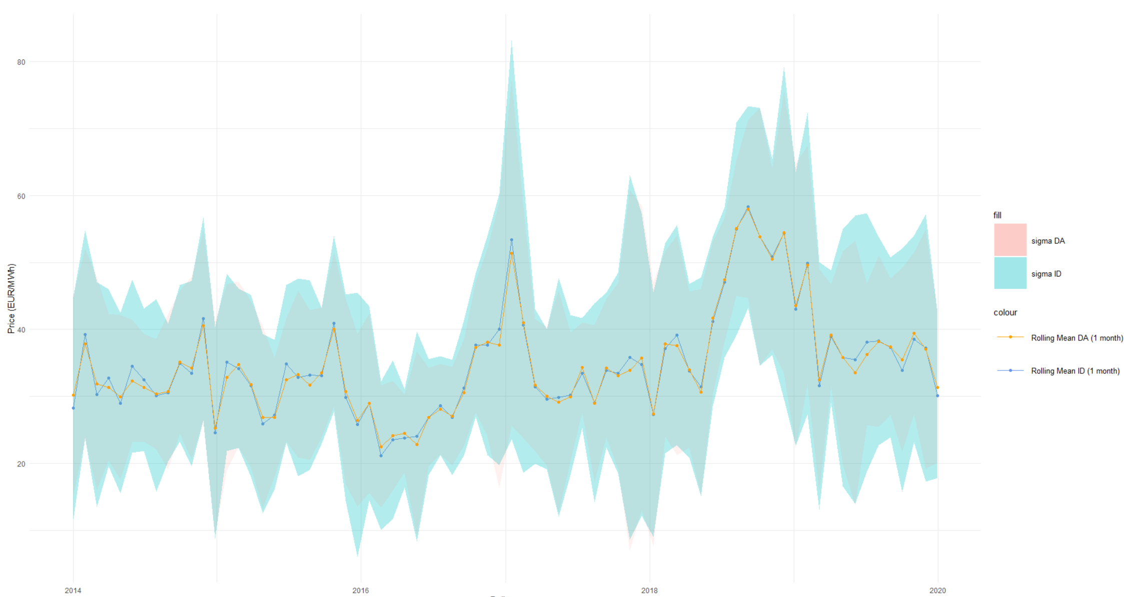

Figure 10: Rolling mean (1 month) of the non-transformed ID and DA time series along with the rolling respective standard deviations from 2014-01-01 to 2019-12-31.

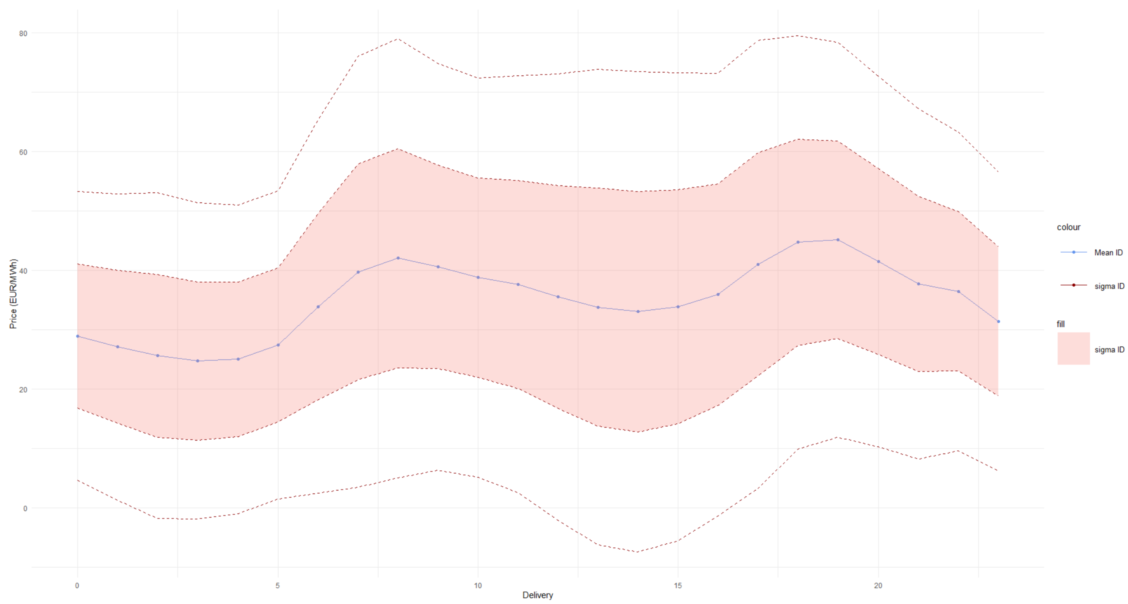

Figure 11: The mean grouped by hour of the non-transformed ID prices along with the standard deviation for the same hour.

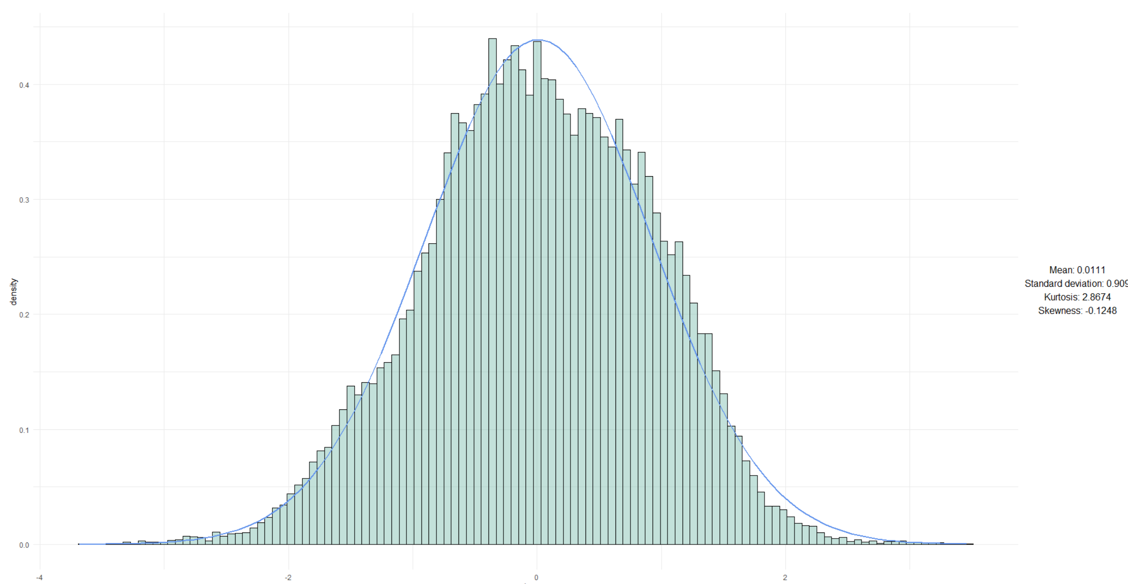

Figure 12: Distribution of transformed ID_3 prices from 2014-01-01 to 2019-12-31. The blue line depicts the normal distribution scaled with the first and second moment of the intraday prices.

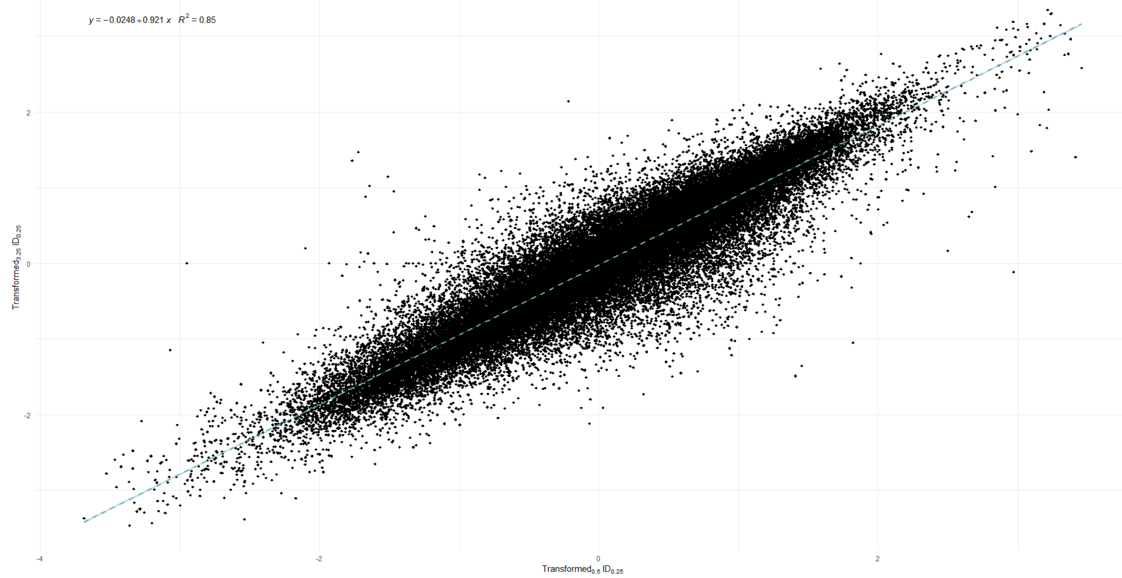

Figure 13: Scatterplot of ID_3 transformed prices against ${}_{0.5}ID_{0.25}$ prices (VWAP of first 15 minutes of trading in 3.25 hours). The green line depicts the linear relationship between the two variables in the form $ID_3 = \alpha + \beta_{0.5}ID_{0.25}$.

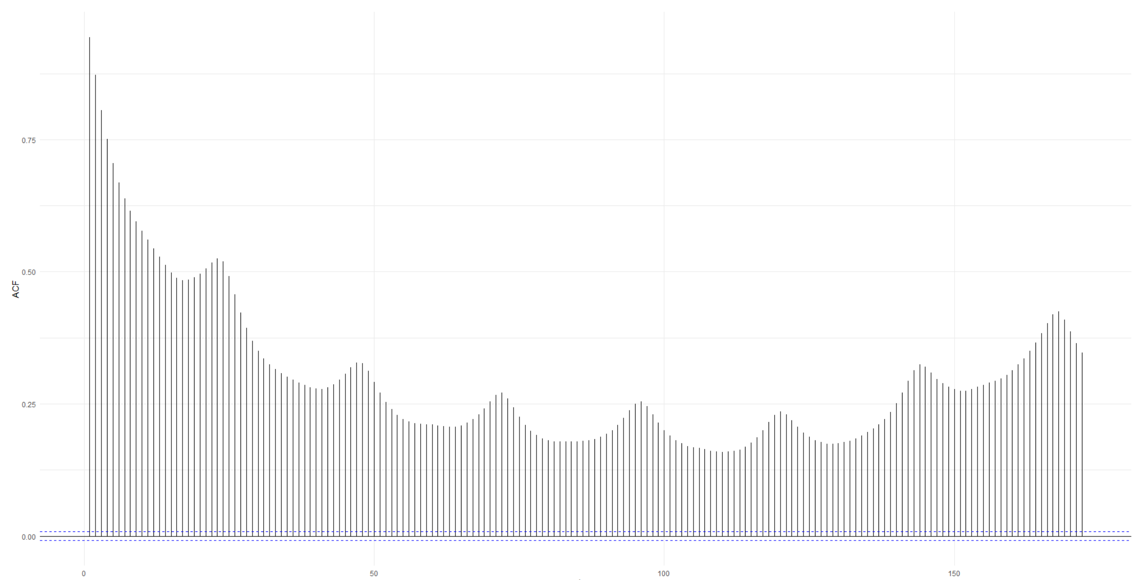

Figure 14: Autocorrelation of the transformed ID_3 prices up to lag 172. The transformation smoothes out the autocorrelation and removes peaks in lag $12 \cdot x$ where $x \in (1, 3, 5, 7, ...)$.

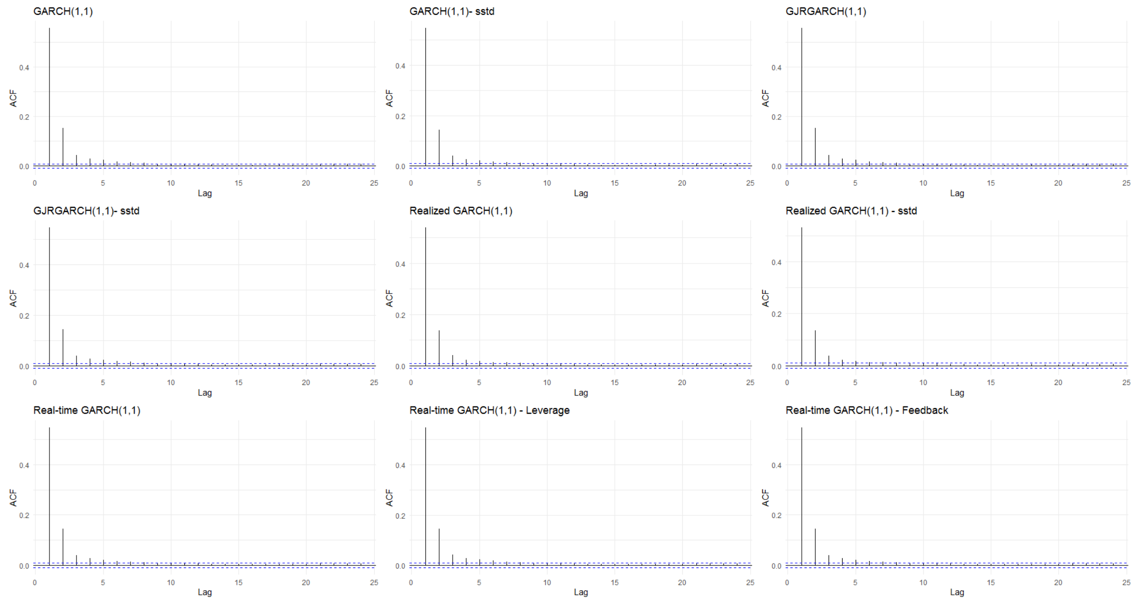

Figure 15: Autocorrelations of the squared residuals from the models complemented with GARCH-like components to capture the conditional heteroskedasticity.

Tables

$q = 0.33$, Exact Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	6364	0.144	7941.903	33970	3779	2585	0.100	0.406	0.144	3246.744
GJR-GARCH(1,1)	6501	0.147	7651.562	33771	3841	2660	0.102	0.409	0.147	3289.869
Realized GARCH(1,1)	17397	0.394	802.896	19695	7021	10375	0.263	0.596	0.394	4927.055
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1) - sstd	5507	0.125	9919.834	35144	3462	2045	0.090	0.371	0.125	2646.127
GJR-GARCH(1,1) - sstd	5623	0.127	9635.108	34991	3499	2124	0.091	0.378	0.127	2756.266
Realized GARCH(1,1) - sstd	17315	0.393	757.749	19792	7006	10308	0.261	0.595	0.393	4933.809
Real-time GARCH										
Real-time GARCH(1,1)	12985	0.294	258.586	24973	6155	6829	0.198	0.526	0.294	4541.428
Real-time GARCH(1,1) - Leverage	13372	0.303	146.231	24476	6265	7106	0.204	0.531	0.303	4552.872
Real-time GARCH(1,1) - Feedback	13714	**0.311**	73.714	24010	6389	7324	0.210	0.534	0.311	4472.065

Table 6: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.33$ and using the exact price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.835}(1) = 1.928$.

$q = 0.33$, Approximate Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	7066	0.160	6524.486	33002	4045	3021	0.109	0.428	0.160	3621.635
GJRGARCH(1,1)	7192	0.163	6288.068	32812	4109	3083	0.111	0.429	0.163	3621.911
Realized GARCH(1,1)	18095	0.410	1238.670	18898	7120	10974	0.274	0.606	0.410	4925.891
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1)- sstd	6145	0.139	8420.394	34245	3723	2422	0.098	0.394	0.139	3015.408
GJRGARCH(1,1)- sstd	6205	0.141	8287.519	34164	3744	2461	0.099	0.397	0.141	3059.154
Realized GARCH(1,1) - sstd	17946	0.407	1137.944	19075	7092	10853	0.271	0.605	0.407	4949.326
Real-time GARCH										
Real-time GARCH(1,1)	13887	0.315	46.481	23848	6378	7508	0.211	0.541	0.315	4643.643
Real-time GARCH(1,1) - Leverage	14255	0.323	9.423	23364	6494	7760	0.218	0.544	0.323	4588.288
Real-time GARCH(1,1) - Feedback	14583	0.331	0.066	22920	6610	7972	0.224	0.547	0.331	4493.144

Table 7: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.33$ and using the approximate price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.835}(1) = 1.928$.

$q = 0.10$, Exact Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	1139	0.026	3722.895	42131	843	296	0.020	0.260	0.026	975.281
GJRGARCH(1,1)	1202	0.027	3545.821	42026	885	317	0.021	0.264	0.027	1023.571
Realized GARCH(1,1)	9154	0.208	4470.147	30517	4442	4712	0.127	0.515	0.208	5747.390
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1)- sstd	620	0.014	5500.693	43011	482	138	0.011	0.223	0.014	563.230
GJRGARCH(1,1)- sstd	622	0.014	5492.486	43005	486	136	0.011	0.219	0.014	548.426
Realized GARCH(1,1) - sstd	7057	0.160	1520.294	33493	3563	3494	0.096	0.495	0.160	5544.596
Real-time GARCH										
Real-time GARCH(1,1)	5288	0.120	183.178	35639	3186	2101	0.082	0.397	0.120	3205.082
Real-time GARCH(1,1) - Leverage	4568	0.104	6.113	36713	2832	1735	0.072	0.380	0.104	2903.824
Real-time GARCH(1,1) - Feedback	4523	0.103	3.114	36794	2796	1726	0.071	0.382	0.103	2939.311

Table 8: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.10$ and using the exact price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.95}(1) = 3.84$

$q = 0.10$, Approximate Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	1514	0.034	2763.026	41499	1100	414	0.026	0.273	0.034	1193.754
GJRGARCH(1,1)	1591	0.036	2591.476	41387	1135	456	0.027	0.287	0.036	1324.937
Realized GARCH(1,1)	9971	0.226	5960.913	29396	4746	5225	0.139	0.524	0.226	5822.410
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1)- sstd	840	0.019	4668.483	42620	653	187	0.015	0.223	0.019	654.745
GJRGARCH(1,1)- sstd	876	0.020	4544.417	42558	679	197	0.016	0.225	0.020	678.789
Realized GARCH(1,1) - sstd	7808	0.177	2422.104	32366	3939	3869	0.108	0.496	0.177	5430.016
Real-time GARCH										
Real-time GARCH(1,1)	6052	0.137	614.792	34524	3537	2514	0.093	0.415	0.137	3514.613
Real-time GARCH(1,1) - Leverage	5308	0.120	191.406	35607	3198	2109	0.082	0.397	0.120	3203.365
Real-time GARCH(1,1) - Feedback	5263	0.119	173.133	35681	3169	2093	0.082	0.398	0.119	3215.185

Table 9: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.10$ and using the approximate price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.95}(1) = 3.84$

$q = 0.05$, Exact Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	532	0.012	1900.196	43173	408	124	0.009	0.233	0.012	556.649
GJRGARCH(1,1)	562	0.013	1814.198	43123	428	134	0.010	0.238	0.013	594.758
Realized GARCH(1,1)	6810	0.154	6671.341	33843	3460	3350	0.093	0.492	0.154	5475.127
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1)- sstd	252	0.006	2903.732	43656	205	47	0.005	0.187	0.006	254.268
GJRGARCH(1,1)- sstd	255	0.006	2890.478	43650	208	47	0.005	0.184	0.006	251.863
Realized GARCH(1,1) - sstd	4657	0.106	2204.261	36943	2513	2144	0.064	0.460	0.106	4616.247
Real-time GARCH										
Real-time GARCH(1,1)	3437	0.078	622.927	38460	2216	1220	0.055	0.355	0.078	2460.411
Real-time GARCH(1,1) - Leverage	2728	0.062	121.467	39558	1827	900	0.044	0.330	0.062	2030.093
Real-time GARCH(1,1) - Feedback	2664	0.060	94.271	39665	1784	879	0.043	0.330	0.060	2021.247

Table 10: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.05$ and using the approximate price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.975}(1) = 5.03$

$q = 0.05$, Approximate Transformation	Unconditional Coverage			Independence						
	T_1	$\hat{\pi}$	LR_{UC}	T_{11}	T_{10} / T_{01}	T_{00}	π_{01}	π_{11}	$\hat{\pi}_2$	LR_{ind}
Benchmark Models with normal distributed errors										
GARCH(1,1)	734	0.017	1379.260	42824	555	179	0.013	0.244	0.017	712.080
GJRGARCH(1,1)	782	0.018	1273.444	42740	591	191	0.014	0.244	0.018	737.296
Realized GARCH(1,1)	7551	0.171	8607.221	32739	3823	3728	0.105	0.494	0.171	5422.095
Benchmark Models with skewed student-t distributed errors										
GARCH(1,1)- sstd	334	0.008	2564.820	43505	274	60	0.006	0.180	0.008	286.315
GJRGARCH(1,1)- sstd	344	0.008	2526.491	43490	279	65	0.006	0.189	0.008	314.127
Realized GARCH(1,1) - sstd	5214	0.118	3175.939	36129	2770	2444	0.071	0.469	0.118	4870.938
Real-time GARCH										
Real-time GARCH(1,1)	4044	0.092	1308.102	37535	2534	1509	0.063	0.373	0.092	2788.232
Real-time GARCH(1,1) - Leverage	3277	0.074	479.642	38696	2140	1136	0.052	0.347	0.074	2322.341
Real-time GARCH(1,1) - Feedback	3177	0.072	398.614	38864	2072	1104	0.051	0.348	0.072	2325.734

Table 11: Coverage statistics for the out-of-sample volatility forecasts. These statistics are computed with $q = 0.05$ and using the approximate price transformation. These criteria are computed over time period 2014-01-01 to 2019-12-31 ($T = 44114$). The critical value for both LR_{UC} and LR_{ind} is $\chi^2_{0.975}(1) = 5.03$.