P10 - The Impact of Variable Renewable Energy Sources Forecast Errors on Electricity Price Spreads and Asymmetric Volatility Dynamics

Aalborg Universitet
Mathematics Economics

AALBORG UNIVERSITET
STUDENTERRAPPORT

Department of Mathematical Science
Mathematics-Economics,
Thomas Manns Vej 23,
9220 Aalborg Oest
http://math.aau.dk/

Title:
The Impact of Variable Renewable Energy Sources Forecast Errors on Electricity Price Spreads and Asymmetric Volatility Dynamics - P10

Project Period:

Spring Semester 2026

Project Group:

MO10-15

Authors:

Marcus Basse

Supervisor:

Eduardo & Mikkel

A Special Thanks: A special thanks to my supervisors Eduardo and Mikkel for guidance throughout the project.

The application of AI: Throughout this project AI has been used for grammatical- and text- enhancement, and debugging of the code.

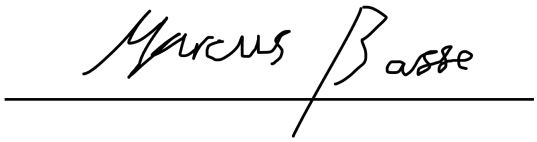
Page number: 66

Abstract:

This thesis investigates how day-ahead Variable Renewable Energy Sources (VRES) production forecast errors propagate into both the level and conditional volatility of the price spread. Using hourly data, initial unit root and Johansen trace tests confirm that the price and production series are stationary in levels ($I(0)$). Ordinary Least Squares regression and Wald hypothesis tests reject perfect market efficiency ($\alpha = 0, \beta = 1$), indicating the existence of systematic premiums. While VRES forecast errors exert a statistically significant linear impact on the price spread level, the static model exhibits a low explanatory power ($R^2 = 0.1445$) with residuals characterized by leptokurtosis and volatility clustering. To parameterize these non-linear risk dynamics, GARCH, EGARCH, and TGARCH frameworks are fitted under a Skewed Student's t-distribution. The empirical findings uncover performance trade-offs: the standard GARCH model exhibits structural instability ($\alpha_1 + \beta_1 > 1$), and the TGARCH model severely underestimates out-of-sample risk. Conversely, the EGARCH(1,1) architecture emerges as the superior of the three, maintaining covariance stationarity and capturing asymmetric leverage effects while demonstrating out-of-sample robustness with a 96.4% empirical coverage rate against a 95% target. Furthermore, seasonal analysis reveals a physical-financial paradox where winter exhibits the highest absolute forecast errors but the lowest price spread volatility due to thermal baseload buffering. Ultimately, while the principle of parsimony dictates the rejection of higher-order exogenous models due to out-of-sample overfitting, the baseline EGARCH(1,1) may offer actionable utility for risk management and ancillary service procurement.

Aalborg Universitet, May 26, 2026

Signatures



Marcus Basse

Table of contents

1 Introduction

The non-storability of electricity makes it the only commodity, where the production and consumption thereof have to be match at all times, this means that electricity generation and consumption must occur simultaneously. The introduction of Variable Renewable Energy Sources (VRES) into the electricity mix, introduces significant stochasticity into the market due to the inherent difficulty in forecasting wind and solar production. In the European electricity market, spot prices are determined through the day-ahead auction, which simultaneously sets prices for each quarter-hour of the following day by matching supply and demand curves. This is a fundamental point, as it highlights that spot prices are essentially based on expectations of both consumption and production, while consumption is relatively easy to predict, production is not, as weather forecast are difficult to make and they directly impact the production [19]. Therefore to ensure real-time balance, the balance-market is used to up- or down-regulate the production and consumption. Like the day-ahead market, an imbalance price is determined every quarter hour. The system is short, if total demand exceeds actual generation, and up-regulation is needed to cover the deficit, which is done by increasing production and/or reducing demand. Conversely, the system is long, if total production exceeds actual consumption, and down-regulation is needed to remove excess energy, which is done by decreasing production and/or increasing consumption.

Problem Statement:
As any deviation between the day-ahead forecast $\hat{Q}_t$ and actual production Q_t forces the system into the imbalance market. Consequently this project investigates how production forecast errors impact the spread between spot price S_t^{spot}, which is entirely based on expectations, and the imbalance price S_t^{imb}, which represents the cost of correcting these errors. This project seeks to quantify the relationship between these two domains. Specifically, we address two primary questions:

1. Long-term Equilibrium: Do the price series (S^{spot}, S^{imb}) and production series $(Q, \hat{Q})$ share a stable dynamic?

2. Error Propagation: How does the production forecast error $(v_t = \hat{Q}_t - Q_t)$ drive the dynamics of the price spread $(u_t = S_t^{spot} - S_t^{imb})$?

The goal is to estimate a dynamic relationship $u_t = f(v_t) + \varepsilon_t$ capturing how production shocks influence both the level and volatility of price spreads, using GARCH model specifications to account for time-varying variance in energy markets. The ultimate goal is to estimate a functional relationship $u_t = f(v_t) + \varepsilon_t$ that captures how production "shocks" impact both the level and the risk of price spreads, utilizing the GARCH framework to account for the time-varying variance inherent in energy markets.

2 Energy Market

This chapter is based on [20], [23], [15].

The chapter outlines the key components of liberalized power markets, including the roles of market operators, producers, and consumers, as well as the distinction between wholesale and balancing markets. Due to the non-storability of electricity at scale, power systems require continuous balancing of supply and demand. As a result, prices are determined in near real time based on expectations of production and consumption.

2.1 Overview of the Energy Market

In 1996 the European liberalization of the electricity market started. It was part of a bigger trend for liberalizing traditional monopolies, as it were believed that regulation of these markets, were the solution to the 1980s financial crisis. The liberation of the electricity market involved, the splitting of the monopolies which were in place, as they controlled the entire supply chain namely, the production, the transmission, and the distribution of electricity. This directive took place in Denmark in 1999. It meant the consumer freely could choose their electricity supplier.

Particular the transmission of electricity is a natural monopoly, as it does not make sense that each house has multiple electricity cables connected. Today this is in fact the case, the transmission is still monopolies. In Denmark the transmission is split into two. The first is the high voltage transmission cables owned and operated by Energinet, which is a state owned company and highly regulated. The second is the local monopolies which owns and operates the transmission cables which connects to the consumers, these companies are owned by the consumers. Both monopolies are highly regulated and forbidden from participating in the rest of the electricity sector, with laws restricting pricing and how profits are used [15]. The difference in how the market worked before liberalization and how it currently works and the different roles are depicted in Figure 2.1 [20].

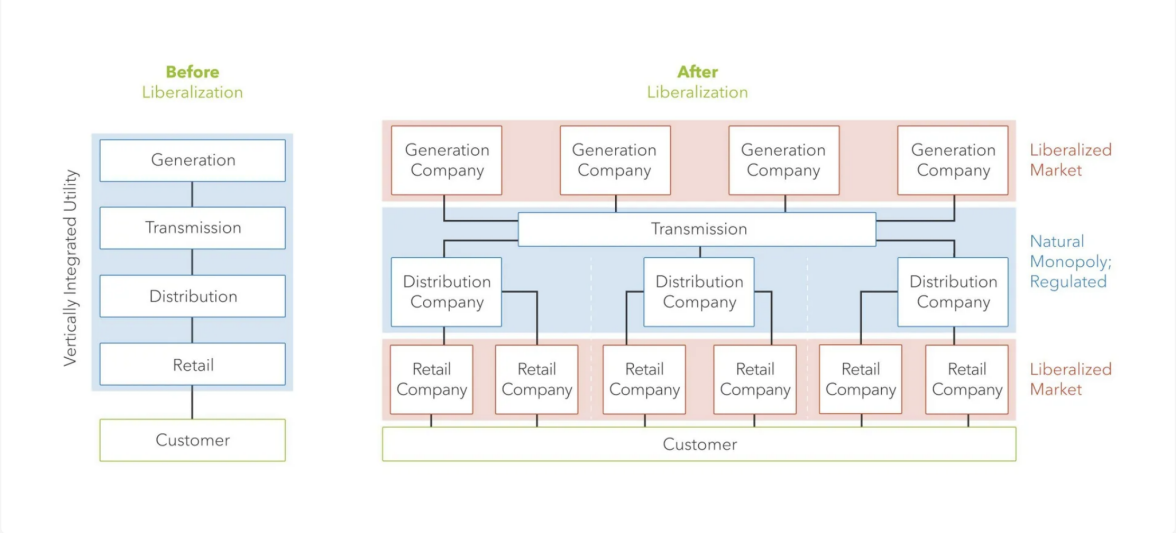


Figure 2.1: Liberalization of the European Energy Market [20]

2.2 Actors in Balancing Markets

The balancing market represents the final timeframe of the electricity market, where the physical equilibrium between supply and demand is maintained in real-time. While the wholesale markets (Day-Ahead and Intraday) allow participants to trade based on forecasts, the balancing market addresses the inevitable discrepancies that arise due to unforeseen events, such as generator outages or weather related fluctuations in renewable production. To manage these deviations and ensure grid stability at 50 Hz, three primary actors interact within a regulated framework: the Transmission System Operator (TSO), the Balance Responsible Parties (BRP), and the Balancing Service Providers (BSP).

2.2.1 TSO

This section is based on [24].
The Transmission System Operator (TSO) holds the ultimate responsibility for balancing the physical electricity system. This involves managing the real-time balance between production/imports and consumption/exports to ensure the grid maintains a constant frequency of 50 Hz. To achieve this, the TSO procures a sufficient amount of balancing capacity through the BSP market role. The TSO activates these services whenever necessary to stabilize the grid. Financially, the TSO pays the BSPs for these services and recovers the costs from the BRPs that deviate from their submitted schedules.

The TSO is a natural monopoly and is strictly prohibited from participating in neither the production nor the consumption of electricity. An exception is made for the TSO's own facilities and administrative buildings, which require a power supply to operate.

2.2.2 BRP

A Balancing Responsible Party is a company, which takes on the role of balancing the consumption and production at metering points, as well as trading physical electricity. A BRP actor participates in both day-ahead market, the intraday market, and the Balancing market. Therefore, the BRP participates in both the wholesale electricity markets and the system balancing mechanism used to manage real-time deviations. The BRP actors are financial responsible for any deviation in the schedules and the actual production/consumption, at each metering point. This incentives the actor to balance their portfolio towards a 'net-zero' position, and thereby avoiding the costs associated with imbalance settlement.

2.2.3 BSP

A Balancing Service Provider is a company, capable of delivering one or multiple of frequency based ancillary service systems such as FFR, FCR, and FCR-D, these are described in Section 2.3. These BSP are directly contracted by the TSO (Energinet), which activates the BSP to delivery a rapid frequency response whenever the system is in imbalance, thereby maintaining grid stability.
Under this specific model, the BSP is contracted solely to deliver a balancing services rather than bulk electricity delivery. Consequently, they do not require a contract with a BRP. The costs of such services are initially paid by the TSO and subsequently passed onto the BRPs, who have deviation in their portfolios. This financial reconciliation process is known as imbalance settlement.

2.3 Types of balancing Markets

This section is based on [5].
To maintain grid stability, the TSO has three primary tools available. These three address different types of balancing needs, categorized by their response time and activation method. The necessity of this tiered reserve structure arise from the technical difficulty of perfect forecasting. While the day-ahead market targets an equilibrium in 15 minute increments, real-time deviations occur in a spectrum of temporal scales. Where short-term stochastic noise requires autonomous sub-second stabilization, large structural imbalance, often driven by VRES forecast errors, require manual activation. This hierarchy ensures that the TSO can maintain a 50 Hz frequency while minimizing procurement costs associated with different response speeds.

1. **Frequency Containment Reserve (FCR):** Is the primary reserve that ensures grid-wide stability. This product requires a very fast response time (seconds); however, the total volume needed is relatively low. The product is activated automatically based on local measurements of the system frequency. The TSO (Energinet) procures the capacity, which is their only responsibility regarding FCR. The financial settlement associated with the activation is conducted through the imbalance settlement process, ensuring that the power delivered is accounted for in the provider's balance [4].

2. **Automatic Frequency Restoration Reserves (aFRR):** Acts as the secondary reserve to resolve larger imbalances and bring the frequency back to exactly 50.00 Hz. The TSO procures sufficient aFRR capacity the day prior to delivery. Providers are paid a capacity price for being available and, if activated, they receive the marginal price for the power delivered. The product is activated by a signal sent every four seconds by the TSO [2], [3].

3. **Manual Frequency Restoration Reserves (mFRR):** Ensures that large, persistent imbalances are stabilized, such as significant deviations in forecasted wind production or the outage of a power plant. Like aFRR, it consists of a capacity market and an activation market. The activation of mFRR is performed manually by the TSO, who sends a set-point signal to the supplier [6], [7].

These three reserves operate in a coordinated sequence. FCR provides an immediate automatic response to frequency deviations, whereas aFRR and mFRR are activated to restore system balance and release FCR capacity for subsequent contingencies.

2.4 Structure of the Balancing Market

This section is based on [26].
The balancing market consist of three distinct phases: balancing service provision, real-time system balancing, and balance settlement.

A schematic of the balancing market is presented in Figure 2.2, where the TSO is labeled as SO. Black arrows represent information flow, while white arrows indicate financial flows

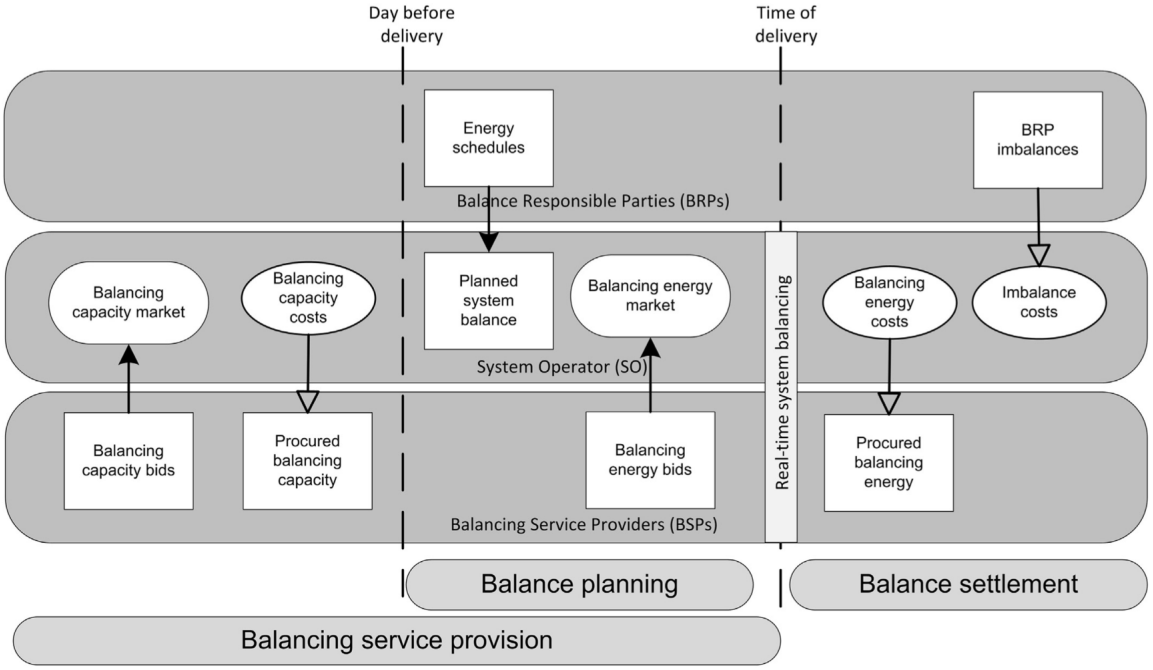


Figure 2.2: Schematic of the Balancing Market Structure [26]

In the first phase, balancing service procurement, the TSO procures balancing capacity from BSPs via a merit order of the lowest bids. Following this, in the day-ahead market, BRPs submit their schedules to the TSO, detailing their planned production and consumption. Simultaneously, BSPs submit their balancing power bids. In Denmark, the gate closure for these submissions is 12:00 CET. On the day of delivery, BRPs may trade on the intraday market to adjust their portfolios as new information becomes available.

The second phase is real-time system balancing. Here the TSO is responsible for ensuring that the physical electricity is balanced at all times, which is done by activating upward or downward regulation. This power is purchased from the BSPs based on the submitted bids.

The final phase is the balance settlement, where the financial cost of balancing the system is resolved. BSPs that delivered balancing services are compensated at the balancing power price. Conversely, BRPs whose actual metered volumes deviated from their schedules are settled against the imbalance price.

2.5 Price Formation

This section is based on [1], [12], [21], and [22].

In the short run, the price of electricity is only depended on the variable cost of production. As the production capacity neither can be increased nor decreased. The power plant should, therefore, be willing to sell electricity just over the variable cost of production. Thus the

capital cost of constructing power plants should not influence the short run price. Thus in the short run marginal cost of production of electricity must equal the variable cost of generation. In the day-ahead market the spot price is formed, which is entirely based on forecast production and load. The price of electricity is determined though supply and demand. The demand side is price inelastic. while production is elastic. The price is determined in quarter hour increments, thus for each increment there is a new supply and demand function. Deriving the supply curve for electricity is done by stacking the power plant by merit. That is a lineup of the power plants wanting to contribute to the electricity market is ordered based on their variable cost. The cheapest power plant supplies electricity first, with progressively more expensive power plants being activated as needed. The last activated power plant i.e. the power plant with the highest variable cost needed to meet demand sets the market price, which all the power plants gets. In short The merit order ensures that the demand is always met in the cheapest possible manner. The price formation can be seen in Figure 2.3.

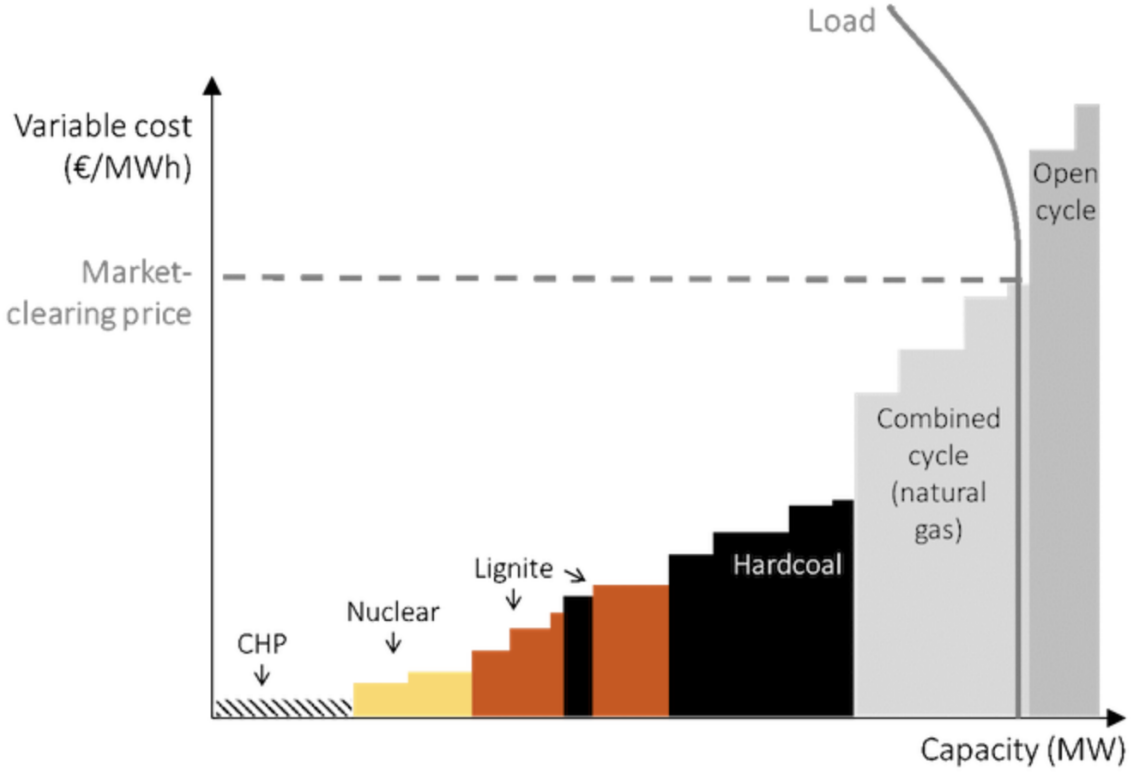


Figure 2.3: Marginal pricing, based on the merit order

The markets overall deviation from the scheduled is handled in the balancing market. The balancing market can be thought of as two markets, depending on whether the market is short or long. This aligns with the procurement process, for which there are two, one for upward regulation and one for downward regulation. This is further emphasized as different products can only act in one way, wind for example can bid in the downward regulation, but not upward, while aggregated batteries can do both. The supply curve for positive imbalance energy is shifted by the activation price of upward reserves, while the supply curve for negative

imbalance energy is shifted by the activation price of downward reserves.

The impact of VRES on the spread between the spot price and imbalance price is quite significant. First the marginal cost of VRES is low, and thus the first on the merit order, which pushes the price in the spot market down. However unlike fossil production plants which have very predictable production, only deviating if problems occur like the plant breaks down. The VRES production is directly linked to the weather, which is difficult to to predict, thus making VRES production prediction difficult.

The impact of renewable on the merit order is quite significant, as they often have extremely low marginal costs. Consequently pushing the price down. With the ever increasing renewable capacity, they more often set the price. The production of a coal of gas power plant is very predictable, as they are able to produce exactly what they say they are. Contrarily VRES are not, due to the difficulty of predicting the weather precisely, which the forecasted production is directly based of. Thus with both low marginal cost and high error between forecasted and actual production, VRES greatly impact the imbalance price.
VRES are seasonally in production capacity, with solar producing more in the summer months and wind producing more in the winter months. This seasonality is also reflected in the imbalance price as a percentage forecast error is different in absolute terms deepening on the season. This difference is important as the power market only cares about the absolute error, as that is the electricity which needs upward or downward regulation.

Variable Renewable Energy Sources (VRES) have a significant impact on both spot and imbalance prices. Due to their very low marginal costs, renewable producers are typically placed early in the merit order, which lowers spot market prices and increasingly allows renewable generation to become price-setting as installed capacity expands. Unlike conventional fossil-fuel power plants, whose production is highly predictable under normal operating conditions, VRES production depends directly on weather conditions. Since weather forecasts are inherently uncertain, renewable production forecasts are also uncertain, leading to deviations between scheduled and actual production. These forecast errors increase the need for upward or downward regulation in the balancing market and thereby contribute to volatility in imbalance prices. In addition, renewable generation exhibits strong seasonal patterns. Solar production is generally higher during summer months, whereas wind generation tends to be higher during winter months. Consequently, a given percentage forecast error may correspond to very different absolute production deviations depending on the season. This distinction is important, as balancing markets respond to the absolute energy imbalance that must be regulated in real time.

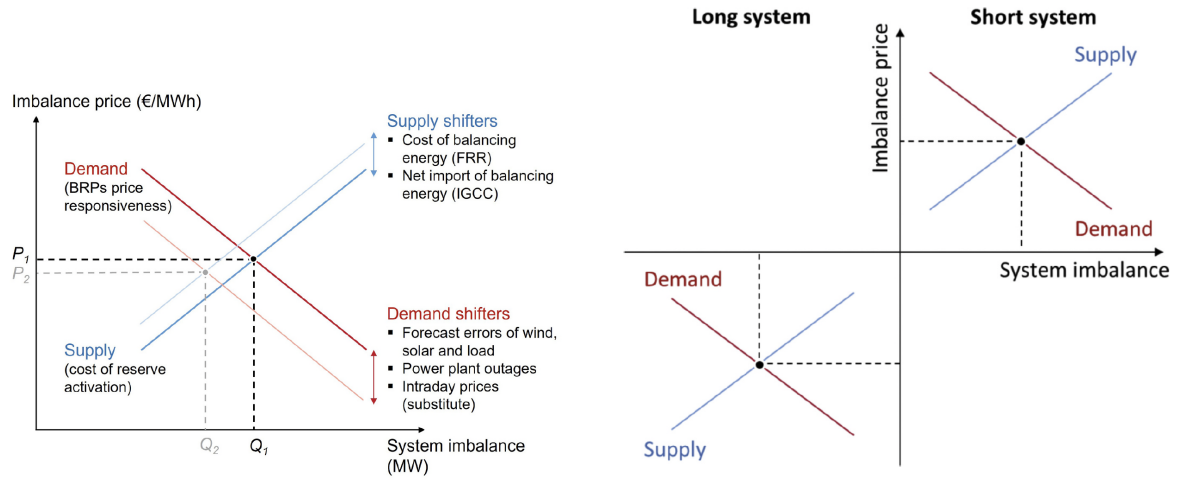
Building on the merit order framework, the pricing dynamics of the imbalance market can be interpreted as a real-time structural equilibrium. As illustrated in Figure 2.4a, the clearing imbalance price is determined by the intersection of the imbalance energy demand and supply curves at any given interval.

The demand curve reflects the real-time balancing requirements of the grid, which are fundamentally driven by exogenous physical shocks. These shifters primarily encompass real-time disturbances such as generation forecast errors from VRES and unexpected load fluctuations.

Conversely, the supply curve represents the marginal cost of reserve capacity available to the TSO, which is determined by domestic balancing power bids and real-time power imports from interconnected neighboring countries. When the grid faces a physical deficit of power, meaning the system is structurally short, the demand for balancing energy shifts outward and upward, resulting in a higher clearing price. Conversely, in a long system regime characterized by a supply surplus, the demand curve shifts downward and to the left, resulting in a collapsed clearing price.

Crucially, this optimization problem behaves differently depending on the sign of the aggregate network deviation. As stylized in Figure 2.4b, the balancing mechanism is economically distinct across long and short system states. In a long system, an unexpected supply surplus forces the equilibrium into the negative quadrant, causing the imbalance price to collapse beneath the spot price, generating a positive price spread. In a short system, a physical deficit shifts the equilibrium into the positive quadrant. Because activating emergency reserves is expensive, the imbalance price spikes vertically above the spot price, driving the price spread into negative territory.

This structural divergence underpins the non-linear volatility transmission observed in the data: the grid's financial response to a physical supply shock is entirely contingent upon whether the system must clear a surplus or manage a deficit along the real-time balancing curves.



(a) Imbalance Market Shifters **(b)** Long vs. Short Regimes

Figure 2.4: The imbalance system equilibrium determined by the intersection of supply and demand curves. Panel (a) depicts structural equilibrium shifts at exemplary times t_1 and t_2. Panel (b) illustrates the distinct market clearing dynamics across long (surplus) and short (deficit) system states.

3 Theory

3.1 Cointegration

This section is based on [13] and [18].

The four series spot prices, S_t^{spot}, imbalance prices, S_t^{imb}, forecasted production, $\hat{Q}_t$, and actual production ,Q_t, are all stationary. Even though all four series are $I(0)$, we still test for a co-integration relationship between them, to avoid the risk fo spurious regression and to formally investigate the long-run equilibrium relations. The co-integration relation occurs if a linear combination of non-stationary processes is stationary, $I(0)$. The primary purpose of co-integration is modelling the long-run economic relations. The important thing is not only that co-integration creates a stationary process, but that when variables drift apart, there is a force pulling them together. This force is called error correction. Thus co-integration allows us to separate the long-run equilibrium, the stationary combination, from the short-run dynamics, the deviation from equilibrium. Furthermore, co-integration proves that a relation between two non-stationary processes is not statistical fluke but a shared underlying trend. This is useful since regression one $I(1)$ on another can lead to high R^2 values, but meaningless results, this type of scenario is called spurious regression. First we define an integrated process as:

Definition 3.1.
A series with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $x_t \sim I(d)$.

With the integration process defined, we can formalize the concept of co-integration. Cointegration describes a relationship between multiple series that are individually non-stationary but share a common stochastic trend, preventing them from drifting apart indefinitely. It is defined as follows

Definition 3.2.
The components of the vector x_t are said to be co-integrated order d,b, denoted $x_t \sim CI(d,b)$, if (i) all components of x_t are $I(d)$; (ii) there exists a vector $\beta(\neq 0)$ so that $z_t = \beta^\top x_t \sim I(d-b)$, $b > 0$. The vector β is called the co-integrating vector.

The Granger representation theorem links the error correction model and cointegration. The p-dimensional vector autoregressive model, x_t,

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t. \tag{3.1}$$

We assume the ε_t are i.i.d with mean zero and variance Ω. Π, $\Gamma_1, \ldots, \Gamma_{k-1}$, Ω, and Φ are parameters, and d_t are deterministic terms, like trend and season. The VAR(p) model is the

error correction model if the characteristics polynomial of x_t have a unit root, and that the matrix Π can be written as $\Pi = \alpha\beta^\top$, where α is the $p \times r$ matrix containing the coefficient for speed of correction, and β is the matrix containing the r linearly independent row of Π, which is the co-integrating vectors. The (3.1) model then becomes:

$$\Delta x_t = \alpha\beta^\top x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t. \tag{3.2}$$

A sufficient condition for (3.2) to be co-integrated of order 1. can be formulated using the characteristic polynomial.

Condition 3.1. The $I(1)$ condition
We assume that $\det(\Pi(z)) = 0$ implies that $|z| > 1$ or $z = 1$ and assume that

$$det(\alpha_\perp^\top \Gamma \beta_\perp) \neq 0.$$

This condition is needed to avoid solutions which have seasonal roots or explosive roots, and solutions which are integrated of order 2 or higher. The condition is equivalent to the condition that the number of roots of $\det \Pi(z) = 0$ is $p - r$.

An equivalent condition, when considering the case where x_t have two components and where $d = b = 1$, a sufficient condition for the series being co-integrated is: that the coherence between the two series is one at zero frequency.

By establishing these I(1) conditions and identifying the cointegrating vectors, we can derive the stationary residuals u_t and v_t, which serve as the basis for analyzing the impact of production shocks on price spreads.

We define u_t and v_t as the deviation from the long-run equilibrium. By normalizing the cointegrating vector β with respect to the spot price and forecasted production respectively, we obtain the pricing error and production error

$$u_t = S_t^{spot} - \beta_1 S_t^{imb}$$
$$v_t = \hat{Q}_t - \beta_2 Q_t,$$

which represents the spread between the two markets.

3.2 GARCH

This section is based on [16] and [25].
We saw the residuals seems to experience heteroskedasticity. Therefor we will use the GARCH model. The GARCH model is an extension of the ARCH framework that allows the conditional variance to depend on its own past values. Assuming we model the return of y_t using linear regression

$$y_t = x_t^\top \xi + \varepsilon_t, \quad t = 1, \ldots, T$$

where x_t is a vector of endogenous variables, d_t represents a vector of deterministic components (such as a constant or a time trend), and ε_t is the structural error term. The GARCH framework can characterize the distribution of the stochastic error ε_t conditional on the past information set of the system $\Omega_{t-1} = \{x_{t-1}, x_{t-2}, \ldots\}$. The conditional variance of the error series, $\sigma_t^2 = \mathrm{Var}(\varepsilon_t|\Omega_{t-1})$, is expressed as a function of past squared errors, lagged conditional variances, and potential predetermined or exogenous variables:

$$\begin{aligned} \sigma_t^2 &= \sigma^2(\varepsilon_{t-1}, \sigma_{t-1}^2, \varepsilon_{t-2}, \sigma_{t-2}^2, \ldots, t, \xi, b) \\ \varepsilon_t &= \sigma_t Z_t \end{aligned}$$

where $Z_t \sim i.i.d$ with $\mathbb{E}[Z_t] = 0$, $\mathbb{E}[Z_t^2] = 1$. By definition, ε_t is serially uncorrelated with mean zero, but the conditional variance of ε_t equals σ_t^2, which may be changing through time. The GARCH(p,q) model is specified as:

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

where, ω, α_i, and β_j are non-negative constants, ensuring that the conditional variance remains positive at all times. The GARCH model effectively decomposes volatility into two distributed lags: lagged squared residuals (the ARCH term), which captures "news" or high-frequency shocks, and lagged conditional variance (the GARCH term), which captures the persistence of volatility over the long term.

The GARCH is a robust model, it is however symmetrical, as it assumes that shocks of either sign have the same influence. This is often not the case as these markets often are asymmetrical. We often observe that negative shock is followed by higher volatility compared to a positive shock. In the electricity market we would assume that asymmetry occurs, as it often is VRES that cause the descreptenty between forecasted production and actual production, which are easier to stop from producing, than to find more electricity.

3.3 EGARCH

This section is based on [16] and [25].
The Exponential GARCH (EGARCH) model accounts for the asymmetric relationship between shocks and conditional volatility. The variance equation is specified as:

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i Z_{t-i} + \gamma_i(|Z_{t-i}| - \mathbb{E}[|Z_{t-i}|]) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2,$$

where α_i represents the sign effect, capturing the asymmetric response to positive versus negative innovations. For the purpose of this study, α_i is expected to be positive, implying that a prediction resulting in a shortfall of electricity, and thus a positive price shock, will impact the price spread volatility more significantly than a prediction resulting in a surplus of electricity. The magnitude effect, γ_i, determines the impact of the shock's absolute size, while

β_j represents the persistence of volatility (the GARCH effect). Z_{t-i} denotes the standardized residuals, defined as $Z_{t-i} = \varepsilon_{t-i}/\sigma_{t-i}$. The term $(|Z_{t-i}| - \mathbb{E}[|Z_{t-i}|]$ utilizes $\mathbb{E}[|Z_{t-i}|]$ as a benchmark; innovations exceeding this expected value result in increased conditional volatility. For this study, we utilize the Skew-Student-t distribution, and the normal distribution serving as a benchmark. Following the re-parameterized framework of Fernandez and Steel [14], the expectation $\mathbb{E}[|Z_{t-i}|]$ is calculated as a function of the skewness and tail parameters, ensuring the innovations are properly centered. Finally, the condition for log-variance to be stationary is that the sum of beta coefficients are smaller than one, $\sum_{j=1}^{q} \beta_j < 1$.

3.4 TGARCH

Another model capable of capturing asymmetrical volatility is the Threshold GARCH (TGARCH) model. The conditional variance is specified as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} (\alpha_i + \gamma_i N_{t-i}) \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where N_{t-i} is an indicator function for negative innovations, defined as:

$$N_{t-i} = \begin{cases} 1 & \text{if } \varepsilon_{t-i} < 0, \\ 0 & \text{if } \varepsilon_{t-i} \geq 0 \end{cases}$$

To ensure that the conditional variance σ_t^2 remains strictly positive, the following constraints are imposed: $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\alpha_i + \gamma_i \geq 0$. The parameter γ_i captures the asymmetry in the volatility response. In a financial context, $\gamma_i > 0$ implies that negative shocks exert a greater impact on future volatility than positive shocks of the same magnitude. For the electricity market, this effect is expected to be reversed ($\gamma < 0$), as the system is more sensitive to positive price shocks. This 'inverse leverage effect' stems from the fact that electricity is non-storable and characterized by an extremely inelastic demand. When supply becomes constrained, due to unexpected low VRES production, or unplanned plant outages, prices spike upward rapidly, triggering significantly higher future volatility than a downward price movement of the same magnitude.

For the process to be covariance stationary, the sum of the persistence parameters must be less than unity:

$$\alpha_1 + \beta_1 + \gamma_1 \kappa < 1$$

where $\kappa = \frac{1}{1+\xi^2}$ represents the probability of a negative innovation, determined by the skewness parameter ξ.

3.5 Parameter Estimation

The parameters of the GARCH models are estimated using Maximum Likelihood Estimation (MLE). Let the parameter vector be denoted by

$$\theta = (\xi, \omega, \alpha, \beta, \gamma, \dots),$$

where the included parameters depend on the specific volatility specification.

Conditional on the information set $\mathcal{F}_{t-1}$, the return innovation is assumed to follow

$$\varepsilon_t = \sigma_t Z_t,$$

where Z_t follows a specified standardized distribution. In this study, the innovations are assumed to follow a Skewed Student-t distribution, allowing for both heavy tails and asymmetry in the conditional distribution.
Given the conditional density $f(\varepsilon_t \mid \mathcal{F}_{t-1}; \theta)$, the log-likelihood function for a sample of size T is

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} \ln f(\varepsilon_t \mid \mathcal{F}_{t-1}; \theta).$$

The parameter estimates are obtained by maximizing the log-likelihood function:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta).$$

Because the conditional variance σ_t^2 depends recursively on past residuals and past variances, the likelihood function has no closed-form solution and must be optimized numerically. The estimation is therefore performed using iterative numerical optimization algorithms implemented in the `rugarch` package in `R` [16].

4 Application

The Data is gathered from Energinet [8], [9], [10], and [11].

To determine the order of integration for the four time series, spot price (S_t^{spot}), imbalance price (S_t^{imb}), actual production (Q_t), and forecast production ($\hat{Q}_t$), we conduct Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS), and Philips Perron (PP) tests. The ADF test is specified with both intercept and trend, the lag length was is determined using BIC, with an upper bound of 168 lags, meaning 168 hours or a full week. For consistency both the KPSS test and the PP test are specified with trend. The results, summarized in Table 4.1, show a disagreement between the three tests for the series in levels. While both the ADF and PP test found stationarity, the KPSS test rejects the null hypothesis of stationarity for all variables.
Given that the data is aggregated on an hourly basis, we further investigate whether these findings are a byproduct of high-frequency data characteristics. To test this, all three tests was reapplied to the series aggregated to a daily frequency. The results remained consistent, with the ADF test and pp test rejecting the null hypothesis for all four series at the daily level, and KPSS rejecting the null hypothesis of stationarity.

Given this contradiction, we inspect the Autocorrelation Function (ACF) plots (see Appendix A). The ACF plots for the series in levels exhibit characteristics of non-stationary processes, though their behaviors differ. For the spot price and both production series, the ACF shows a very slow linear decay with no sharp initial drop, alongside clear, seasonal cycles, a strong indication of a unit root. In contrast, the imbalance price exhibits a more pronounced drop after the initial lags; however, the spikes remain above the significance threshold and maintain a seasonal pattern thereafter.

Upon applying first-differencing, the ADF, KPSS, and PP tests reach a consensus, all finding stationarity for the differenced series. Although the ADF and PP tests suggest stationarity in levels, the KPSS test rejects the null hypothesis of stationarity for all variables. This discrepancy, combined with the slow linear decay observed in the ACF plots, suggests that the series might possess unit-root characteristics or long-memory dynamics. To avoid the risk of spurious regression and to formally investigate the long-run equilibrium relations between these variables, we proceed by testing for a co-integration relation.

While the ADF and PP tests indicate stationarity in levels, the KPSS test provides conflicting evidence by rejecting the null hypothesis of stationarity. On balance, the evidence suggests that the variables are stationary. Nevertheless, to ensure the robustness of our findings and account for the possibility of persistent dynamics, we conduct co-integration tests to assess whether a long-run equilibrium relationship exists between the variables and to ensure that our subsequent analysis is not affected by potential spurious regression arising from overlooked non-stationarity.

Variable Name	ADF (trend) p-val	ADF (trend) Stat.	ADF (trend) Lags BIC	KPSS (trend, 44 lags) p-val	KPSS (trend, 44 lags) Stat.	PP (trend) p-val	PP (trend) Stat.
Level							
(S_t^{imb}) Imbalance Price	0.01*	−19.17	23	0.01*	3.42	0.01*	−15702.15
(S_t^{spot}) Spot Price	0.01*	−7.72	168	0.01*	5.57	0.01*	−1367.94
(Q_t) Actual Production	0.01*	−11.47	48	0.01*	1.34	0.01*	−772.38
($\hat{Q}_t$) Forecast Prod.	0.01*	−14.98	97	0.04	0.97	0.01*	−873.84
Differenced							
Δ (S_t^{imb}) Imbalance	0.01*	−33.94	47	0.10**	0.0004	0.01*	−18628.67
Δ (S_t^{spot}) Spot	0.01*	−17.24	167	0.10**	0.0007	0.01*	−5937.07
Δ (Q_t) Actual Prod.	0.01*	−21.44	73	0.10**	0.0007	0.01*	−5647.14
Δ ($\hat{Q}_t$) Forecast	0.01*	−18.57	96	0.10**	0.0008	0.01*	−3904.23

* Note: Reported p-value is the lower bound ($p < 0.01$).
** Note: Reported p-value is the upper bound ($p > 0.10$).

Table 4.1: Unit Root Test Results: ADF, KPSS, and PP.

4.1 Co-integration

Since two out of three test found the series to be stationary in levels, we do a robustness check to ensure that we do not find spurious regression. Therefore, we propose the hypothesis that there exists a co-integration relation between S_t^{spot} and S_t^{imb}, as well as between $\hat{Q}_t$ and Q_t. Meaning we test for the number of r linear independent stationary relations in:

$$u_t = S_t^{spot} - \beta_1 S_t^{imb}$$
$$v_t = \hat{Q}_t - \beta_2 Q_t.$$

We employ the Johansen trace test to identify the cointegrating rank for both the price series and production series systems. The hypotheses which are tested are $\mathcal{H}_0 : r = 0$ and $\mathcal{H}_0 : r \leq 1$, where r is the number of co-integrating relations.
For the price pair S_t^{spot} , S_t^{imb}, the trace statistics (876.98 and 239.96) exceeds the 1% critical values, leading us to reject the null hypotheses of $r = 0$ and $r \leq 1$. Thus the Johansen trace test finds a full-rank system. This result suggests that no co-integrating relation exists, because the underlying series are likely stationary in levels.

Similarly for the production pair $\hat{Q}_t$, Q_t, the trace statistics (333.29 and 128.87) exceeds the 1% critical values, leading us to reject the null hypotheses of $r = 0$ and $r \leq 1$. Similarly to price series, the Johansen trace test finds a full-rank system. Again this result suggests that no co-integrating relation exists, because the underlying series are likely stationary in levels.

4.2 Analysis of Systematic Bias and Market Integration

Since the Johannes trace test indicated full rank for the two pairwise series, we will investigate whether a linear relation exist between S_t^{spot} , S_t^{imb} and between $\hat{Q}_t$, Q_t. We therefore make

the following two models:

$$S_t^{spot} = \alpha_1 + \beta_1 S_t^{imb}$$
$$\hat{Q}_t = \alpha_2 + \beta_2 Q_t$$

To account for the significant autocorrelation identified by the Durbin-Watson tests in the residuals (DW : Price series $= 0.35; DW$: Production series $= 0.17$), we employed Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors. This is done to ensure that our inference remains valid, and the p-values are not artificially inflated by temporal dependence in the residuals.
We perform a joint Wald test for the null-hypothesis $\mathcal{H}_0 : \alpha = 0$, $\beta = 1$, together with two individual test: $\mathcal{H}_0 : \alpha = 0$ and $\mathcal{H}_0 : \beta = 1$.
The Results are given below in Table 4.2, and the estimated coefficients can be seen in Table 4.3.

Hypothesis	P-value (HAC Adjusted)	
	Price Model $S_t^{spot}=\alpha_1+\beta_1 S_t^{imb}$	Production Model $\hat{Q}_t=\alpha_2+\beta_2 Q_t$
$\mathcal{H}_0 : \alpha = 0$ and $\beta = 1$	$p < 0.0001$	$p < 0.0001$
$\mathcal{H}_0 : \alpha = 0$	$p < 0.0001$	$p < 0.0001$
$\mathcal{H}_0 : \beta = 1$	$p < 0.0001$	$p < 0.0001$

Table 4.2: Hypotheses for the Linear Regression Models.

(y)	(x)	Intercept (α)	Slope (β)	R^2
S_t^{spot}	S_t^{imb}	$65.5989^{\cdots}$	$0.1525^{\cdots}$	0.1445
		(0.4036)	(0.0027)	
$\hat{Q}_t$	Q_t	$329.5722^{\cdots}$	$0.6987^{\cdots}$	0.8635
		(4.9848)	(0.0020)	

Note: Standard errors are reported in parentheses. $^{\cdots}$ denotes significance at the 0.1% level ($p < 0.001$). Total observations $N = 18{,}546$.

Table 4.3: OLS Regression Estimates for Prices and Production

The rejection of all three null hypotheses for both the price and production pairs.
For the price pair, the rejection of $\beta = 1$ indicates that the spot and imbalance prices are not perfectly related but exhibit a non-proportional relationship. Similarly, for the production pair, the rejection of $\beta = 1$ implies that the forecasting errors are informationally inefficient rather than completely independent.
Collectively, these findings provide the empirical justification for modeling the price spread ($S_t^{spot} - S_t^{imb}$) and the production forecast error ($\hat{Q}_t - Q_t$) directly. Which for the remaining part of the paper will be denoted u_t and v_t respectively.

4.3 Linear Regression model

The first relation between price spread, u_t, and production forecast error, v_t, we investigate is the simple linear regression

$$u_t = \alpha + \beta v_t + \varepsilon_t.$$

We test the null-hypothesis, $\mathcal{H}_0 : \beta = 0$, which means that the production error have no linear effect on the price. Below in Figure 4.1 is a scatterplot over price spread against production error, with the red line being the OLS regression. Furthermore the model i specified as follows:

$$u_t = -5.586 - 0.0125 v_t + \varepsilon_t$$

where both the coefficient associated with v_t and the intercept are statistically significant at the 1% level. The estimated values can be seen in Table 4.4.

Variable	Estimate	Std. Error	t-value	p-value
Intercept	-5.586	0.968	-5.768	8.13×10^{-9}
v_t	-0.0125	0.00159	-7.857	4.15×10^{-15}

Table 4.4: Estimated coefficients for the linear regression model.

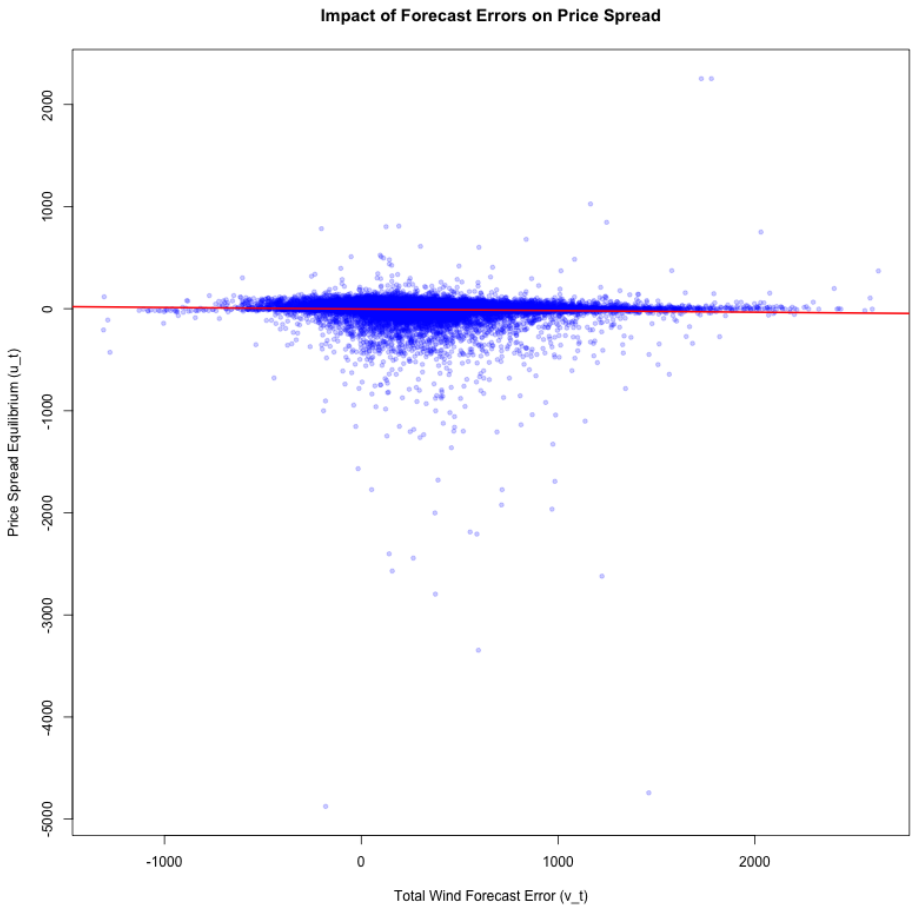
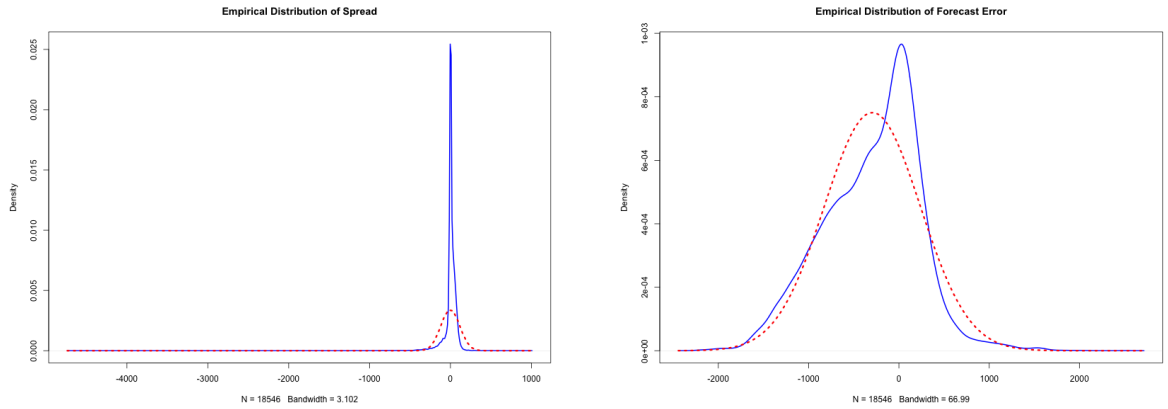


Figure 4.1: Scatter plot of the price spread (u_t) against production error (v_t), with the OLS regression line.

The null-hypothesis is rejected, with a p-value of $p < 0.0001$. This indicates a statistical relation between the production error and price spread. The β value is -0.0156 which suggest that a 1 MW over-forecast in production leads to a 0.0156 Euro narrowing of the spread between spot and imbalance prices. However, the R^2 value is 0.0033, which indicates that this simple model is insufficient at capturing the full market dynamics.
The median residual is 7.3 while the maximum and minimum are 1008.5 and -4734.4 respectively. The maximum residual means the spot price was 1008.5 Euro higher than the imbalance price, and vice versa for the minimum which means the spot was 4734.4 Euro lower than the imbalance price, which suggest that the market experience extreme shocks, i.e. the market have fat-tails. Figure 4.1 indicates a clear asymmetry in the distribution of shocks, where extreme negative deviations in the price spread are significantly more frequent and severe than positive ones. This 'downward leakage' confirms the presence of fat tails in the market's price dynamics. Furthermore, the plot suggests heteroskedasticity; the residuals are tightly concentrated around the regression line when forecast errors are small but exhibit much higher variance as the magnitude of v_t increases. This violation of the constant variance assumption, combined with the low explanatory power of the static model, necessitates a transition to a more robust, dynamic framework capable of modeling conditional heteroskedasticity, specifically the GARCH, EGARCH, and TGARCH specifications.

Before constructing the GARCH model, we examine the distributional properties of the data. Preliminary inspection of the scatterplot in Figure 2.2 indicates a distribution characterized by significant kurtosis and skewness. To formalize this, we calculate the moments, perform the Jarque-Bera test, and visualize the empirical distribution against a theoretical normal curve. Using the moments library in R, we find that the price spread u_t exhibits an extreme kurtosis of 295.4 and a negative skewness of 11.78. In comparison, the production forecast error v_t shows a kurtosis of 3.3 and a slight negative skewness of 0.22. The Jarque-Bera test for both series yields a p-value of $p < 0.0001$, resulting in a rejection of the null hypothesis of normality. A visual inspection of the price spread density in Figure 4.2a reveals a leptokurtic profile, featuring a sharp peak near zero and heavy tails. While the forecast error v_t appears more well-behaved, it fundamentally departs from a standard Gaussian curve. Crucially, because these empirical distributions aggregate the entire sample statically, they reflect the unconditional properties of the series. In energy time series, such excess kurtosis is typically a structural symptom of uncaptured temporal dependencies and volatility clustering, where shocks persist across consecutive periods. Therefore, failing to account for lagged dynamics would result in a model whose innovations are not independently and identically distributed (i.i.d.), violating a key requirement of a correctly specified conditional meanâĂŞvariance framework. While these preliminary diagnostics strongly suggest that a Skewed Student's t-distribution (sstd) will ultimately be required to accommodate the remaining tail fatness, this severe non-normality primarily necessitates the introduction of conditional lag structures, such as an AR(p)-GARCH(p,q) framework, to filter out the time dependent memory of the variance process before the final innovation distribution can be formally evaluated.



(a) Price Spread (u_t) **(b)** VRES Forecast Error (v_t)

Figure 4.2: Empirical Densities vs. Normal Distributions. Plot (a) displays the leptokurtosis of the price spread (u_t). Plot (b) shows the relatively well-behaved, yet non-normal, distribution of the VRES forecast error (v_t). Both plots use solid blue lines for empirical density and dashed red lines for a normal curve with matching moments.

4.4 GARCH model

To examine the dynamic interaction between production uncertainty and price spreads, we employ an AR(1)-GARCH(1,1)-X model. This framework allows the production forecast error, v_t, to enter as an exogenous regressor in both the conditional mean and conditional variance equations, thereby explicitly parameterizing the temporal lag structures and volatility clustering motivated in Section 4.3. To model the innovation process, we utilize a sstd. While the kurtosis in the unconditional data was partially driven by uncaptured ARCH effects, an inspection of the standard residuals' sstd-QQ plot (Figure 4.3) confirms that the filtered innovations still retain heavy tails and significant skewness. Even after accounting for conditional variance dynamics through our lag structures, negative spikes reaching -4000 still deviate from the theoretical distribution. To justify this distribution empirically, the baseline specification is also estimated utilizing a normal distribution, yielding an AIC score of 11.724. Based on information criteria, the Gaussian framework is inferior to the model constructed under the sstd, which achieves a lower AIC score of 9.9697. Consequently, because the dynamic lag structure alone cannot entirely eliminate the structural tail fatness of the electricity market, the sstd is maintained as the innovations distribution for both the subsequent EGARCH and TGARCH specifications to ensure a robust cross-model comparison. The comprehensive diagnostic plots for all three models under the normal distribution baseline are compiled in Appendix C. The specification of the AR(1)-GARCH(1,1)-X model is estimated as follows:

$$u_t = -0.000245 + 0.466699u_{t-1} + \varepsilon_t$$
$$\sigma_t^2 = 0.000001 + 0.999440\varepsilon_{t-1}^2 + 0.430402\sigma_{t-1}^2$$

Parameter	Estimate	Std. Error	t-value	Pr(>|t|)
Mean Model				
μ	-0.000245	0.001735	-0.141116	0.887778
$ar1$	0.466699	0.189620	2.461238	0.013846*
$mxreg1$	0.000000	0.000282	-0.001731	0.998619
Variance Model				
ω	0.000001	0.000004	0.162136	0.871199
α_1	0.999440	0.277432	3.602467	0.000315***
β_1	0.430402	0.190676	2.257241	0.023993*
$vxreg1$	0.000000	0.000554	0.000013	0.999989

Table 4.5: Full Parameter Estimates for the AR(1)-GARCH(1,1) Model

In the conditional mean equation, the AR(1) term accounts for the inherent persistence in the price spread. This specification is statistically significant with a p-value of $p < 0.05$ under robust standard errors, yielding an estimated coefficient of 0.467. This suggests that approximately 47% of the previous hour' price spread carries over to the current hour. The adequacy of this structure is confirmed by the Weighted Ljung-Box test on standardized residuals with a p-value of $p = 0.999$, which fails to reject the null hypothesis of no serial correlation, see Table 4.6. This is further supported by the ACF plot of the standardized residuals 4.3, where all lags are within the significance bounds.

Diagnostic Tests	Statistic	p-value
Log-Likelihood	-92440.01	
AIC	9.9697	
BIC	9.9735	
Ljung-Box (Resid.) $Q(1)$	7.415e-07	0.9993
Ljung-Box (Sq. Resid.) $Q(1)$	0.0002050	0.9886
Sign Bias	1.12802	0.2593
Negative Sign Bias	0.02193	0.9825
Positive Sign Bias	0.87836	0.3798
Sign Bias (Joint Effect)	1.70323	0.6362

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors applied.

Table 4.6: Diagnostic Results for the sGARCH(1,1) Model

While the raw data scatter plot indicates a visual relationship, the exogenous regressor in the mean equation mxreg1 is not statistically significant in this specification ($p = 0.998$). This suggests that once the autoregressive memory of the market is accounted for, the marginal linear impact of a single hour's forecast error on the spread level is negligible. Similarly the exogenous regressor in the variance equation vxreg1 also remains insignificant ($p = 0.999$), indicating that v_t does not provide additional explanatory power for the price spread's uncertainty beyond the internal GARCH dynamics.

More critically, the GARCH parameters highlight the extreme nature of electricity market volatility. The ARCH (α_1) and GARCH (β_1) coefficients are estimated at 0.999 and 0.430, respectively. Their sum ($\alpha_1 + \beta_1 = 1.429$) violates the standard stationarity condition ($\alpha + \beta < 1$), suggesting an explosive variance process. The shocks trigger persistent regimes of extreme volatility rather than decaying toward a long-run mean. The ACF of squared

standardized residuals 4.3, shows that all lags are wihtin the significane levels. This means that there are no ARCH effect left in the residuals, thus the model has successfully captured the volatility clustering.

Finally, the news impact curve 4.3 displays a symmetric U-shaped curve, reflecting the assumption of the GARCH framework, of symmetric impact of both negative and positive shocks. This contradicts the physical reality of power systems. Due to the technical constraints of the grid, it is mechanically easier to down regulate VRES production. Conversely, up regulating production is often an expensive endavour. This physical asymmetric flexibility suggests that the volatility response to forecast errors should be lopsided, thus a model which can handle this dynamic is employed, namely, the EGARCH model.

Consequently, while the GARCH seems to be whitening the residuals. The stationarity assumption is violated, suggesting an explosive variance process. Together with the economic reasoning of the of the power market, we implement an EGARCH.

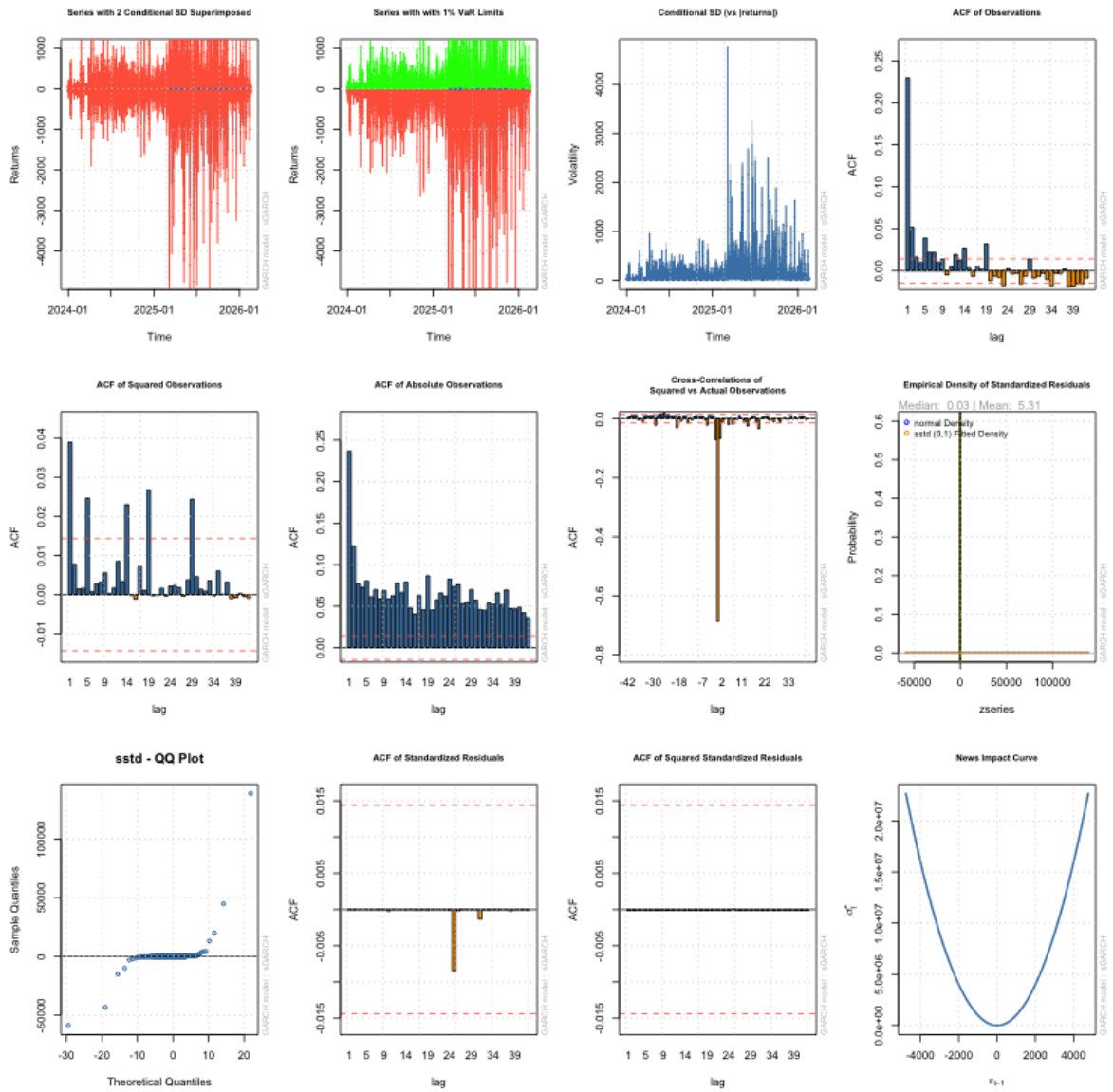


Figure 4.3: GARCH Model: All the plots

4.5 EGARCH

To address the structural limitations and non-stationarity of the standard GARCH framework, we transition to an AR(1)-EGARCH(1,1)-X specification. The model is estimated as follows:

$$u_t = -1.021158 + 0.344651 u_{t-1} - 0.001489 v_t + \varepsilon_t$$

$$\ln \sigma_t^2 = 1.338967 + 1.788926 Z_{t-1} + 4.073074(|Z_{t-1}| - \mathbb{E}[|Z_{t-1}|]) + 0.878343 \ln \sigma_{t-1}^2 + 0.000010 v_t$$

In the mean equation the AR(1) term is statistically significant, with an estimated coefficient of 0.345, confirming a strong autoregressive process within the price spread. However, the Weighted Ljung-box test on standardized residuals, seen in Table 4.8, yields a p-value of 0.0454. At a 5% significance level, this indicates the presence of lingering serial correlation in the mean process. While this typically implies that a higher order autoregressive structure is warranted, exploratory testing from lag 1 through 24 revealed that, revealed non of those 24 lags could produce a p-value above the 5% significance level, subsequently the AIC score of each model was investigated. The AR(1) specification produced an AIC value of 10.09368, while the best AIC score was achieved by an AR(6) specification yielding a score of 10.08872, with the difference being so small (-0.005), the choice landed on an AR(1), to avoid both over-parametrization and computational complexity, which becomes particular relevant for the forecasting section. Furthermore, the ACF plot of standardized residuals, show only a single lag which is significant. The residual autocorrelation is likely an effect of the extreme volatility, which makes the linear AR terms struggle to absorb the structural shocks. Furthermore, the exogenous mean regressor, $mxreg1$, yields a p-value of 0.123 with an estimated coefficient of 0.001489. Because this parameter fails to achieve statistical significance at standard confidence level of ($p > 0.05$), the model indicates that physical production errors do not exert a reliable, linear directional impact on the price spread baseline level. Similar to the standard GARCH specification, the exogenous regressor in the conditional variance process (vxreg1) is not statistically significant (p=0.760). Furthermore, the intercept ω is statistically relevant (p<0.001) with an estimate of 1.339, indicating a solid baseline log-volatility level. The ARCH size effect parameter α_1 and the asymmetry leverage parameter γ_1 are both significant (p<0.001), yielding coefficients of 1.789 and 4.073, respectively. Crucially, the persistence parameter β_1 is estimated at 0.878 and is significant (p<0.001). In the EGARCH framework, covariance stationarity is solely determined by the persistence parameter, requiring that $|\beta| < 1$. Because the estimated β sits strictly within this unit circle, the model successfully satisfies the stationarity requirement. This confirms that while real-time volatility shocks to the price spread are highly persistent, they correctly decay over a long-term horizon rather than exploding into non-stationarity as observed in the standard GARCH model.

Parameter	**Estimate**	**Std. Error**	**t-value**	**Pr(>|t|)**
Mean Model				
μ (Intercept)	-1.021158	0.766898	-1.331540	0.183010
$ar1$	0.344651	0.080573	4.277520	0.000019***
$mxreg1$	-0.001489	0.000966	-1.541720	0.123143
Variance Model				
ω	1.338967	0.326382	4.102460	0.000041***
α_1	1.788926	0.305838	5.849260	0.000000***
β_1	0.878343	0.029254	30.024590	0.000000***
γ_1	4.073074	0.408542	9.969790	0.000000***
$vxreg1$	0.000010	0.000034	0.305210	0.760205

Table 4.7: Full Parameter Estimates for the AR(1)-eGARCH(1,1) Model

The Weighted Ljung-box test on squared residuals, with a p-value of 0.65, resulting in not rejecting the null hypothesis, together with the ACF plot of squared standardized residuals seen in Figure 4.4, confirms no remaining ARCH effects. This demonstrates that the EGARCH(1,1) successfully "whitens" the variance process, capturing the market's volatility clustering.

The distributional and shock-response dynamics are visually captured via the diagnostic panel in Figure 4.4. The standardized residual QQ-plot (bottom row, left) reveals a pronounced heavy lower tail, where extreme negative pricing anomalies deviate substantially from the theoretical skewed Student-t distribution. Conversely, the upper tail conforms tightly to the theoretical baseline, with a few exception, highlighting that extreme market risk is fundamentally lopsided. This structural asymmetry is mirrored by the News Impact Curve (bottom row, right), which displays a severe right-sided exponential skew. This profile demonstrates that positive shocks to the spread $\varepsilon_{t-1} > 0$, corresponding to widening spot-to-imbalance differentials, trigger a vastly more aggressive, non-linear volatility cascade than equivalent negative shocks. Crucially, while these diagnostics showcase massive underlying asymmetry, the formal Sign Bias tests presented in Table 4.8 remain statistically insignificant (p>0.05). Supporting the structural adequacy of the EGARCH specification; the model's asymmetric parameters (α_1 and γ_1) have successfully absorbed and internalized the majority of directional leverage effects, leaving the standardized residuals free of untreated sign bias.

Diagnostic Tests	**Statistic**	**p-value**
Log-Likelihood	-93588.72	
AIC	10.0940	
BIC	10.0980	
Ljung-Box (Resid.) $Q(1)$	4.0020	0.0454*
Ljung-Box (Sq. Resid.) $Q(1)$	0.2083	0.6481
Sign Bias	0.3349	0.7377
Negative Sign Bias	0.5320	0.5947
Positive Sign Bias	1.4723	0.1410
Sign Bias (Joint Effect)	2.9393	0.4011

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors applied.

Table 4.8: Diagnostic Results for the eGARCH(1,1) Model

To objectively determine the overall statistical superiority between the evaluated frameworks, we examine the information criteria. The standard GARCH specification yields an AIC of 9.9697 and BIC of 9.9735 (Table 4.6). In contrast, the EGARCH model registers a slightly higher AIC of 10.0940 and a BIC of 10.0980 (Table 4.8). While classical model selection strictly favors the lower values of the GARCH framework, this statistical artifact must be interpreted with caution in high-frequency electricity markets. The lower penalty score of the GARCH model is a direct mathematical consequence of its parsimonious parameter structure. However, as demonstrated by its explosive variance process, this parsimony comes at the expense of structural validity. The minor information criteria penalty incurred by the EGARCH specification is a minor trade-off for capturing the significant asymmetric parameters. By accommodating the exponential leverage effects of real-time balancing shocks, the EGARCH framework trades marginal information criteria efficiency for strict covariance stationarity, cementing its operational superiority for risk management.

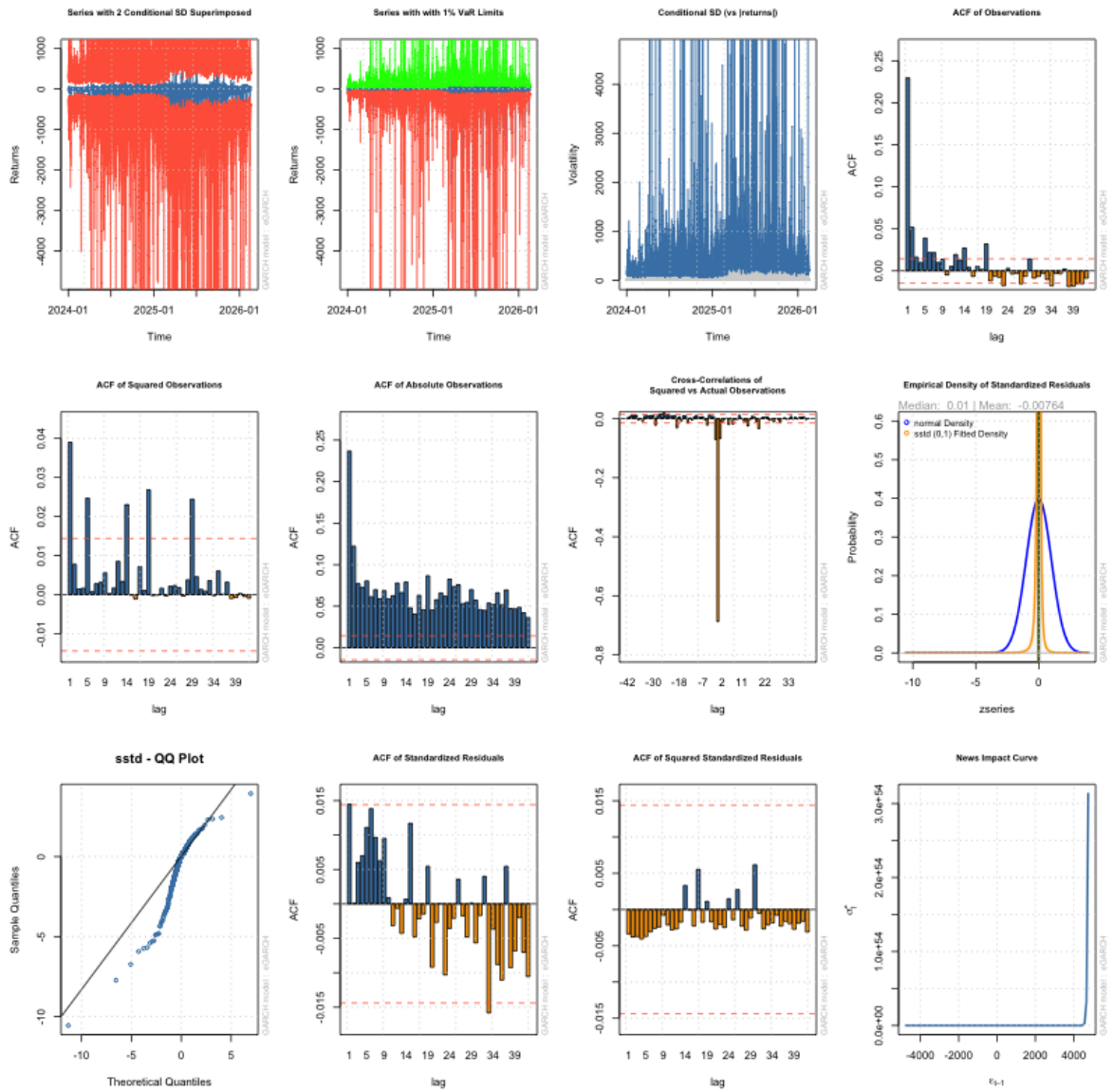


Figure 4.4: All the figures

4.6 TGARCH

While the EGARCH specification performs well overall and captures key aspects of volatility asymmetry, it still appears to understate the magnitude of extreme volatility episodes observed in the data. In particular, periods of large shocks are not fully reflected in the conditional variance dynamics. We therefore consider an alternative asymmetric GARCH framework, namely the Threshold GARCH (TGARCH) model. Specifically, we estimate an AR(1)-TGARCH(1,1)-X specification, where the exogenous variable is the production forecast error. The estimated model is given by:

$$u_t = -0.008650 + 0.366180u_{t-1} + 0.000020v_t + \varepsilon_t$$

$$\sigma_t = 0.133500 + 1.000000\left(|\varepsilon_{t-1}| - (-0.170170)\varepsilon_{t-1}\right) + 0.525090\sigma_{t-1} + 0.000000v_t$$

Parameter	Estimate	Std. Error	t-value	Pr(>|t|)
Mean Model				
μ	-0.008650	0.003068	-2.819773	0.004806**
$ar1$	0.366180	0.019934	18.369060	0.000000***
$mxreg1$	0.000020	0.000008	2.578719	0.009917**
Variance Model				
ω	0.133500	0.051253	2.604680	0.009196**
α_1	1.000000	0.068025	14.700557	0.000000***
β_1	0.525090	0.031939	16.440131	0.000000***
γ_1	-0.170170	0.019567	-8.696996	0.000000***
$vxreg1$	0.000000	0.000141	0.000004	0.999997

Table 4.9: Full Parameter Estimates for the ARFIMA(1,0,0)-TGARCH(1,1) Model

As shown on the Weighted Ljung-Box test on standardized residuals, with a p-value of $p = 0.8755$, seen in Table 4.10, suggesting that the null hypothesis of no serial correlation can not be rejected at the 5% level. However, an inspection of the ACF of standardized residuals seen in Figure 4.5 reveals a remaining significant spike at five lags, particular lag 24. This suggest that while the AR(1) process captures the vast majority of temporal dependency, a degree of daily seasonality persists in the residuals. The ACF of squared standardized residuals shows minor exceedances at higher lags. Given that Weighted Ljung-Box test yields a p-value (p=0.8147), we conclude that the TGARCH model has adequately captured the volatility dynamics, and the remaining spikes are likely due to isolated extreme events in the electricity price series rather than systematic model failure.

Diagnostic Tests	Statistic	p-value
Log-Likelihood	-91331.19	
AIC	9.8502	
BIC	9.8545	
Ljung-Box (Resid.) $Q(1)$	0.02457	0.8755
Ljung-Box (Sq. Resid.) $Q(1)$	0.05491	0.8147
Sign Bias	1.4108	0.1583
Negative Sign Bias	0.9628	0.3357
Positive Sign Bias	2.9403	0.003283**
Sign Bias (Joint Effect)	10.4483	0.015116*

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors applied.

Table 4.10: Results for the TGARCH(1,1) Model with ARFIMA(1,0,0) Mean Specification

The estimated coefficients for α_1 and β_1 are 1.000000 and 0.525090, respectively, while the threshold asymmetry parameter γ_1 is -0.170170. In a TGARCH specification under a Skewed Student-t distribution, the requirement for covariance stationarity dictates that the persistence

metric must be strictly less than one ($\alpha_1 + \beta_1 + \gamma_1\kappa < 1$). Because the innovations follow a heavy-tailed, left-skewed density, the expected value of negative shocks (κ) falls between 0.75 and 0.85. This yields an empirical persistence value of approximately 1.65 to 1.67. Consequently, the model is mathematically non-stationary, breaching the stability threshold. This lack of stationarity is structurally driven by the parameter α_1 hitting its maximum optimization boundary of 1.000000. This boundary constraint indicates that the TGARCH model is severely restricted by its own linear architecture, choking on the true magnitude of high-frequency power market anomalies.

In the mean equation the exogenous variable mxreg1 is significant with a p-value of $p < 0.001$ and a coefficient of -0.0002. This suggest that for every MW increase of VRES forecast error, there is a reduction in the price spread of 0.0002 EUR. The exogenous regressor in the variance equation vxreg1 remains insignificant, with a p-value of $p = 0.9999$.

In contrast to both the GARCH and EGARCH model, the TGARCH have sign bias. We reject the null hypothesis of no sign bias for positive shocks, with a p-value of $p = 0.0033$. This indicates that the TGARCH framework under-predicts volatility following a positive shock, i.e. the system is short. The sstd-QQ plot in Figure 4.5 provides further evidence of this; with very little of data being close to the theoretical distribution line, and the extreme tails, particularly the negative suggesting that the market's reaction to scarcity is more aggressive than the TGARCH can accommodate. The news impact curve in Figure 4.5 bottom right, is "U" shaped with more weight to the right of zero. Both the positive sign bias and joint sign bias effect are statistical relevant with p-values of 0.003283 and 0.015116, seen in Table 4.6, respectively.

With an AIC value of 9.8502 and BIC value of 9.8545, it performs slightly better than the both the GARCH model and EGARCH model, indicating that it provides a marginal improvement in in-sample fit for this specific dataset.

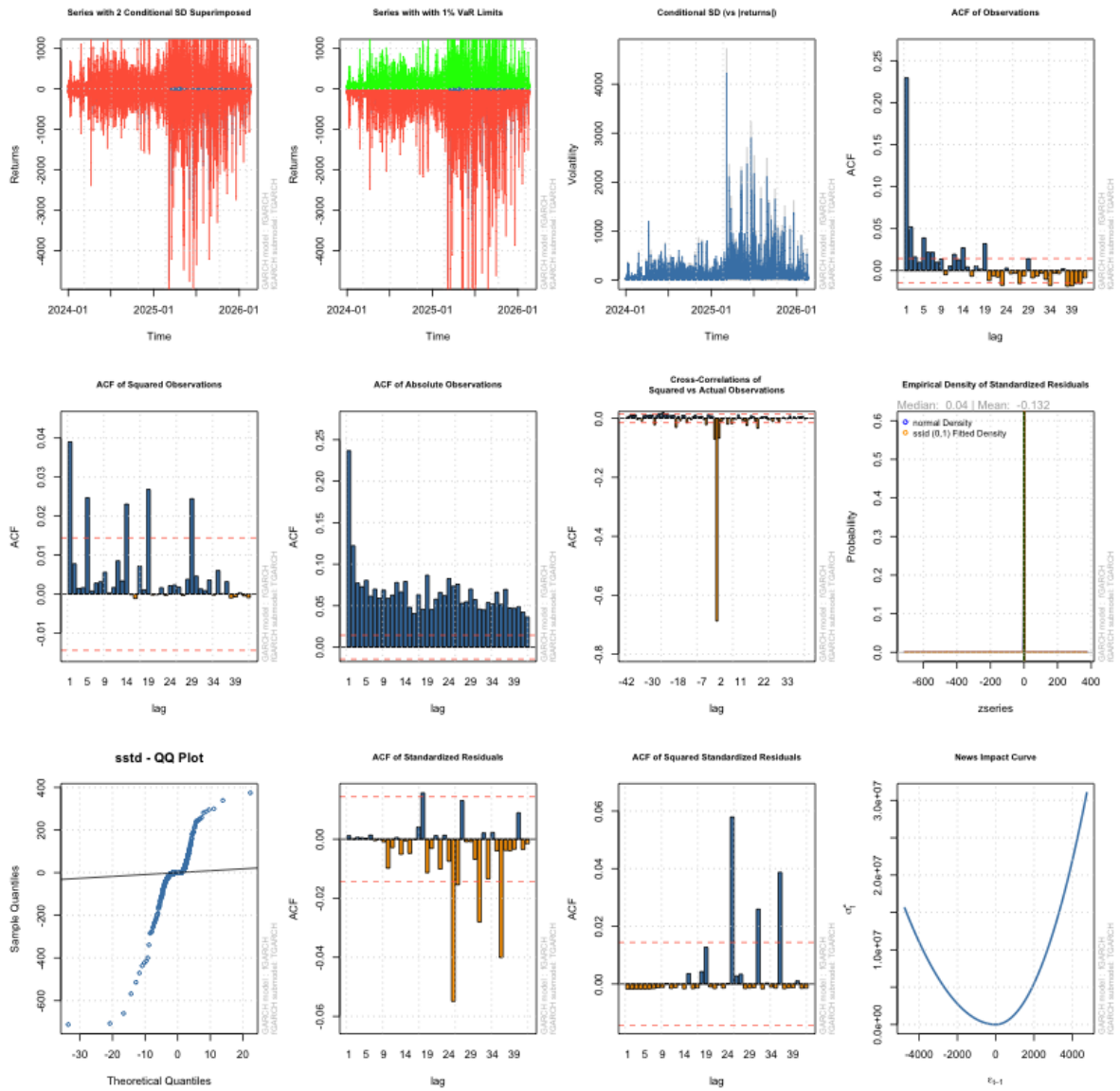


Figure 4.5: TGARCH figures

4.7 Seasonality impact

VRES production is seasonal, solar produce more electricity during the summer, as there simply are more sunny hours each day, while, wind produces more electricity during the winter as there are more wind during these months. We hypothesize that the forecast production error is the same percentage wise year round, however crucially different in levels depending on the season. This would mean that a 5% forecasting production error would deviate, in levels, with two different amount depending on weather it is summer or winter. Since the electricity market does not care about the percentage error but only the absolute level of missing electricity. Thus the demand for imbalance power would be greater, during the high VRES production months, the resulting price impact would be that we would move up on the

merit order and thus a higher marginal price. To check this we first check if there is season in the VRES production, thus we find the mean level of electricity during each of the four season. With the null-hypothesis that winter has the highest level of electricity production.
As seen in Table 4.11, the mean production levels confirm the seasonal nature of wind energy, with winter having the highest average output of 2363 MWh, followed by spring and autumn, and summer having the lowest at 1886 MWh. This supports the hypothesis: that winter have the highest production level. The next hypothesis is: That the absolute production error is largest during the winter. This hypothesis is not true as the absolute forecasting error is greatest during the autumn. The last hypothesis on the production would be: The percentage forecast error is the same across seasons. This hypothesis is also not true, with the lowest percentage error being summer with 27% and spring having the highest percentage error of 37%.
Lastly we look at the price spread. We investigate weather the mean price spread is the same, with the hypothesis that it is largest during winter, and the volatility of the price spread, for which we again expect that the percentage volatility of the price spread is largest during the winter. The mean price spread is across all four seasons are close to each other, with the notable difference being that winter is positive, indicating the that the market is on average long during the winter. The volatility of the price spread is greatest during spring then closely followed by summer, then fall and lastly winter, this is in contrast to our expectations, indicating that during winter there is more balance power available.

Season	Mean Production (MW)	Mean Price spread (EUR)	Production Pct. Error (σ%)	Production Abs. (MW) Error	Standard Deviation Price Spread
Winter	2363	0.42	31%	527.35	67.89
Spring	1992	-3.93	37%	504.40	143.82
Autumn	1969	-2.53	33%	544.89	107.40
Summer	1886	-2.17	27%	479.33	136.86

Table 4.11: Seasonal Comparison of Production and Price Spread.

Contradicting initial expectations, price spread volatility does not peak during the winter. While winter indeed exhibits the highest mean production levels, it conversely registers the lowest standard deviation in the price spread. Therefore, despite clear seasonality in production, the hypothesis that physical errors linearly drive price risk is reversed, with price spread volatility being smallest during winter. An explanation for this paradox is that because base load and generation are higher during winter months, flexible imbalance power and online spinning reserves are more readily available to stabilize the grid when forecasting errors occur.

When these seasons are used in the GARCH models, with winter as baseline, all 3 seasons are significant on a 10% level with summer and fall being significant on a 5% level. An argument for why all season are at least somewhat relevant, might be that Denmark only have two seasons, in regards to power. Namely, a heating season and a non-heating season, and a high tariff season and low tariff season. These two season overlap [17], and we will therefore use the specific dates for when the tariff changes, the summer months being from 1. April to 30. September and winter months being from 1. October to 31. marts.

Similar to the four-season scenario, we expect the colder months to feature the highest mean production, which the data confirms. Under this two-season specification, we also observe that

physical production forecasting errors are greatest during the winter season, both in absolute terms and percentage-wise. Yet, in contrast to these physical risk metrics, the standard deviation of the financial price spread is minimized during winter. This decoupling suggests that winter's higher baseload capacity provides a natural cushion of flexible imbalance power, preventing physical supply shocks from translating into explosive price volatility. The data can be seen in Table 4.12.

Season	Mean Production (MW)	Mean Price spread (EUR)	Volatility Pct. Error (σ%)	Volatility Abs. (MW) Error	Volatility Price Spread (σ spread)
Winter (October - Marts)	2188	1.91	0.345	532.70	99.96
Summer (April - September)	1892	-1.89	0.278	476.83	135.96

Table 4.12: Seasonal Comparison of Production and Price Spread.

As a robustness check to ensure macro-level temporal dynamics are captured, dummy variables representing seasonal components, first utilizing a four-season specification and subsequently a parsimonious two-season (winter vs. summer) setup, were introduced into the mean and conditional variance equations. The resulting information criteria are detailed in Table 4.13. The inclusion of seasonal regressors yields negligible changes to both the AIC and BIC compared to the baseline models. Thus when applying the models in a forecasting setup, the baseline models will be utilized.

Model Specification	**AIC**	**BIC**
Four Seasons Setup		
GARCH	9.9718	9.9781
EGARCH	10.0850	10.0920
fGARCH	9.8506	9.8573
Two Seasons Setup		
GARCH	9.9713	9.9759
EGARCH	10.0870	10.0930
fGARCH	9.8503	9.8562
Baseline Setup		
GARCH	9.9697	9.9735
EGARCH	10.0940	10.0980
TGARCH	9.8502	9.8545

Table 4.13: AIC and BIC Comparison Across Seasonal Variations and Baseline Models

4.8 Forecasting

We apply the three models to a prediction task using an 80/20 train-test split. We chose a 336-hour (two-week) forecasting horizon to test the models' structural integrity over the medium term. Furthermore, the models are re-estimated every 24 hours. This interval mirrors the power market's daily trading cycle, and simultaneously reducing the total computational load.

To assess performance, we first examine the point forecasts for prices. While not our primary focus, these provide a baseline for model behavior. We then evaluate each models' ability to capture volatility, by checking the number of actual price spread values which ended up

outside the predicted confidence bands. This allows us to determine if the GARCH framework captures risk more effectively than a standard 95% confidence interval. To ensure these intervals remain informative, we evaluate their precision (width), as excessively wide intervals provide little value. Finally, the GARCH models are bench marked against a static AR(1) model with a fixed 95th percentile interval to quantify the improvement gained from dynamic volatility modeling. The aggregated results can be seen in Table 4.14.

As anticipated, the standard GARCH model is the least effective at price prediction, yielding the highest average MAE and RMSE, whereas the EGARCH model emerges as the top performer. Furthermore, while the GARCH model achieves the highest 'coverage,' this metric is misleading; it significantly overestimates risk by capturing 99.8% of observations within an intended 95% confidence interval. This suggests that the standard GARCH produces excessively wide and uninformative bands compared to the more precise EGARCH and TGARCH alternatives.
The primary failure of the GARCH model lies in its lack of precision; the average width of its confidence interval is $2.39 \cdot 10^{25}$, a figure so vast, it surpasses the global M2 money supply, that it renders the results meaningless. This extreme divergence is a direct consequence of model instability, as the persistence parameters $\alpha + \beta > 1$ lead to an explosive variance process and an inflated maximum sigma.
The EGARCH model provides the most robust results, achieving the lowest accuracy measures, while maintaining a 96.4% coverage rate, which closely aligns with the 95% target. With an average interval width of 202.19, EGARCH offers a realistic representation of electricity market volatility and maintains the numerical stability that the simpler GARCH specification lacks.
Conversely, while the TGARCH model produces the narrowest average width of 121.11, its intervals are overly restrictive. A coverage rate of only 82.8% suggests that it underestimates true market volatility, failing to capture nearly 17% of price movements. From a risk management perspective, such a model would likely lead to frequent, unexpected losses.
Finally, both the EGARCH and TGARCH models significantly outperform the static baseline in terms of accuracy. The dynamic nature of the GARCH family intervals offers a clear advantage over the static approach; while the static baseline achieves high coverage, its confidence bands are more than twice the size of those produced by the EGARCH model. This confirms that a time varying volatility framework is essential for achieving precision without sacrificing reliability.

Model	**Avg MAE**	**Avg RMSE**	**Avg Coverage**	**Avg Width**	**Max Sigma**
Baseline	49.063	107.688	0.973	440.81	$1.12 \cdot 10^2$
EGARCH	33.630	43.543	0.964	202.19	$1.32 \cdot 10^2$
TGARCH	40.972	50.013	0.828	121.11	$6.11 \cdot 10^1$
GARCH	49.079	98.367	0.998	$2.39 \cdot 10^{25}$	$5.02 \cdot 10^{27}$

Note: GARCH results exhibit numerical divergence over the 14-day horizon.

Table 4.14: Global Model Performance Comparison (336-Hour Horizon)

In the table below, Table 4.15, the MAE, the coverage of the confidence intervals and average width of the confidence interval is shown, for each of the three models, grouped in short (first 24 hours), medium (from hour 25 to 168), and long (from hour 169 to 336) term periods. This

is done in order to see if the performance drops with prediction further into the future.

The transition from short-term (24h) to long-term (336h) forecasting highlights a critical divergence in model utility. The EGARCH specification manages to preserve the statistical integrity of its confidence intervals, maintaining 95% coverage even as the MAE rises. In contrast, the TGARCH exhibits a clear decay in risk-capturing ability, with its coverage falling nearly 14% below the target by the end of the 14-day period. This suggests that for power market participants managing medium-term risk, the EGARCH's structural stability offers a superior balance between predictive accuracy and reliable uncertainty quantification.

	MAE (Error)			**Coverage (95% Target)**			**Average Width**		
Model	**Short**	**Med.**	**Long**	**Short**	**Med.**	**Long**	**Short**	**Med.**	**Long**
EGARCH	**30.67**	**31.19**	**38.22**	0.96	0.97	0.95	198.69	202.68	202.43
TGARCH	33.13	37.90	43.93	0.87	0.85	0.81	**116.40**	**121.42**	**121.65**
GARCH	42.68	45.79	51.90	0.99	1.00	1.00	10^{25}	10^{25}	10^{25}

Short: 1–24h; Medium: 25–168h; Long: 169–336h. Bold denotes best performance in category.

Table 4.15: Performance Decay Across Forecast Horizons (Short vs. Medium vs. Long Term)

To supplement the statistical metrics, Figures 4.6, 4.7, and 4.8 illustrate the models' performance over the final 14-day window of the test set. These visualizations provide a granular view of the conditional nature of the forecasts. Specifically, they highlight the EGARCH model's superior ability to dynamically adjust its confidence envelope in response to realized price shocks, whereas the TGARCH plot reveals frequent coverage breaches and the GARCH plot demonstrates the numerical instability discussed in Section 4.4.
The following three figures, Figure 4.6, 4.7, and 4.8 are 14 days in the test data set, for which GARCH, EGARCH, and TGARCH are used to predict. The line in each figure is the forecasted price, the shaded area is the confidence interval, and the red dotted line is the actual price, furthermore all plots are set to the same size on the y-axis, such that the actual prices are the visual the same size across all three plots, for easier comparisons, the same plots but unscaled can be seen in Figure B.1, B.2, and B.3.

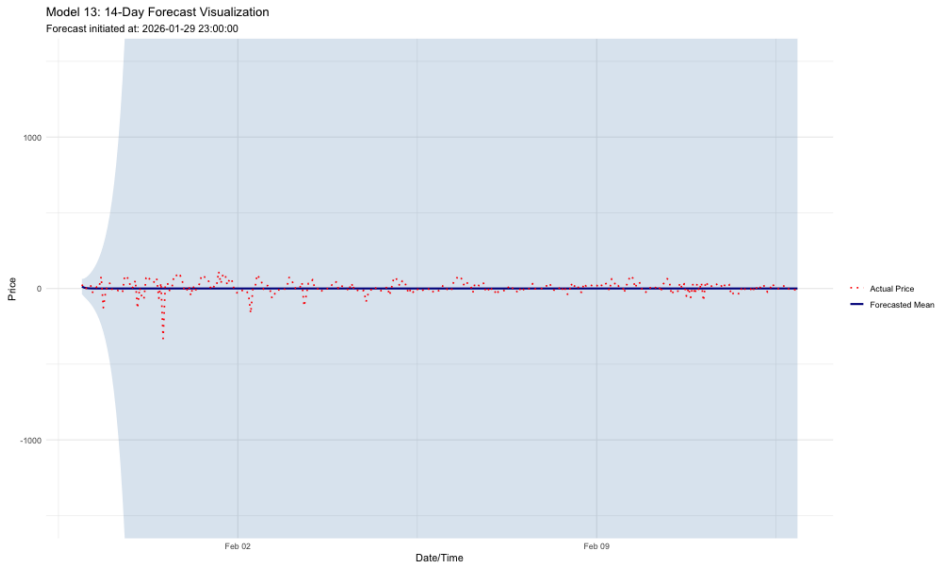


Figure 4.6: GARCH Model: Last 14 days predicting

The structural instability of the standard GARCH(1,1) specification, arising from non-stationary parameter estimates where $\alpha_1 + \beta_1 > 1$, is visually apparent in Figure 4.6. The corresponding forecast confidence intervals exhibit explosive behavior, rapidly expanding beyond the plot dimensions. As illustrated by the unscaled axis in Figure B.1, the variance forecast trends toward extreme values on the order of $\pm 2 \times 10^{27}$. Ultimately, the chart yields no predictable insights into the underlying price spread dynamics, serving instead as a textbook demonstration of an unstable, integrated variance process.

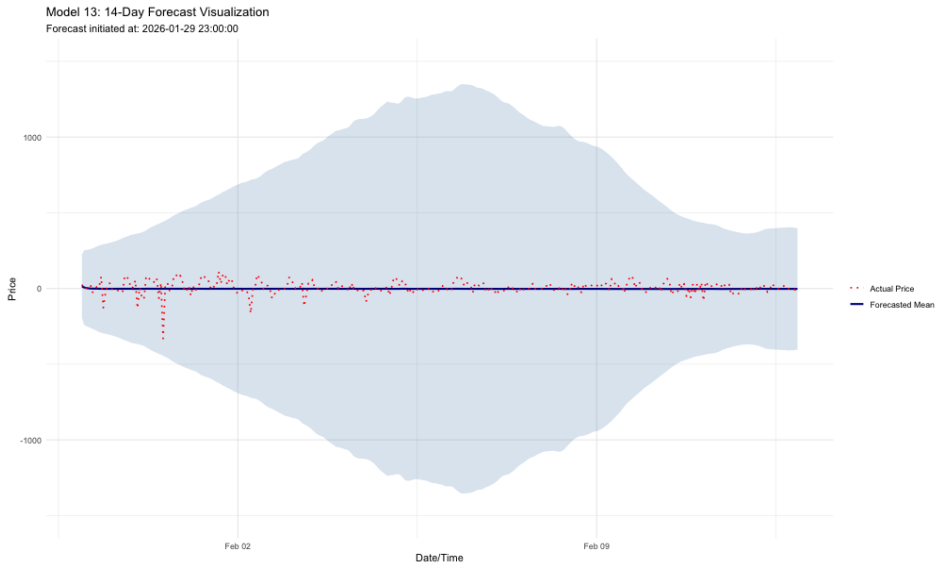


Figure 4.7: EGARCH Model: Last 14 days predicting

The forecast performance of the EGARCH specification, seen in Figure 4.7, stands in stark contrast to the instability observed in the standard GARCH model. The EGARCH model

produces a stable volatility prediction, which remains economically meaningful over the 2 week horizon, and with all actual values staying inside these confidence bands.

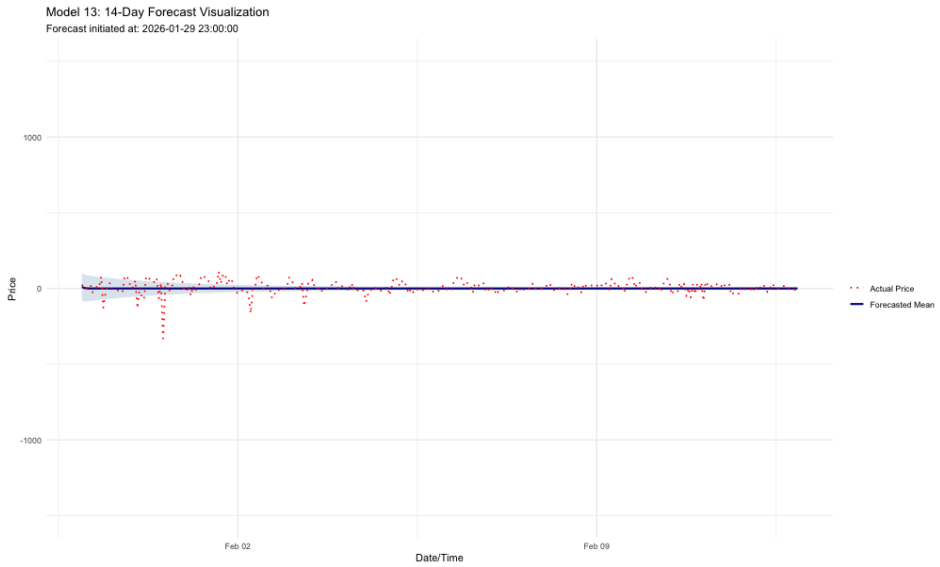


Figure 4.8: TGARCH Model: Last 14 days predicting

The predictive performance of the TGARCH specification, as seen in Figure 4.8, shows a predicted conditional mean that remains entirely horizontal over the 14-day horizon. Furthermore, the predicted conditional confidence bands are excessively narrow, failing to capture a large number of the actual price spread observations. Most notably, the specification completely underpredicts the system's tail risk, leaving several extreme negative price spikes entirely outside the shaded variance boundaries.

The GARCH model is more or less completely useless as it is unstable and explodes, this is seen both in the average width of the predicted confidence bands, and supported by a visual example in Figure 4.6. In contrast, both the EGARCH and TGARCH specifications maintain stable volatility paths, demonstrating that modeling the asymmetric leverage effect is vital when analyzing electricity price spreads. A distinct trade-off emerges between these two asymmetric variants: the EGARCH model generates larger, more conservative confidence bands that successfully capture a greater proportion of actual price realizations, whereas the TGARCH model enforces narrower, more tight boundaries at the expense of missing a higher number of extreme tail events. Ultimately, the EGARCH specification emerges as the superior model for capturing the stylized facts of the electricity price spread the best. While TGARCH dampens its volatility response due to rigid parameter constraints, EGARCH's logarithmic formulation allows it to accommodate the exponential, non-linear nature of real-time balancing shocks. By generating wider, dynamic confidence bands that successfully encompass the extreme tail events, the EGARCH model proves to be the most robust framework for reflecting the true financial and physical risks of the market.

4.9 Higher-Order EGARCH and TGARCH Specification

Initial estimation results indicated that the standard EGARCH(1,1) specification was insufficient for fully capturing the volatility dynamics present in the data. In particular, residual diagnostics and model selection criteria suggested that additional lag structures may improve the model fit and better account for persistence and asymmetry in volatility. Therefore, several alternative EGARCH and TGARCH specifications are estimated using different combinations of volatility terms. We evaluate all combinations of lag $p = 1, \ldots, 6$ and $q = 1, \ldots, 6$. The competing models are evaluated using the AIC, allowing for a comparison between model fit and model complexity. The same procedure was subsequently mirrored to the TGARCH framework in order to examine whether threshold effects and asymmetric responses to shocks provide a better representation of volatility dynamics in electricity price spreads.
Based on the AIC, the best performing specification overall is the AR(1)-EGARCH(6,5)-X model. However, this model is non-stationary, as the sum of its autoregressive volatility parameters (β) exceeds unity. It achieved a p-value of 0.5893 on the Weighted Ljung-Box test, meaning we fail to reject the null hypothesis of no serial correlation. In fact, the highest-ranked specification that successfully maintains covariance stationarity (i.e., $\sum_{i=1}^{p} \beta_i < 1$) is the AR(1)-EGARCH(2,6)-X, exhibiting a persistence of 0.9825. Furthermore, this specification achieved a Weighted Ljung-Box test p-value of 0.7172, successfully capturing remaining autocorrelations, a notable improvement over the baseline (1,1) specification. Its AIC value of 9.9999 is only marginally higher than the model with the best AIC (9.9281). Indeed, all 36 specifications fall within a narrow 0.3 AIC-point band, indicating that statistical improvements from higher-order models are relatively modest. The complete grid search results are presented in Table 4.16.

GARCH (p)	ARCH (q)	AIC	BIC	Persistence	Stationary	Ljung-Box p-val
6	5	9.9281	9.9382	1.0003	No	0.5893
5	3	9.9327	9.9412	1.0003	No	0.5104
6	4	9.9338	9.9435	1.0006	No	0.5686
5	6	9.9357	9.9454	1.0006	No	0.3536
6	6	9.9366	9.9471	1.0010	No	0.3165
5	4	9.9406	9.9494	1.0007	No	0.5401
6	2	9.9407	9.9496	1.0011	No	0.2818
5	5	9.9409	9.9502	1.0012	No	0.4419
4	4	9.9438	9.9518	1.0008	No	0.6042
3	3	9.9440	9.9507	1.0005	No	0.4505
3	4	9.9445	9.9517	1.0005	No	0.3564
5	2	9.9456	9.9536	1.0009	No	0.4589
4	3	9.9456	9.9532	1.0006	No	0.4412
4	5	9.9457	9.9541	1.0008	No	0.4685
3	6	9.9458	9.9539	1.0005	No	0.7153
4	6	9.9472	9.9560	1.0012	No	0.7771
6	1	9.9502	9.9586	1.0000†	Yes	0.8124
5	1	9.9548	9.9624	1.0000†	Yes	0.8494
4	2	9.9564	9.9636	1.0025	No	0.7132
4	1	9.9580	9.9647	1.0000†	Yes	0.7323
3	2	9.9721	9.9784	1.0018	No	0.9662
3	1	9.9731	9.9790	1.0000†	Yes	0.9734
2	**6**	**9.9999**	**10.0070**	**0.9854**	**Yes**	**0.7175**
2	5	10.0083	10.0151	0.9848	Yes	0.5342
2	4	10.0122	10.0185	0.9817	Yes	0.9811
2	3	10.0152	10.0211	0.9850	Yes	0.9784
2	2	10.0172	10.0227	0.9854	Yes	0.9364
2	1	10.0194	10.0245	0.9886	Yes	0.9380
1	6	10.0383	10.0446	0.9453	Yes	0.0265
1	5	10.0448	10.0507	0.9389	Yes	0.0260
1	4	10.0545	10.0600	0.9284	Yes	0.0153
3	5	10.0626	10.0702	0.8504	Yes	0.1608
1	3	10.0669	10.0720	0.9124	Yes	0.0115
1	2	10.0792	10.0838	0.8974	Yes	0.0192
1	1	10.0937	10.0979	0.8784	Yes	0.0441
6	3	10.1667	10.1760	1.0001	No	0.0000

Specifications are ordered by ascending AIC. All models are evaluated using an underlying AR(1) mean process. † Indicates estimated parameters that hit the numerical upper bound limit (0.99999999) assigned by the optimizer.

Table 4.16: Complete EGARCH Grid Search Results ($p, q \in [1,6]$)

The estimated parameters for the optimal stationary AR(1)-EGARCH(2,6)-X specification are detailed in Table 4.17, with the underlying process estimated as follows:

$$u_t = -1.850165 + 0.437461u_{t-1} - 0.001602v_t + \varepsilon_t$$

$$\begin{aligned}\ln(\sigma_t^2) = {} & 0.158047 + 0.418531\big(|Z_{t-1}| - \mathbb{E}|Z_{t-1}|\big) + 4.506769Z_{t-1} \\ & + 0.763049\big(|Z_{t-2}| - \mathbb{E}|Z_{t-2}|\big) - 2.217514Z_{t-2} \\ & + 1.000000\ln(\sigma_{t-1}^2) - 0.094229\ln(\sigma_{t-2}^2) - 0.018590\ln(\sigma_{t-3}^2) \\ & - 0.026310\ln(\sigma_{t-4}^2) + 0.002265\ln(\sigma_{t-5}^2) + 0.1119353\ln(\sigma_{t-6}^2) \\ & - 0.000027v_t\end{aligned}$$

Parameter	Estimate	Std. Error	t-value	Pr($>\|t\|$)
Mean Model				
μ	-1.850165	0.677602	-2.73046	0.006325**
$ar1$	0.437461	0.040737	10.73869	0.000000***
$mxreg1$	-0.001602	0.000415	-3.86146	0.000113***
Variance Model				
ω	0.158047	0.138406	1.14191	0.253492
α_1	0.418531	0.261006	1.60353	0.108817
α_2	0.763049	0.348065	2.19226	0.028361*
β_1	1.000000	0.000026	38632.63	0.000000***
β_2	-0.094229	0.000026	-3628.98	0.000000***
β_3	-0.018590	0.012423	-1.49638	0.134555
β_4	-0.026310	0.001510	-17.41795	0.000000***
β_5	0.002265	0.003658	0.61928	0.535732
β_6	0.119353	0.000273	437.71716	0.000000***
γ_1	4.506769	0.351102	12.83607	0.000000***
γ_2	-2.217514	0.575741	-3.85158	0.000117***
$vxreg1$	-0.000027	0.000011	-2.35995	0.018277*

Table 4.17: Full Parameter Estimates for the AR(1)-eGARCH(2,6) Model

Similar to the baseline AR(1)-EGARCH(1,1)-X model, the higher-order AR(1)-EGARCH(2,6)-X specification successfully mitigates significant sign bias. This is confirmed by the diagnostics presented in Table 4.18, where all respective p-values exceed the 10% significance level.

Diagnostic Tests	Statistic	p-value
Log-Likelihood	-92711.87	
AIC	9.9999	
BIC	10.0071	
Ljung-Box (Resid.) $Q(1)$	0.1319	0.7165
Ljung-Box (Sq. Resid.) $Q(1)$	0.2217	0.6378
Sign Bias	0.3989	0.6900
Negative Sign Bias	0.6753	0.4995
Positive Sign Bias	1.5456	0.1222
Sign Bias (Joint Effect)	2.8904	0.4088

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors applied.

Table 4.18: Results for the EGARCH(2,6) Model with AR(1) Mean Specification

Again we apply the model in a prediction setting. Forecasting a 14 day period or 336 hours, which can be seen in Figure 4.9. In this forecasting window, we see that all actual values are within the predicted confidence bands, and that the bands gets narrower the further into the future we predict.

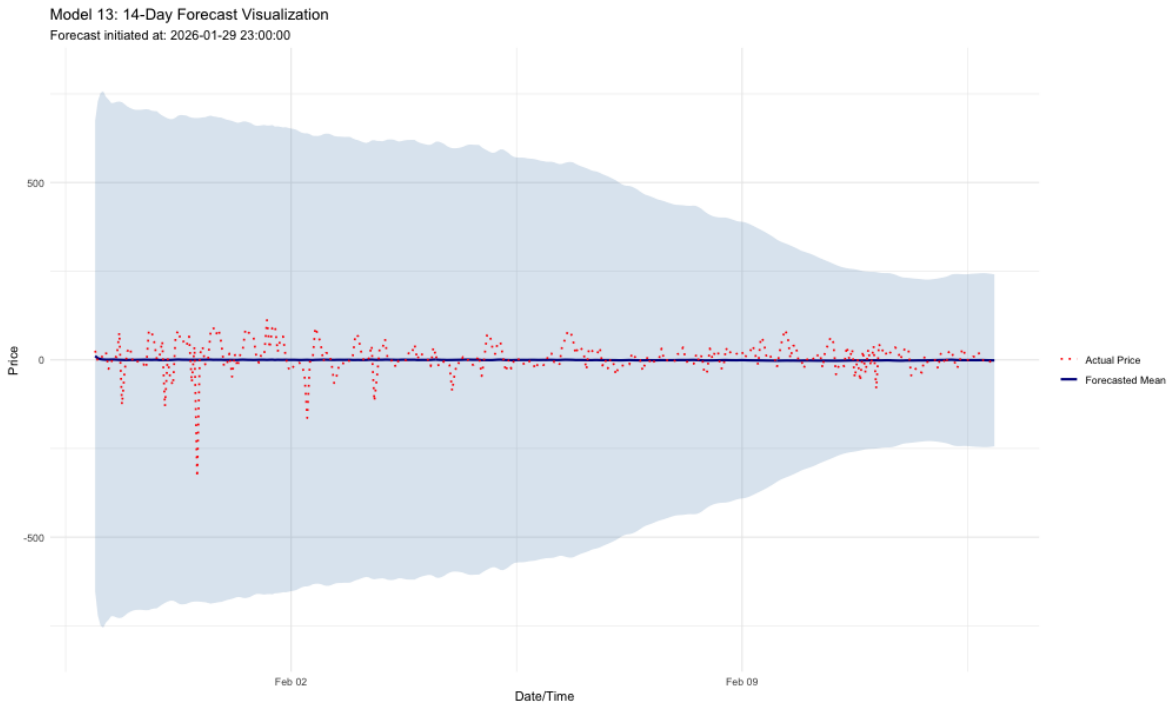


Figure 4.9: EGARCH(2,6) 14 day Forecast

Table 4.19 and Table 4.20 present the performance metrics evaluated across the full 14-day forecast horizon, alongside their respective decay dynamics across short-, medium-, and long-term horizons. This evaluation covers all five specifications: the three benchmark GARCH frameworks introduced in previous sections, alongside the newly implemented EGARCH and TGARCH models.

Model	**Avg MAE**	**Avg RMSE**	**Avg Coverage**	**Avg Interval Width**	**Stability (Max σ)**
EGARCH(1,1)	33.630	43.543	0.964	202.197	132.816
EGARCH(2,6)	49.344	98.277	0.974	575.877	2135.330
TGARCH(1,1)	40.972	50.013	0.828	121.108	61.121
TGARCH(5,4)	49.080	98.367	0.827	38953.470	929,774.300
GARCH	49.079	98.367	0.998	23.88×10^{24}	5.02×10^{27}

Table 4.19: Model Performance Comparison: Predictive Accuracy, Coverage, and Stability

Interestingly, the higher-order models demonstrate inferior point-forecasting accuracy compared to their simpler (1,1) counterparts, yielding higher MAE and RMSE metrics overall. While empirical coverage levels remain largely stable, the average interval width for the highly parameterized models more than doubles, indicating significantly higher forecast uncertainty and reduced precision in capturing future price movements.

Model	Avg. MAE			Avg. Coverage			Avg. Interval Width		
	Short (1–24h)	**Medium** (25–168h)	**Long** (169–336h)	**Short** (1–24h)	**Medium** (25–168h)	**Long** (169–336h)	**Short** (1–24h)	**Medium** (25–168h)	**Long** (169–336h)
EGARCH(1,1)	25.970	31.810	37.500	0.980	0.970	0.950	196.340	202.640	202.410
EGARCH(2,6)	42.120	47.510	51.490	0.980	0.970	0.970	474.480	542.910	611.290
TGARCH(1,1)	31.690	40.210	42.900	0.880	0.830	0.820	115.590	121.560	121.580
TGARCH(5,4)	46.800	47.130	50.930	0.820	0.830	0.830	45,462.42	47,166.65	31,689.50
GARCH(1,1)	46.600	47.320	50.840	1.000	1.000	1.000	1.51×10^{25}	2.56×10^{25}	2.41×10^{25}

Table 4.20: Performance Decay Across Short-, Medium-, and Long-Term Forecast Horizons

The Story remains consistent across the short-, medium-, and long-term forecast horizons. Meaning similar to their $(p = 1, q = 1)$ counterpart their performance also decay as the forecast horizon increases. Furthermore they are all best at the short-term forecast horizon, which is to be expected.

A similar estimation procedure was conducted for the TGARCH specification, with the comprehensive results compiled in Appendix D. The empirical narrative closely aligns with that of the AR(1)-EGARCH(2,6)-X specification; despite the added parameterization, the higher-order TGARCH(5,4) model exhibits slightly inferior out-of-sample forecasting performance relative to the TGARCH(1,1) baseline.

Âť

5 Discussion

5.1 Limitations of Methodology

5.1.1 The Chosen Univariate Models

By defining the dependent variable price spread as the difference between the spot price and the imbalance price, we are forcing to different market clearing mechanisms into a univariate series. The spot market and the balancing market are governed by entirely separate clearing timelines, different sets of market participants, and distinct liquidity constraints. Collapsing these two complex price streams into a single series restricts the models from capturing independent volatility paths or decoupled variance shocks occurring simultaneously in both markets.

5.1.2 Aggregating to hourly interval

New regulation of the electricity market across Europa were enforced when we entered the new year 2026. The regulatory changes, were a shift from hourly spot price and imbalance price intervals to a quarter hourly interval. However, to ensure a continuity in the data, and to ensure that a whole calender year were included, aggregation of the data were necessary. Aggregation of production data were also necessary as it was in 5 min intervals. However, Aggregating these series acts as a smoothing filter, as an event could occur where a sudden shock would incurred in the one 15 min interval, but is smoothed out over an hour. The consequence is that the true risk faced by a BRP could be much worse.

5.1.3 Time-Invariant Parameters vs. Structural Regime Shifts

The GARCH models implement in this paper uses continuous, time-invariant parameters, meaning the α, β, and γ coefficients are fixed numbers computed as an average across the entire dataset. However, the extreme volatility of the electricity market warrants the investigation of whether a regime shifting parameter would improve the models, as there is a difference between a stable market where volatility still is high, and an unstable market were the electricity prices can explode more than a hundred fold. The latter could be coursed by a sudden bottleneck, like an interconnector fails or reached capacity, which forces the TSO to activate power along the inelastic "elbow" of merit order curve.

5.1.4 Time varying exogenous parameters

The model considers the absolute level of the forecast production error v_t, but does not take in to consideration other metrics, such available balancing power, where net power import is included. This is important as large production shortfall, does have to very different impacts on whether there is plenty of available balancing power or not. Another factor that could affect available balancing power is time of day, whether it is peak or off-peak. The volume of available spinning reserves and market depth changes drastically between peak business hours and overnight periods. The inclusion of time-of-day dummy variables or ramp velocity metrics, could improve the model, as currently the model assumes the market's physical elasticity

is constant throughout the 24-hour cycle.

5.2 Model performance

5.2.1 LM model

The baseline linear model identified a statistically significant relationship between the production forecast error and the price spread. However, because the data exhibits leptokurtosis and heteroskedasticity, the model was unable to achieve a high degree of explanatory power, as visually supported by Figure 4.1 and the low R^2 value of 0.0033. Crucially, the presence of these unresolved residual dynamics, specifically time-varying heteroskedasticity and volatility clustering, violates standard OLS assumptions and directly motivates the deployment of a GARCH framework.

5.2.2 GARCH model

The introduction of the standard GARCH(1,1) specification confirmed that accounting for conditional heteroscedasticity is vital when modeling the price spread. However, the GARCH framework proved to be highly unstable and mathematically non-stationary, as evidenced by an explosive variance parameter sum of:

$$\alpha_1 + \beta_1 = 0.999440 + 0.430402 = 1.429442.$$

The explosive behavior of the model, resulted in predicted confidence intervals so large that the became uninformative about risk, rendering the model impractical for any application.

The core reason for this failure is the mechanical symmetry enforced by the standard GARCH architecture. The GARCH model relies on a quadratic shock term, meaning it treats positive and negative shocks identically, assuming they exert the exact same impact on future volatility. In real-time balancing power markets, this symmetry is physically and economically impossible due to the uneven technical constraints of the grid's merit-order curve. By attempting to capture sudden, one-sided vertical leaps using a symmetric quadratic function, the GARCH parameters overcompensated and exploded, misinterpreting directional structural spikes as a permanent cascade toward infinite volatility.

The forecast production error were not significant in either the mean part or the variance part of the model. Meaning for a GARCH model production forecast error does not improve the inference of the price spread.

5.2.3 EGARCH model

The EGARCH specification emerged as the superior econometric framework for capturing the price spread's empirical properties. Unlike the standard GARCH, it achieved covariance stationarity, and it significantly outperformed alternative models during out-of-sample forecasting by generating dynamic, reliable risk boundaries. This superiority stems from two core advantages:

Logarithmic Volatility Formulation: By parameterizing the natural logarithm of the conditional variance ($\ln \sigma_t^2$), the model is freed from the rigid non-negativity parameter constraints that bind the GARCH and TGARCH algorithms. This allows the variance to expand rapidly during extreme price spikes without hitting artificial algorithmic walls.

Exponential Shock Response: Modeling the log-variance establishes an exponential relationship between market shocks and conditional volatility. This perfectly mirrors the physical shape of the grid's inelastic merit-order curve, where small forecast errors are easily absorbed, but extreme errors push the system onto the vertical "elbow" of the supply curve, triggering an explosive volatility cascade.

The overwhelming statistical significance of the size effect ($\gamma_1 = 4.073$) and sign parameter ($\alpha_1 = 1.789$) confirms that directional leverage effects dominate market variance, a reality visually validated by the severe right-sided skew of the News Impact Curve.
However, the baseline AR(1)-EGARCH(1,1)-X model exhibited a slight statistical vulnerability, yielding a Weighted Ljung-Box p-value of 0.0454 and indicating lingering serial correlation in the standardized residuals. To resolve this, an iterative lag-optimization routine was executed across alternative orders of the GARCH dependencies (p and q lags). This refined lag architecture successfully pushed the Ljung-Box diagnostics back above the 5% significance threshold, eliminating remaining serial correlation and validating the optimized EGARCH framework as a robust, fully stable risk management tool. Similiar to the GARCH model, the forecast production error were not significant in the EGARCH specification.

5.2.4 TGARCH

The AR(1)-TGARCH(1,1)-X model highlights how volatile the electricity market is, by pushing the α_1 parameter to the software's limits, by hitting the ceiling value at exactly 1.00000. Combined with the $\beta_1 = 0.52090$ coefficient and asymmetry parameter $\gamma_1 = -0.170170$, under the sstd which results in the covariance stationarity condition not being met.
The model did not perform well in a forecasting setting, as a large number of actual values fell outside the predicted confidence intervals. Furthermore it achieved the second best MAE and RMSE values. Ultimately, while the TGARCH achieved a lower penalty score and marginally superior AIC (9.8502) and BIC (9.8545) values by tightly fitting the market's quiet baseline hours, its boundary constraints and non-stationary architecture render it hazardous for robust tail-risk forecasting compared to an unconstrained exponential framework. In contrast to the other two volatility models, forecast production was significant in the mean part of the model.

5.2.5 Evaluation of Higher-Order Asymmetric Specifications

To assess whether capturing long-memory dependencies in conditional variance could yield predictive gains, the empirical analysis extended the baseline frameworks to higher-order lag structures: an AR(1)-EGARCH(2,6)-X and an AR(1)-TGARCH(5,4) specification.

In-sample, both models performed slightly better that their (1,1) counterparts. As documented in the diagnostic estimations, the EGARCH(2,6) and TGARCH(5,4) models successfully minimize the AIC, and for the EGARCH specification eliminate standard serial correlation, as

confirmed by the non-significant Ljung-Box test p-value of 0.7175. Furthermore, the inclusion of multiple asymmetry parameters (η) is significant across both setups, validating the presence of pronounced time-varying leverage effects in electricity price spreads. Furthermore, in the EGARCH(2,6) specification the forecast production error was significant in both the mean and variance process, which was not the case for the TGARCH(5,4) specification.

However, when subjected to out-of-sample forecasting, both specifications perform worse, reinforcing the econometric principle of parsimony. Both models had higher MAE and RMSE throughout the short-, medium-, and long-term forecasting horizons. Furthermore, the models predicted confidence bands become more than twice the size, resulting in less precise bands.

Additionally, the TGARCH(5,4) specification fails to completely resolve asymmetric volatility shocks, as evidenced by the remaining significant positive sign bias ($p = 0.0051$). Consequently, while the higher-order lag structures excel at mining historical noise to optimize in-sample fit, they overfit the data. Out-of-sample, they are consistently outperformed by the simpler EGARCH(1,1) and TGARCH(1,1) baselines, which generate more narrow predictive bands.

5.2.6 Impact of the Exogenous Regressor

The production forecast error exhibits a unique role across the estimated specifications. Interestingly, this exogenous regressor is statistically significant in both the conditional mean and conditional variance processes for only a single model: the AR(1)-EGARCH(2,6)-X. Notably, this specification stands out as the sole model that simultaneously satisfies the conditions for parameter stationarity and eliminates residual autocorrelation.

Finally, these findings are supported by the linear regression model, which confirms a statistically significant baseline relation between the production forecast error and the price spread. Together, these results imply that while external supply-side shocks linearly pressure the spread, their capacity to systematically drive conditional volatility is sensitive to the chosen GARCH architecture.

5.3 Practical Implications and Use Cases for Market Participants

The transition toward Variable Renewable Energy Sources (VRES) fundamentally alters the risk landscape of modern power systems. Therefore, the choice between the estimated volatility frameworks carries significant financial implications. The findings established in this study offer practical utilities for BRPs and TSOs.

BRPs carry the financial responsibility for ensuring that their schedules are balanced and any deviation there from. Because VRES deviations are inherently asymmetric, where unexpected wind deficits force exposure to severe positive price spikes while surpluses can be cheaply curtailed, the EGARCH model captures this asymmetry, and therefore provides commercial

utility. The EGARCH(2,6) allows the BRPs to better manage their risk and dynamically adjust the risk profile embedded in their forward contracts based on the direcional profile of forecasted VRES production.

For the TSO, whose objective is system stability rather than generate a profit, the volatility metrics can be used for determining ancillary service capacity requirements. This prevents the TSO from procuring to much expensive reserve capacity, which inflates consumer tariffs, while ensuring that grid stability margins are robust against extreme asymmetric weather events.

5.4 Future work

The empirical limitations encountered in this project, gives an indication for future research. To capture the full complexity of high-frequency electricity price spreads, subsequent investigations should pivot towards different GARCH models and the VAR framework.

As demonstrated by the non-stationarity of our baseline specifications, electricity price spreads switch violently between structural states. When the grid is well balanced, prices remain stable and follow a low variance, mean-reverting path. However, when the system experiences severe shortages, it transitions into an explosive, inelastic state. This inelastic state reflects the physical constraints of the power grid, namely, interconnecting cables hitting their maximum capacity. Once transmission is saturated, price coupling breaks down, forcing local prices to absorb the entire shock, which manifests mathematically as the explosive variance observed in our standard GARCH frameworks.
Future research should employ a Markov-Switching GARCH (MS-GARCH) model to resolve this issue. Unlike standard frameworks that force a single parameter space across the entire dataset, MS-GARCH allows the parameters to switch discretely based on an unobserved state variable governed by a transition matrix. This would isolate the stable baseline variance from the explosive vertical spikes, eliminating the parameter boundary inflation observed in our linear threshold models and providing stable, state dependent forecasting horizons.

6 Conclusion

This thesis has investigated the empirical and dynamic relationships between physical production forecast errors of Variable Renewable Energy Sources (VRES) and the financial clearing dynamics of the electricity market. Specifically, we analyzed how day-ahead VRES forecast errors (v_t) propagate into both the level and conditional volatility of the price spread (u_t), defined as the difference between the day-ahead spot price (S_t^{spot}) and the real-time imbalance settlement price (S_t^{imb}). By evaluating these interactions through a series of increasingly sophisticated econometric frameworks, this project has answered the two primary research questions established in the problem statement.

1. Long-term Equilibrium and Market Integration
Regarding the first research question, the empirical evidence did not support modelling either the price pair (S_t^{spot},S_t^{imb}) or the production pair ($\hat{Q}_t$,Q_t) through a cointegrating framework. Although the unit root diagnostics produced mixed results, Johansen trace tests were conducted as a robustness check and indicated full-rank systems in both cases. These findings suggest that the variables are more appropriately treated as stationary processes rather than non-stationary series connected through a common stochastic trend. Consequently, no error-correction mechanism was required for the subsequent analysis. To evaluate market related dynamics directly, linear relationships between the spot and imbalance prices and between actual and forecast production were estimated using HAC-robust inference. The joint Wald hypothesis of perfect related ($\alpha = 0$,$\beta = 1$) was rejected for both systems, indicating the presence of systematic deviations between expected and realized market outcomes. In the case of prices, this finding suggests that the balancing market does not merely replicate day-ahead price formation but instead reflects additional costs and constraints associated with real-time system balancing. The subsequent spread model identified a statistically significant relationship between VRES forecast errors and the price spread. However, the explanatory power of the static specification was limited, while residual diagnostics revealed non-normality, heavy tails, and volatility clustering. These results motivated the transition toward conditional volatility models capable of capturing the dynamic risk structure of electricity markets.

2. Error Propagation and Volatility Architecture
Regarding the second research question, the findings demonstrate that forecast errors affect not only the level of the price spread but also its conditional volatility, with the magnitude and persistence of these effects depending strongly on the volatility specification employed.

The empirical findings revealed sharp performance trade-offs among the architectures:

- **GARCH(1,1):** Suffered from structural instability. Because it forces a symmetric, quadratic response to shocks, the parameter estimates were non-stationary ($\alpha_1+\beta_1 > 1$), leading to explosive variance forecasts out-of-sample.
- **TGARCH(1,1):** Failed to meet covariance stationarity under the sstd specification and underestimated market risk during out-of-sample evaluation, yielding a subpar interval coverage rate of 82.8%.

- **EGARCH(1,1):** Emerged as the best of the three frameworks for practical deployment. By modeling log-variance ($\ln \sigma_t^2$), it bypassed non-negativity constraints, maintained covariance stationarity, and successfully isolated asymmetric leverage effects. Out-of-sample, the EGARCH(1,1) baseline generates confidence bands that achieved a 96.4% empirical coverage rate against a 95% target.

Crucially, the role of the physical forecast error (v_t) as an exogenous volatility driver was found to be architecture dependent. While it was statistically significant in both the conditional mean and variance processes for the complex in-sample EGARCH(2,6) model, the econometric principle of parsimony ultimately dictated the rejection of higher-order models due to out-of-sample overfitting, wider confidence intervals, and elevated predictive errors (MAE and RMSE).

Seasonality and Practical Implications
Seasonal analysis uncovered a compelling physical-financial paradox: while winter exhibits the highest average VRES generation and absolute forecast errors, it paradoxically features the lowest price spread volatility. This phenomenon can be explained by the presence of larger thermal baseload capacities and online spinning reserves during winter, which act as a physical buffer against stochastic VRES shocks.

The results suggest that BRPs may benefit from incorporating asymmetric volatility estimates derived from the EGARCH framework into risk management, hedging, and scheduling decisions. For the TSO, these volatility metrics may contribute to more informed ancillary service procurement and reserve planning.

Overall, the results demonstrate that renewable generation forecast errors influence electricity balancing markets through both price formation and risk transmission mechanisms. While their impact on the price spread is statistically detectable, their effect on market risk is fundamentally non-linear and driven by asymmetric volatility dynamics. Future research could extend this framework by incorporating regime-switching models, such as MS-GARCH, to capture structural shifts associated with transmission congestion, reserve scarcity, and other system-level constraints.

7 Bibliography

[1] Eicke, A., Ruhnau, O., and Hirth, L. (2021). Electricity balancing as a market equilibrium: An instrument-based estimation of supply and demand for imbalance energy. *Energy Economics*, 102:105455.

[2] Energinet (2024a). afrr-energiaktiveringsmarked. `https://energinet.dk/el/balancering-og-systemydelser/markeder-og-udbud/afrr-energiaktiveringsmarked/`. Accessed: 2024-05-22.

[3] Energinet (2024b). afrr-kapacitetsmarked. `https://energinet.dk/el/balancering-og-systemydelser/markeder-og-udbud/afrr-kapacitetsmarked/`. Accessed: 2024-05-22.

[4] Energinet (2024c). Fcr (frequency containment reserve). `https://energinet.dk/el/balancering-og-systemydelser/markeder-og-udbud/fcr/`. Accessed: 2024-05-22.

[5] Energinet (2024d). Introduktion til systemydelser. Technical report, Energinet.

[6] Energinet (2024e). mfrr-energiaktiveringsmarkedet. `https://energinet.dk/el/balancering-og-systemydelser/markeder-og-udbud/mfrr-energiaktiveringsmarkedet/`. Accessed: 2024-05-22.

[7] Energinet (2024f). mfrr-kapacitetsmarkedet. `https://energinet.dk/el/balancering-og-systemydelser/markeder-og-udbud/mfrr-kapacitetsmarkedet/`. Accessed: 2024-05-22.

[8] Energinet (2025a). Electricity production 5-min realtime. `https://www.energidataservice.dk/tso-electricity/ElectricityProdex5MinRealtime`. Accessed: 2025-03-09. Contains data used pursuant to 'Conditions for use of Danish public-sector data' from the Energi Data Service portal.

[9] Energinet (2025b). Forecasts 5-min (wind and solar power production). `https://www.energidataservice.dk/tso-electricity/Forecasts_5Min`. Accessed: 2025-03-09. Contains data used pursuant to 'Conditions for use of Danish public-sector data' from the Energi Data Service portal.

[10] Energinet (2025c). Imbalance prices (from 2025-03-04 13:00). `https://www.energidataservice.dk/tso-electricity/ImbalancePrice`. Accessed: 2025-03-09. Contains data used pursuant to 'Conditions for use of Danish public-sector data' from the Energi Data Service portal.

[11] Energinet (2025d). Regulating balance power data (prior to 2025-03-04 13:00). `https://www.energidataservice.dk/tso-electricity/RegulatingBalancePowerdata`. Accessed: 2025-03-09. Contains data used pursuant to 'Conditions for use of Danish public-sector data' from the Energi Data Service portal.

[12] Energinet (n.d.). Energinet position on full cost balancing. Accessed: 2026-05-22.

[13] Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

[14] Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

[15] Forsyningstilsynet (2024). Kerneopgaver og regulering af energimarkedet. `https://forsyningstilsynet.dk/om-os/kerneopgaver`. BesÃÿgt: 23. februar 2026.

[16] Ghalanos, A. (2022). *Introduction to the rugarch package (Version 1.4-9)*. R package vignette.

[17] HjÃÿrring Varmeforsyning (2024). Typisk fordeling af forbrug pr. mÃěned. `https://www.hjvarme.dk/alt-om-varmeregningen/typisk-fordeling-af-forbrug-pr-maaned/`. BesÃÿgt: 20. maj 2024.

[18] Johansen, S. (2004). Cointegration: An overview. *Department of Applied Mathematics and Statistics, University of Copenhagen*. Preprint/Working Paper.

[19] Mühlenpfordt, J., Sioshansi, R., and Wozabal, D. (2022). Intermittency or uncertainty? impacts of renewable energy in electricity markets. CESifo Working Paper 9902, CESifo.

[20] Next Kraftwerke (2024). Liberalization of energy markets: Background and developments. `https://www.next-kraftwerke.com/knowledge/liberalization-energy-markets`. BesÃÿgt: 23. februar 2026.

[21] Nord Pool (n.d.). How is the electricity price calculated? Accessed: 2026-05-22.

[22] Open Electricity Economics (n.d.). Electricity markets. Accessed: 2026-05-22.

[23] RÃijdiger, M. (2024). Liberaliseringen af den danske el- og gassektor. `https://danmarkshistorien.lex.dk/Liberaliseringen_af_den_danske_el`. BesÃÿgt: 23. februar 2026.

[24] TenneT (n.d.). Balancing markets. Accessed: 2026-05-22.

[25] Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons, Hoboken, New Jersey, 3rd edition. Available at: https://blog.nus.edu.sg/dist/0/6796/files/2017/03/analysis-of-financial-time-series-copy-2ffgm3v.pdf.

[26] van der Veen, R. A. and Hakvoort, R. A. (2016). The electricity balancing market: Exploring the design challenge. *Utilities Policy*, 43:186–194.

A ACF and PACF plots

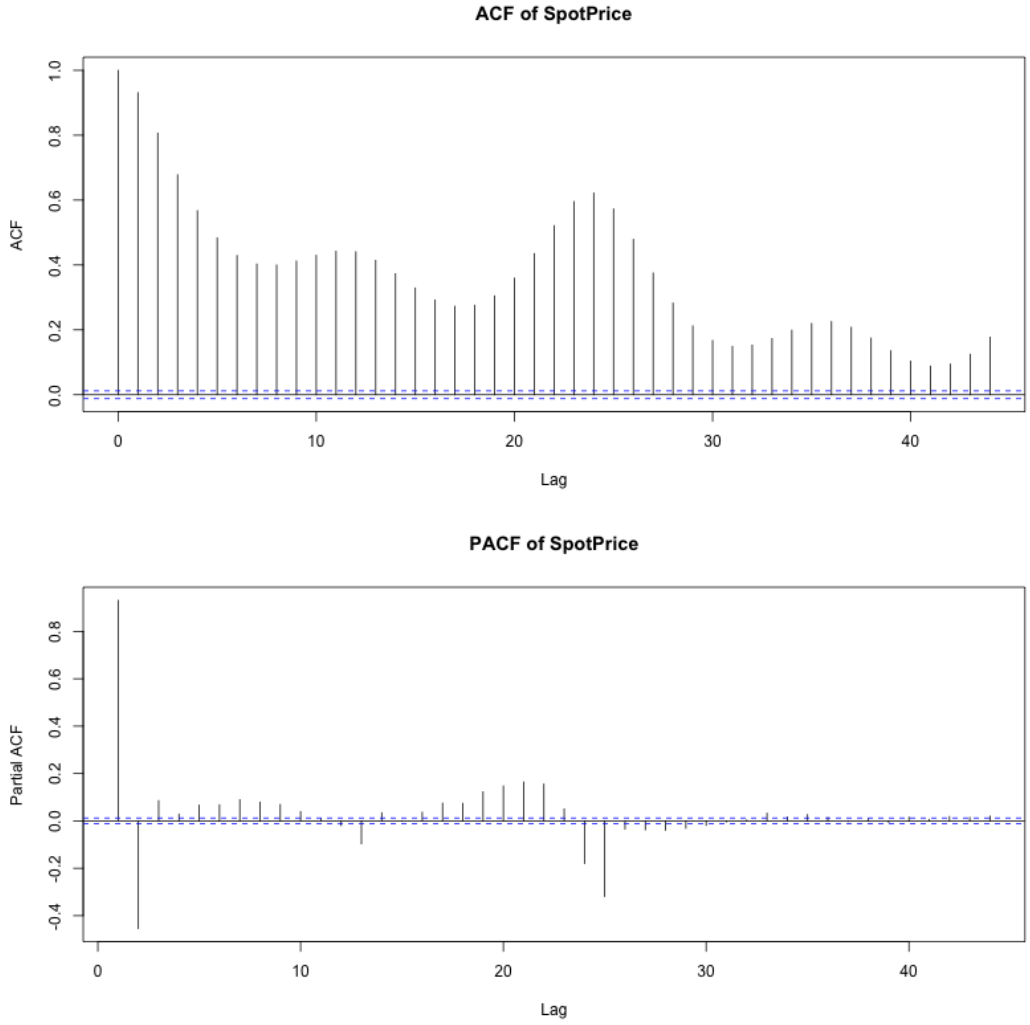


Figure A.1: ACF&PACF Spot Price

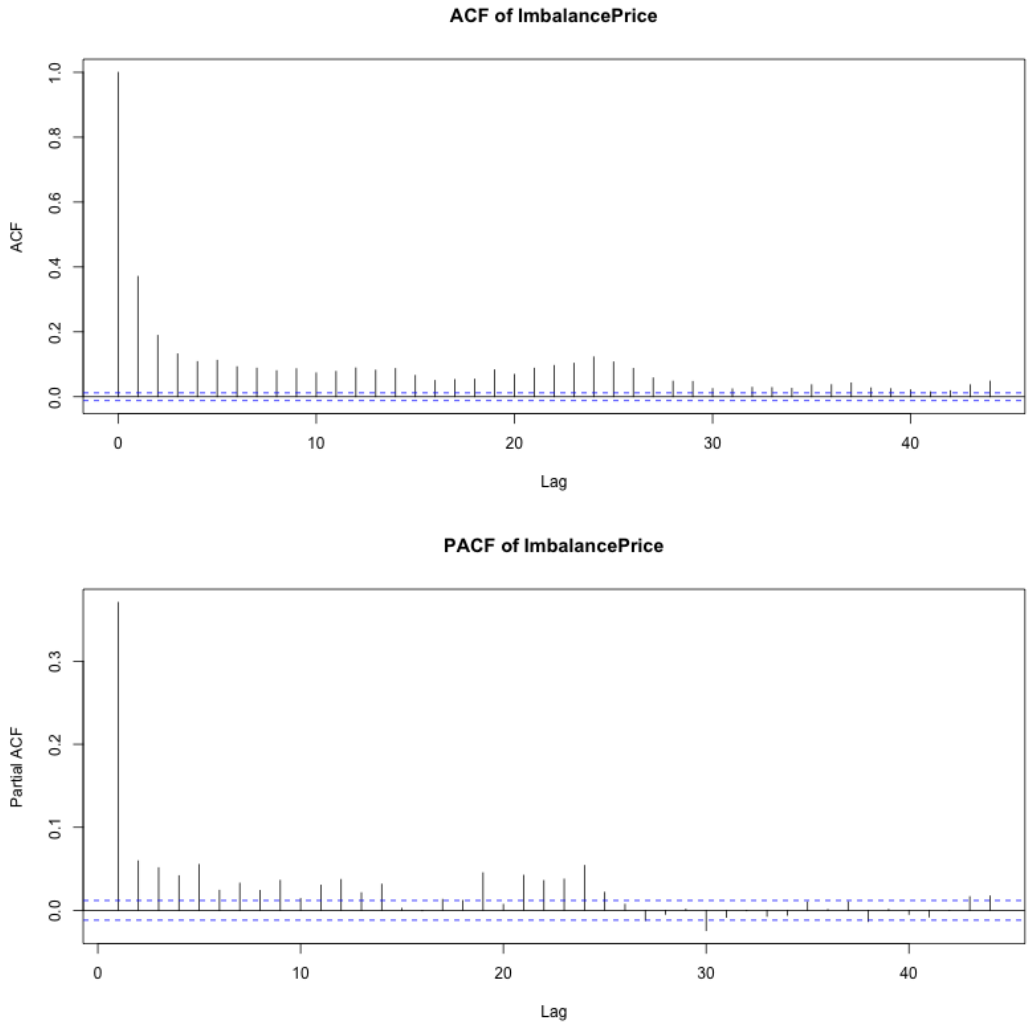


Figure A.2: ACF&PACF Imbalance Price

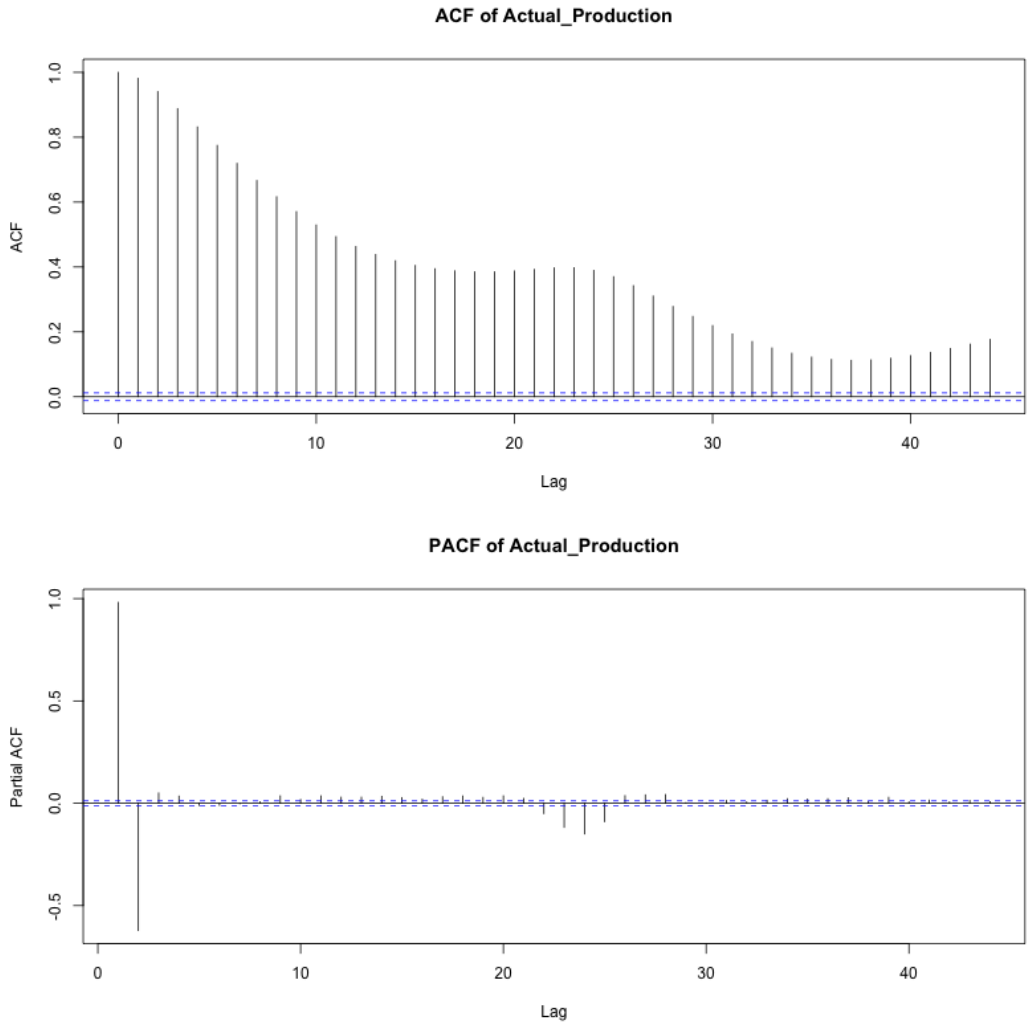


Figure A.3: ACF&PACF Actual Production

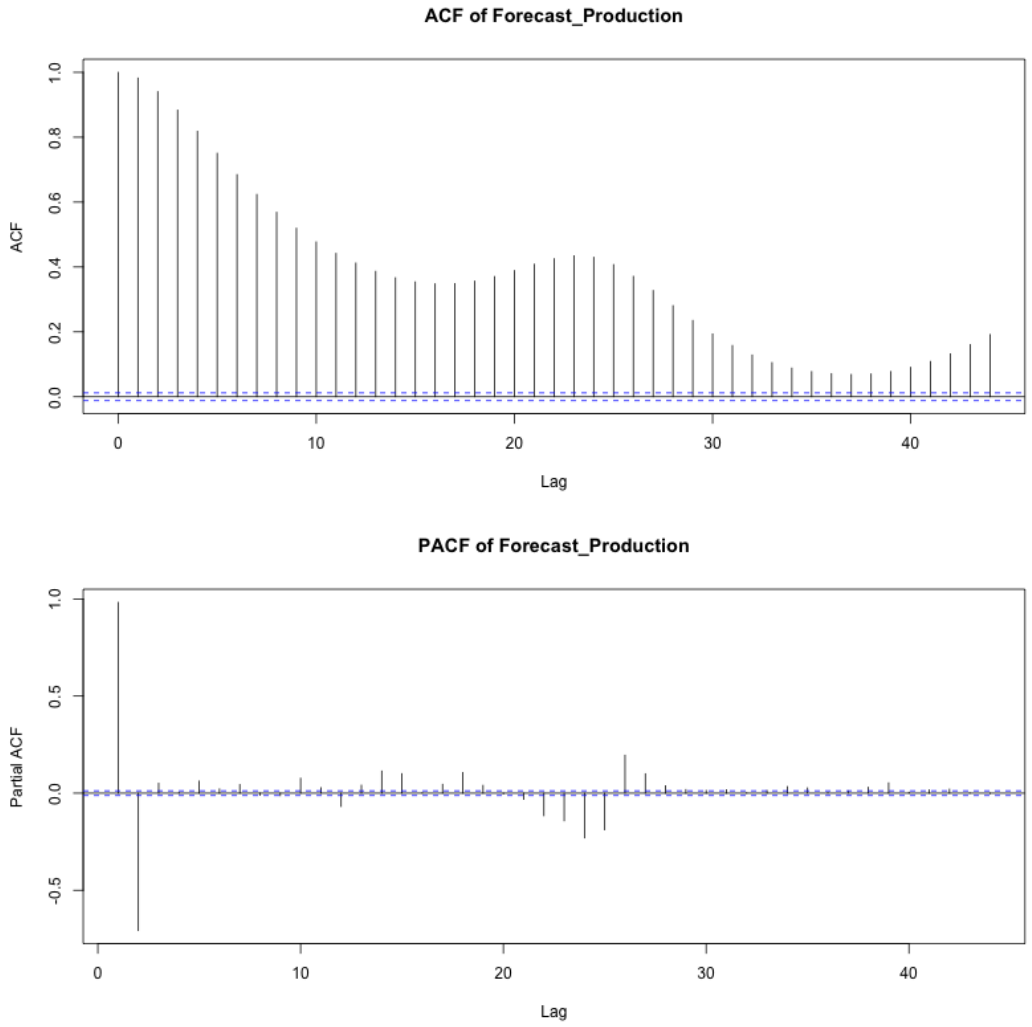


Figure A.4: ACF&PACF Forecast Production

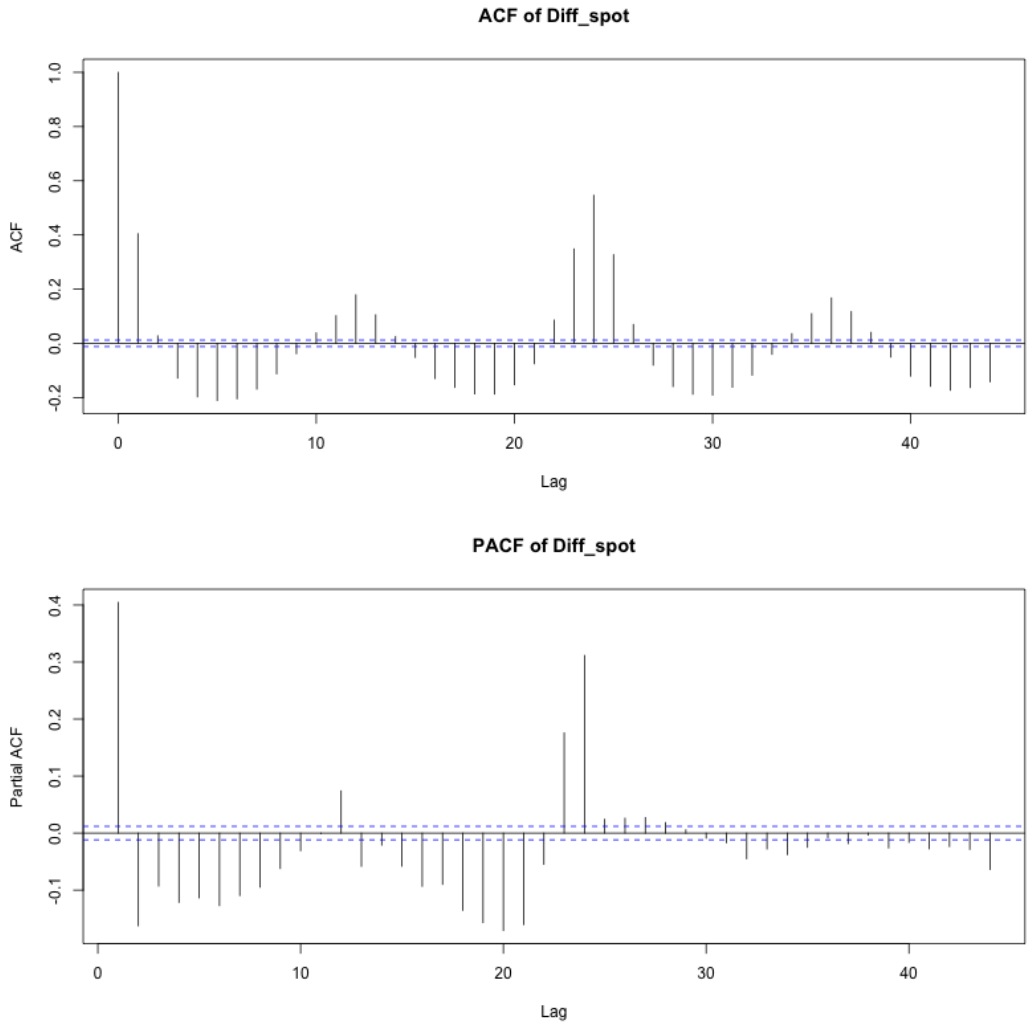


Figure A.5: I(1) ACF&PACF Difference Spot Price

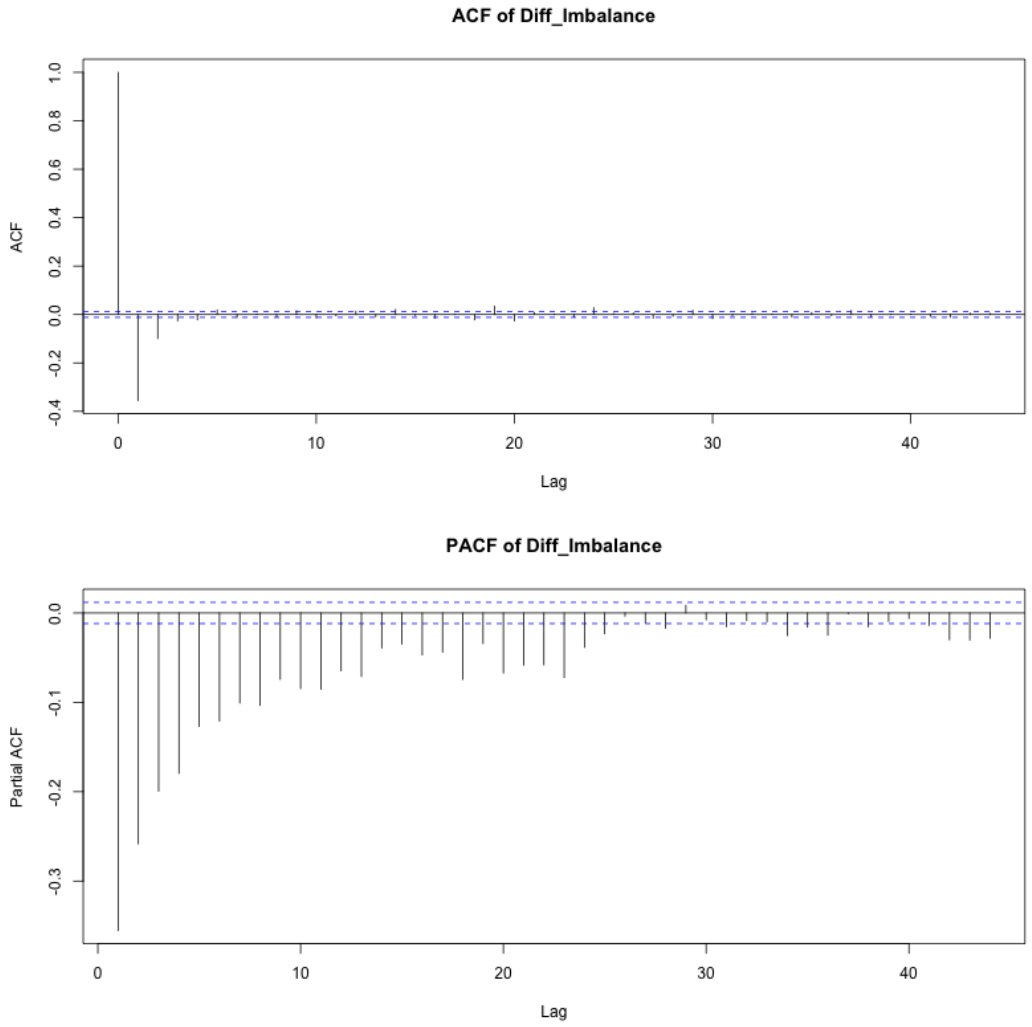


Figure A.6: I(1) ACF&PACF Difference Imbalance Price

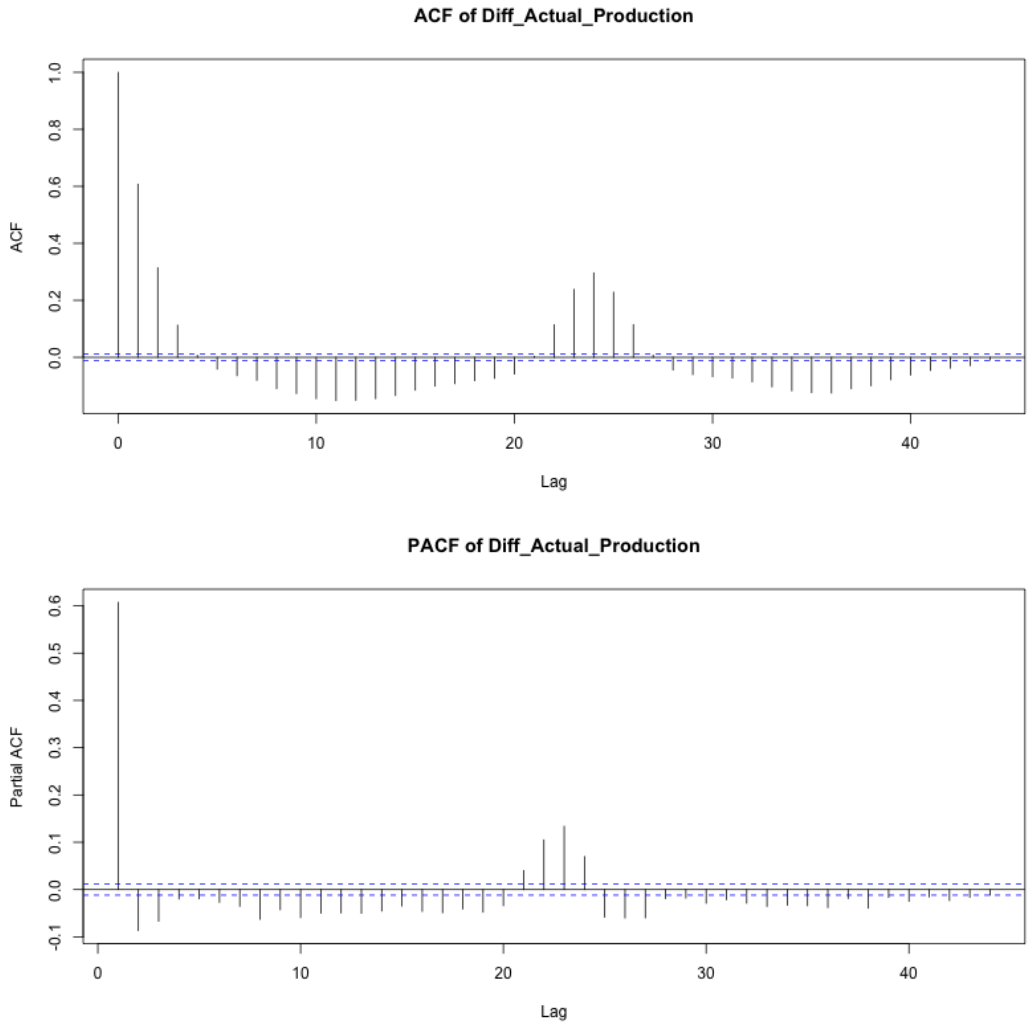


Figure A.7: I(1) ACF&PACF Difference Actual Production

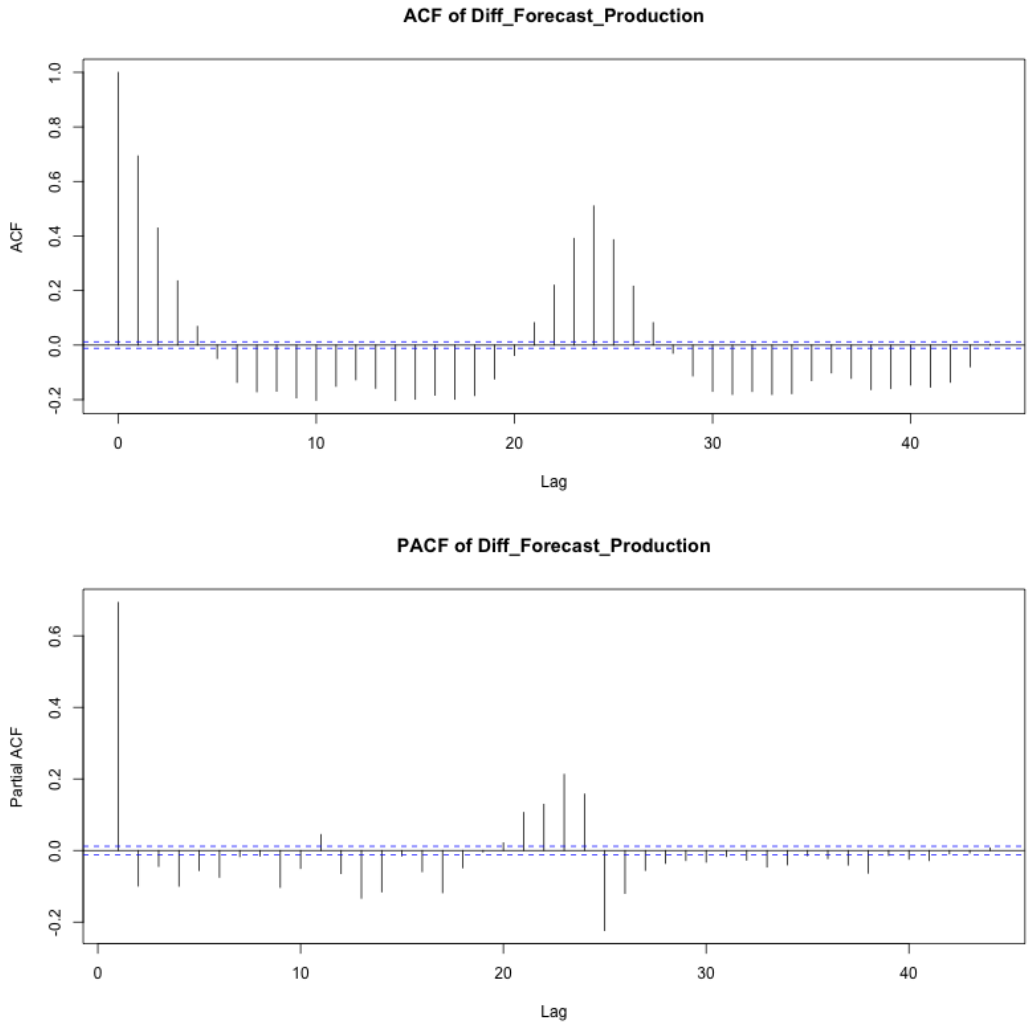


Figure A.8: I(1) ACF&PACF Difference Forecast Production

B Unscaled 14 day Forecast

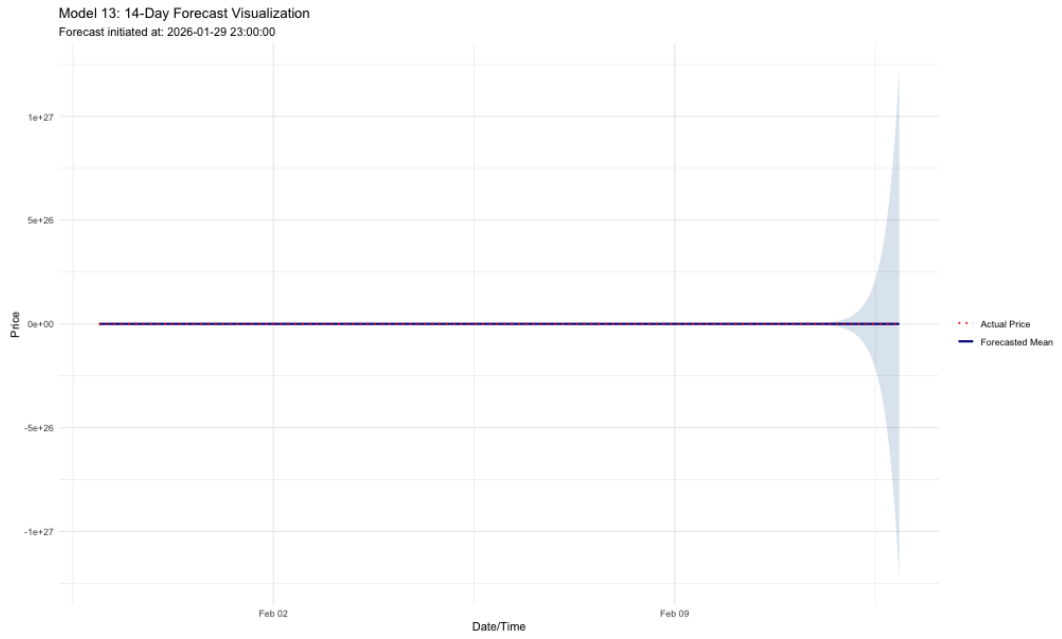


Figure B.1: GARCH(1,1) not scaled

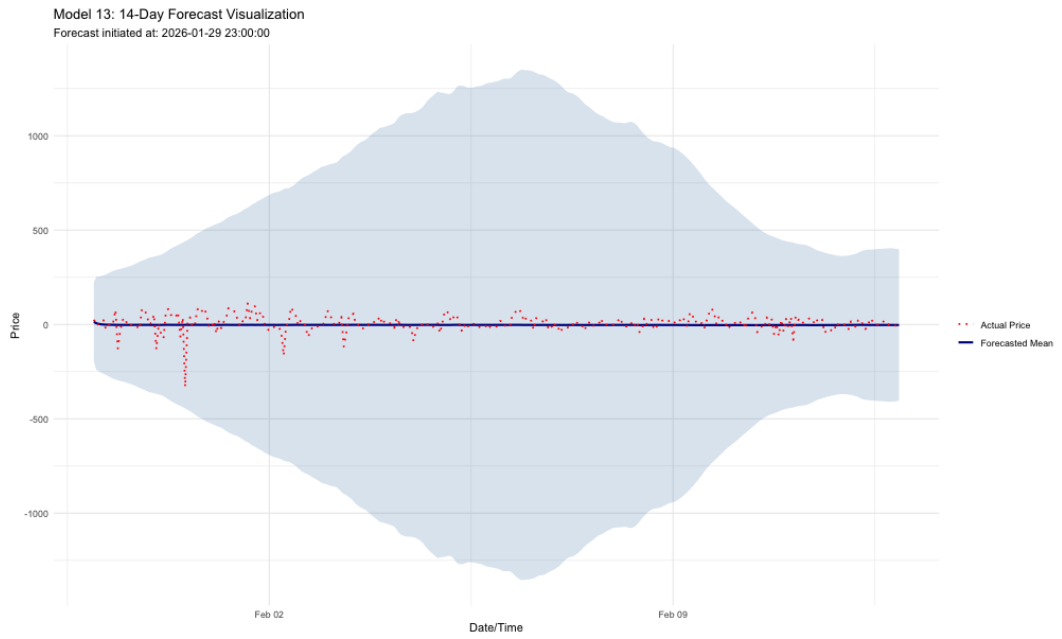


Figure B.2: EGARCH(1,1) not scaled

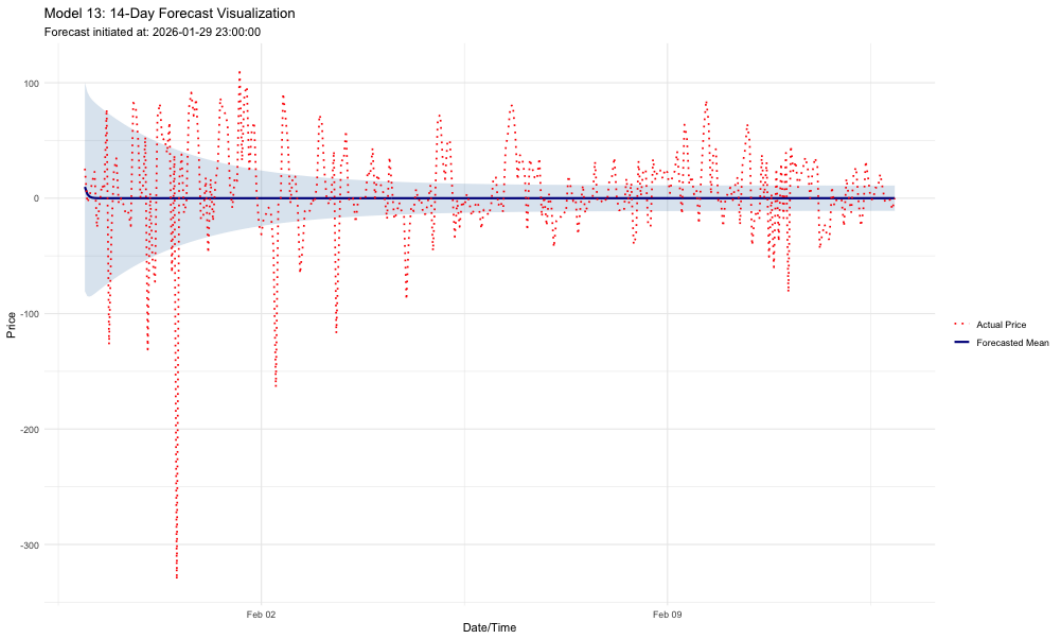


Figure B.3: TGARCH(1,1) not scaled

C Volatility Models Fitted with Normal Distribution

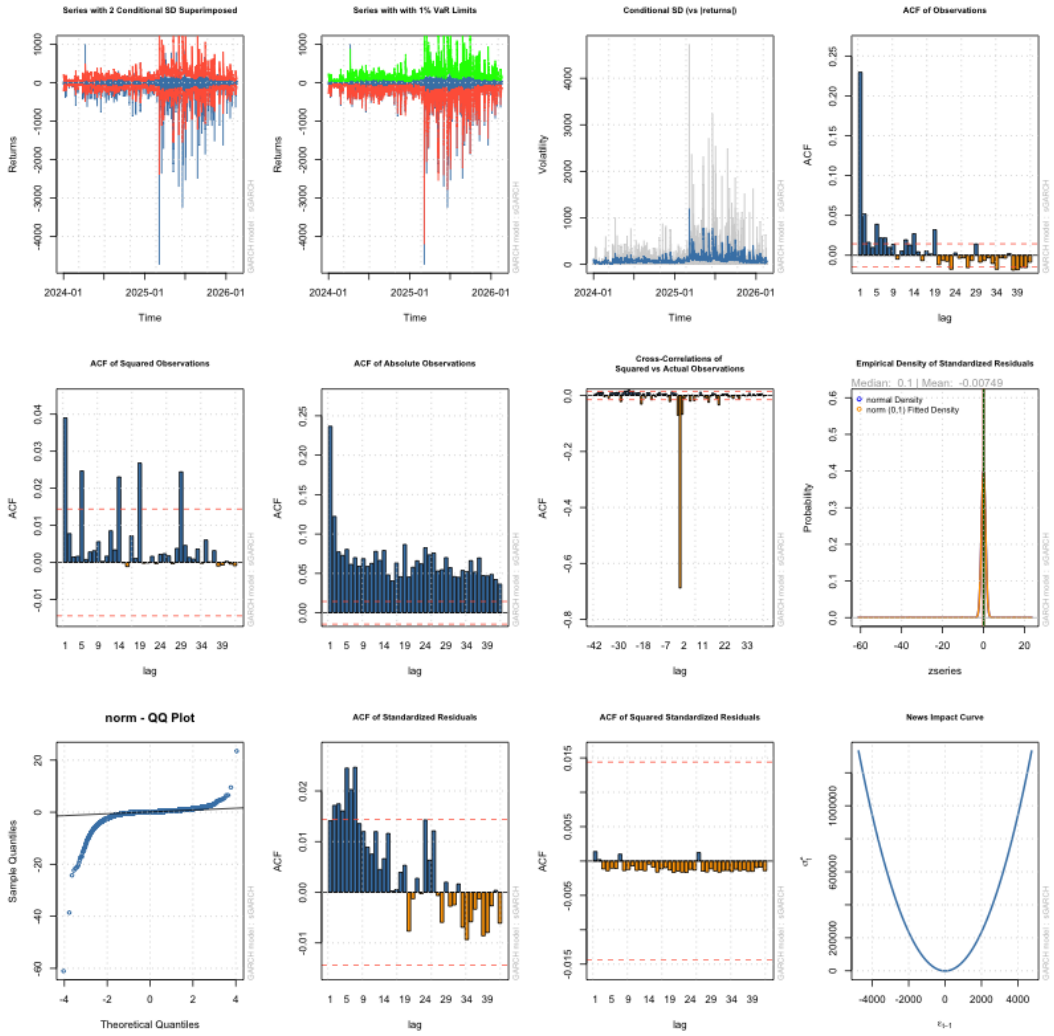


Figure C.1: GARCH(1,1) Model Fitted using a Normal Distribution

The AIC value GARCH(1,1) specification fitted utilizing a normal distribution is 11.724.

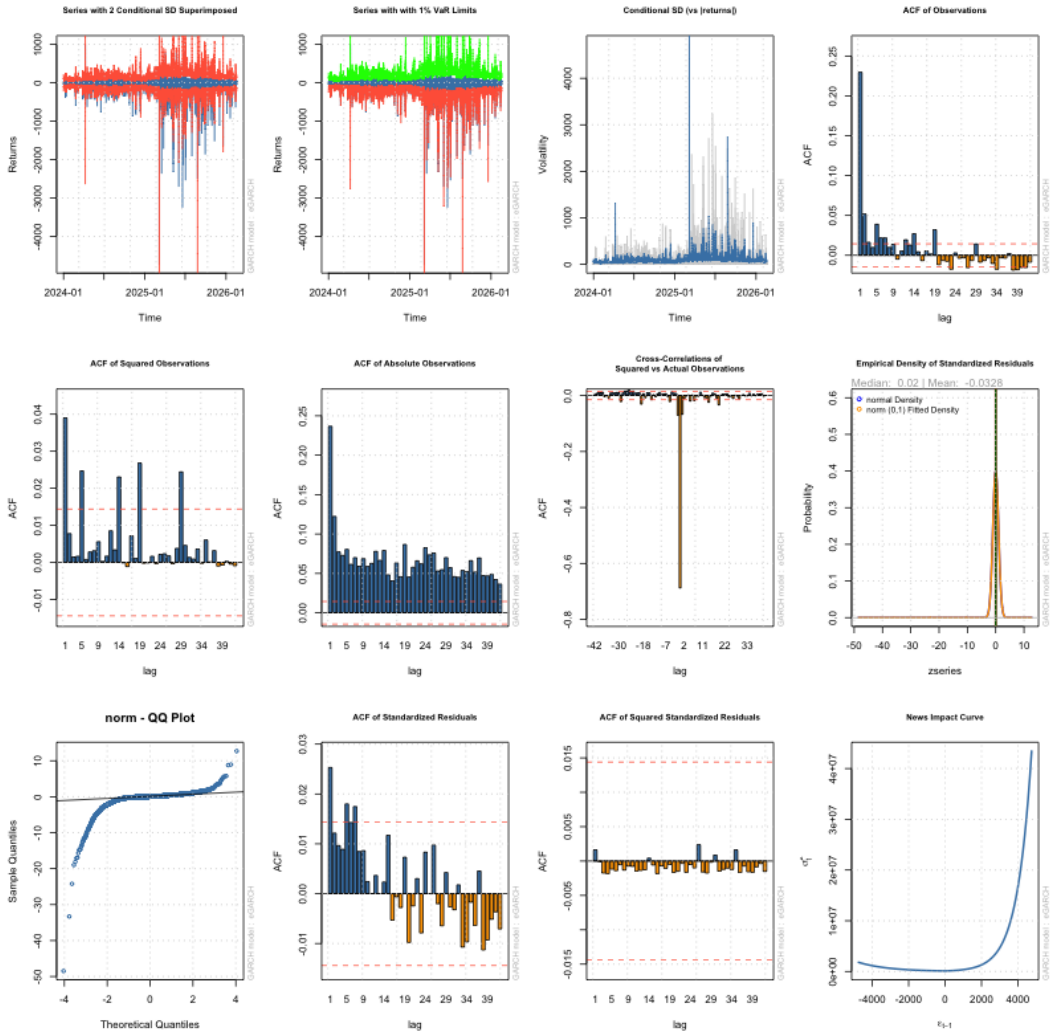


Figure C.2: EGARCH(1,1) Model Fitted using a Normal Distribution

The AIC value EGARCH(1,1) specification fitted utilizing a normal distribution is 11.623.

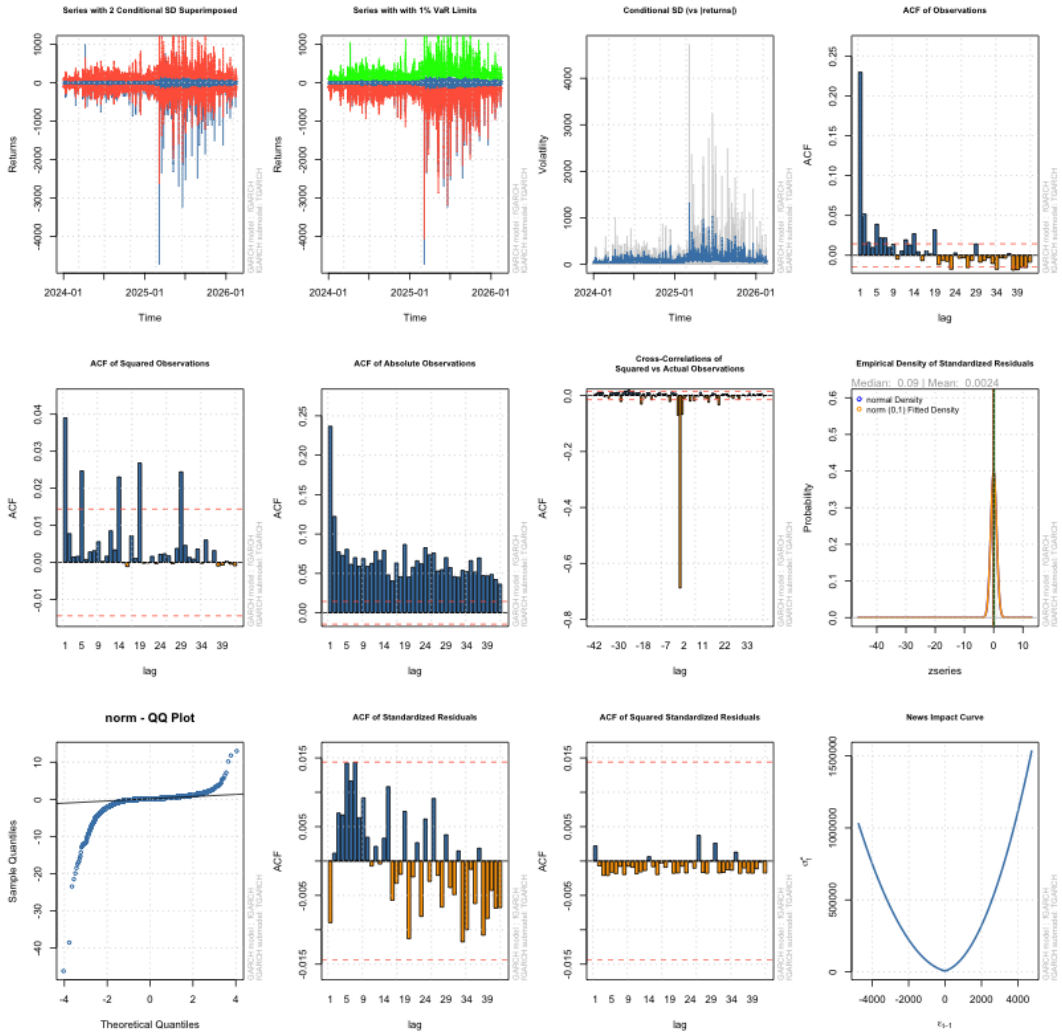


Figure C.3: TGARCH(1,1) Model Fitted using a Normal Distribution

The AIC value TGARCH(1,1) specification fitted utilizing a normal distribution is 11.558.

D TGARCH(5,4)

GARCH (p)	ARCH (q)	AIC	BIC	Persistence	Stationary	Ljung-Box p-val
4	1	9.8355	9.8422	1.0349	No	0.9186
4	2	9.8356	9.8428	1.0349	No	0.9187
4	3	9.8357	9.8433	1.0349	No	0.9187
5	1	9.8357	9.8433	1.0349	No	0.9187
4	4	9.8358	9.8438	1.0351	No	0.9189
5	2	9.8358	9.8438	1.0349	No	0.9187
6	1	9.8359	9.8444	1.0349	No	0.9187
5	3	9.8359	9.8444	1.0349	No	0.9186
4	5	9.8359	9.8444	1.0349	No	0.9187
6	2	9.8360	9.8449	1.0349	No	0.9187
4	6	9.8360	9.8449	1.0349	No	0.9187
6	3	9.8361	9.8454	1.0350	No	0.9188
5	5	9.8361	9.8454	1.0355	No	0.9188
5	6	9.8362	9.8460	1.0349	No	0.9187
6	4	9.8362	9.8460	1.0349	No	0.9187
6	5	9.8364	9.8465	1.0349	No	0.9187
6	6	9.8365	9.8470	1.0283	No	0.9179
5	**4**	**9.8372**	**9.8461**	**0.9725**	**Yes**	**0.9108**
1	1	9.8502	9.8545	0.9798	Yes	0.8755
1	2	9.8503	9.8550	0.9798	Yes	0.8755
2	1	9.8504	9.8555	0.9844	Yes	0.8762
1	3	9.8504	9.8555	0.9798	Yes	0.8755
1	4	9.8505	9.8560	0.9798	Yes	0.8755
2	2	9.8505	9.8560	0.9821	Yes	0.8758
2	3	9.8506	9.8566	0.9867	Yes	0.8766
1	5	9.8506	9.8566	0.9798	Yes	0.8755
3	1	9.8506	9.8566	0.9822	Yes	0.8759
2	4	9.8507	9.8570	0.9846	Yes	0.8762
1	6	9.8508	9.8571	0.9798	Yes	0.8755
3	2	9.8508	9.8571	0.9867	Yes	0.8766
2	5	9.8509	9.8576	0.9825	Yes	0.8759
3	3	9.8509	9.8576	0.9827	Yes	0.8759
3	4	9.8509	9.8581	0.9854	Yes	0.8763
2	6	9.8510	9.8581	0.9820	Yes	0.8759
3	5	9.8511	9.8587	0.9840	Yes	0.8762
3	6	9.8512	9.8592	0.9803	Yes	0.8756

Specifications are ordered by ascending AIC. All models are evaluated using an underlying ARFIMA(1,0,0) mean process. † Indicates estimated parameters that hit the numerical upper bound limit (0.99999999) assigned by the optimizer.

Table D.1: Complete TGARCH Grid Search Results ($p, q \in [1,6]$)

Parameter	Estimate	Std. Error	t-value	Pr($>\|t\|$)
Mean Model				
μ	-0.005451	0.001949	-2.79769	0.005147**
$ar1$	0.380022	0.016906	22.47912	0.000000***
$mxreg1$	0.000012	0.000008	1.56649	0.117234
Variance Model				
ω	0.133497	0.043678	3.05640	0.002240**
α_1	1.000000	0.037146	26.92112	0.000000***
α_2	0.001284	0.000057	22.40808	0.000000***
α_3	0.000054	0.000002	34.55343	0.000000***
α_4	0.159656	0.027341	5.83946	0.000000***
α_5	0.000002	0.000000	22.34889	0.000000***
β_1	0.445537	0.046365	9.60939	0.000000***
β_2	0.000000	0.069117	0.00000	0.999998
β_3	0.000000	0.063430	0.00000	0.999999
β_4	0.000000	0.033565	0.00000	0.999999
η_{11}	-0.104138	0.019341	-5.38431	0.000000***
η_{12}	-0.998906	0.027183	-36.74740	0.000000***
η_{13}	-0.998609	0.020582	-48.51771	0.000000***
η_{14}	-0.999993	0.186845	-5.35198	0.000000***
η_{15}	-0.998204	0.016709	-59.74180	0.000000***
$vxreg1$	0.000000	0.000135	0.00000	1.000000
Distribution Parameters				
Skew	0.863847	0.009288	93.00969	0.000000***
Shape (ν)	2.295894	0.012036	190.74656	0.000000***

Table D.2: Full Parameter Estimates for the AR(1)-TGARCH(5,4) Model

Diagnostic Tests	Statistic	p-value
Log-Likelihood	-91199.53	
AIC	9.8372	
BIC	9.8461	
Ljung-Box (Resid.) $Q(1)$	0.01255	0.9108
Ljung-Box (Sq. Resid.) $Q(1)$	0.04769	0.82714
Sign Bias	1.3630	0.172960
Negative Sign Bias	0.9240	0.355498
Positive Sign Bias	2.7990	0.005129***
Sign Bias (Joint Effect)	9.5470	0.022840**

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Robust standard errors applied.

Table D.3: Results for the TGARCH(5,4) Model with ARFIMA(1,0,0) Mean Specification

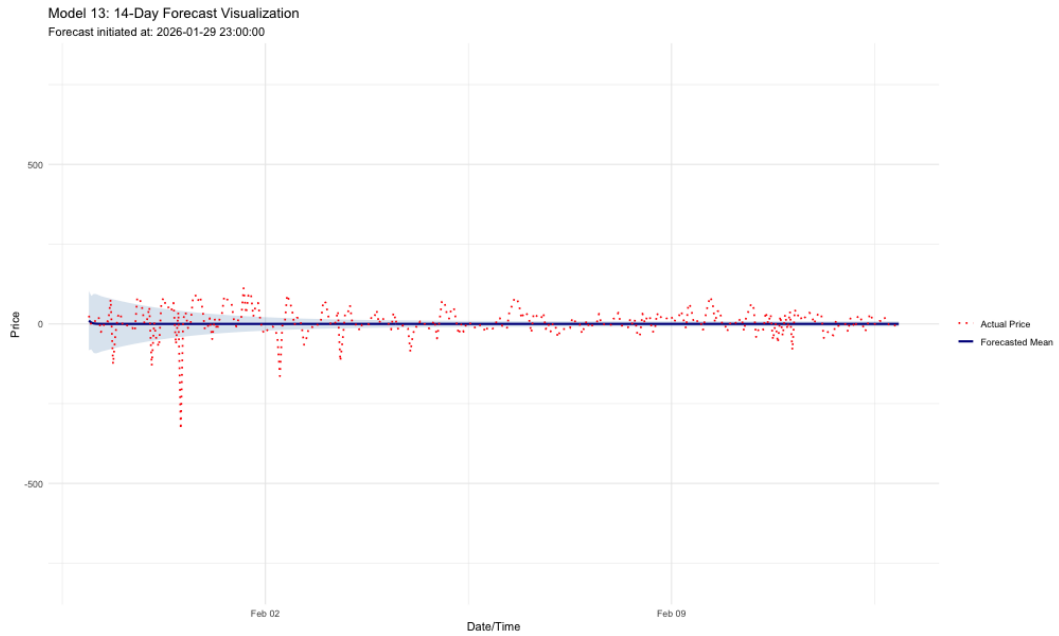


Figure D.1: TGARCH(5,4) 14 day Forecast

E Link to Code

Below is a link to my github, where the code for the project can be found:
https://github.com/Marcusbp1/10.semester_project