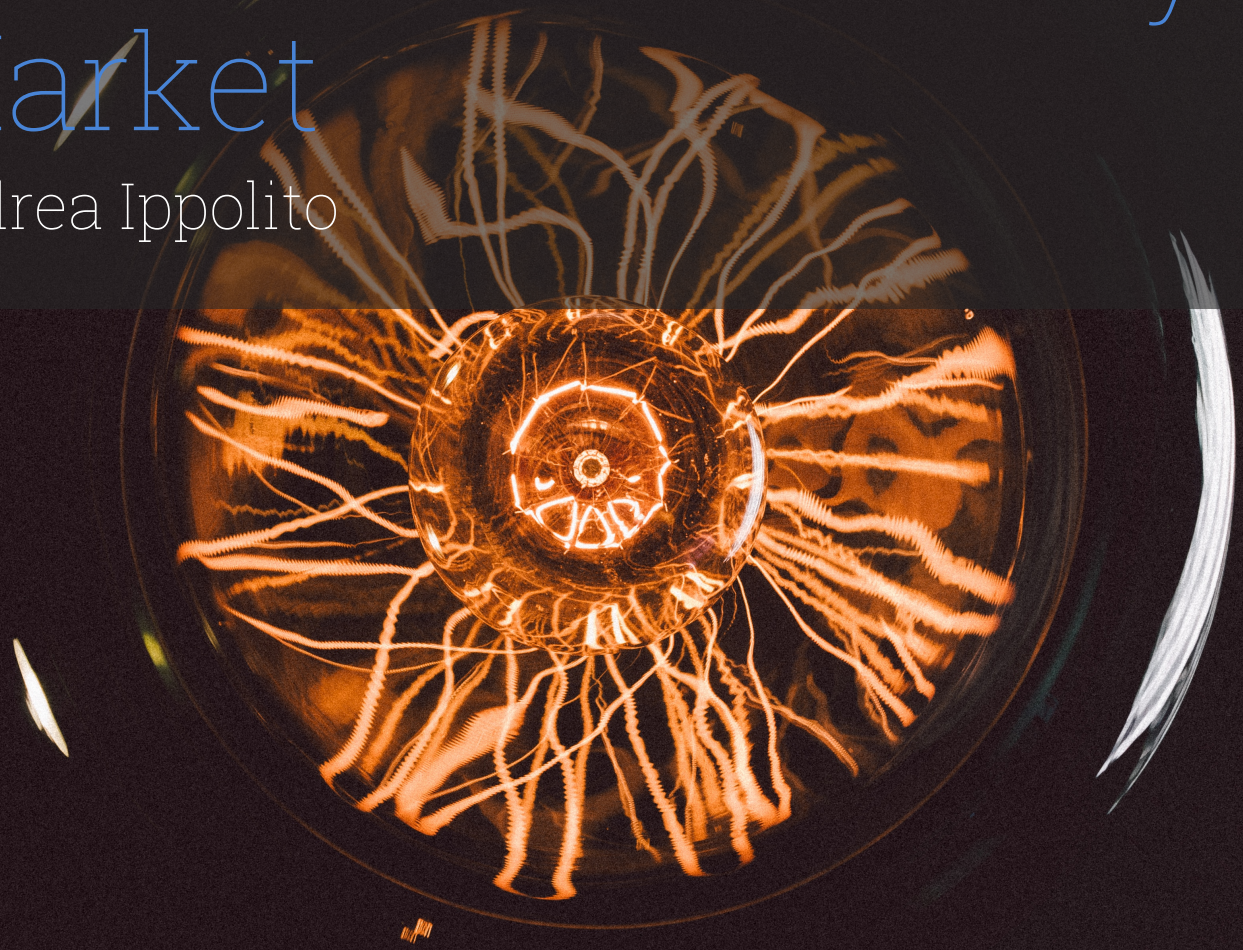
Stochastic Process Model for Energy Prices in the Intraday Market

Andrea Ippolito



Delft University of Technology

TU Delft

Stochastic Process Model for Energy Prices in the Intraday Market

by

Andrea Ippolito

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Wednesday, July 5, 2023.

Thesis Committee:	Prof. Dr. A. P. Papapantoleon	TU Delft	Supervisor
	A. Stawska	Priogen Energy B.V.	Supervisor
	Dr. N. Parolya	TU Delft	
Faculty:	Applied Mathematics		
Student Number:	5147547		
Duration of the Project:	June 1st 2022- July 5th 2023.		

TU Delft

Abstract

The increasing number of Renewable Sources (RES) in the European electric grid has resulted in the necessity for producers to adjust their position with respect to the change in weather forecasting. Therefore, the European Power Exchange (EPEX SPOT) has seen an expansion of the Intraday Market (ID) where market participants can trade up until a few minutes before delivery. However, literature in Electricity Price Forecasting (EPF) primarily concerns the Day-ahead market (DA).

The goal of this thesis is to develop a model to predict the prices of hourly contracts in the ID market using stochastic processes. This would facilitate the obtainment of a probability density function (pdf) for the prices of the contracts at any time during the day. This thesis focuses on the German Energy market which is the most liquid among the European countries.

I then propose 4 different Levy processes for the forecasting of the process: a simple Brownian Motion, a Jump Diffusion Process, a Normal Inverse Gaussian (NIG), and a Generalised Hyperbolic model (GH). The Normal Inverse Gaussian model is selected according to the Akaike Information Criterion and the Bayesian Information Criterion.

In order to integrate other variables into the model, I then conducted a fundamental analysis. I find a low correlation between the volume traded during the ID phase and the change in Wind, solar, and consumption forecasting. Moreover, the volatility of the contract shows a low correlation with the changes in forecasting and the traded volume.

Additionally, I conduct an augmented Dickey-Fueller test and a Mann-Kendall test on the price of the contracts to verify the presence of either mean reversion or drift. Results show that the latter cannot be refused for most days.

Finally, alongside the NIG model (noise model) I propose one model including a drift component (noise-drift model) and a second model adding both a drift and a volatility component (noise-drift-volatility model). The three models are compared in the accuracy of forecasting the tail probabilities of the ID_1 and ID_3 indexes. It is found that the noise model and the noise-drift model cast a better prediction than the noise-drift-volatility model. This suggests that the volatility during the first part of a day of trading is a poor predictor of volatility in the hours close to delivery.

Acknowledgments

I want to thank in the first place Priogen Energy for giving me the opportunity to conduct my research. In particular, I would like to thank Anna Stawska for her constant support, guidance and help during my days there.
I would like also to thank Professor Papapantoleon for his feedback and insights.
Last but not least I would like to thank my family for their support, my cousin Davide for his care, and my friend Zoe for the long talks and good advice.

Andrea Ippolito
Delft, June 2023

Contents

1 Introduction

Energy Market in Europe has been liberalised since the 1990s, allowing multiple independent buyers and sellers in the energy market. The transfer of electric energy happens on a grid connecting producers and consumers, and it is designed to work at 50 Hz in Europe. To maintain this frequency, the production of electric energy must always equal the consumption. The balance is achieved through the interplay of the Balance Responsible Parties (BRPs) and the Transmission System Operator (TSO). BRPs represent producers and consumers whose role is to maintain a balanced position: producers need to sell the energy that will be produced, and consumers need to buy energy to compensate for the consumption. Instead, the TSO is the authority responsible for balancing the grid and, consequently, the preservation of the frequency. It achieves so by, for example, creating incentives for the BRPs to maintain a balanced position or providing for reserves in case of imbalances.
The exchange of contracts can take place during three phases, the future market, in which over-the-counter (OTC) contracts are exchanged; the Day-ahead (DA) market, which closes the day before delivery and the Intraday Market (ID) opens right after the closure of the DA market and closes a few minutes before the moment of delivery.
To prevent Global warming and climate change, government and political institutions have invested in and incentivised the production of electrical energy through renewable sources (RES), particularly wind and solar power. The forecasting of energy production depends on weather forecasting; wind turbine production depends on wind speed, and solar power depends on the sun's intensity. However, these factors cannot be controlled by humans and can change during the day, causing an amount of energy production different from the expected one. This difference then needs to be bought or sold by the producer before the moment of delivery, and for this reason, the share of energy trading in the short-term market, i.e. ID, has become more relevant.
As in every other market, operating under fewer uncertainties allows for better resource management and achieving this goal; accurate forecasting is paramount. Forecasting deals with many components in the energy market: wind and solar power, consumption, and prices [11]. Electricity Price Forecasting (EPF) is a discipline at the intersection of Electrical Engineering, Finance, Statistics and Applied Probability that deals with the forecasting of energy prices. Despite the increasing importance of the ID market, the literature covering EPF in the ID market is scarce, and more than 90% of this literature refers to the DA market [12], [26]. [15] and [17] investigated the fundamentals behind the price of 15-minute contracts in the ID market of Germany, finding the phenomena causing the change in prices, such as the change in Solar and Wind power forecasting, the change in the price of nearby contracts or the mean reverting effect, i.e. the trend of the price of the contract to return to its average. Other attempts to study the price formation in the ID market can be found in [22] and [10]. In [19], the authors use probabilistic methods to predict the price of hourly contracts in the last three hours of the German ID market, [25] use deep learning techniques to predict the price of the contracts.
This paper aims to expand the literature on Energy Price Forecasting in the Intraday market utilising Stochastic Processes. Stochastic processes are mathematical objects emerging from Probability and are used to model a random behaviour in time. In other words, a stochastic process represents a variable whose value in the future is described by a set of trajectories. The values that the trajectories assume at a future time t are distributed according to a Probability Distribution, which depends on a

set of parameters defined at present time s. Parameters are calibrated to capture different behaviours and phenomena. For example, a Brownian Motion $\mathbb{W}_\cdot$ starting at $s = 0$ with initial value $\mathbb{W}_s = 0$ is a Stochastic process normally distributed at time t. The parameters of a Brownian Motion determine the mean μt and the variance ($\sigma^2 t$) for every given time. Brownian Motion is used for describing the log-return of Stocks in the Black and Scholes models, and higher volatility (σ) means that the price of a Stock change very quickly. Another kind of process is the Ornstein-Uhlenbeck process, and it is used to model the mean reversion effect of a process, that is, its tendency to move back to its average. The parameters of this process are the speed of mean reversion and the average of the process. The choice of parameters, then, reflects a possible scenario, and different sets of parameters can reflect different sets of hypotheses on the future behaviour of the process, returning for each one a probability distribution. This comes useful, for example, when a trader needs to include her personal beliefs in the market. If she expects the price of a Stock to rise, she will choose a positive μ for the Brownian Motion. Also, the stochastic process can include different phenomena like the aforementioned mean reversion, which can be found in the power market [17], [6].

The goal is to create a naive model that would predict the changes in prices only through the historical values of the contracts. This first model can then be expanded by adding the fundamentals governing the change in the price of a contract which can be found, for example, in [17] or [10]. The subject of study of this dissertation is the price of the 60-minute contracts exchanged in Germany, which represents the most traded and, therefore, liquid contract in Europe. The high liquidity means a larger amount of available data allowing a better analysis.

The dissertation will proceed as follows, in section 2, an introduction to stochastic processes is given; in area 3, there is an explanation of the EPEX Spot Market and the ID market, which is followed by the mathematical formulation of the problem. Finally, in section 5, the formulation and testing of the model are made.

2

Lévy Processes

The goal of this section is to give the mathematical background that is used in the following sections. The arrival point is the presentation of Lévy processes to the extent that is relevant for this thesis; an interested reader may find more information in [1], [3], or [24]. Lévy processes are a particular case of stochastic processes, so the latter is introduced.

2.1. Stochastic Processes and Martingales

Definition 2.1.1 *A stochastic process is a collection of random variables, indexed by some set $\mathbb{T}$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, assuming values on a measurable space S with some σ-algebra Σ. Usually S is $\mathbb{R}$ or $\mathbb{R}^n$ and Σ is the Borel σ-algebra, $\mathcal{B}(\mathbb{R}^n)$.*

Usually, the set $\mathbb{T}$ is either a subset of $\mathbb{R}^+$, and then the stochastic processes are called time-continuous, or $\mathbb{N}$, and then time-discrete. We now define the concept of filtration for a probability space.

Definition 2.1.2 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{F}_\cdot$ denote a sub-σ-algebra of $\mathcal{F}$ and T a set of indexes. Then $(\mathcal{F}_t)_{t\in T}$ is called a filtration if $\mathcal{F}_t \subseteq \mathcal{F}_s$, $\forall t \leq s, t, s \in T$. $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in T})$ is called a filtered probability space. If for all t holds that $\mathcal{F}_t = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$, then the filtration is right-continuous.*

Definition 2.1.3 *Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in T})$, the stochastic process $(X_t)_{t\in T}$ is called adapted to the filtration $(\mathcal{F}_t)_{t\in T}$ if X_t is $\mathcal{F}_t$-measurable $\forall t \in T$.*

In other words, a process is adapted if at any time it can be measured with the information available at that moment. For example, a stochastic process representing the price of a stock is adapted because the values it can assume at that time reflect only the information available up to that time.

Definition 2.1.4 *A Stochastic Process $Y : T\times\Omega \to S$ taking values in a Banach Space S is a martingale with respect to a filtration $\mathcal{F}_\cdot$ and probability measure $\mathbb{P}$ if*

- *$\mathcal{F}_\cdot$ is a filtration of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$*
- *Y is adapted to the filtration $\mathcal{F}_\cdot$, i.e., for each t in the index set T, the random variable Y_t is an $\mathcal{F}_t$ measurable function.*
- *for each t, Y_t lies in $L^1(\Omega, \mathcal{F}, \mathbb{P})$*
- *$Y_s = \mathbb{E}_\mathbb{P}(Y_t|\mathcal{F}_s)$ for all $s < t$*

This means that the expected value of a Martingale in the future, given all the information available at the present time, is its current value. A Stochastic Process can be either continuous or discontinuous in the same way as a function.

Definition 2.1.5 *A function $f : [0, T] \to \mathbb{R}$ is called continuous at a point $t \in [0, T]$ if its right limit and left limit, that is*

$$f(t+) = \lim_{s\to t, s>t} f(s), \quad f(t-) = \lim_{s\to t, s<t} f(s), \tag{2.1}$$

exist and coincide. Otherwise, it is called discontinuous. A function is called right continuous at t if $f(t) = f(t+)$ *and left continuous if* $f(t) = f(t-)$*. A function is called a càdlàg ("continue à droite, limite à gauche", meaning right continuous with left limit) function in t if the left limit exists and it is right continuous.*

A discontinuity in the path will be called a jump, and the stochastic processes will be càdlàg processes. This is a fact induced by the nature of prices: a jump can be only known only after it has happened and its value cannot be anticipated. See for example figure 2.1.

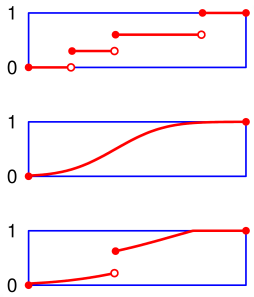


Figure 2.1: Examples of càdlàg functions.

Now, I will illustrate the most known and used example of a continuous stochastic process, the Brownian motion.

Definition 2.1.6 *A Brownian Process* $\{\mathbb{W}_t, t \in \mathbb{R}^+\}$ *is a Stochastic Process defined as follows:*

$$\begin{aligned} &\mathbb{W}_0 = 0 \\ &\mathbb{W} \textit{ has independent increments, i.e., } \mathbb{W}_t - \mathbb{W}_s \perp\!\!\!\perp \mathbb{W}_u \, \forall u \leq s \leq t \\ &\mathbb{W} \textit{ has Gaussian increments, i.e., } \mathbb{W}_t - \mathbb{W}_s \sim \mathcal{N}(0, t-s) \\ &\mathbb{W} \textit{ has continuous paths in time.} \end{aligned} \tag{2.2}$$

From this definition it directly follows that $\mathbb{W}_t \sim \mathcal{N}(0,t)$.
I now introduce a large class of stochastic processes that can allow

Definition 2.1.7 ***Lévy process*** *A càdlàg stochastic process* $(X_t)_{t\geq 0}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in* $\mathbb{R}^d$ *such that* $X_0 = 0$ *is called a Lévy process if it possesses the following properties:*

1. *Independent increments: for every increasing sequence of times* $t_0, ..., t_n$ *the random variables* $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_n}, X_{t_{n-1}}$ *are independent.*
2. *Stationary increments: the law of* $X_{t+h} - X_t$ *does not depend on t*
3. *Stochastic continuity* $\forall \epsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \epsilon) = 0$.

The third condition is imposed in order to exclude those processes where jumps happen at fixed moments in time.
The first property implies that if a Lévy process is sampled at regular intervals $0, \Delta, 2\Delta, ..., n\Delta$, then $X_\Delta, X_{2\Delta-\Delta}, ..., X_{n\Delta-(n-1)\Delta}$ are independent, and from property 2 it follows that they are distributed as X_Δ. By choosing $n\Delta = t$ then X_t can be represented as the sum of n $i.i.d.$ random variables following the distribution of X_{t_n}. A distribution having this property is said to be infinitely divisible.

Definition 2.1.8 *A probability distribution* F *is said to be infinitely divisible if, for any integer* $n \geq 2$*, there exists* n *i.i.d. random variables* $Y_1, ..., Y_n$ *such that* $Y_1 + ... + Y_n$ *has distribution F.*

An example of an infinite divisible distribution is the Normal distribution where a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ can be seen as the sum of n i.i.d. random variables each distributed as $\mathcal{N}(\frac{\mu}{n}, \frac{\sigma^2}{n})$. Note that, in general, Y_i and $Y_1 + ... + Y_n$ do not belong to the same class of distributions; if it is the case then the class of distribution is called *closed under convolution*.
From these considerations, it follows that

Proposition 2.1.1 *Let $(X_t)_{t\geq 0}$ be a Lévy process. Then for every t, X_t has an infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution then there exists a Lévy process X_t such that the distribution of X_1 is given by F.*

The direct implications have been shown above, while the proof of the converse can be found in [24]. It can be shown that the characteristic function of a Lévy process in $\mathbb{R}$ is then given by

$$\phi_{X_t}(z) = \mathbb{E}[e^{izX_t}] = \mathbb{E}[e^{izX_1}]^t = e^{t\psi(z)}, \quad z \in \mathbb{R}. \tag{2.3}$$

where

$$\psi(z) = -\frac{1}{2}\gamma z^2 + iAz + \int_{-\infty}^{\infty} (e^{izx} - 1 - izx\mathbb{I}_{|x|\leq 1})\nu(dx) \tag{2.4}$$

Equation 2.3 is called the Lévy-Khinchin representation and the triplet (A, γ, ν) is called the characteristic triplet; a Lévy process is uniquely defined by a positive definite matrix A, a vector γ and a positive measure ν.
The real value A determines the drift component of the process, the real value γ is the variance of the Brownian Motion, and ν is called Lévy measure and determines the jump process of the Lévy process. The process corresponding to the triplet (A,0,0) for some $A \in \mathbb{R}$ is the only deterministic process and it corresponds to a line in the real plane. Indeed, the characteristic exponent reads

$$\phi_{X_t}(z) = e^{iAzt} \tag{2.5}$$

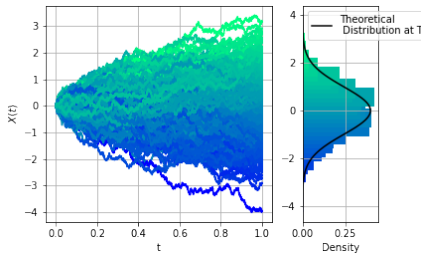
and therefore, the process X_t has the form

$$X_t = At. \tag{2.6}$$

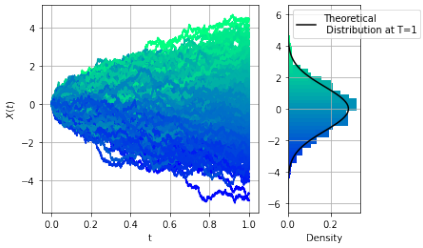
Brownian Motion is the Lévy process corresponding to the triplet $(0, \gamma, 0)$ for some $\gamma \in \mathbb{R}$. γ determines the variance of the Brownian Motion and equation 2.3 reads

$$\phi_{X_t}(z) = e^{t\psi(z)} = e^{-\frac{1}{2}z\gamma t} \tag{2.7}$$

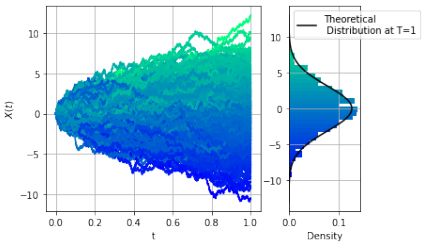
which is the characteristic function of a normal distribution with variance γt. In 2.2 examples of paths of Brownian Motion for different values of γ are presented. For each value of γ, 1000 paths are simulated and the empirical variance for the values of the paths at time $t = 1$ is presented. It can be seen that for higher values of the γ parameter, the paths assume with higher probability values far away from the average which is 0.



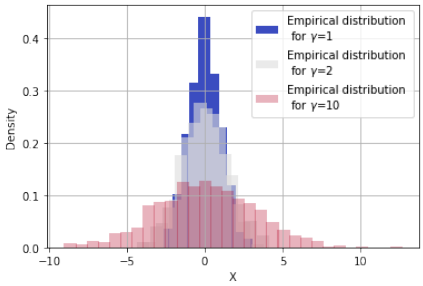
(a) Paths and empirical distribution for $\gamma = 1$.



(b) Paths and empirical distribution for $\gamma = 2$.



(c) Paths and empirical distribution for $\gamma = 10$.



(d) Comparison of the empirical distributions for $\gamma = 1, 2, 10$.

Figure 2.2: Each figure represents the result of the simulations of 10^4 Brownian Motion paths between $t = 0$ and $T = 1$ for different values of γ. Higher values correspond to a higher probability for the process to assume values far away from the average. In figure 2.2d the empirical distributions are compared.

The triplet $(A, \gamma, 0)$ corresponds to a Brownian Motion with drift, meaning that at time t the process follows a Normal distribution with average At and variance γt. In figure 2.3 simulations for this kind of process for $(A = 1, \gamma = 1, 0)$ are presented.

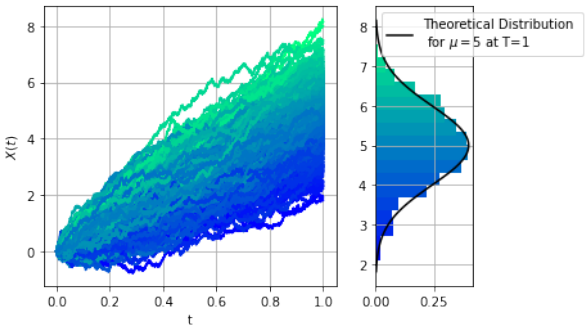


Figure 2.3: Simulations of Lévy process with the triplet (5,1,0). On the right the empirical distribution of the simulated processes at time t=1.

The Lévy measure describes the jump behaviour of the Lévy process through a positive measure on the real numbers. For a real set, S is defined as the expected number of jumps whose size is in S in a unit of time. To define this formally let us first define for the process $(X_t)_{t\geq 0}$ the measure

$$\mu^X(B, t) = \sum_{s \leq t} \Delta X_s \mathbb{I}_{\Delta X_s \in B} \tag{2.8}$$

where $\Delta X_s = X_s - X_{s^-}$. This means that μ counts the number of jumps up to time t whose size is in B, if $B = \{1\}$, then $\mu^X(1, t)$ counts the jumps of size 1 up to time t, if $B = [1, 2]$ it counts the jumps whose size is a real number between 1 and 2, and so on. The sets B are subsets of $\mathbb{R}$ that exclude 0 that corresponds to exclude continuities in the path. Then, the Lévy measure is defined by

$$\nu(B) = \mathbb{E}\left[\mu^X(B, 1)\right] \tag{2.9}$$

which counts the expected number of jumps whose size is in B in one unit of time. The Lévy measure satisfies

$$\nu(0) = 0 \qquad \int_{\mathbb{R}} (1 \wedge x^2)\nu(dx) < \infty \tag{2.10}$$

where the integrand is the minimum between 1 and x^2. This means that the measure has no mass at the origin (considering a jump of size 0 would not be sensible), while the integral says that only a finite number of big jumps can occur. This constraint is relevant because it implies that, if $\nu(\mathbb{R}) = \infty$, it means that infinitely many jumps occur but these need to have a small size. A Lévy Process such that $\nu(\mathbb{R}) = \infty\,(< \infty)$ is called of infinite (finite) activity. Let us start by defining a simple stochastic jump process, the Compound Poisson process, which is defined as

$$X_t = \sum_{k=1}^{N_t} J_k \tag{2.11}$$

where N_t is a Poisson random process, i.e. a stochastic process whose distribution at time t follows a Poisson random variable of intensity λt where λ is usually called the jump intensity of the process. N_t, then, determines the number of jumps that occurred before time t according to the following probability

$$\mathbb{P}(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

The J_k are *i.i.d.* random variables that determine the size of each jump. We want to find

$$\nu(dx) = \mathbb{E}[\mu^X(dx, 1)]. \tag{2.12}$$

To do so, we make use of the characteristic function of the Compound Poisson process and of the following theorem

Theorem 2.1.1 $(X_t)_{t\geq 0}$ *is a compound Poisson process if and only if it is a Lévy process and its sample paths are piece-wise constant functions.*

The proof of this theorem can be found, for example, in chapter 3 of [7]. This theorem states that Compound Poisson processes are equivalent to piece-wise constant Lévy processes. Therefore, in order to find the Lévy measure, it is sufficient to find the characteristic function of the process in the form of equation 2.3. So, if the jump-size distribution J has associated measure $f(dx)$, the characteristic function reads

$$\begin{aligned}\mathbb{E}[e^{izX_t}] &= \mathbb{E}\left[e^{iz\sum_{k=0}^{N_t} J_k}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{iz\sum_{k=0}^{N_t} J_k}|N_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[(e^{izJ})\right]^{N_t}\right] = \\ &= \sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t\mathbb{E}\left[(e^{izJ})\right])^n}{n!} = e^{\lambda t(\mathbb{E}\left[(e^{izJ})\right]-1)} = \\ &= e^{\lambda t\int_{\mathbb{R}}(e^{izx}-1)f(dx)} = e^{t\int_{\mathbb{R}}(e^{izx}-1)\lambda f(dx)}.\end{aligned} \tag{2.13}$$

Combining equation 2.13 and Theorem 2.1.1, the Lévy measure is thus given by $\nu(dx) = \lambda f(dx)$. Roughly, the expected number of jumps of a certain size in a unit of time is given by the jump intensity multiplied by the probability density of the jump size. In figure 2.4 an example of the Poisson Compound process is presented, where jumps are distributed as a Standard Normal Distribution.

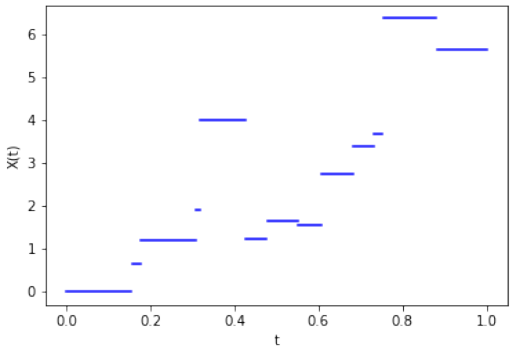


Figure 2.4: Simulation of Compound Poisson process with $\lambda = 20$ and $J \sim \mathcal{N}(0, 1)$

A Lévy triplet of the form $(a, \gamma, \lambda f)$ for some real-valued a, γ, and λ, and a probability density f is called a jump-diffusion process. It is the result of the combination of Brownian Motion, a drift and a Compound Poisson process. In figure 2.5 an example of a jump-diffusion process with no drift is shown.

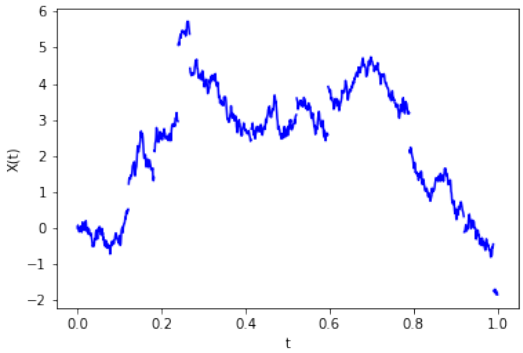
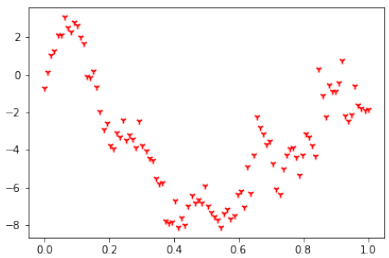
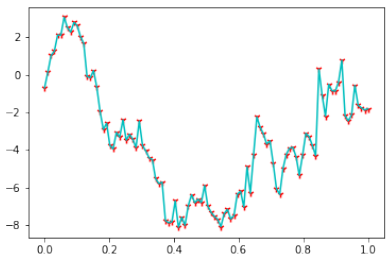


Figure 2.5: Simulation of jump-diffusion process with $A = 0$, $\gamma = 1$, $\lambda = 7$, and $J \sim \mathcal{N}(0, 1)$

Infinite activity processes are processes that, in any given interval, present an infinite number of jumps. In the following subsections, I discuss two examples of infinite activity Jump processes, the Normal Inverse Gaussian process and the Generalised hyperbolic jump process. Figure 2.6 presents a simulation of an infinite activity process. The reader should be aware that a graphic representation of an infinite activity process is not possible: at every time scale the process should present an infinite number of jumps, but this is of course computationally not feasible.



(a) Scatter plot of the cumulative sum of the increments at every interval.



(b) Interpolation of the cumulative sum of the increments of the infinite activity process.

Figure 2.6: Discrete representation of a Normal Inverse Gaussian Infinite activity process on the left and the interpolation of the jumps on the right.

2.1.1. Generalised Hyperbolic Model

The generalised hyperbolic model is a Jump process that at time $t = 1$ follows a Generalised Hyperbolic distribution. This distribution was first introduced in [2] and it has been widely used in different fields since it allows the presence of semi-heavy tails and it has been used, for example, to model returns in Finance. See for example [8] or [9] for applications and [23] for a complete dissertation on the argument. The distribution is defined as the normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian distribution. That is, if

$$Y = \alpha + \beta V + \sigma\sqrt{V}X \tag{2.14}$$

with X and V independent random variables, V following a Generalised Inverse Gaussian distribution and X a normal distribution, then Y follows a Generalised Hyperbolic distribution. This distribution is defined by 5 parameters and has a probability density function given by

$$f_X(x) = \frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)}\, e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\sqrt{\delta^2+(x-\mu)^2}/\alpha\right)^{1/2-\lambda}}.$$

$(\lambda, \alpha, \beta, \mu, \delta)$ are real-valued parameters, $\gamma = \sqrt{\alpha^2 - \beta^2}$, and K_ν is the Modified Bessel function of the second kind with index ν. The first two moments are given by

$$\begin{aligned} \mathbb{E}[X] &= \mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} \\ \mathbb{V}ar(X) &= \frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left(\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \frac{K^2_{\lambda+1}(\delta\gamma)}{K^2_\lambda(\delta\gamma)}\right) \end{aligned} \tag{2.15}$$

β is the asymmetry parameter and determines the skewness of the distribution; a left-skewed distribution gives more weight to negative values, while a right-skewed distribution gives more weight to positive values. If $\beta = 0$ the distribution is symmetric as it can be seen in figure 2.7.

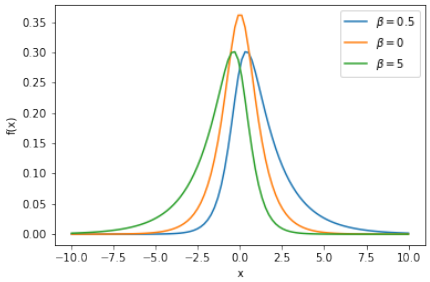
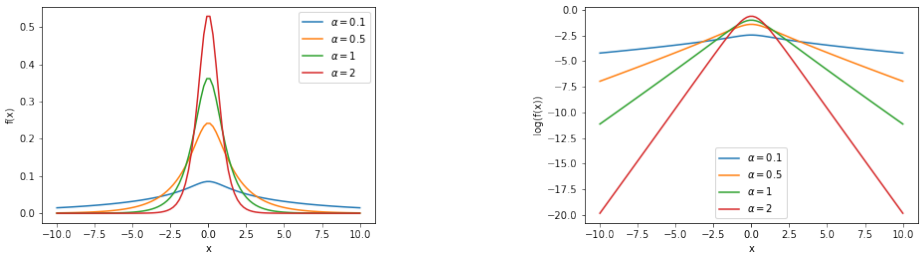


Figure 2.7: Probability distribution function for the generalised hyperbolic distribution for different values of β. The other parameters are set to $(\lambda = 0.5, \alpha = 1\, \mu = 0, \delta = 1)$

α is a positive parameter such that $0 \leq |\beta| < \alpha$ and it determines the heaviness of the tail, more specifically the smaller the parameter, the heavier the tail. See figure 2.8.

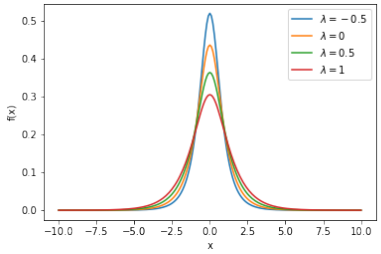


(a) Probability density function for different values of α. The smaller the value, the higher the probability in the tails.

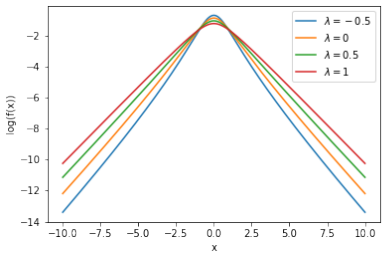
(b) Logarithm of the probability density function for different values of α

Figure 2.8: In figure 2.8a the pdf for different values of α is shown, in figure 2.8b the log probability.

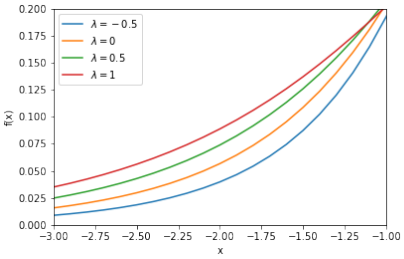
λ determines the shape of the distribution, a higher λ yields a lower peak, figure 2.9a and lighter tails, figure 2.9b, but it allocates more density to medium values, figures 2.9c and 2.9d.



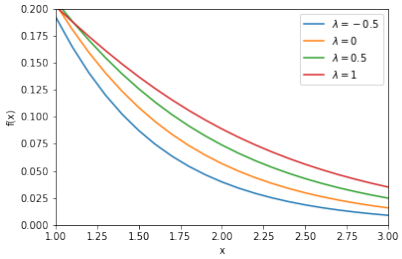
(a) Probability density function for different values of λ



(b) Logarithm of the probability density function for different values of λ.



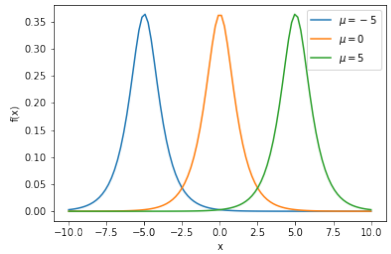
(c) Zoom-in of the probability density function between $x = -3$ and $x = -1$ for different values of λ



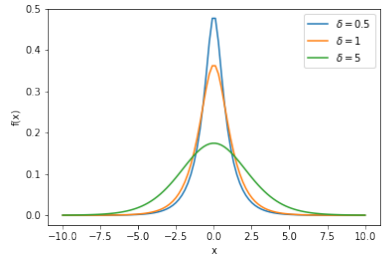
(d) Zoom-in of the probability density function between $x = 1$ and $x = 3$ for different values of λ

Figure 2.9: Probability density function 2.9a, the log-plot 2.9b, and the magnification of the intervals $[-3, -1]$, 2.9c, and $[1, 3]$, 2.9d for different values of the parameters λ. The other parameters are set to $(\alpha = 1, \beta = 0, \mu = 0, \delta = 1)$.

δ and μ are respectively the scale and location parameters; roughly, the location parameter determines around which value the distribution is centred, while the scale parameter how values are spread. Note that for $\beta = 0$ the expected value is given by $\mathbb{E}[X] = \mu$; if the distribution is symmetric the expected value is equal to the location parameter. Moreover, $\mathbb{V}ar(X)$ depends both on δ and δ^2.



(a) Probability density function for different values of μ with $\delta = 1$



(b) Logarithm of the probability density function for different values of δ and $\mu = 0$.

Figure 2.10: Figure 2.10a shows different probability distribution functions for different values of μ while 2.10b the pdf for different values of δ. The other parameters are set to $(\lambda = 0.5, \alpha = 1, \beta = 0)$.

This distribution is infinitely divisible, see for example [2], and therefore, from theorem 2.1.1, there is a Lévy process that at a given time has a Generalised Hyperbolic distribution. A Generalised Hyperbolic model X_t is a Lévy process that at time $t = 1$ follows a Generalised Hyperbolic distribution and it is defined by the Lévy triplet $(E[X_1], 0, \nu^{GH})$, where

$$\nu^{GH}(dx) = \frac{e^{\beta x}}{|x|}\left(\int_0^\infty \frac{e^{-2\sqrt{2y+\alpha^2}|x|}}{\pi^2 y(J^2_{|\lambda|}(\delta\sqrt{2y}) + Y^2_{|\lambda|}(\delta\sqrt{2y}))} K_1(\alpha|x|)dy + \lambda e^{-\alpha|x|}\mathbb{I}_{\lambda\geq 0}\right)dx. \tag{2.16}$$

The Lévy process is given by

$$X_t = t\mathbb{E}[X_1] + \int_0^t \int_{\mathbb{R}} x(\mu^X - \nu^{GH})(dxds) \tag{2.17}$$

The Generalised Hyperbolic distribution, however, is not close under convolution, meaning that the sum of independent random variables distributed as a Generalised Hyperbolic distribution is not, in general, a Generalised Hyperbolic distribution. This means that if the Lévy process at time $t = 1$ follows a GH distribution, then at other time scales a closed form for the distribution is not available in closed form.

2.1.2. Normal Inverse Gaussian Model

The Normal Inverse Gaussian distribution arises by fixing $\lambda = -\frac{1}{2}$ in the Generalised Hyperbolic distribution. This class of distributions still presents semi-heavy tails and the meaning of the parameters is the same as the Generalised Hyperbolic Distribution. The Normal Inverse Gaussian Process at time $t = 1$ follows a Normal Inverse Gaussian distribution, it presents infinite activity, and the Levy measure is given by

$$\nu^{\mathcal{NIG}}(dx) = e^{\beta x} \frac{\delta\alpha}{\pi|x|} K_1(\alpha|x|)dx \tag{2.18}$$

The related Levy process is then given by

$$L_t = t\mathbb{E}\left[L_1\right] + \int_0^t \int_{\mathbb{R}} x(\mu^L - \nu^{\mathcal{NIG}}])(dx, ds) \tag{2.19}$$

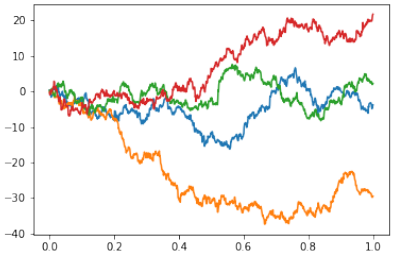
where

$$\mathbb{E}\left[L_1\right] = \left(\mu + \frac{\delta\beta}{\sqrt{a^2 - \beta^2}}\right) \tag{2.20}$$

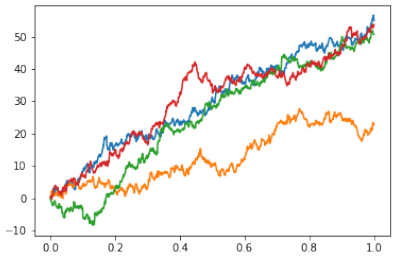
and is the expected value of the Normal Inverse Gaussian distribution. The Levy triplet of the Normal Inverse Gaussian distribution is $\left(\mathbb{E}\left[L_1\right], 0, \nu^{\mathcal{NIG}}\right)$ This process can be used to model returns that are distant from a Normal. In figure 2.11, different simulations for different values of the parameters are available. α influences the tail heaviness of the distribution; that is, the smaller the value of α, the size of jumps becomes more extreme. β is the asymmetry parameter, positive (negative) values imply that the distribution assigns more (less) density to values greater than the average, μ and δ are, respectively, the location parameter and the scale parameter and are connected respectively to the mean and the variance of the distribution. It can be seen that positive values of β and μ cause a drift to the Levy process and higher values of δ and smaller values of α imply larger probabilities for values far away from the mean. In financial terms, a positive parameter β represents the belief that positive returns will be more frequent than negative returns, a smaller α that extreme returns might happen with greater probability, and higher values of δ imply that values around the average return are less probable. Values of β or μ different from 0 cause a drift in the price of the contract, i.e. that the expected value of the contract at a future time t will have an average given by

$$\mathbb{E}\left[X_t\right] = X_0 + \mathbb{E}\left[L_t\right] = X_0 + t\mathbb{E}\left[L_1\right] = X_0 + t\left(\mu + \frac{\delta\beta}{\sqrt{a^2 - \beta^2}}\right). \tag{2.21}$$

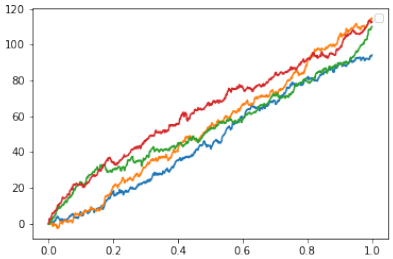
Where the last equality comes from the fact that at time $t = 1$, the infinite activity $\mathcal{NIG}$ process follows a $\mathcal{NIG}$ distribution with parameters $(\alpha, \beta, \mu, \delta)$.



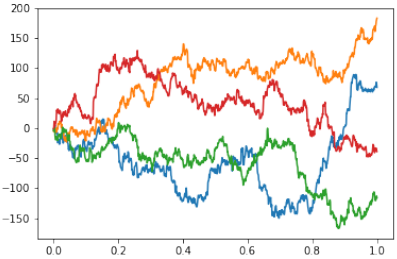
(a) These NIG processes were obtained with the set of parameters $(\alpha = 1, \beta = 0, \mu = 0, \delta = 0.5)$



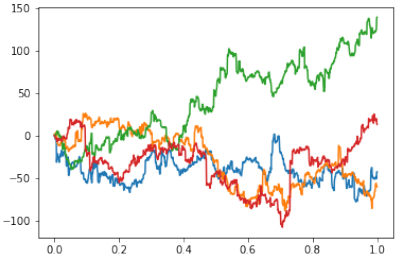
(b) These NIG processes were obtained with the set of parameters $(\alpha = 1, \beta = 0.1, \mu = 0, \delta = 0.5)$. To a positive β it corresponds a non-zero expected value; thus, the processes show a trend in the direction of the expected value.



(c) These NIG processes were obtained with the set of parameters $(\alpha = 1, \beta = 0, \mu = 0.1, \delta = 0.5)$. As in 2.11b, a positive μ corresponds to a non-zero expected value, and thus the processes show a trend in the direction of the average.



(d) These NIG processes were obtained with the set of parameters $(\alpha = 1, \beta = 0, \mu = 0, \delta = 4)$. Jumps with greater sizes happen more often, and at each time, the process presents more extreme values.



(e) These NIG processes were obtained with the set of parameters $(\alpha = 0.1, \beta = 0, \mu = 0, \delta = 0.5)$. A smaller α corresponds to heavier tails meaning that more extreme values happen more often, leading to more extreme values in the paths.

Figure 2.11: Simulations of infinite activity NIG Levy processes between in the interval $[0, 1]$ for different parameters.

2.2. Stochastic Calculus

We now define the stochastic integral with respect to Brownian Motion between time t and time T as follows

Definition 2.2.1 *Let $(\mathbb{W}_t)_{t\in[0.T]}$ be a Brownian Motion, let π_n be a sequence of partitions where the mesh, i.e., $\sup_{i\in\{1,\dots,n\}} |t_i - t_{i-1}|$ goes to 0 and such that $t_0 = t$ and $t_n = T$, let S_t be a càdlàg, locally bounded, square-integrable process adapted to the filtration generated by $(\mathbb{W}_t)_{t\in[0.T]}$, Then the Itô integral is defined as*

$$\int_0^T S_t \, d\mathbb{W}_t := \lim_{n\to\infty} \sum_{t_{i-1}, t_i \in \pi_n} S_{t_{i-1}} (\mathbb{W}_{t_i} - \mathbb{W}_{t_{i-1}}) \tag{2.22}$$

Note that this integral evaluates the function on the left side of the interval (t_{i-1}) and is defined for càdlàg processes. An example of these processes is the Itô diffusion process. As a stochastic integral,

it reads

$$X_T = x + \int_0^T \sigma_t \, d\mathbb{W}_t + \int_0^T \mu_t \, dt \tag{2.23}$$

and in this case, we have that

$$X_T | X_0 \sim \mathcal{N}(\int_0^T \mu_t \, dt + X_0, \int_0^T \sigma^2(t) dt) \tag{2.24}$$

Another relevant notation for stochastic differential equations is the so-called differential form, which is not mathematically founded but is widely used because it allows easier manipulations. The differential notation for an Itô diffusion process would read:

$$\begin{cases} dX_t = \sigma_t d\mathbb{W}_t + \mu_t dt. \\ X_0 = x \quad x \in \mathbb{R} \end{cases} \tag{2.25}$$

where x is the initial condition of the process. ‘

3

Problem Formulation

‘ This research aims to find a stochastic process that would represent the probability distribution of the changes in prices of an hour contract in the Intraday Market (in Germany. To do so, first, an introduction to the functioning of the German energy market is given, and then, the formulation of the stochastic differential equation modelling the problem is presented.

3.1. Energy Market Context

Energy markets are used to trade energy in the form of contracts which represent an obligation to receive (to supply) energy during a period of time. Here, I only focus on 15, 30 or 60-minute contracts. In this section, I will describe the functioning of the German Energy Market.

The energy market can be subdivided into two phases, the long-term and the short-term market. The short-term market is composed of the Day-Ahead (DA) and Intraday (ID) markets; the former opens 45 days beofore and closes at 12:00 of the day before delivery. Prices in this phase are determined by a blind-side auction. The blind-side auction collects offers and bids until its expiration when these are aggregated into curves. The price and the volume at the intersection are called market clearing price and market clearing volume. Bids that are higher and offers that are lower than the market clearing price are cleared at the market clearing price, also called the Day-Ahead price. In Germany, only hourly contracts are exchanged during this phase.

The ID market opens at 15:00 of the day before delivery and closes 5 minutes before delivery. In this phase, contracts are continuously traded. Bids and offers are collected in an order book and matched according to a price and temporal order, meaning that if there are multiple bids (offers) matching an offer (bid), the first bid (offer) in temporal order is the one that is executed. In Germany, during this phase, it is possible to trade 15, 30 and 60-minute contracts. For 15-minute contracts, a blind-sided auction closes at 15:00 of the day before delivery, but it is out of the scope of this thesis; see for example [20] for more information. Table 3.1 shows the total traded volumes for Day-Ahead and Intraday markets.

	2019	2020
Total Volume Germany (GWh)	280,191.2	279848.3
Day-Ahead Volume Germany (GWh)	226,409.8	216,221.3
Intraday Volume Germany (GWh)	53,781.4	63,627.0
Total Volume EPEX Spot (Gwh)	593,432.8	614,839.3
Day-Ahead Volume EPEX Spot (GWh)	501,568.8	503,650.8
Intraday Volume EPEX Spot (GWh)	91,864.0	111,188.5

Table 3.1: Volumes of energy traded in the EPEX Spot Energy Market. The first three rows show the value for the fraction of energy traded in the German market and the last three the amount of energy traded across all the European countries. The total traded volume in the spot market has decreased from 2019 to 2020 due to a decline in the volume traded in Day-Ahead. Across all markets, the total power increased seeing a larger growth in the Intraday Market.

The Transmission System Operator (TSO) is the authority responsible for maintaining the frequency at the same level in every settlement period. A day is subdivided into settlement periods, and for each settlement period, the TSO ensures that the total amount of production is equal to the electricity consumption.
The current regulation provides that producers and consumers, designated as Balance Responsible Parties (BRPs), take part in maintaining the grid in balance through the self-balancing of their positions. For each settlement period, they must submit a nomination which consists of the forecasting of produced or consumed energy and the energy contracts held in their portfolio. The law enforces the balancing of the nomination, meaning that the total amount of consumed and sold energy must be equal to the total amount of energy bought and produced by the BRP.
An example of BRP is a power supplier in the form of a Wind Park that needs to forecast its energy production and find buyers. Similarly, a household energy provider has to forecast its users' consumption and find an adequate supply.
Between the closure of the DA market and the moment of delivery, the position of a BRP may be subjected to unexpected changes that can be traded in the ID market: shortages can be bought, and surpluses can be sold. For example, if a coal power plant faces unexpected outages and cannot provide the traded energy then the corresponding position needs to be balanced by buying energy during the ID market. Conversely, a wind park would need to sell excess energy if the wind speed at the moment of delivery was higher than the forecast one on the day before. To push the market participants into trading their imbalances in the ID market, the TSO imposes the imbalance price on eventual disparities in the position.
A BRPs imbalance position is the difference between the nominated position after the closure of the ID market and the actual net exchange of electric energy with the grid in real-time and can be short (negative net balance), long (positive net balance) or balanced. The sum of all imbalance positions determines the System Imbalance (SI) [4] and to keep the system balanced, the TSO can activate the Restoration Reserves and a price is determined. The imbalance price, i.e. the price that parties with a long position receive and those with a short position pay, is increased if the SI is negative and decreased if positive. This works as an incentive for the parties to trade their positions during the ID market. A value of reference for this penalty is the "imbalance price spread", which is the difference between the imbalance price and the index ID_3, the volume weighted average of the last three hours of trading in the Intraday Market [4], [16].

3.2. Mathematical Formulation

Now that the functioning of the intraday energy market has been explained, the mathematical formulation is presented. I model the price of a contract X_t at a given time t depending on the value of the contract at time $t = 0$ and the realisation of the Lévy process L. Lévy processes allow choosing amidst

a rich family of distributions that can represent different tail behaviours and incorporate the presence of jumps. The model is then formulated as

$$X_t = X_0 + L_t.$$

which implies that the return $X_t - X_0$ is distributed as L_t. The goal of the research then boils down to finding the Lévy triplet (A, γ, ν) that most accurately represents the distribution of the returns. To do so, it is possible to first choose if including a drift A, of Brownian Motion component γ and to which family of distributions the Lévy measure belongs. This procedure simplifies the model selection because it allows excluding beforehand those triplets that would not capture some of the characteristics of the price of the contract. For example, a process with sparse large jumps can exclude a Lévy measure that implies infinite activity or Brownian motion. After this choice, it is necessary to find the set of parameters that better fits the historical returns for that triplet.
After having determined different possible triplets and a set of parameters for each of them using historical data, it is relevant to test the models to compare their performances. For example, by assessing the accuracy in predicting the distribution of the price increments in the test set according to different criteria. In the next section, the statistical methods to achieve this goal are introduced.

4

Statistics

In this section, I will explain the statistical tools that will apply to the study of the data of the Intraday Market. First, since it is used to estimate the parameters of each Lévy triplet, an explanation of the maximum likelihood estimation is given. Lévy processes can differ in the number of parameters describing the Lévy triplet, and thus, to quantify and compare the performance in describing the data, it is necessary to introduce quantitative measures. With this regard, the Akaike Information Criterion and the Bayesian Information Criterion will be used, and their details are explained in the second subsection. Moreover, the data analysis needs quantitative measures to accept or refuse the hypothesis. Specifically, data representing the prices of a contract will be in the form of a time series. So some of the relative statistical tests are introduced in the third subsection.

4.1. Maximum Likelihood and Fisher Information Matrix

Given a set of observations $x = (X_1, ..., X_n)$ where the random variables have joint probability distribution $f_\theta(x)$ for some set of parameters θ, the likelihood function is the function

$$L(\theta; x) = f_\theta(x) \tag{4.1}$$

that maps a real value to each set of parameters θ in the parameter space Θ. For example, if $(x_1, ...x_n)$ are the realization of n Bernoulli random variables $X \sim \mathit{Be}(p)$ the set of parameters is $\theta = p$ and $\Theta = [0, 1]$. If the variables are independent and identically distributed and follow a certain density $f_\theta(x)$, then the likelihood can be written as

$$L(\theta; x) = \prod_{i=1}^{n} f_\theta(x_i). \tag{4.2}$$

In this context, the $\log(L(\theta; x))$, called the log-likelihood, is often taken into consideration because it allows reducing the products to a sum of logarithms which often leads to more straightforward calculations, see Example 4.1.1. The maximum likelihood estimator $\hat{\theta}$ is defined as the estimator such that

$$\hat{\theta} := \max_{\theta \in \Theta} L(\theta; x) \tag{4.3}$$

Since the logarithm is a continuous and increasing function, the $\hat{\theta}$ is also a maximizer for the log-likelihood.

Example 4.1.1 *Given a sample* $X_1 = x_1, ...X_n = x_n$ *where* $X_i \sim \mathcal{N}(\mu, \sigma^2)$ $\theta = (\mu, \sigma^2)$ *and* $\Theta = \mathbb{R} \times (0, +\infty)$*, the likelihood is given by*

$$L(\theta; x) = L(\mu, \sigma^2; x) = \prod_{i=1}^{n} f_{X_i}(x_i; \mu, \sigma^2) = \prod_{i=0}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{4.4}$$

and the log-likelihood by

$$l(\theta; x) = \log(\prod_{i=1}^{n} f_{X_i}(x_i; \theta)) = \sum_{i=1}^{n} \log(f_{X_i}(x_i; \theta)) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \tag{4.5}$$

The maximum likelihood estimator is found as the couple of parameters $(\hat{\mu}, \hat{\sigma})$ *such that*

$$\frac{\partial l(\theta; x)}{\partial \mu}\Big|_{\mu=\hat{\mu}} = 0 \tag{4.6}$$

$$\frac{\partial l(\theta; x)}{\partial \sigma}\Big|_{\sigma=\hat{\sigma}} = 0 \tag{4.7}$$

Which leads to

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4.8}$$

and

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}. \tag{4.9}$$

However, the maximum-likelihood parameters are not available in closed form for other distributions, and other techniques are necessary. For example, the Normal Inverse Gaussian presents the following log-likelihood

$$l(\alpha, \beta, \mu, \delta; x) = n\log(\alpha) + n\log(\delta) + n\delta\sqrt{\alpha^2 - \beta^2} + \sum_{i=1}^{n} \beta(x_i - \mu) + \log(K_1(\alpha\sqrt{\delta^2 + (x_i - \mu)^2}) \\ -n\log(\pi) - \frac{1}{2}\log(\delta^2 + (x_i - \mu)^2). \tag{4.10}$$

The modified Bessel Function of the second order leads to first derivatives with respect to α, β, δ whose zeros are not representable in closed form. Some techniques are available to work around this problem. In [13], for example, an EM-type algorithm is implemented.

4.1.1. Fisher Information

For a distribution D with parameters θ in some parameter space Θ, the Fisher Information Matrix $I(\theta)$ is defined as the matrix having as elements

$$(I(\theta))_{i,j} = \mathbb{E}\left[\frac{\partial}{\partial\theta_i} l(X; \theta) \frac{\partial}{\partial\theta_j} l(X; \theta)\right] \tag{4.11}$$

l is the log-likelihood, and the covariance is taken under the probability measure induced by the distribution D for a given set of parameters θ. Note that the Fisher Information Matrix does not rely on any particular observations but considers the full support of the distribution. It can be shown that if the following conditions are satisfied,

1. $\frac{\partial f(X;\theta)}{\partial\theta}$ exists almost everywhere, i.e. it fails to exist on a set of measure 0 w.r.t. the Lebesgue measure.
2. The support of $f(X; \theta)$ does not depend on θ.
3. The integral of $f(X; \theta)$ can be derived under the sign of the integral sign with respect to θ.

then,

$$(I(\theta))_{i,j} = -\mathbb{E}\left[\frac{\partial}{\partial\theta_i\partial\theta_j} l(X; \theta)\right]. \tag{4.12}$$

Given a set of independent observations $(x_1, ..., x_n)$ assumed to be drawn from the same population, the observed information matrix is defined as

$$(J(\theta))_{i,j} = \frac{\partial l(x_1, ..., x_n; \theta)}{\partial \theta_i \partial \theta_j}|_{\theta=\theta^*}. \tag{4.13}$$

For a set of parameters θ and an MLE $\hat{\theta}$, if certain regularity conditions are satisfied, and the Fisher Information Matrix is non-singular, then

$$\sqrt{n}(\theta - \hat{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}(\theta)). \tag{4.14}$$

where $\xrightarrow{\mathcal{D}}$ indicates the convergence in distribution. This property is called the asymptotic normality of the MLE and states that the estimator will converge in distribution to the true parameter. This property is used to derive confidence intervals for the estimated parameters based on the Fisher Information Matrix and the number of samples. The Observed Information Matrix for the *Normal Inverse Gaussian* matrix is given by

$$J(\theta) = \begin{bmatrix} \frac{\partial^2}{\partial \alpha^2} l(x) & \frac{\partial^2}{\partial \alpha \partial \beta} l(x) & \frac{\partial^2}{\partial \alpha \partial \mu} l(x) & \frac{\partial^2}{\partial \alpha \partial \delta} l(x) \\ \frac{\partial^2}{\partial \alpha \partial \beta} l(x) & \frac{\partial^2}{\partial \beta^2} l(x) & \frac{\partial^2}{\partial \beta \partial \mu} l(x) & \frac{\partial^2}{\partial \beta \partial \delta} l(x) \\ \frac{\partial^2}{\partial \alpha \partial \mu} l(x) & \frac{\partial^2}{\partial \alpha \partial \mu} l(x) & \frac{\partial^2}{\partial \mu^2} l(x) & \frac{\partial^2}{\partial \mu \partial \delta} l(x) \\ \frac{\partial^2}{\partial \alpha \partial \delta} l(x) & \frac{\partial^2}{\partial \beta \partial \delta} l(x) & \frac{\partial^2}{\partial \mu \partial \delta} l(x) & \frac{\partial^2}{\partial \delta^2} l(x) \end{bmatrix} \tag{4.15}$$

where $\theta = (\alpha, \beta, \mu, \delta)$ and $l(x)$ is given by equation (4.10). We define $v = \alpha\sqrt{\delta^2 + (x-\mu)^2}$. and

therefore

$$\frac{\partial^2}{\partial\alpha^2}l(x) = -\frac{n}{\alpha^2} - \frac{n\delta\beta^2}{(\alpha^2-\beta^2)^{\frac{3}{2}}} + \sum_{i=1}^{n}\frac{1}{K_1(v)^2}\left(K_1(v)\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\alpha}\right)^2 - \left(\frac{\partial^2 K_1(v)}{\partial v^2}\right)^2\left(\frac{\partial v}{\partial\alpha}\right)^2\right) \tag{4.16}$$

$$\frac{\partial^2}{\partial\beta^2}l(x) = \frac{-\alpha^2 n\delta}{(\alpha^2-\beta^2)^{\frac{3}{2}}} \tag{4.17}$$

$$\begin{aligned}\frac{\partial^2}{\partial\mu^2}l(x) =& \sum_{i=1}^{n}\frac{-\delta^2+(x-\mu)^2}{(\delta^2+(x-\mu)^2)^2} + \\ &+\frac{1}{K_1(v)^2}\left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\mu}\right)^2 + \left(\frac{\partial^2 v}{\partial\mu^2}\right)\frac{\partial K_1(v)}{\partial v}\right)K_1(v) - \left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\right)^2\left(\frac{\partial v}{\partial\mu}\right)^2\right)\right)\end{aligned} \tag{4.18}$$

$$\begin{aligned}\frac{\partial^2}{\partial\delta^2}l(x) =& \sum_{i=1}^{n} -\frac{1}{\delta^2} + \frac{\delta^2-(x-\mu)^2}{(\delta^2+(x-\mu)^2)^2} + \\ &+\frac{1}{K_1(v)^2}\left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\delta}\right)^2 + \left(\frac{\partial^2 v}{\partial\delta^2}\right)\frac{\partial K_1(v)}{\partial v}\right)K_1(v) - \left(\frac{\partial K_1(v)}{\partial v}\right)^2\left(\frac{\partial v}{\partial\delta}\right)^2\right)\end{aligned} \tag{4.19}$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}l(x) = \frac{\alpha\beta\delta n}{(\alpha^2-\beta^2)^{\frac{3}{2}}} \tag{4.20}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha\partial\mu}l(x) =& \sum_{i=1}^{n}\frac{1}{K_1(v)^2}\left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\alpha}\right)\left(\frac{\partial v}{\partial\mu}\right) + \right.\right. \\ &\left.\left.\left(\frac{\partial^2 v}{\partial\alpha\partial\mu}\right)\frac{\partial K_1(v)}{\partial v}\right)K_1(v) - \left(\frac{\partial K_1(v)}{\partial v}\right)^2\left(\frac{\partial v}{\partial\alpha}\frac{\partial v}{\partial\mu}\right)\right)\end{aligned} \tag{4.21}$$

$$\begin{aligned}\frac{\partial^2}{\partial\alpha\partial\delta}l(x) =& \sum_{i=1}^{n}\frac{\alpha}{\sqrt{\alpha^2-\beta^2}} + \frac{1}{K_1(v)^2}\left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\alpha}\frac{\partial v}{\partial\delta}\right) + \right.\right. \\ &\left.\left.+\left(\frac{\partial^2 v}{\partial\alpha\partial\delta}\right)\frac{\partial K_1(v)}{\partial v}\right)K_1(v) - \left(\frac{\partial K_1(v)}{\partial v}\right)^2\left(\frac{\partial v}{\partial\alpha}\frac{\partial v}{\partial\delta}\right)\right)\end{aligned} \tag{4.22}$$

$$\frac{\partial^2}{\partial\beta\partial\mu}l(x) = -n \tag{4.23}$$

$$\frac{\partial^2}{\partial\beta\partial\delta}l(x) = \frac{-n\beta}{\sqrt{\alpha^2-\beta^2}} \tag{4.24}$$

$$\begin{aligned}\frac{\partial^2}{\partial\mu\partial\delta}l(x) =& \sum_{i=1}^{n}\frac{-2\delta(x-\mu)}{(\delta^2+(x-\mu)^2)^2} + \frac{1}{K_1(v)^2}\left(\left(\frac{\partial^2 K_1(v)}{\partial v^2}\left(\frac{\partial v}{\partial\mu}\frac{\partial v}{\partial\delta}\right) + \right.\right. \\ &\left.\left.+\left(\frac{\partial^2 v}{\partial\mu\partial\delta}\right)\frac{\partial K_1(v)}{\partial v}\right)K_1(v) - \left(\frac{\partial K_1(v)}{\partial v}\right)^2\left(\frac{\partial v}{\partial\mu}\frac{\partial v}{\partial\delta}\right)\right),\end{aligned} \tag{4.25}$$

Where

$$\frac{\partial}{\partial v}K_1(v) = -K_0(v) - \frac{1}{v}K_1(v) \tag{4.26}$$

and

$$\frac{\partial^2}{\partial v^2}K_1(v) = \frac{(v^2+2)}{v^2}K_1(v) + \frac{K_0(z)}{z} \tag{4.27}$$

4.1.2. Akaike Information Criterion and Bayesian Information Criterion

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are defined as

$$AIC := 2k - 2\log(L(\theta)) \tag{4.28}$$

$$BIC := k\log n - 2\log(L(\theta)) \tag{4.29}$$

where k is the number of the parameters of the distribution, n is the number of observations, $\theta = (\theta_1, ...\theta_k)$ is the set of parameters, and L is the likelihood.
Both these criteria consider the goodness of fit of a model, represented by the negative log-likelihood and the number of parameters it uses. These criteria estimate the relative quality of different statistical models by creating a trade-off between accuracy and complexity.

4.2. Time series

The prices of the contracts are available as a time series, i.e. a collection of points in time; therefore, in this section, some statistical tools that will be later used are introduced.

4.2.1. Mann-Kendall test

The Mann-Kendall test, [14] and [18], is used to determine the presence of a trend in a time series. It makes use of the Mann-Kendall test, defined as

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sign(x_j - x_i) \tag{4.30}$$

where

$$sign(x_j - x_i) = \begin{cases} 1 & \text{if } x_j > x_i \\ 0 & \text{if } x_j = x_i \\ -1 & \text{if } x_j < x_i \end{cases} \tag{4.31}$$

S is thus the summation of the sign of the difference between a current value x_i and all the future increments x_j. If this statistic is 0, then the probability of a trend in the time series is low; conversely, if it is larger (smaller), then 0 the probability that a positive (negative) trend is present is higher. The variance of the Kendall-Mann statistic is defined as

$$\mathbb{V}ar(S) = \frac{1}{18}(n(n-1)(2n+5) - \sum_{k=1}^{p} q_k(q_k-1)(2q_k+5)). \tag{4.32}$$

The statistics

$$Z = \begin{cases} \frac{S-1}{\sqrt{\mathbb{V}ar(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{\mathbb{V}ar(S)}} & \text{if } S < 0 \end{cases} \tag{4.33}$$

is approximately Normally Distributed with mean 0 and variance 1. The null-hypotheses is given by

$$H_0 : \text{No trend is present.} \tag{4.34}$$

and it is not refused at significance level α if $|Z| < Z_{1-\frac{\alpha}{2}}$, where Z_a is the a-th quantile of the Standard Normal Distribution.

4.2.2. Dickey-Fueller Test

The Dickey-Fueller test determines if an autoregressive model $AR(1)$ contains mean-reversion or not. The model is given by the following equation

$$y_t = \rho y_{t-1} + u_t \tag{4.35}$$

with u_t representing white noise and ρ a coefficient.

$$\Delta y_t := y_t - y_{t-1} = (\rho - 1)y_t + u_t = \delta y_t + u_t \tag{4.36}$$

The hypotheses of the test are

$$H_0 : \{\delta = 0\} \quad \text{vs.} \quad H_1 : \{\delta \neq 0\}. \tag{4.37}$$

If $\delta = 0$ then the increment Δy_t is not influenced by the current value of the process and indicates that mean-reversion can be excluded. This test is also available for the models

$$\Delta y_t = a_0 + \delta y_t + u_t \tag{4.38}$$

and

$$\Delta y_t = a_0 + a_1 t + \delta y_t + u_t. \tag{4.39}$$

a_0 in equations (4.38) and (4.39) represents a constant while $a_1 t$ in equation (4.39) a time-dependent trend in the time series. Also in these cases, $\delta = 0$ means that there is no correlation between the value of the process and the value of the following increment.

5

Data Analysis and Model Selection

This section aims to find a Lévy process to model the returns of the hourly contracts in the German EPEX Spot Market. As illustrated in Section 3, the first objective is to find possible Lévy triplets. This is achieved by analysing the data and choosing those processes that are a better fit for the behaviour of the returns of the prices, e.g. a jump-diffusion model or an Infinite activity process. Then, for each Lévy triplet, find the set of parameters that produce a better fit of the data. Finally, select one or more of these models based on the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC).

The first step consisted of analysing the data concerning the trades of energy contracts in the EPEX SPOT market to find possible priors for the Lévy triplet. I.e., a real value θ representing the drift, a positive value σ representing the standard deviation structure of the Brownian Motion, and a Lévy measure ν, which accounts for the jump activity of the process. The data description is available in the appendix [1]. The analysis focused on the 'BUY' sides of the trades for the hourly contracts. Contracts have then been subdivided by their delivery hour (1 through 24), and for each trade, the execution time, the price and the volume were available.

The execution times, i.e. the timestamp when a trade happens, allow for determining how trades are distributed during a trading day. These are relevant because they are a determinant factor in the jump properties and the presence of Brownian motion in the model; if trades happened at a considerable distance in time, then a jump process with very low activity, that is, a low expected number of jumps in a unit of time, would be more suited. On the other hand, very close trades would correspond to a high activity process or a Brownian Motion because they can capture the close-in-time changes in price. Moreover, clustered in-time trades would mean that a noise effect significantly influences the observed price change; see for examples [5]. The results of this analysis can be seen in figure 5.1. The distances smaller than 10 seconds account for about 83% of all the distances, and those smaller than 1 second are responsible for more than 80% alone. Furthermore, distances smaller than 1 millisecond account for more than 90% of time distances up to 100 milliseconds.

[1] appendix

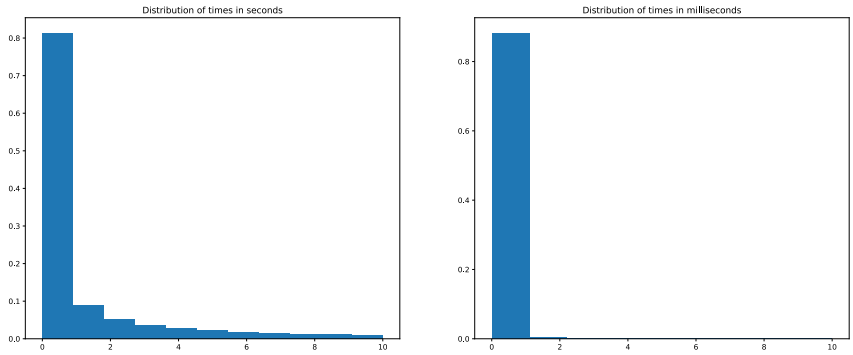


Figure 5.1: Time distance distributions in two different scales; on the left, the distribution of the trades whose distance is up to 10 seconds, on the right up to 10 milliseconds.

This distribution of time distances suggests most trades happen in clusters very close in time. Figure 5.2 shows the average traded volume during a year for each minute. It can be seen that during a trading day, the traded volume increases and it is the most in the last hours. These two facts combined show that trades are irregularly distributed during a trading day with many clusters happening in the last hours and fewer in the first hours.

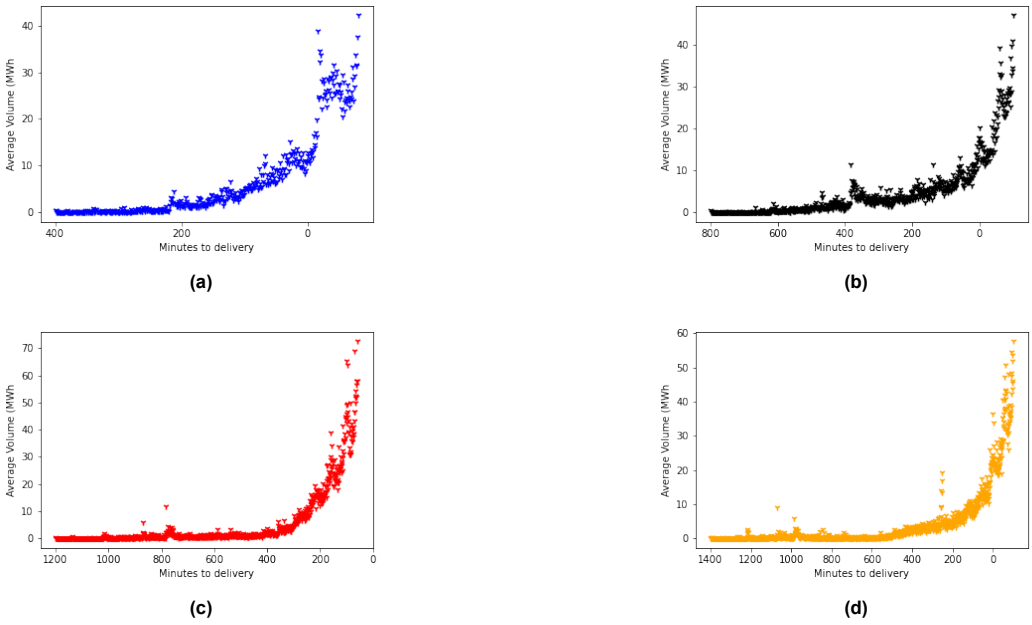


Figure 5.2: Average volume traded in each minute for the year 2020 for different contracts. Figure 5.2a for hour 1, figure 5.2b for hour 8, figure 5.2c for hour 12, figure 5.2d for hour 18.

However, to find a triplet and a distribution for the Lévy Process at one-time scale Δ, returns of the form $X_{t+\Delta} - X_t = L_\Delta$ are necessary. In other words, returns need to be equally distanced in time. To achieve this, data are aggregated into trading periods, and for each trading period, the volume-weighted average price is found through the following formula

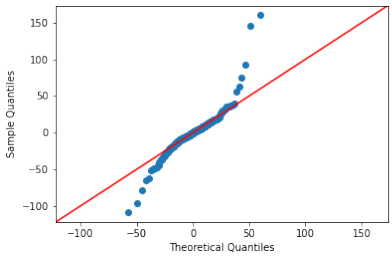
$$\text{AvgPrice} = \sum_{i=0}^{n} \frac{p_i v_i}{\sum_{j=0}^{n} v_j}. \tag{5.1}$$

If there were no trades during one trading period, then the price is equal to the previous trading period. I aggregated data in trading periods of 1, 5, 15, 30, and 60 minutes; the average price was found for

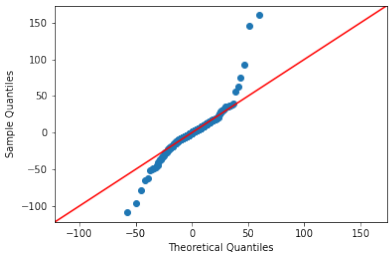
each. Then, the increments between each trading period are collected, and the empirical distributions are found. The distribution shows heavier tails, i.e. more mass far away from the median of the distribution, and a higher peak, i.e. more probability mass around the mode of the distribution than a normal distribution. The empirical kurtosis, an indicator of tail heaviness, yields values far larger than 0, the kurtosis of the normal distribution. Therefore, the underlying distribution can be searched among the heavy and semi-heavy tail distributions. Figure (5.4) presents the QQ-plot for the prices regarding hour 12. It shows that the Normal Inverse Gaussian distribution and Generalised Hyperbolic are a good fit for the tail behaviour of the data, suggesting that the actual underlying distribution presents semi-heavy tails. Indeed, the fitted Normal Distribution, figure (5.4c), and the Student's t-distribution, (5.4d), show a worse fitting of the empirical data and therefore light and heavy tails can be excluded. The other contracts show similar behaviour. The parameters obtained by maximum likelihood estimation (MLE) for the student's t-distribution yield an infinite kurtosis, which may be compatible with some of the highest kurtosis values. Concerning the remaining distributions fitted through MLE, there is not a clear choice in terms of kurtosis. In some cases, the Generalised Hyperbolic provides a better fit, while in others, the Normal Inverse Gaussian is a more suited solution. For extremely high values, neither can replicate the empirical findings. Results are available in table 5.1. In figure 5.3, the observed distribution for the noon contracts and the fitted distributions are available.

Hour	5 minutes	15 minutes	30 minutes	60 minutes
1	[265.83 68.4 128.44]	[40.63 40.06 79.49]	[22.85 30.12 54.56]	[9.8 12.93 23.6]
2	[110.81 62.52 112.68]	[64.54 51.85 77.24]	[34.22 39.71 53.21]	[19.28 23.45 30.26]
3	[296.08 73.64 115.31]	[44.93 48.22 74.48]	[35.12 37.62 59.2]	[20.47 17.95 29.8]
4	[120.44 57.01 109.73]	[55.6 45.85 76.64]	[37.21 37.22 57.51]	[17.87 19.37 30.91]
5	[267.53 72.51 152.76]	[99.54 57.14 102.89]	[70.34 46.99 75.61]	[49.64 26.77 43.12]
6	[1126.54 70.77 214.23]	[722.3 62.9 148.98]	[882.66 50.8 113.37]	[229.82 32.32 57.5]
7	[2446.06 70.61 243.67]	[1665.58 2.81 200.12]	[655.55 70.42 139.41]	[173.49 44.49 74.5]
8	[897.55 61.4 180.68]	[464.2 49.59 143.29]	[264.54 51.72 122.99]	[149.26 32.6 73.25]
9	[171.9 44.18 128.38]	[199.39 39.22 114.93]	[84.89 33.58 100.79]	[35.39 22.19 61.34]
10	[107.63 56.06 113.8]	[64.7 45.17 105.58]	[89.28 39.17 96.06]	[253.32 33.51 63.13]
11	[89.28 46.25 94.43]	[53.49 39.72 91.73]	[42.36 35.88 86.1]	[39.72 30.86 63.31]
12	[184.01 64.63 105.8]	[68.11 46.34 97.47]	[46.92 41.18 90.07]	[25.42 30.59 67.2]
13	[113.32 58.35 101.52]	[76.29 52.53 92.55]	[57.76 50.17 86.93]	[39.78 32.33 67.17]
14	[124.8 63.09 110.65]	[65.7 55.79 97.35]	[43.54 46.83 88.5]	[31.79 34.24 68.31]
15	[146.11 63.53 125.22]	[87.66 60.75 116.14]	[63.44 5.05 99.4]	[48.99 32.93 72.04]
16	[227.09 59.16 149.45]	[135.36 66.2 127.38]	[55.35 52. 116.09]	[38.63 35.06 91.2]
17	[154.6 78.67 204.89]	[119.26 74.55 192.55]	[77.74 69.34 165.8]	[55.93 44.56 120.04]
18	[5227.95 111.43 303.78]	[699.47 111.71 312.16]	[356.69 82.71 263.36]	[209.75 57.71 177.39]
19	[1002.46 87.43 225.]	[460.4 76.65 203.31]	[150.75 58.26 177.91]	[40.45 41.03 116.12]
20	[258.46 66.31 189.79]	[106.58 62.18 167.89]	[58.79 50.29 135.41]	[46.23 37.56 100.65]
21	[729.72 63.16 167.42]	[425.63 53.71 152.99]	[275.25 47.4 141.]	[131.69 34.78 92.56]
22	[542.13 66.65 142.68]	[528.98 70.55 127.51]	[177.09 68.98 115.52]	[93.85 50.4 86.13]
23	[732.93 23.8 165.82]	[302.71 69.84 122.43]	[108.28 4.1 99.33]	[46.42 28.55 60.54]
24	[161.68 56.68 162.31]	[61.45 56.87 100.52]	[35.11 35.18 62.96]	[13.8 20.53 31.52]

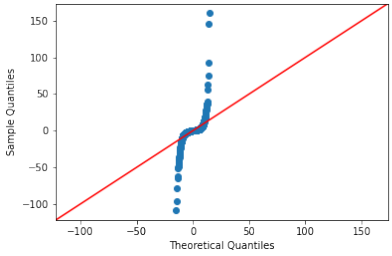
Table 5.1: Each entry shows kurtosis values for the empirical distribution (left), the fitted Generalised Hyperbolic Distribution (centre) and the Fitted Normal Inverse Gaussian (right) for each delivery hour (row) and trading intervals (column). Empirical Kurtosis is way greater than 0 suggesting that the underlying distribution is not Normal.



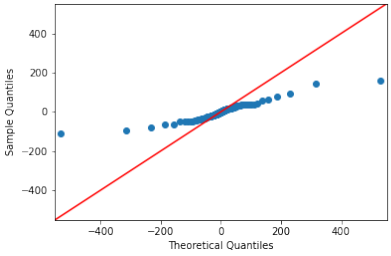
(a) QQ-plots for the Normal Inverse Gaussian distribution for the 15-minute increments of the hour 12 contract



(b) QQ-plots for the Normal Inverse Gaussian distribution for the 15-minute increments of the hour 12 contract



(c) QQ-plots for the Normal distribution for the 15-minute increments of the hour 12 contract



(d) QQ-plots for the Student's t-distribution for the 15-minute increments of the hour 12 contract

Figure 5.4: The QQ-plots of the Normal Inverse Gaussian distribution (5.4a), the Generalised Hyperbolic distribution (5.4b), the Normal distribution (5.4c), and the Student's t-distribution (5.4d) vs the 15-minute increments of the hour 12 contract are presented.

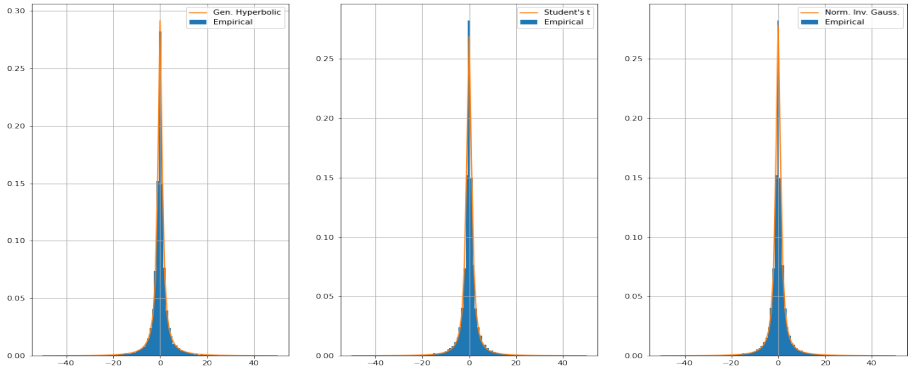


Figure 5.3: An example of fitting the distributions of the increments in the average weighted price between 15-minute trading periods.

5.1. Model Selection

As stated in chapter 4 of [7], a model choice can be made between the jump-diffusion process, a combination of Brownian Motion and finite activity jumps, and infinite activity processes, which instead do not contain Brownian Motion. The goal is two-fold, find a process that would at the same time replicate the high frequency of the trades and that be able to replicate the heavy tails of the distribution of the average-weighted price increments for 1, 5, 15, 30 or 60 minutes. A jump-diffusion model would be easy to simulate. Still, finding parameters, such as jump size distribution and jump intensity, that would be a good fit for the distribution might be challenging. On the other hand, an infinite activity jump model would easily replicate the chosen distribution.

The considered models are, therefore:

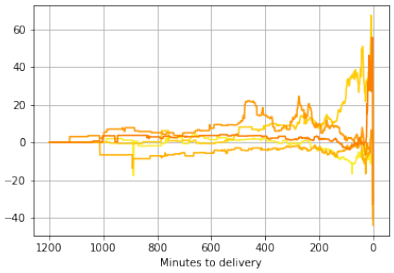
- Model 1: Diffusion Process
- Model 2: Jump-Diffusion Process with jumps sized as a normal distribution
- Model 3: Infinite activity process with NIG distribution.
- Model 4: Infinite activity process with GH distribution.

5.1.1. Diffusion Model

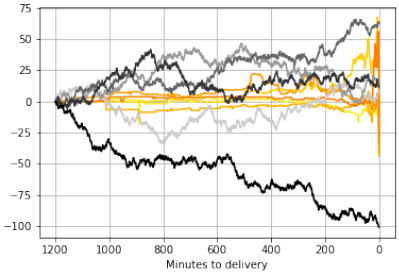
The diffusion model corresponds to the Lévy process

$$L_t = \gamma t + \sigma \mathbb{W}_t.$$

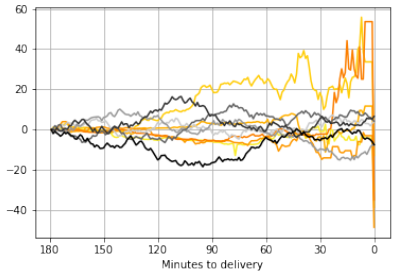
This model could not replicate the excess kurtosis of the price increments distribution. However, for its simplicity, evaluating its performance in modelling returns is sensible. The model is calibrated by taking the standard deviation of the increments and their average that would maximise the likelihood of the Normal Distribution. Table (5.2) shows the score for the AIC, while table (5.3) for the BIC. Notably, this value is infinite for most contracts and short time intervals. This is because some price increments fall in the extreme of the distribution, and numerically the likelihood value is 0; hence the negative of its logarithm is $+\infty$. However, the price changes are less heavily tailed for larger time increments, and therefore the likelihood is never 0. In picture 5.5, the realisation of the price behaviour for hour-12 contract for different days is compared with five simulations of the diffusion process whose parameters are inferred from the 1-minute interval. The prices of the contracts are set to start from 0, and then the cumulative sum of the increments is shown for the different contracts. It can be noted that, as discussed, this model cannot replicate the large increments in the prices both for the whole-day simulation, picture (5.5b) and the simulation in the last 180-minutes, (5.5c).



(a) 1-minute volume-weighted average of the price of the hour 12 contract from June 29^{th} 2020 to July 3^{rd} 2020/



(b) Simulations of the diffusion process in the scale of grey vs the realisation of five contracts



(c) Cumulative increments of the contract in the last 3 hours of trading vs the simulation of five diffusion processes.

Figure 5.5: Cumulative increments of the hour 12 contract volume-weighted prices between 29^{th} 2020 to July 3^{rd} 2020 vs the simulation of diffusion process in a whole day 5.5b and the last three hours of trading 5.5c.

Table 5.2: AIC scores for the Diffusion model for the different hours and intervals

	1	5	15	30	60
1	inf	inf	inf	46106.92	21363.3
2	inf	inf	85892.75	45795.66	21400.3
3	inf	inf	92801.78	49781.8	24149.1
4	inf	inf	102321.89	54138.92	26457.8
5	inf	inf	108761.2	56369.1	28235.85
6	inf	inf	inf	70139.98	33499.46
7	2499632.96	510525.08	151446.77	75798.85	36447.52
8	inf	inf	inf	inf	inf
9	inf	inf	inf	inf	inf
10	inf	inf	inf	inf	inf
11	inf	inf	inf	inf	45512.26
12	inf	inf	173882.1	inf	inf
13	inf	inf	179711.28	98058.41	51323.68
14	inf	inf	185925.62	99539.36	52373.03
15	inf	inf	199252.88	108537.01	56061.83
16	inf	inf	inf	114114.14	60503.68
17	inf	inf	228906.57	125543.44	64493.66
18	inf	inf	inf	inf	inf
19	inf	inf	inf	inf	inf
20	inf	inf	inf	inf	inf
21	inf	inf	inf	inf	73732.81
22	inf	inf	inf	inf	inf
23	inf	inf	inf	inf	inf
24	inf	inf	inf	138395.05	70545.27

Table 5.3: BIC scores of the diffusion model.

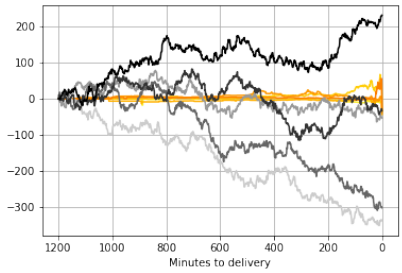
	1	5	15	30	60
1	inf	inf	inf	58512.92	27199.3
2	inf	inf	114280.75	59623.66	27948.3
3	inf	inf	124187.78	65107.8	31445.1
4	inf	inf	136627.89	70924.92	34483.8
5	inf	inf	145987.2	74615.1	36991.85
6	inf	inf	inf	89845.98	42985.46
7	3155898.96	641191.08	194512.77	96964.85	46663.52
8	inf	inf	inf	inf	inf
9	inf	inf	inf	inf	inf
10	inf	inf	inf	inf	inf
11	inf	inf	inf	inf	58648.26
12	inf	inf	231548.1	inf	inf
13	inf	inf	240297.28	127984.41	65919.68
14	inf	inf	249431.62	130925.36	67699.03
15	inf	inf	265678.88	141383.01	72117.83
16	inf	inf	inf	148420.14	77289.68
17	inf	inf	301172.57	161309.44	82009.66
18	inf	inf	inf	inf	inf
19	inf	inf	inf	inf	inf
20	inf	inf	inf	inf	inf
21	inf	inf	inf	inf	94168.81
22	inf	inf	inf	inf	inf
23	inf	inf	inf	inf	inf
24	inf	inf	inf	184381.05	93171.27

5.1.2. Jump Diffusion Model

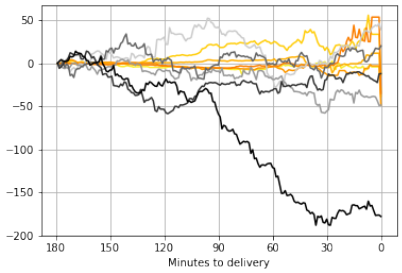
The Jump diffusion model is a Lévy process of the form

$$L_t = \gamma t + \sigma \mathbb{W}_t + \sum_{i=1}^{N_t} J_i \tag{5.2}$$

To estimate the process parameters, I proceed as follows. The diffusion term is constrained to a standard normal distribution with a mean of 0 and a variance of 1. Thus, by increasing the number of free parameters by 1, namely the jump intensity, it was possible to improve the simple diffusion model in terms of score in the AIC (table (5.4)) and the BIC (table 5.5)). The Jump size is a normal distribution whose parameters are the maximum likelihood estimators for the variance and the mean. To find the jump intensity, the expected number of jumps in a unit of time, the domain for lambda is restricted to the set of integers between 1 and 10 and the value that maximises the likelihood is chosen. The density function is approximated through the COS method; see [21] chapter 6 for an extended discussion on the topic. In a few words, it approximates the density function by transforming the characteristic function, known in closed form for jump-diffusion processes. However, the price distribution could not capture the tail behaviour of the price increments for most cases yielding, therefore, infinite values for the AIC and BIC score. The number of times this happens is larger than the diffusion model (109 vs 80) due to the COS method's slow convergence for the approximation of the values in the tail of the distribution making its use limited by this factor.



(a) In the scale of grey, the simulated paths of the Jump Diffusion process vs some realisations of the hour-12 contract for the whole trading day.



(b) In the scale of grey, the simulated paths of the Jump Diffusion process vs some realisations of the hour-12 contract for the whole trading day.

Figure 5.6: Simulations of the Jump-Diffusion process for different windows of time.

Table 5.4: AIC scores for the jump-diffusion model.

	1	5	15	30	60
1	inf	inf	inf	inf	18224.4
2	inf	inf	inf	37391.14	19387.43
3	inf	inf	inf	40540.33	21166.9
4	inf	inf	inf	inf	22780.08
5	inf	inf	inf	46572.89	24828.03
6	inf	inf	inf	inf	27762.82
7	inf	inf	110992.63	60478.4	32040.52
8	inf	inf	inf	inf	inf
9	inf	inf	inf	inf	inf
10	inf	inf	inf	inf	inf
11	inf	inf	inf	inf	38121.61
12	inf	inf	inf	inf	inf
13	inf	inf	inf	inf	inf
14	inf	inf	142032.94	inf	inf
15	inf	inf	inf	inf	inf
16	inf	inf	inf	inf	inf
17	inf	inf	inf	inf	inf
18	inf	inf	inf	inf	inf
19	inf	inf	inf	inf	inf
20	inf	inf	inf	inf	inf
21	inf	inf	inf	inf	inf
22	inf	inf	inf	inf	inf
23	inf	inf	inf	inf	inf
24	inf	inf	inf	inf	inf

Table 5.5: BIC scores for the jump-diffusion model.

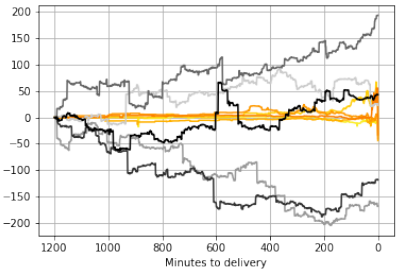
	1	5	15	30	60
1	inf	inf	inf	inf	26978.4
2	inf	inf	inf	58133.14	29209.43
3	inf	inf	inf	63529.33	32110.9
4	inf	inf	inf	inf	34819.08
5	inf	inf	inf	73941.89	37962.03
6	inf	inf	inf	inf	41991.82
7	inf	inf	175591.63	92227.4	47364.52
8	inf	inf	inf	inf	inf
9	inf	inf	inf	inf	inf
10	inf	inf	inf	inf	inf
11	inf	inf	inf	inf	57825.61
12	inf	inf	inf	inf	inf
13	inf	inf	inf	inf	inf
14	inf	inf	237291.94	inf	inf
15	inf	inf	inf	inf	inf
16	inf	inf	inf	inf	inf
17	inf	inf	inf	inf	inf
18	inf	inf	inf	inf	inf
19	inf	inf	inf	inf	inf
20	inf	inf	inf	inf	inf
21	inf	inf	inf	inf	inf
22	inf	inf	inf	inf	inf
23	inf	inf	inf	inf	inf
24	inf	inf	inf	inf	inf

5.1.3. Normal Inverse Gaussian Model

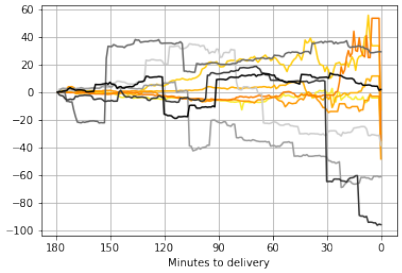
This model describes the distribution of the change in price as

$$L_t = t\mathbb{E}\left[L_1\right] + \int_0^t \int_{\mathbb{R}} x(\mu^L - \nu^{\mathcal{NIG}})(dx, ds) \tag{5.3}$$

where ν^{NIG} is the Lévy measure described by equation (2.18). The parameters of the increments were inferred through maximum likelihood estimation implemented in the package scipy.stats.norminvgauss.fit. This model uses four free parameters and allows for heavier tails than the two aforementioned. Indeed, the log-likelihood for this model is never infinite, and it achieves the highest number of selections for both the AIC and the BIC. Moreover, this model is closed under convolution, meaning that the sum of two independent random variables distributed as a $\mathcal{NIG}$ distribution with the same α and β, and possibly different μ and δ, then the resulting distribution is still $\mathcal{NIG}$ distributed with parameters $(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$. Therefore defining the distribution for a time scale automatically defines the distribution at every time scale: a 5 minute increment can be seen as the sum of 5 independent and identically distributed increments of 1 minute. Table 5.6 shows the AIC scores, and table 5.7 the BIC scores. Figure (5.7) shows the simulations of five paths generated by the model and the realisation of 5 contracts, and it can be seen that this model captures the large increment behaviour of the prices.



(a) In the scale of grey, the simulated paths of the Normal Inverse Gaussian process vs some realisations of the hour-12 contract for the whole trading day.



(b) In the scale of grey, the simulated paths of the Normal Inverse Gaussian process vs some realisations of the hour-12 contract for the last three hours of a trading day.

Figure 5.7: Simulations of the Normal Inverse Gaussian process for different windows of time.

Table 5.6: AIC scores for the Normal Inverse Gaussian model.

	1	5	15	30	60
1	231432.19	132941.44	59825.85	33909.54	17987.87
2	204076.14	134621.36	62719.38	36255.84	19404.33
3	190474.48	141264.8	67059.72	39033.56	21013.41
4	178564.96	146527.47	70960.69	41465.01	22451.16
5	188134.75	155414.26	76061.7	44564.17	24622.02
6	209303.23	167248.66	83329.	49614.51	27314.33
7	272955.91	188989.39	93469.71	55429.35	30989.24
8	322288.95	204175.36	102095.72	60655.93	33819.22
9	361302.56	216623.29	108007.53	64678.22	35986.21
10	338022.84	212524.84	107515.59	64852.8	36573.42
11	318362.29	212752.43	107596.67	65447.82	37288.98
12	310223.07	216981.02	110107.63	67683.79	38603.37
13	336025.22	225984.99	114035.09	69573.36	40053.26
14	328315.37	232085.7	117633.22	71939.28	41515.24
15	329680.7	238853.37	121109.92	73845.41	42636.6
16	313326.98	239439.43	123761.25	76298.51	44382.97
17	294053.01	241016.6	127097.6	78866.23	46293.58
18	297593.78	251708.74	133990.02	83515.58	49084.84
19	322478.07	261014.47	138735.45	87535.53	51602.48
20	377346.07	276336.	144953.95	90720.02	53652.96
21	1608423.47	272922.53	145313.89	90677.72	54047.74
22	222558.41	253525.39	140863.59	89285.89	53530.45
23	190651.52	252257.74	142679.67	91014.87	54657.05
24	139521.88	240414.61	135523.13	85734.3	51357.7

Table 5.7: BIC scores for the Normal Inverse Gaussian process.

	1	5	15	30	60
1	1018364.19	289153.44	110917.85	58721.54	29659.87
2	1076212.14	307877.36	119495.38	63911.84	32500.33
3	1152606.48	332516.8	129831.72	69685.56	35605.41
4	1228296.96	355299.47	139572.69	75037.01	38503.16
5	1325466.75	381706.26	150513.7	81056.17	42134.02
6	1434235.23	411060.66	163621.	89026.51	46286.33
7	1585487.91	450321.39	179601.71	97761.35	51421.24
8	1722420.95	483027.36	194067.72	105907.93	55711.22
9	1849034.56	512995.29	205819.53	112850.22	59338.21
10	1913354.84	526416.84	211167.59	115944.8	61385.42
11	1981294.29	544164.43	217088.67	119459.82	63560.98
12	2060755.07	565913.02	225439.63	124615.79	66335.37
13	2174157.22	592436.99	235207.09	129425.36	69245.26
14	2254047.37	616057.7	244645.22	134711.28	72167.24
15	2343012.7	640345.37	253961.92	139537.41	74748.6
16	2414258.98	658451.43	262453.25	144910.51	77954.97
17	2482585.01	677548.6	271629.6	150398.23	81325.58
18	2573725.78	705760.74	284362.02	157967.58	85576.84
19	2686210.07	732586.47	294947.45	164907.53	89554.48
20	2828678.07	765428.	307005.95	171012.02	93064.96
21	4147355.47	779534.53	313205.89	173889.72	94919.74
22	2849090.41	777657.39	314595.59	175417.89	95862.45
23	2904783.52	793909.74	322251.67	180066.87	98449.05
24	2941253.88	799586.61	320935.13	177706.3	96609.7

5.1.4. Generalised Hyperbolic Model

This model describes the distribution of the changes in price as

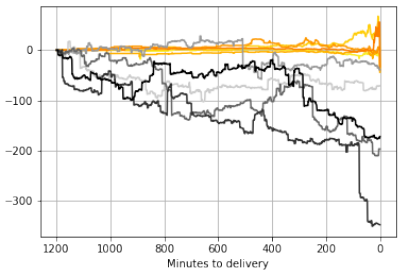
$$L_t = t\mathbb{E}\left[L_1\right] + \int_0^t \int_{\mathbb{R}} x(\mu^L - \nu^{GH})(dx, ds) \tag{5.4}$$

where ν^{GH} is the Lévy measure described by equation (2.16. The parameters of the increments were inferred through maximum likelihood estimation implemented in the package scipy.stats.genhyperbolic.fit. This model uses five parameters and to find the maximum likelihood estimation of λ it is necessary to calculate a large amount of Bessel functions which is computationally expensive [23]. This class of distributions is closed under affine transformations meaning that if

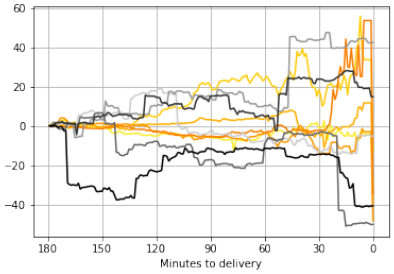
$$Y = aX + b \quad a \in \mathbb{R}^+, \quad b \in \mathbb{R}, \tag{5.5}$$

and X follows a Generalised Hyperbolic distribution, then also Y belongs to this class of distributions. Also, in this case, the AIC scores (table (5.8)) and BIC scores (table (5.9)) are not infinite, and table (5.10) shows that this model is to be preferred to the Normal Inverse Gaussian model in some scenarios. In figure 5.8, simulations of the process are shown in comparison with some of the realisation of the hour 12 contract. Even in this case, the model captures the tails of the short-time behaviour of the price increments. However, in this case, the full-day simulation yields increments far away from the

actual realisations.



(a) In the scale of grey, the simulated paths of the Generalised Hyperbolic process vs some realisations of the hour-12 contract for the whole trading day.



(b) In the scale of grey, the simulated paths of the Generalised Hyperbolic process vs some realisations of the hour-12 contract for the last three hours of a trading day.

Figure 5.8: Simulations of the Generalised Hyperbolic process for different windows of time.

Table 5.8: AIC scores for the Generalised Hyperbolic model

	1	5	15	30	60
1	195702.7	131260.49	59707.81	33942.28	18042.49
2	179417.68	133263.98	62810.19	36381.16	19533.57
3	161886.32	138985.7	66811.01	39023.57	21046.36
4	148612.86	145309.11	70700.47	41445.91	22486.31
5	165918.21	154190.31	79985.86	44614.77	24717.82
6	187203.28	166222.76	83238.76	49700.43	27397.69
7	259563.39	188865.6	94321.41	56001.17	31330.11
8	298424.69	202947.63	101859.7	60700.61	36036.13
9	337903.74	216210.24	107916.94	64705.91	36052.01
10	325618.42	213230.28	107238.89	64881.46	36655.7
11	307977.97	213004.19	107583.52	65621.05	37453.83
12	305768.75	218485.55	110118.81	67960.07	41556.94
13	332748.08	227475.04	114162.4	69660.76	43810.3
14	308057.72	233218.61	117552.52	72018.69	41609.1
15	302846.33	239248.72	120833.4	73849.25	42686.06
16	295944.31	241150.07	123298.28	76290.96	44461.42
17	273919.73	247325.88	127148.24	79084.91	52430.32
18	274192.62	253657.07	134318.99	83816.16	49255.35
19	331767.85	264770.27	140508.19	88549.02	52262.26
20	362882.84	278173.66	146186.45	91187.2	54003.47
21	348669.92	278675.07	147908.65	91862.6	54625.83
22	235405.4	260685.07	143610.67	90507.29	54165.22
23	194542.73	260619.09	144888.01	91890.79	55063.5
24	128444.01	250378.94	139893.36	87753.05	52097.38

Table 5.9: BIC scores of the Generalised Hyperbolic distribution

	1	5	15	30	60
1	1179367.7	326525.49	123572.81	64957.28	32632.49
2	1269587.68	349833.98	133780.19	70951.16	35903.57
3	1364551.32	378050.7	145276.01	77338.57	39286.36
4	1460777.86	406274.11	156465.47	83410.91	42551.31
5	1587583.21	437055.31	173050.86	90229.77	46607.82
6	1718368.28	470987.76	183603.76	98965.43	51112.69
7	1900228.39	515530.6	201986.41	108916.17	56870.11
8	2048589.69	551512.63	216824.7	117265.61	63401.13
9	2197568.74	586675.24	230181.94	124920.91	65242.01
10	2294783.42	605595.28	236803.89	128746.46	67670.7
11	2386642.97	627269.19	244448.52	133136.05	70293.83
12	2493933.75	654650.55	254283.81	139125.07	76221.94
13	2630413.08	685540.04	265627.4	144475.76	80300.3
14	2715222.72	713183.61	276317.52	150483.69	79924.1
15	2819511.33	741113.72	286898.4	155964.25	82826.06
16	2922109.31	764915.07	296663.28	162055.96	86426.42
17	3009584.73	792990.88	307813.24	168499.91	96220.32
18	3119357.62	821222.07	322283.99	176881.16	94870.35
19	3286432.85	854235.27	335773.19	185264.02	99702.26
20	3427047.84	889538.66	348751.45	191552.2	103268.47
21	3522334.92	911940.07	357773.65	195877.6	105715.83
22	3518570.4	915850.07	360775.67	198172.29	107080.22
23	3587207.73	937684.09	369353.01	203205.79	109803.5
24	3630609.01	949343.94	371658.36	202718.05	108662.38

	NIG	GH	Diffusion	Jump Diffusion
Selection by AIC	75	44	0	1
Selection by BIC	91	1	19	9

Table 5.10: Number of times a model is selected according to the AIC and the BIC. A model is selected if its score in the criterion is the lowest for a particular choice of trading intervals and hours.

Table 5.10 summarises the times a model is selected according to the AIC and the BIC. In both cases, the Normal Inverse Gaussian Model represents the best model confirming once again the tail-heaviness behaviour of the process governing the energy contract price. Moreover, its closure under convolution allows for a forecast of a price distribution for every time scale.
The predictions of this model, however, are based exclusively on the past behaviour of the contracts and do not consider any daily peculiarity. To overcome this, in the next section, the relationship between the volatility of a contract and the change in weather forecasting is presented. Moreover, the presence of mean-reversion or trend-following in the price of a contract is analysed.

6 Fundamental Analysis

Before applying the NIG model to the prediction of the index ID_1 and ID_3, it is sensible to analyse possible relationships between the change in the forecast of Energy Production and Consumption and the behaviour of the market. In this way, the model can be adapted to the daily behaviours of the contract and possibly yield better forecasting. Namely, in the first subsection, a relationship between the volatility of the contract and the changes in the forecasting is looked for, while in the second one, the presence of either mean-reversion or trend following is analysed.

6.1. Energy Production Forecasting and Market Behaviour

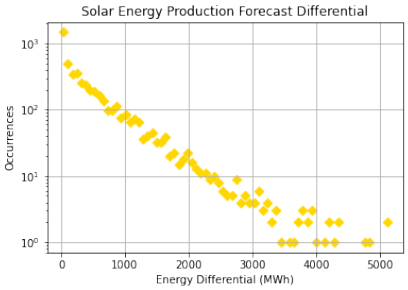
The ID market depends on the imbalances arising after the closure of the Day-Ahead market. Imbalances can depend on different factors, and here I present the analysis regarding the forecasting of solar and wind production and the forecasting of consumption. Table 6.1 presents a summary of the available data description. Wind forecasting and Solar forecasting are given as the total amount of energy produced by these sources for 15-minute intervals in the whole country; consumption forecasting is the forecasting of the total consumed energy in one hour of time.

Data	Interval	Update	Country
Solar Day-Ahead	15 minutes	Last update day before	Germany
Wind Day-Ahead	15 minutes	Last update day before	Germany
Consumption Day-Ahead	60 minutes	Last update day before	Germany
Solar Intraday	15 minutes	15 minutes	Germany
Wind Intraday	15 minutes	15 minutes	Germany
Consumption Intraday	60 minute	60 minute	Germany

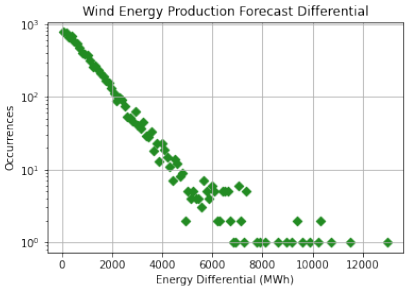
Table 6.1: Description of the data of the available forecast, the column interval shows the granularity of the data (15 minutes means that the forecast is available for a period of 15 minutes), update how often the forecast is updated.

As discussed in the previous sections, in the ID market phase, Balance Responsible Parties (BRPs) try to sell their surplus or compensate for shortages. The position of a BRP that produces Electric Energy through Renewable Sources (RES) is submitted after the closure of the DA market, but the production of energy depends on the weather forecasting for the coming day. However, as it can be seen in figure (6.1), this can change significantly, and so, the producer needs to buy power for shortages and sell the excesses. In the same way, a supplier for households has to forecast the consumption of

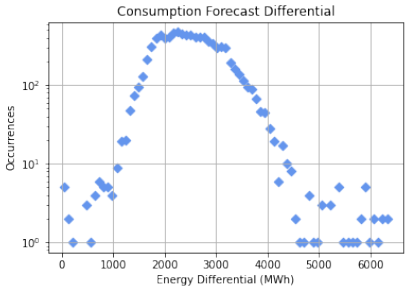
its customers. This leads to the fact that most of the energy, for all hours, is traded in the last three hours of trading. Picture 6.2 shows this fact for all the hours of 2020.



(a) Solar forecast differential between the Day-Ahead forecast and the forecast three hours before delivery.

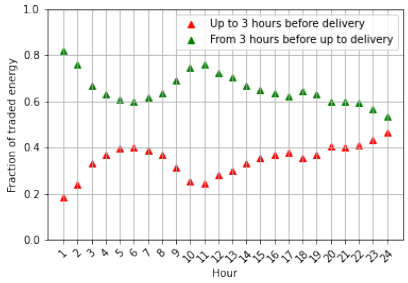


(b) Wind forecast differential between the Day-Ahead forecast and the forecast three hours before delivery.

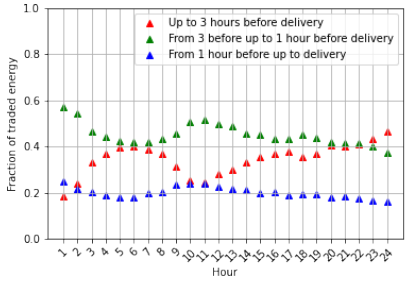


(c) Consumption forecast differential between the Day-Ahead forecast and the forecast three hours before delivery.

Figure 6.1: Forecast differentials between Day-Ahead and three hours before delivery for Solar (6.1a), Wind ((6.1b)) and Consumption (6.1c).



(a) Share of energy traded up to three hours before delivery and one hour before.



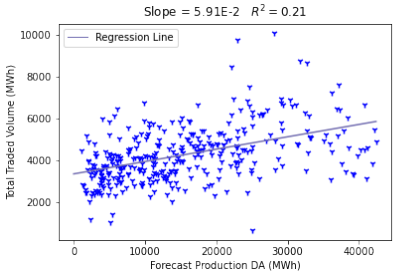
(b) Share of energy traded up to three hours before delivery, from three hours up to one hour before delivery, and in the last hour of trading.

Figure 6.2: Shares of energy traded for different sub-intervals of a trading day.

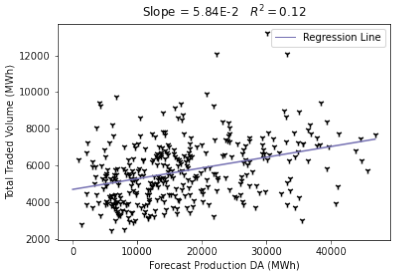
This suggests that there might be an association between the amount of volume traded, the market behaviour, and the differential in forecasting. In particular, I focus on the volatility of the contract and the total traded volume. I started by aggregating the quarter-hourly intervals for Wind (WF_{hq}) and Solar forecasting (SF_{hq}) for both Intraday and Day-ahead in hour intervals using the following formula

$$WF_h = \sum_{q=1}^{4} WF_{hq}. \tag{6.1}$$

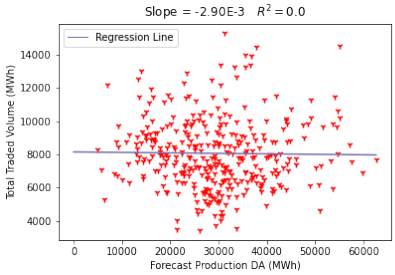
Thus, now Solar and Wind forecasting concern periods that are 1 hour long. The first possible relationship I examined was the one between the total forecast volume of Solar and Wind production in the DA market ($PF = SF_h^{DA} + WF_h^{DA}$), and the total traded volume in the last three hours of the ID market for each hour. In figure 6.3 the results are shown for hour 1, hour 8, hour 12, and hour 18.



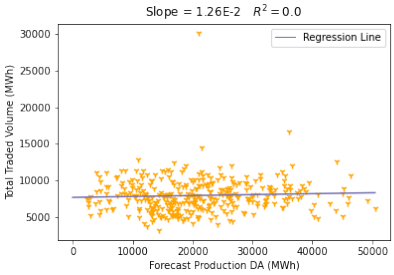
(a) Total forecast volume and total traded volume for hour 1.



(b) Total forecast volume and total traded volume for hour 8.



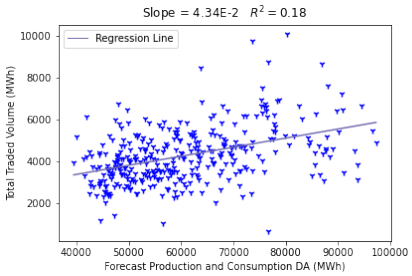
(c) Total forecast volume and total traded volume for hour 12.



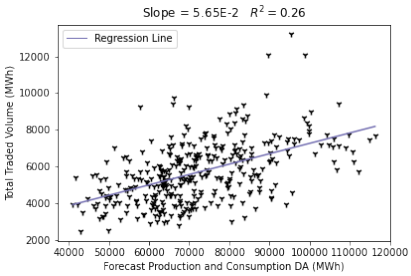
(d) Total forecast volume and total traded volume for hour 18.

Figure 6.3: Each figure shows the forecast energy production as the sum of Wind and Solar (x-axis) and the total volume traded in the last three hours (y-axis) for an hour of the day for the year 2020. hour 1 is shown in figure 6.3a, hour 8 in figure 6.3b, hour 12 in figure 6.3c, and hour 18 in figure 6.3d.

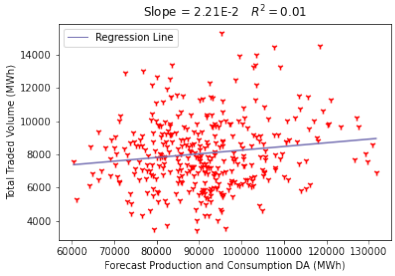
For each hour, the correlation coefficient between the two variables is close to 0, and the R^2 is as high as 0.21 for hour 1 but 0 for hours 12 and 18. It can be concluded that the total forecast volume is not an explanatory variable for the total traded volume. By adding the DA forecast consumption CF_h^{DA} to the forecast production, yielding $CF_h^{DA} + SF_h^{DA} + WF_h^{DA}$. The results can be seen in figure (6.4).



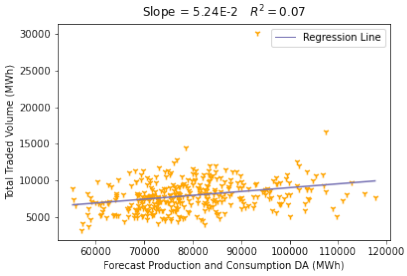
(a) Total forecast and total traded volume for hour 1.



(b) Total forecast and total traded volume for hour 8



(c) Total forecast and total traded volume for hour 12



(d) Total forecast and total traded volume for hour 18.

Figure 6.4: Each figure shows the forecast traded energy as the sum of the wind, solar, and consumption forecasting (x-axis) and the total volume traded in the last three hours (y-axis) for an hour of the day for the year 2020. hour 1 is shown in figure 6.4d, hour 8 in figure 6.4d, hour 12 in figure 6.4d, and hour 18 in figure 6.4d.

Also in this case the linear relationship appears weak, suggesting that the volume traded in the last three hours of trading does not depend on the forecast of the day before. So, a possible explanatory

variable could be the differential between the forecasting available during the day ahead and those available during the ID market. To examine this, I find the total imbalance (TI) calculated as

$$TI_h = |(WF_h^{ID} - WF_h^{DA}) + (SF_h^{ID} - SF_h^{DA}) - (CF_h^{ID} - CF_h^{DA})| \tag{6.2}$$

The considered forecast for the ID market is the forecast available three hours before the start of the delivery. If the value is not available, then the value used as a reference for the ID forecast is the forecast available at the moment closest to the three hours before delivery. For example, if the delivery starts at noon, then the forecast is the one at 9:00 if it is available. Otherwise, if there is a forecast available at 9:01 and one at 8:46, then the former is chosen. When close to delivery, the forecast is accurate, so this should not influence the final result. The total imbalance is then compared to the Total traded volume in the last three hours. Results are available in figure 6.5.

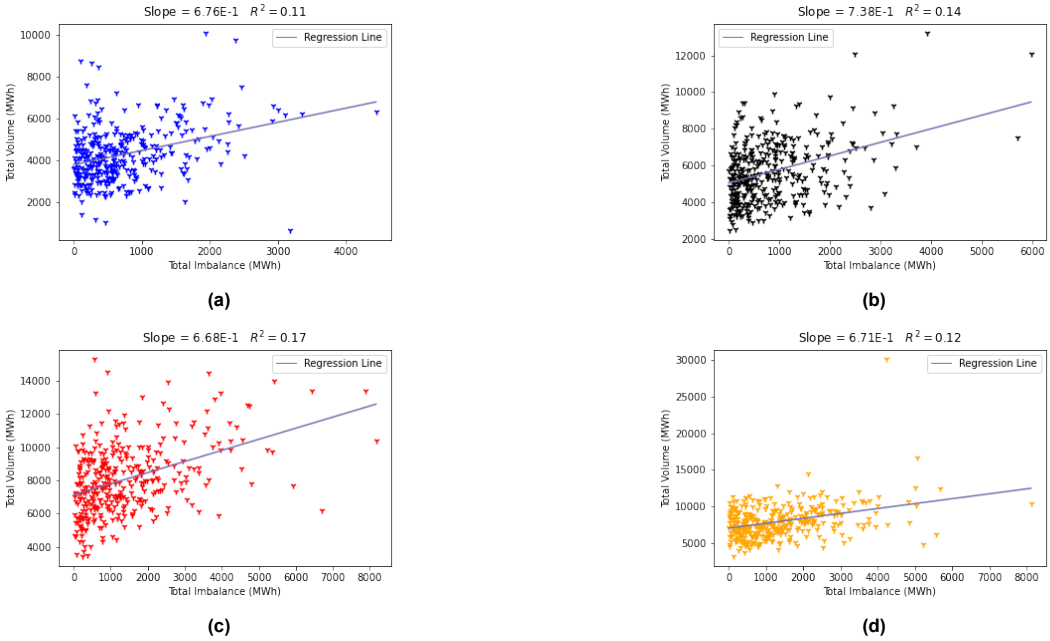


Figure 6.5: Each figure shows the total imbalance between DA and ID (x-axis) and the total volume traded in the last three hours (y-axis) for an hour of the day for the year 2020. hour 1 is shown in figure 6.5a, hour 8 in figure 6.5b, hour 12 in figure 6.5c, and hour 18 in figure 6.5d.

.

Again, data suggest a low correlation between the two compared quantities, and therefore a linear relationship between the total traded volume and the forecast or the total imbalance can be excluded. Finally, in figure (6.6) and figure (6.7), I show the relationship between the standard deviation of the increments of the prices in the last three hours and, respectively, the volume traded in the last three hours and the total imbalance. This relationship is analysed in order to find a possible bound between the volatility of the contract in the last three hours of trading and the fundamentals. the close to 0 slopes and R^2 make us refuse the linear relationship between the two quantities in all the cases.

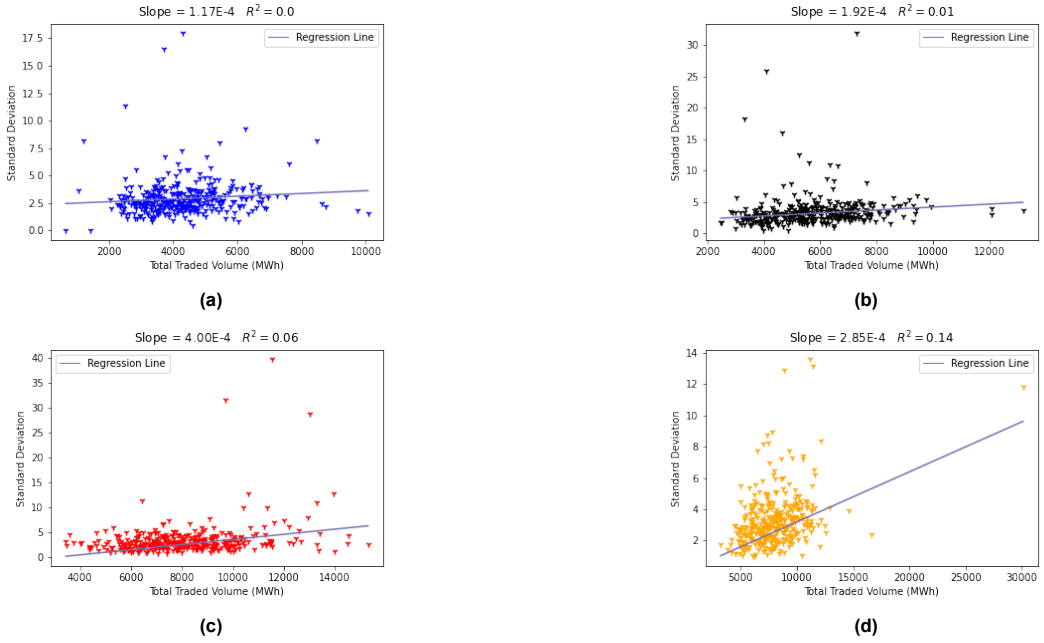


Figure 6.6: Each figure shows the total traded volume in the last three hours (x-axis) and the standard deviation of the 1-minute increment of the prices (y-axis) for an hour of the day for the year 2020. hour 1 is shown in figure 6.6a, hour 8 in figure 6.6b, hour 12 in figure 6.6c, and hour 18 in figure 6.6d.

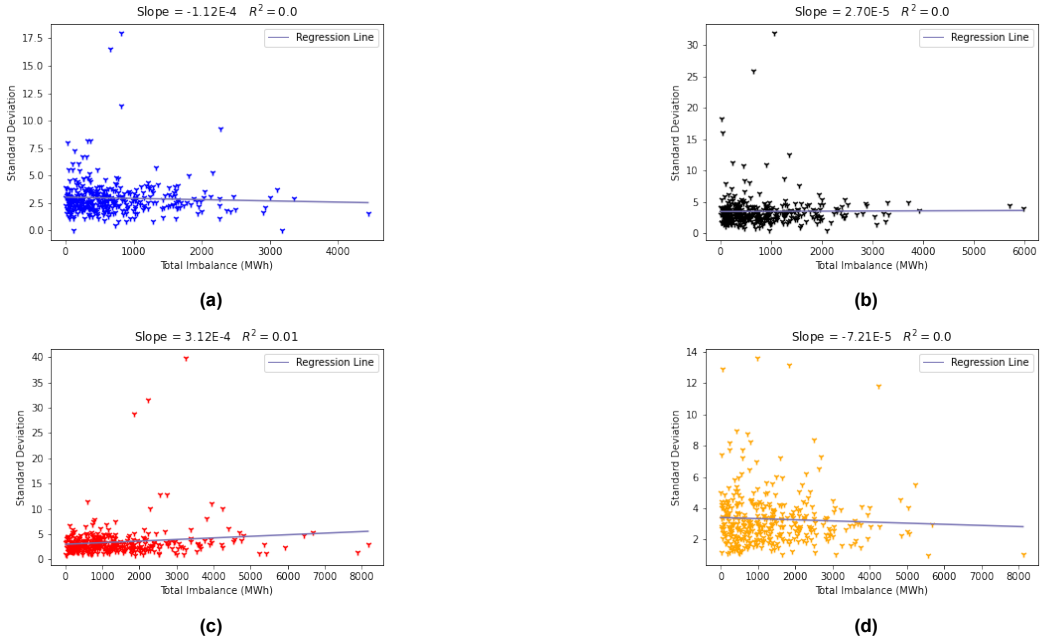


Figure 6.7: Each figure shows the total imbalance (x-axis) and the standard deviation for the price of a contract (y-axis) for an hour of the day for the year 2020. hour 1 is shown in figure 6.7a, hour 8 in figure 6.7b, hour 12 in figure 6.7c, and hour 18 in figure 6.7d.

6.2. Mean Reversion and Trend Following

To improve the model's forecasting power, the presence of a trend or mean-reversion can be looked for in the price of a contract. These behaviours can be easily integrated into the model by adding a drift

component μ to the Stochastic Differential Equation, i.e.

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t dL_s. \tag{6.3}$$

Mean Reversion describes the property of a stochastic process to oscillate around a value corresponding to its average. At the same time, the presence of a trend implies that the expected value of the future price of the contract will be higher (lower) if the trend is positive (negative).

Mean Reversion

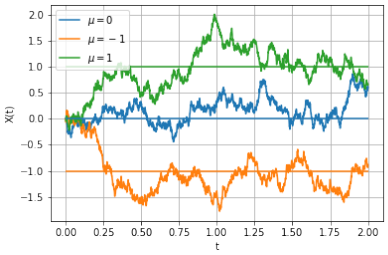
Mean reversion mathematically translates to a stochastic drift term of the form

$$\mu(t, X_t) = \kappa(\theta - X_{t-}).$$

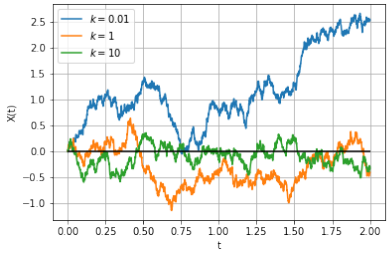
where θ is the long-term mean of the process and κ is the speed of mean reversion. These two parameters define, respectively, which value the process oscillates around and at which speed the process returns to its average. The meaning of this formulation is that if the process assumes values smaller than its average, then the drift term is positive, implying that the process will more likely return to its mean rather than depart from it. Conversely, if the values the stochastic process assumes are higher, the drift term is negative. Figure 6.8 shows an example of the process

$$\begin{cases} dX_t = k(\theta - X_{t-})dt + \gamma d\mathbb{W}_t \\ X_0 = 0 \end{cases} \tag{6.4}$$

for different parameters μ, figure 6.8a, and k, figure 6.8b .



(a) Simulations of the process described by the SDE 6.4 for different values of the mean μ. $k = 4$ for each of the processes.



(b) Simulations of the process described by the SDE 6.4 for different values of the speed k. $\mu = 0$ for each of the processes. Note that the lower the speed of mean reversion, the more the process diverts from the average.

Figure 6.8: Simulations of the SDE 6.4 for different values of the mean (6.8a) and speed (6.8b).

In the case of this thesis, the problem consists of verifying if the process governing the price is of the form

$$\begin{cases} dX_t = \kappa(\theta - X_{t-})dt + dL_t \\ X_0 = D \quad D \in \mathbb{R} \end{cases} \tag{6.5}$$

where L_t is the Levy process governing the jump activity of the process. To conduct this analysis for every contract and each day, I proceeded as follows. I took the volume-weighted average price for each contract for one minute. The last recorded price is used if no trade happens during an interval. Then, I performed the Augmented Dickey-Fuller test for each contract. This test considers a time series and, as the null hypothesis, the independence of an increment from the last value assumed by the time series. Table (6.2) shows that the hypothesis for which increments are independent of previous values cannot be refused in most cases, meaning that the price behaviour cannot be explained through mean-reversion.

Confidence Interval	1-minute Interval	5-minute Interval	15-minute Interval
0.1	2431 (27.68%)	1570 (17.87%)	1901 (21.64%)
0.05	1996 (22.72%)	1158 (13.18%)	1362 (15.51%)
0.01	1221 (13.90%)	597 (6.80%)	726 (8.27%)

Table 6.2: Results of the augmented Dickey-Fueller test for different confidence intervals (row) and trading intervals

6.2.1. Trend Following

The trend following behaviour corresponds to a drift component of the form

$$\mu(t, X_t) = \theta \quad \theta \in \mathbb{R} \tag{6.6}$$

resulting in the following process

$$\begin{cases} dX_t = \theta dt + dL_t \\ X_0 = D \quad D \in \mathbb{R} \end{cases} \tag{6.7}$$

The expected value of the process at time t is then given by

$$\mathbb{E}[X(t)] = D + \theta t + \mathbb{E}\left[\int_0^t dL_t\right] \tag{6.8}$$

To test the hypothesis of the presence of the drift, I performed the Mann-Kendall test for all hours in the year 2020 and different lengths of trading intervals. Table (6.3) shows the results for the 15-minute interval. The 'No Trend' column represents the total number of contracts for which the null hypothesis is accepted at that confidence interval. Conversely, 'Increasing' and 'Decreasing' is the totality of the contracts for which the alternative hypothesis is accepted. 'Total Trend' is the sum of contracts showing an increasing or decreasing trend. These results indicate that it is sensible to consider a daily trend in the data and integrate the model selected in the previous section with a drift component of the form shown in equation (6.6).

Confidence Interval	No Trend	Increasing	Decreasing	Total Trend
0.1	1736	3792	3256	7048 (80.24%)
0.05	2054	3616	3114	6730 (76.62%)
0.01	2683	3299	2802	6101 (69.46%)

Table 6.3: Results of the Mann-Kendall test for various confidence intervals and 15-minute intervals. Most of the contracts indicate the presence of a trend

6.3. Drift Model

After concluding that the trend-following hypothesis cannot be refused in most of the cases for the year 2020, in this section, I proceed to integrate the finding into the Normal Inverse Gaussian model by adding a drift component. Formally, the model will be of the form

$$X_t = X_0 + \theta t + \int_0^t \sigma dL_t \tag{6.9}$$

where θ is the drift, σ, the volatility and the noise parameter dL_t with distribution the Normal Inverse Gaussian distribution. Two different assumptions are made regarding the noise: the first considers a model where dL_t is assumed to be centralised with mean 0 and variance 1 (Model Noise-Drift-Variance); the other, where only the mean is constrained and set equal to 0 (Model Noise-Drift).

To infer the parameter of the Noise-Drift-Variance model, I proceed as follows. As discussed in the previous section, I discretize every trading day into trading intervals which leads to a process of the form

$$X_i = X_{i-1} + \theta \Delta t + \sigma \Delta L_i \quad i \in \{1, ..., n\} \tag{6.10}$$

where Δt is the length of the trading interval, X_i is the volume-weighted average price of the trades that happened in the $i-th$ interval, and $\Delta L \overset{iid}{\sim} \mathcal{NIG}(\alpha, \beta, \mu, \delta)$ is the noise of the price. The last registered volume-weighted average price is used if no trade happens during a given interval. To find an estimator of σ the fact that

$$\sigma \Delta L_i = X_i - X_{i-1} - \theta \Delta t \tag{6.11}$$

is used. Since the ΔL_i are assumed to possess mean 0 and variance 1, then

$$\begin{aligned} \sigma^2 = \sigma^2 Var(\Delta L_i) = Var(\sigma \Delta L_i) = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{i-1} - \theta\Delta t)^2 = \\ = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{i-1} - \theta\Delta t)^2 = \frac{1}{n}\sum_{i=1}^{n}(\Delta X_i - m)^2. \end{aligned} \tag{6.12}$$

where n+1 is the number of trading intervals in a trading day for a particular contract. The first equality comes from the ΔL_i are *iid*, and in the last passage, I defined $m = \theta \Delta t$. Then I proceeded by finding the value m that would minimise the variance, and thus, I want to find where

$$\frac{\partial \frac{1}{n}\sum_{i=1}^{n}(\Delta X_i - m)^2}{\partial m} = 0 \tag{6.13}$$

is achieved. Simple calculations lead to

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n}(\Delta X_i) = \frac{X_n - X_0}{n} \tag{6.14}$$

This is a point of minimum for the variance because

$$\frac{\partial^2 \frac{1}{n}\sum_{i=1}^{n}(\Delta X_i - m)^2}{\partial m^2} = 1 > 0 \quad \forall m \in \mathbb{R} \tag{6.15}$$

and thus, the empirical variance is a convex function in m. The estimate of the variance is thus

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\Delta X_i - \hat{m})^2. \tag{6.16}$$

To estimate the noise parameters, I make use of the fact that the Normal Inverse Gaussian Distribution is closed under affine transformations (this applies to the Generalised Hyperbolic Distribution as a whole class). Specifically, the available data regards the increment

$$Y_i = X_{i+1} - X_i = m + \sigma \Delta L_i \tag{6.17}$$

So, if $\Delta L_i \sim \mathcal{NIG}(\alpha, \beta, \mu, \delta)$, then $Y_i \sim \mathcal{NIG}(\tilde{\alpha}, \tilde{\beta}, \tilde{\mu}, \tilde{\delta})$ with

$$\tilde{\alpha} = \frac{\alpha}{\sigma} \quad \tilde{\beta} = \frac{\beta}{\sigma} \quad \tilde{\mu} = \mu\sigma + m \quad \tilde{\delta} = \sigma\delta \tag{6.18}$$

The procedure is then the following, for each day, the increments $X_{i+i} - X_i$ are aggregated, and the values of $\hat{m}$ and $\hat{\sigma}$ are found. Hence, the normalised increments are found as $\Delta L_i = \frac{\Delta X - \hat{m}}{\hat{\sigma}}$. Finally, the increments of the whole training set are aggregated into a vector of observations, and for these, the parameters $(\alpha, \beta, \mu, \delta)$ are looked for.
Since the noise is assumed to have 0 mean and variance equal to 1, parameters μ and δ are constrained as follows

$$\mathbb{E}[\Delta L_i] = \mu + \frac{\beta\delta}{\gamma} = 0 \quad \mathbb{V}ar(\Delta L_i) = \frac{\delta\alpha^2}{\gamma^3} = 1 \tag{6.19}$$

which imply

$$\mu = -\frac{\beta\delta}{\gamma} \quad \delta = \frac{\gamma^3}{\alpha^2} \tag{6.20}$$

To estimate the parameter of ΔL_t, the scipy.stats.norminvgauss.fit routine is applied to the observations. This method finds the initial estimates of the parameters via the method of moments and then, since there is no closed form for the maximum likelihood estimators of the NIG distribution, applies an optimisation method with objective function the negative log-likelihood. Finally, the inference of the parameters $(\tilde{\alpha}, \tilde{\beta}, \tilde{\mu}, \tilde{\delta})$ is done daily by finding the daily values for $(\hat{m}, \hat{\sigma})$, and applying equation (6.18). To infer the parameters of the Noise-Drift model, the same procedure is followed but with $\sigma = 1$ and only μ constrained. In the next section, comparisons between the three models are applied to predicting the ID_1 and ID_3 indexes.

7

Results

In the previous sections, I analysed the Germany Intraday Data to find a model to capture the price behaviour for different trading intervals. The Normal Inverse Gaussian distribution was selected most according to the Akaike and Bayesian Information Criteria. This suggests that the semi-heavy tails explain the tail behaviour of the price of a contract.
The model, however, models the change in price on a noise component which is exclusively determined by historical data, i.e. the maximum likelihood estimators of the distribution of the 2020 data. Hence, in the previous section, I looked at the historical data on possible relationships between the fundamentals (Renewable Energy and Consumption Forecasting), but no significant correlation was found. Data, however, suggest the presence of a drift in the price behaviour according to the Mann-Kendall test and the founding was integrated into the model. Thus, we now have three different solutions for modelling the price of a contract at time t (X_t): the Noise model. i.e.,

$$X_t = X_0 + \int_0^t dL_s, \tag{7.1}$$

the Noise-Drift model where

$$X_t = X_0 + \theta t + \int_0^t dLs, \tag{7.2}$$

and the Noise-Drift-Variance model with

$$X_t = X_0 + \theta t + \sigma \int_0^t dL_s. \tag{7.3}$$

where θ and σ are the drift and the variance and dL_s is a Lévy process that follows a Normal Inverse Gaussian distribution when $t = 1$. Each of these models forecasts the price of a contract at a future time t as a Normal Inverse Gaussian distribution with distinct parameters. Thus, it is relevant to compare them to find which forecasting is more reliable. Specifically, the models are tested on predicting the ID_1 and ID_3, the volume-weighted average of the prices of the transactions in the last hour and the last three hours before delivery, respectively. The goal is to give a trader a quantification of the probability of incurring losses of a certain extent; if they have a short position on a particular contract, upward movements in the price of the contract will damage them, causing a loss and, conversely, a long position would be damaged by the downward of the price. To determine which model better predicts these losses, I compare the model estimation of extreme events with the number of times extreme events happen. Hence, for each day, I compare the number of times the change in price falls beyond or above a certain threshold, and I compare the percentage of this event with the threshold. For example, the threshold at 90% is how often events with a probability greater than 0.9 occur. If the model is accurate, then about 10% of the events should fall in this range; if the number of events is higher, then the model underestimates the quantiles of the actual distribution. Otherwise, it overestimates them. If the model is accurate, a trader can evaluate their position one or three hours before delivery and account for different scenarios.

For example, Table (7.1) shows that the probability that increments estimated to happen in the 0.01% of the cases (> 99.99%) occur about in the 1% of the cases for model Noise calibrated on the 1-minute interval, 6% of the instances for model Noise-Drift, and 30.83% for model Noise-Drift-Variance. Specifically, the increment of the ID_1 with respect to the known price one hour before the delivery and the increment of the ID_3 with respect to the known price three hours before being considered, and then the probability is estimated for each of the models and confronted with the actual realisation. The following tables show the number of times that the realisation of the ID_1 and ID_3 index exceeds the forecast according to the different models calibrated on different trading intervals.
Tables (7.1) and (7.2) show that the model Noise model calibrated on 1-minute contract overestimates the tail probabilities in the left tail of the distributions but underestimates the right-tail. For the 5-minute increments model, Noise still presents this behaviour, but the accuracy of the right tail predictions has greatly increased for all thresholds (tables (7.3) and (7.4)). The calibration on 15-minute increments shown in tables (7.5) and (7.6) yields a good fit for the ID-3 predictions' estimating every percentile close to its empirical realisations. Also, the ID-1 tail predictions perform well for the more extreme percentile (0.01, 0.1, 99.9 and 99.99) but tend to underestimate the others. For example, the increments below the estimated $10 - th$ percentile account for the 35% of the realisations. This behaviour can also be seen for the models calibrated on the 30 and 60-minute intervals, but in both cases, the predictions of the ID-1 percentiles worsen (7.7) and (7.9). The ID-3 predictions for the 30-minute model in the Table (7.8) perform similarly to those of the 15-minute model, while the performance of the 60-minute model is worse for every percentile (7.10).
The model Noise-Drift improves the performance in predicting the percentile when compared with the model calibrated on the 60-minute interval both for ID-1 and ID-3 and with the left percentiles of the 1-minute increments where the Noise Model overestimates their values. It performs similarly to the model Noise when trained on 30-minute intervals. However, it underperforms in the rest of the situations, showing no improvement. The Model Noise-Drift-Variance underestimates the tail probabilities in every scenario by a large margin. These results suggest that the drift and the volatility in the first hours of trading are poor estimators of the price behaviour in the last hours.
In conclusion, the Noise and Noise-Drift models perform well when calibrated at 15-minute and 30-minute intervals in predicting ID-3. Concerning ID-1, the best results are obtained by the Noise model trained on the 5-minute intervals. Table (7.11) presents the parameters for the Noise-model calibrated on the 15-minute intervals and the 95% confidence intervals. The intervals are obtained by applying the convergence in distribution of the Maximum Likelihood Estimator to a Normal distribution with variance the inverse of the observed information matrix. I.e., for every parameter $\varphi \in \{\alpha, \beta, \mu, \delta\}$ and $\hat{\varphi}_i$ the MLE for the $i - th$ parameter, the value

$$\hat{\varphi}_i \pm c\sqrt{J^{-1}(\varphi)_{ii}} \tag{7.4}$$

is reported. c is the confidence interval in the standard normal distribution, and $J^{-1}(\varphi)_{ii}$ is the $i - th$ diagonal element of the inverse of the observed information matrix in equation 4.15.

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	1018 (11.62%)	720 (8.22%)	445 (5.08%)	227 (2.59%)	126 (1.44%)
Model Noise-Drift	2917 (33.30%)	2260 (25.80%)	1404 (16.03%)	819 (9.35%)	532 (6.07%)
Model Noise-Drift-Variance	2956 (33.74%)	2921 (33.34%)	2839 (32.41%)	2763 (31.54%)	2701 (30.83%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	118 (1.35%)	38 (0.43%)	2 (0.02%)	0 (0.00%)	0 (0.00%)
Model Noise-Drift	1011 (11.54%)	748 (8.54%)	415 (4.74%)	210 (2.40%)	134 (1.53%)
Model Noise-Drift-Variance	5499 (62.77%)	5452 (62.24%)	5371 (61.31%)	5274 (60.21%)	5176 (59.09%)

Table 7.1: Empirical realisation of the tail probabilities for ID_1 calibrated on 1-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	3488 (39.82%)	2977 (33.98%)	1762 (20.11%)	815 (9.30%)	487 (5.56%)
Model Noise-Drift	4206 (48.01%)	2673 (30.51%)	1042 (11.89%)	408 (4.66%)	231 (2.64%)
Model Noise-Drift-Variance	4517 (51.56%)	4479 (51.13%)	4394 (50.16%)	4318 (49.29%)	4249 (48.50%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	6 (0.07%)	1 (0.01%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Model Noise-Drift	286 (3.26%)	189 (2.16%)	100 (1.14%)	49 (0.56%)	31 (0.35%)
Model Noise-Drift-Variance	3890 (44.41%)	3841 (43.85%)	3719 (42.45%)	3596 (41.05%)	3516 (40.14%)

Table 7.2: Empirical realisation of the tail probabilities for ID_3 calibrated on 1-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	1613 (18.41%)	934 (10.66%)	298 (3.40%)	77 (0.88%)	40 (0.46%)
Model Noise-Drift	2541 (29.01%)	2001 (22.84%)	1205 (13.76%)	588 (6.71%)	347 (3.96%)
Model Noise-Drift-Variance	2881 (32.89%)	2851 (32.55%)	2758 (31.48%)	2667 (30.45%)	2609 (29.78%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	750 (8.56%)	236 (2.69%)	9 (0.10%)	0 (0.00%)	0 (0.00%)
Model Noise-Drift	1315 (15.01%)	749 (8.55%)	240 (2.74%)	73 (0.83%)	36 (0.41%)
Model Noise-Drift-Variance	5520 (63.01%)	5461 (62.34%)	5324 (60.78%)	5143 (58.71%)	4988 (56.94%)

Table 7.3: Empirical realisation of the tail probabilities for ID_1 calibrated on 5-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	752 (8.58%)	500 (5.71%)	284 (3.24%)	117 (1.34%)	55 (0.63%)
Model Noise-Drift	1861 (21.24%)	1186 (13.54%)	541 (6.18%)	246 (2.81%)	131 (1.50%)
Model Noise-Drift-Variance	3095 (35.33%)	3027 (34.55%)	2908 (33.20%)	2773 (31.66%)	2678 (30.57%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	43 (0.49%)	8 (0.09%)	1 (0.01%)	0 (0.00%)	0 (0.00%)
Model Noise-Drift	312 (3.56%)	164 (1.87%)	49 (0.56%)	19 (0.22%)	9 (0.10%)
Model Noise-Drift-Variance	5028 (57.40%)	4910 (56.05%)	4706 (53.72%)	4488 (51.23%)	4259 (48.62%)

Table 7.4: Empirical realisation of the tail probabilities for ID_3 calibrated on 5-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	2259 (25.79%)	1577 (18.00%)	588 (6.71%)	177 (2.02%)	69 (0.79%)
Model Noise-Drift	2403 (27.43%)	1871 (21.36%)	978 (11.16%)	401 (4.58%)	212 (2.42%)
Model Noise-Drift-Variance	3348 (38.22%)	3208 (36.62%)	2926 (33.40%)	2578 (29.43%)	2277 (25.99%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	2580 (29.45%)	1658 (18.93%)	535 (6.11%)	54 (0.62%)	9 (0.10%)
Model Noise-Drift	2903 (33.14%)	2232 (25.48%)	1068 (12.19%)	395 (4.51%)	161 (1.84%)
Model Noise-Drift-Variance	4353 (49.69%)	4160 (47.49%)	3730 (42.58%)	3299 (37.66%)	2858 (32.63%)

Table 7.5: Empirical realisation of the tail probabilities for ID_1 calibrated on 15-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	855 (9.76%)	506 (5.78%)	185 (2.11%)	73 (0.83%)	35 (0.40%)
Model Noise-Drift	1039 (11.86%)	708 (8.08%)	346 (3.95%)	127 (1.45%)	68 (0.78%)
Model Noise-Drift-Variance	2731 (31.18%)	2492 (28.45%)	2088 (23.84%)	1751 (19.99%)	1497 (17.09%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	896 (10.23%)	486 (5.55%)	91 (1.04%)	10 (0.11%)	1 (0.01%)
Model Noise-Drift	1712 (19.54%)	1134 (12.95%)	503 (5.74%)	157 (1.79%)	62 (0.71%)
Model Noise-Drift-Variance	4101 (46.82%)	3756 (42.88%)	3131 (35.74%)	2457 (28.05%)	1983 (22.64%)

Table 7.6: Empirical realisation of the tail probabilities for ID_3 calibrated on 15-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	2666 (30.43%)	1983 (22.64%)	760 (8.68%)	203 (2.32%)	72 (0.82%)
Model Noise-Drift	2511 (28.66%)	1864 (21.28%)	779 (8.89%)	249 (2.84%)	95 (1.08%)
Model Noise-Drift-Variance	3292 (37.58%)	3083 (35.19%)	2698 (30.80%)	2255 (25.74%)	1903 (21.72%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	3092 (35.30%)	2229 (25.45%)	734 (8.38%)	91 (1.04%)	21 (0.24%)
Model Noise-Drift	2994 (34.18%)	2226 (25.41%)	825 (9.42%)	193 (2.20%)	72 (0.82%)
Model Noise-Drift-Variance	4157 (47.45%)	3931 (44.87%)	3441 (39.28%)	2818 (32.17%)	2297 (26.22%)

Table 7.7: Empirical realisation of the tail probabilities for ID_1 calibrated on 30-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	1167 (13.32%)	688 (7.85%)	234 (2.67%)	72 (0.82%)	28 (0.32%)
Model Noise-Drift	1031 (11.77%)	661 (7.55%)	243 (2.77%)	77 (0.88%)	29 (0.33%)
Model Noise-Drift-Variance	2556 (29.18%)	2315 (26.43%)	1883 (21.50%)	1477 (16.86%)	1176 (13.42%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	1242 (14.18%)	713 (8.14%)	153 (1.75%)	14 (0.16%)	3 (0.03%)
Model Noise-Drift	1650 (18.84%)	992 (11.32%)	325 (3.71%)	64 (0.73%)	25 (0.29%)
Model Noise-Drift-Variance	3780 (43.15%)	3356 (38.31%)	2625 (29.97%)	1917 (21.88%)	1490 (17.01%)

Table 7.8: Empirical realisation of the tail probabilities for ID_3 calibrated on 30-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	2985 (34.08%)	2422 (27.65%)	1091 (12.45%)	294 (3.36%)	98 (1.12%)
Model Noise-Drift	2753 (31.43%)	2033 (23.21%)	715 (8.16%)	155 (1.77%)	55 (0.63%)
Model Noise-Drift-Variance	3054 (34.86%)	2794 (31.89%)	2298 (26.23%)	1757 (20.06%)	1337 (15.26%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	3569 (40.74%)	2820 (32.19%)	1077 (12.29%)	179 (2.04%)	27 (0.31%)
Model Noise-Drift	3194 (36.46%)	2355 (26.88%)	707 (8.07%)	107 (1.22%)	34 (0.39%)
Model Noise-Drift-Variance	3859 (44.05%)	3558 (40.62%)	2887 (32.96%)	2156 (24.61%)	1551 (17.71%)

Table 7.9: Empirical realisation of the tail probabilities for ID_1 calibrated on 60-minute intervals

	>90%	>95%	>99%	>99.9%	>99.99%
Model Noise	1571 (17.93%)	1001 (11.43%)	345 (3.94%)	97 (1.11%)	39 (0.45%)
Model Noise-Drift	1160 (13.24%)	691 (7.89%)	189 (2.16%)	49 (0.56%)	15 (0.17%)
Model Noise-Drift-Variance	2295 (26.20%)	1991 (22.73%)	1490 (17.01%)	1055 (12.04%)	776 (8.86%)
	<10%	<5%	<1%	<0.1%	<0.01%
Model Noise	1752 (20.00%)	1067 (12.18%)	270 (3.08%)	28 (0.32%)	4 (0.05%)
Model Noise-Drift	1680 (19.18%)	921 (10.51%)	212 (2.42%)	26 (0.30%)	6 (0.07%)
Model Noise-Drift-Variance	3134 (35.78%)	2676 (30.55%)	1979 (22.59%)	1310 (14.95%)	867 (9.90%)

Table 7.10: Empirical realisation of the tail probabilities for ID_3 calibrated on 60-minute intervals

Hour	α	β	μ	δ
1	7.005e-02 ± 1.709e-09	-9.039e-03 ± 1.930e-05	2.868e-02 ± 2.735e-04	8.373e-01 ± 3.302e-04
2	7.465e-02 ± 1.742e-09	-2.354e-03 ± 2.073e-05	5.521e-03 ± 2.348e-04	7.841e-01 ± 2.579e-04
3	8.557e-02 ± 2.555e-09	-9.978e-03 ± 2.218e-05	8.880e-03 ± 2.046e-04	7.609e-01 ± 2.266e-04
4	8.587e-02 ± 2.074e-09	-9.949e-03 ± 2.160e-05	1.218e-02 ± 1.740e-04	7.125e-01 ± 1.797e-04
5	7.732e-02 ± 1.124e-09	-7.924e-03 ± 1.897e-05	4.716e-03 ± 1.554e-04	6.698e-01 ± 1.440e-04
6	7.274e-02 ± 7.349e-10	-4.572e-03 ± 1.751e-05	6.146e-03 ± 1.458e-04	6.355e-01 ± 1.176e-04
7	3.509e-02 ± 0.000e+00	-2.393e-03 ± 7.125e-06	-2.959e-03 ± 1.314e-04	6.626e-01 ± 1.179e-04
8	5.670e-02 ± 3.726e-10	-5.233e-03 ± 1.061e-05	4.932e-03 ± 1.226e-04	6.883e-01 ± 1.238e-04
9	5.245e-02 ± 2.118e-10	-2.405e-03 ± 9.291e-06	9.620e-03 ± 1.223e-04	7.070e-01 ± 1.250e-04
10	6.064e-02 ± 3.422e-10	-3.648e-03 ± 1.126e-05	1.161e-02 ± 1.141e-04	6.706e-01 ± 1.131e-04
11	5.252e-02 ± 1.829e-10	-1.294e-03 ± 9.871e-06	1.047e-02 ± 1.117e-04	6.566e-01 ± 1.078e-04
12	6.097e-02 ± 4.275e-10	-2.367e-04 ± 1.101e-05	5.854e-03 ± 1.081e-04	6.614e-01 ± 1.048e-04
13	6.197e-02 ± 6.035e-10	-5.170e-05 ± 1.031e-05	1.134e-02 ± 1.133e-04	7.115e-01 ± 1.191e-04
14	5.518e-02 ± 1.475e-11	-3.777e-03 ± 8.927e-06	5.040e-03 ± 1.068e-04	6.950e-01 ± 1.096e-04
15	6.150e-02 ± 3.353e-10	-3.745e-03 ± 9.587e-06	5.666e-03 ± 1.038e-04	6.996e-01 ± 1.066e-04
16	5.801e-02 ± 2.292e-10	-8.497e-05 ± 9.306e-06	5.426e-03 ± 9.614e-05	6.579e-01 ± 9.118e-05
17	7.476e-02 ± 5.909e-10	-3.881e-03 ± 1.192e-05	7.243e-03 ± 9.454e-05	6.536e-01 ± 8.697e-05
18	6.550e-02 ± 3.059e-10	-5.625e-03 ± 1.045e-05	6.321e-03 ± 9.138e-05	6.203e-01 ± 7.513e-05
19	4.504e-02 ± 7.425e-16	1.536e-05 ± 7.170e-06	6.219e-03 ± 9.104e-05	5.951e-01 ± 6.592e-05
20	4.058e-02 ± 4.426e-11	-3.574e-03 ± 5.791e-06	7.561e-03 ± 8.381e-05	6.265e-01 ± 7.120e-05
21	4.108e-02 ± 0.000e+00	1.642e-03 ± 6.185e-06	2.483e-04 ± 8.631e-05	5.999e-01 ± 6.440e-05
22	5.955e-02 ± 5.923e-14	-7.611e-04 ± 1.027e-05	6.616e-03 ± 1.085e-04	5.480e-01 ± 5.276e-05
23	9.261e-02 ± 2.313e-09	-2.950e-03 ± 1.653e-05	5.995e-03 ± 1.072e-04	5.508e-01 ± 5.261e-05
24	1.199e-01 ± 0.000e+00	-8.172e-03 ± 2.100e-05	2.088e-02 ± 9.672e-05	5.654e-01 ± 5.553e-05

Table 7.11: Parameters of the Noise model with the 95% confidence interval calibrated on the 15-minute intervals.

8

Conclusion

This thesis is part of the Electricity Price Forecasting (EPF) literature, which aims to forecast the price of Electric Energy in both the short and long-term markets. Specifically, it seeks to produce a model for predicting probability distributions of the price of hourly contracts in the ID Germany market. Lévy processes are a rich family of stochastic processes that can, for example, reproduce high values of the excessive kurtosis in the data, include jumps, and represent skewed distributions.

In particular, the contract price has been modelled through a diffusion process, a jump-diffusion process, a Normal Inverse Gaussian Process, and a Generalised Hyperbolic Distribution. A selection was made based on the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC).

The diffusion process models the price as a Normal distribution, but this does not capture the tail distribution of the price increments resulting approximation of the probabilities at 0 for some of them. The jump-diffusion process comprises a diffusion term and a jump component, with the former following a standard normal distribution and the latter a Poisson process with jumps sized as a Normal distribution whose mean and variance are inferred from historical data. Adding one parameter to the model increased the performance in some scenarios. Nevertheless, it was impossible to reproduce the tail behaviour, so the log-likelihood would also result in infinite values in this scenario.

The Normal Inverse Gaussian Model is a 4-parameter model with infinite activity; this model at time $t = 1$ follows a Normal Inverse Gaussian distribution which presents semi-heavy tails and thus can capture the tail behaviour of the prices. Finally, the generalised Hyperbolic model is a general case of the Normal Inverse Gaussian model and, at time $t = 1$ follows a Generalised Hyperbolic distribution.

The Normal Inverse Gaussian model was selected by both the AIC and the BIC for most of the contracts and therefore was chosen as the Lévy process representing the process modelling the price. Moreover, this class of distributions is closed under convolution, and by knowing the distribution at one timescale, the distribution at every timescale is known.

However, the parameters of these models were based exclusively on the Historic data from 2020. Possible relationships between the price volatility and the change in weather forecasting were looked for to model the daily behaviour of the prices. Still, no correlation between these two quantities is present. After that, I conducted the Kendall-Mann test and the Augmented Dickey-Fueller test on the price of the contracts to verify if either a trend-following or a mean-reversion behaviour was present in the data, which resulted in accepting the presence of a trend on most of the days.

This fact was then integrated into the model to increase its Forecast Power by adding a drift component to the price. Consequently, two additional models were proposed, using the Normal Inverse Gaussian model as noise components. Finally, these models were compared on their forecasting accuracy of the ID_1 and ID_3 indexes' quantiles, where the Normal Inverse Gaussian model trained on 15-minute increments was the better performer. This suggested that the volatility and drift in the first hours of the day are poor predictors of price behaviour in the last hours. From a trader's perspective, this model explains the probability that their position can incur losses of a certain extent in the following 15-minute interval and adequately adjust their strategy.

References

[1] David Applebaum. *Lévy Processes and Stochastic Calculus*. English. 2nd edition. Cambridge ; New York: Cambridge University Press, May 2009. ISBN: 978-0-521-73865-1.

[2] Ole Barndorff-Nielsen. “Exponentially decreasing distributions for the logarithm of particle size”. EN. In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* (Mar. 1977). Publisher: The Royal Society London. DOI: 10.1098/rspa.1977.0041. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1977.0041 (visited on 11/30/2022).

[3] Ole E. Barndorff-Nielsen, Thomas Mikosch, and Sidney I. Resnick, eds. *Levy Processes*. English. 1st. Boston: Birkhäuser, Mar. 2001. ISBN: 978-0-8176-4167-2.

[4] Tom Brijs et al. “Interactions between the design of short-term electricity markets in the CWE region and power system flexibility”. en. In: *Applied Energy* 195 (June 2017), pp. 36–51. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2017.03.026. URL: https://www.sciencedirect.com/science/article/pii/S0306261917302428 (visited on 11/17/2022).

[5] Álvaro Cartea and Dimitris Karyampas. “The Relationship between the Volatility of Returns and the Number of Jumps in Financial Markets”. en. In: *SSRN Electronic Journal* (2011). ISSN: 1556-5068. DOI: 10.2139/ssrn.1507673. URL: http://www.ssrn.com/abstract=1507673 (visited on 09/20/2022).

[6] Julien Chevallier and Stéphane Goutte. “Estimation of Lévy-driven Ornstein–Uhlenbeck processes: application to modeling of CO2 and fuel-switching”. en. In: *Annals of Operations Research* 255.1 (Aug. 2017), pp. 169–197. ISSN: 1572-9338. DOI: 10.1007/s10479-015-1967-5. URL: https://doi.org/10.1007/s10479-015-1967-5 (visited on 01/11/2023).

[7] Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes*. English. 1st ed. Financial Mathematics Series. Chapman & Hall/CRC, 2004. ISBN: 1-58488-413-4.

[8] E. Eberlein and F. Ozkan. “Time consistency of Lévy models”. en. In: *Quantitative Finance* 3.1 (Jan. 2003), pp. 40–50. ISSN: 1469-7688, 1469-7696. DOI: 10.1088/1469-7688/3/1/304. URL: http://www.tandfonline.com/doi/abs/10.1088/1469-7688/3/1/304 (visited on 11/17/2022).

[9] Ernst Eberlein and Karsten Prause. “The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures”. en. In: *Mathematical Finance — Bachelier Congress 2000: Selected Papers from the First World Congress of the Bachelier Finance Society, Paris, June 29–July 1, 2000*. Ed. by Hélyette Geman et al. Springer Finance. Berlin, Heidelberg: Springer, 2002, pp. 245–267. ISBN: 978-3-662-12429-1. DOI: 10.1007/978-3-662-12429-1_12. URL: https://doi.org/10.1007/978-3-662-12429-1_12 (visited on 09/15/2022).

[10] Marc Gürtler and Thomas Paulsen. “The effect of wind and solar power forecasts on day-ahead and intraday electricity prices in Germany”. en. In: *Energy Economics* 75 (Sept. 2018), pp. 150–162. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2018.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0140988318302512 (visited on 10/20/2022).

[11] Tao Hong et al. *Energy forecasting: A review and outlook*. en. Tech. rep. WORMS/20/08. Publication Title: WORking papers in Management Science (WORMS). Department of Operations Research, Business Intelligence, Wroclaw University of Science, and Technology, May 2020. URL: https://ideas.repec.org/p/ahh/wpaper/worms2008.html (visited on 05/04/2022).

[12] Arkadiusz Jędrzejewski et al. “Electricity Price Forecasting: The Dawn of Machine Learning”. In: *IEEE Power and Energy Magazine* 20.3 (May 2022). arXiv: 2204.00883, pp. 24–31. ISSN: 1540-7977, 1558-4216. DOI: 10.1109/MPE.2022.3150809. URL: http://arxiv.org/abs/2204.00883 (visited on 05/03/2022).

[13] Dimitris Karlis. “An EM Type Algorithm for Maximum Likelihood Estimation for the Normal Inverse Gaussian Distribution”. In: *Statistics & Probability Letters* 57 (Mar. 2002), pp. 43–52. DOI: 10.1016/S0167-7152(02)00040-8.

[14] M.G. Kendall. *Rank correlation methods*. Rank correlation methods. Oxford, England: Griffin, 1948.

[15] Rüdiger Kiesel and Florentina Paraschiv. “Econometric analysis of 15-minute intraday electricity prices”. In: *Energy Economics* 64.C (2017). Publisher: Elsevier, pp. 77–90. ISSN: 0140-9883. URL: https://econpapers.repec.org/article/eeeeneeco/v_3a64_3ay_3a2017_3ai_3ac_3ap_3a77-90.htm (visited on 04/30/2022).

[16] Christopher Koch and Lion Hirth. “Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany’s electricity system”. en. In: *Renewable and Sustainable Energy Reviews* 113 (Oct. 2019), p. 109275. ISSN: 1364-0321. DOI: 10.1016/j.rser.2019.109275. URL: https://www.sciencedirect.com/science/article/pii/S1364032119304836 (visited on 11/17/2022).

[17] Marcel Kremer, Rüdiger Kiesel, and Florentina Paraschiv. “An econometric model for intraday electricity trading”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2202 (June 2021). Publisher: Royal Society, p. 20190624. DOI: 10.1098/rsta.2019.0624. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2019.0624 (visited on 03/20/2023).

[18] Henry B. Mann. “Nonparametric Tests Against Trend”. In: *Econometrica* 13.3 (1945). Publisher: [Wiley, Econometric Society], pp. 245–259. ISSN: 0012-9682. DOI: 10.2307/1907187. URL: https://www.jstor.org/stable/1907187 (visited on 02/16/2023).

[19] Michał Narajewski and Florian Ziel. “Ensemble forecasting for intraday electricity prices: Simulating trajectories”. en. In: *Applied Energy* 279 (Dec. 2020), p. 115801. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.115801. URL: https://www.sciencedirect.com/science/article/pii/S0306261920312824 (visited on 10/20/2022).

[20] Karsten Neuhoff et al. “Intraday Markets for Power: Discretizing the Continuous Trading?” en. In: *SSRN Electronic Journal* (2016). ISSN: 1556-5068. DOI: 10.2139/ssrn.2723902. URL: http://www.ssrn.com/abstract=2723902 (visited on 11/17/2022).

[21] Cornelis Oosterlee and Grzelak, Lech. *Mathematical Modeling and Computation in Finance*. World Scientific, 2020. ISBN: 978-1-78634-805-0.

[22] Christian Pape, Simon Hagemann, and Christoph Weber. “Are fundamentals enough? Explaining price variations in the German day-ahead and intraday power market”. en. In: *Energy Economics* 54 (Feb. 2016), pp. 376–387. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2015.12.013. URL: https://www.sciencedirect.com/science/article/pii/S0140988316000037 (visited on 04/22/2022).

[23] Karsten Prause. “The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures”. de. In: (), p. 168.

[24] Ken-iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. English. Illustrated edition. Cambridge, U.K. ; New York: Cambridge University Press, Nov. 1999. ISBN: 978-0-521-55302-5.

[25] Christoph Scholz et al. “Towards the Prediction of Electricity Prices at the Intraday Market Using Shallow and Deep-Learning Methods”. en. In: *Mining Data for Financial Applications*. Ed. by Valerio Bitetta et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 101–118. ISBN: 978-3-030-66981-2. DOI: 10.1007/978-3-030-66981-2_9.

[26] Bartosz Uniejewski, Grzegorz Marcjasz, and Rafał Weron. “Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO”. en. In: *International Journal of Forecasting* 35.4 (Oct. 2019), pp. 1533–1547. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2019.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0169207019300123 (visited on 07/29/2022).