Simulating Energy Prices and Energy Yield to Generate Probability Distributions

Bachelor Graduation Thesis
Industrial Engineering and Management
University of Twente
Andrés Iturbide Figge

Author's Declaration & Acknowledgement

Dear reader,

This originally started as an abstract idea with not a clear way to be done. However, after a bunch of hours in front of the computer reading articles, programming, or simply experimenting things started coming together. Now that the final product is done it feels nice to see something concrete, which solves the original problem but still has a lot of room for improvement. You as the reader only see the final product, but it is the result of a process from idea to reality. The uncertainty within this process made it the most challenging part of the project but also the most exciting.

My first acknowledgement goes to my family, particularly my parents. I feel grateful for the opportunities and support you have given to me which allowed me to complete my Bachelor. Secondly, I want to thank all the employees at Company X with whom I had interaction during my internship. This goes specially to my supervisor at Company X who gave me the opportunity for this internship and made sure to make it a nice experience. Thirdly, I want to thank my thesis supervisor Berend Roorda. Many times, in this project I felt a bit lost whoever he helped on putting me back on track. Finally, I want to thank all the people that where part of my life during my Bachelor, thank you for making these student years very special!

Andrés Iturbide Figge

London, June 2024.

A note about confidentiality

The company in which this project was conducted requested to remain anonymous. Instead of referring to the company by its name it will be referred to as Company X. The specific team for which this simulation was developed will be referred as Team Y. All the data seen on pictures about the simulation is fictional and does not represent anything related to Company X's clients.

Management Summary

Introduction

One of Company X main activities is to finance renewable energy projects. These projects are evaluated via a cash flow model, which is then used to determine how much money the project can borrow and at what terms (debt structure). The model relies completely on deterministic data regardless that some inputs are of stochastic nature. This means that the current cash flow model bases itself on one scenario and works with deterministic data. Some inputs in the model such as energy prices are uncertain, this can cause the actual financial performance of a project to fluctuate from what was originally projected. Consequently, the robustness of a debt structure might get compromised, and the project could go into a risk of default.

Problem Statement

Company X would like to extend the current cash flow model to capture the uncertainty in energy prices and energy yield (how much energy a project produces in a year) and be able to evaluate the robustness of a debt structure through different scenarios. This thesis deals with the following problem: "**The current cash flow model does not incorporate the randomness in energy prices and energy yield and its outputs are limited to one basic scenario.** " The selected outputs are: cashflows available for debt service (CFADS), debt service (DS), and debt service coverage ratio (DSCR) as these are the main financial figures used when evaluating a debt structure. This research is focused on solar energy projects in Spain. The main deliverable is a simulation for energy prices and energy yield of Spanish energy projects that can serve as a tool for Team Y to evaluate a debt structure via its possible CFADS, DS, and DSCR. The research questions for building this tool are outlined according to the simulation methodology by Robinson (2004).

Approach

The first step to build the simulation is to replicate the calculations done by the current cash flow model to calculate CFADS, DS, and DSCR. As the current model is extensive and complex some simplifications where made that allow to calculate CFADS, DS, and DSCR for each year in the simulation.

Next step is to find a mathematical model to simulate the average daily Spanish electricity prices for one year. The model used is Cartea & Figueroa (2006)'s jump diffusion model as it is the simplest model that represents mean reversion, seasonality and jumps in energy prices. This model is then calibrated to the Spanish market using R and historical daily price data from 2018 to 2022. Using the mid-price forecast for a particular year as an initial condition and the calibrated Cartea & Figueroa (2006)'s jump diffusion to simulate daily prices allows to get the average daily Spanish energy price for one year. This procedure allows to include the long-term fundamental trends given in the energy price forecast and short-term random fluctuations given by Cartea & Figueroa (2006)'s model in the simulation.

A sampling distribution for the yearly energy yield of a Spanish solar energy project must be found. After a theoretical review and an interview with an energy yield assessment consultant it was determined that energy yield of solar energy projects is normally distributed, however its mean and standard deviation are specific to the project being analysed. It is shown how to calculate the mean and standard deviation for a solar energy project out of the P50 and P90 numbers provided on the energy yield assessment report. With the mean and standard deviation, the annual energy yield of a Spanish energy project can be simulated.

The calibrated jump diffusion model for energy prices and sampling distribution for energy yield can then be used to generate different scenarios. This was combined in a Montecarlo simulation in VBA where their respective CFADS, DS, DSCR can be calculated. To find a suitable number of iterations different number of iterations were evaluated on the trade-off between computation time and standard error. The recommended number of iterations for this simulation is between 2500 as it provides sufficient accuracy with a relatively low computation time and does not put the simulation at risk of crashing.

Applications

The simulation generates a sample of possible scenarios of CFADS, DS, and DSCR, however, this sample needs to provide information that can be used to evaluate debt structures. First, the expected scenario is found by calculating a confidence interval for the mean CFADS, DS, and DSCR on each year. Secondly, the range of possible scenarios for an arbitrary year is shown via the 5 numbers summary. The shape of the underlying distribution is examined via the empirical skewness coefficient and excess kurtosis and provide more insights about the confidence intervals for the mean and 5 numbers summary. Probabilities can be calculated via the empirical CDF. This can be used to evaluate relevant questions such as: What is the probability that DSCR falls out of the preferred range of 1.2-1.4 on an arbitrary year? The robustness of the debt structure is evaluated via the probability of default which expresses the proportion of scenarios where a default occurred on a specific year. A confidence interval for the probability of default is given which serves as the main indicator of the robustness of a debt structure. All these statistics are visualized in a dashboard that can serve as a tool that Team Y can use to evaluate a debt structure.

Conclusion and Points for Future Research

This project provided a solution to the core problem: “The current cash flow model does not incorporate the randomness in energy prices and energy yield and its outputs are limited to one basic scenario. ”

The problem was solved via a Montecarlo simulation which can be used as a tool to evaluate debt structure under different scenarios. The simulation consists of the following:

1) Energy prices are simulated via Cartea & Figueroa (2006)’s Jump Diffusion model calibrated to the Spanish market using historical data from 2018 to 2022.

2) Energy yield is simulated via a normal distribution with mean and standard deviation obtained from the parameters P50 and P90 given in the energy yield assessment report of the specific project being analysed.

3) For a set of simulated energy prices and energy yield the simulation calculates their respective CFADS, DS, and DSCR.

4) Is recommended to run the simulation for 2500 iterations for obtaining good precision with an acceptable computation time.

5) For the generated scenarios different statistics are calculated that help evaluating debt structures.

There are multiple points for future research. The most important are:

1) Changing the assumptions and simplifications in replicating the original cash flow model so the calculations done by the simulation better resemble reality. For example this simulation assumes that energy is sold at a constant contracted price and market price. Additionally, the volumes sold at these prices are divided on constant percentages. A renewable energy project typically sells most of the energy produced at a contracted price in earlier years via a power purchase agreement (PPA). And as time progresses the exposure to market price grows. Next to that, a project might engage in multiple PPAs which might can cause to have different contracted prices during the lifetime of the project.

2) Expanding the scope of the research to different countries by calibrating the Jump Diffusion Model with historical data from other countries. Additionally, different types of renewable energy projects could be covered such as wind by changing the sampling distribution of energy yield to a suitable distribution for that type of source.

3) Including other stochastic inputs into the simulation such as interest rates by finding a suitable mathematical model.

Readers Guide

Chapter 1 – Defines core problem and establishes research questions such that core problem can be solved.

Chapter 2 – Introduces important concepts related to the modelling of energy prices and energy yield.

Chapter 3 – Explains calculations for CFADS, DS, and DSCR preformed by the simulation.

Chapter 4 – Explains Cartea & Figueroa (2006)'s Jump Diffusion model and its calibration to the Spanish market.

Chapter 5 – Determines that the Energy Yield of PV projects follow a distribution an explain how to calculate its mean and standard deviation out of the parameters provided from an Energy Yield assessment report.

Chapter 6 – Describes the Montecarlo simulation and find a suitable number of iterations based on the trade-off between number of iterations and computation time.

Chapter 7 – Explains different statistics that can be calculated out of the sample of simulated CFADS, DS, and DSCR and their application to evaluate a debt structure.

Table of Contents

List of Figures

List of Tables

List of Equations

Equation Number	Description
1	Convergence relationship spot price and futures price (Aïd. 2015)
2	SDE for log energy spot price (Hull. 2009)
3	Log spot price stochastic and seasonal components (Lucia & Schwartz. 2002)
4	SDE for stochastic component of log energy spot price (Lucia & Schwartz. 2002)
5	Log spot price stochastic and seasonal components (Cartea & Figueroa. 2006)
6	SDE for stochastic component of log energy spot price (Cartea & Figueroa. 2006)
7	Poisson process for jump occurrence (Cartea & Figueroa. 2006)
8	Distribution in jump size (Cartea & Figueroa. 2006)
9	SDE for energy spot price (Cartea & Figueroa. 2006)
10	Seasonal component for energy spot price (Cartea & Figueroa. 2006)
11	Approximation of Equation 6 via the Euler Maruyama method
12	Cumulative jump size over a time period t
13	Inverse Poisson function
14	Linear regression for mean reversion rate (Boucher et al. 2023)
15	Log return adjusted to length of time interval (Eydeland & Wolyniec. 2003)
16	P90 expressed in standard normal distribution
17	Standard Error (Hull. 2009)
18	Empirical Skewness Coefficient (Poortena et al. 2021)
19	Empirical Excess Kurtosis (Poortena et al. 2021)
20	Empirical CDF (Poortena et al. 2021)

List of Abbreviations

Term	Abbreviation
Auto correlation function	ACF
Cashflow available for debt service	CFADS
Cumulative Distribution Function	CDF
Debt service	DS
Debt service coverage ratio	DSCR
Direct normal solar irradiance	DNI
Earnings before interests, taxes, depreciation, and amortization	EBITDA
Generalized Autoregressive Conditional Heteroskedasticity models	GARCH
Inner Quartile Range	IQR
Megawatt Hour	MWh
Photovoltaic Cells	PV
Power purchase agreement	PPA
Probability of exceedance at 50, 75, 90 level	P50, P75, P90
Quantile – Quantile plot	QQ plot
Standard Error	SE
Stochastic differential equation	SDE

Chapter 1- Introduction

One of the main activities of Company X is to provide loans for projects related to infrastructure and renewable energy generation, this activity is referred to as project finance. The assignment will be carried on for Team Y of Company X. Team Y is the front force for achieving deals related to renewable energy projects. It is responsible for acquiring and managing relationships with potential clients, structure finance, perform due diligence and execute potential deals. Within a deal Company X will build a debt structure for the renewable energy project that needs financing. A debt structure determines the amount of money that can be borrowed and what are its terms for use and repayment based on the cashflows of the project. This chapter will provide an outline for this research project. First, the different problems a debt structure might experience are presented in a problem cluster. The problems in the problem cluster are used to formulate a core problem. Then the norm and reality of the problem are compared. Finally, a methodology for solving the core problem is introduced in which the multiple research questions covered on this thesis are raised.

1.1 Problem Cluster & Core problem

During a deal the Team Y will work closely with clients and other departments of Company X. Besides that, external consultants are hired to supervise different areas such as the legal and technical aspects of the project. An inhouse developed cash flow model is used to predict the future cashflows and financial development of the project. Using this Team Y can obtain different figures about future financial performance that are used to build the debt structure. This allows Company X to provide the correct type of financing to its clients and such that the structure attains to Company X's risk preference. The current problems faced by Company X are shown in the problem cluster of Figure 1. This outlines the problems discussed during meetings with members of Team Y. The starting problem is that the borrower might default on its loans, meaning that a renewable energy project is not able to repay the money it has borrowed from Company X. One of the causes for this problem is that the current debt structures built by Team Y may not be robust when performance of the project deviates from the one outlined in the cash flow model. On one hand this is caused because the financing / product offering that Team Y is working with is moving away from a predictable revenue stream. In the

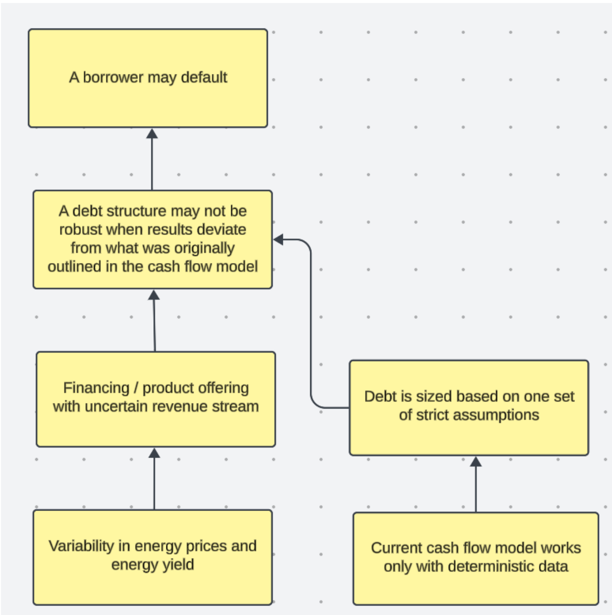


Figure 1 – Problem Cluster

past the Company X provided loans to renewable energy projects that obtained subsidies from governments. This provided a constant influx of cash every year, making the cashflows of the project predictable. Nowadays the Company X is giving loans to projects with less government subsidies that have a bigger variability in potential payoffs. This uncertainty in cashflows is due to multiple factors. As indicated by members of Team Y the two main factors are energy prices and energy yield. A renewable energy project has exposure to constant changes in energy prices due to market movements, which could have a drastic effect in revenue. Additionally, energy yield measures the amount of energy produced by a project over a year. This naturally may vary (e.g. a solar panel project could produce low energy levels due to bad weather), and the amount of energy a project produces directly impacts the energy available to be sold.

Another cause for a debt structure not being robust is due to the way it is built. The amount of money a renewable energy project can borrow and the terms for use and repayment are based on the projections made by Company X's cash flow model. For running the model various assumptions about future performance must be done. Therefore, the financial figures used to size the debt is limited to the outlook given by these assumptions. Company X's current cash flow model only works with deterministic data, which means that when determining the outlook of a renewable energy project the information is limited to strictly one scenario. As it can be seen in Figure 1 there are two main causes that lead to the problem that a borrower may default on its debts. One cause is that energy prices and energy yield are uncertain. The other cause is that the current cash flow model is restricted to deterministic data. The combination of these two causes lead to the core problem: **The current cash flow model does not incorporate the randomness in energy prices and energy yield and its outputs are limited to one basic scenario.** The motivation for choosing this core problem is that by modelling the randomness in energy prices and energy yield the uncertainty in potential payoffs can be represented within the cash flow model. This allows to obtain possible scenarios of outputs that a renewable energy project might experience. With this Team Y can evaluate a debt structure and check on which scenarios the robustness of the debt structure is compromised.

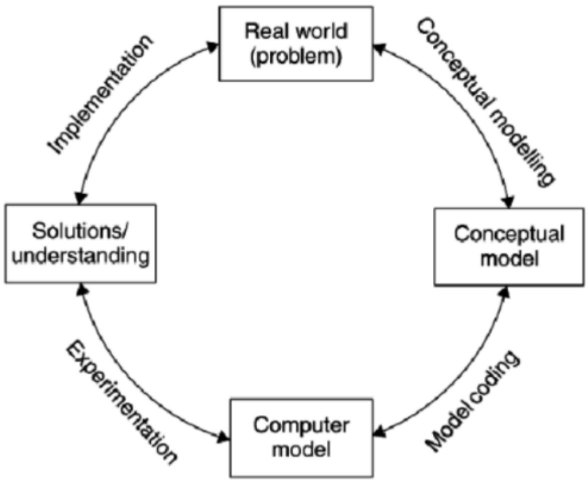
1.2 Norm and reality

The current cash flow model has a diverse set of inputs and outputs. For this assignment Company X has indicated to only focus on the following outputs: cashflows available for debt service (CFADS), debt service, and debt coverage ratio. These outputs are some of the most important financial figures used in building a debt structure. The motivation for choosing CFADS is that it is the main indicator of the profitability of the project. Debt service indicates the amount of cashflows that are destined to paying back the debt acquired to finance the project. Therefore, debt service is the main indicator of how debt is being repaid. Debt coverage ratio is defined as: total debt service divided by CFADS. Company X likes to maintain this ratio between 1.2 to 1.4 in all their projects. DSCR gives a measure of how well covered debt service is from the cash flows produced by the project. The cash flow model just does a simple calculation and gives a value to each output based on the provided inputs. Currently an estimate in energy prices is given by an external consultant. The consultant will give 3 different projections for the energy price: high, middle, and low. Company X will use either the low projection or an average between the middle and low as an input for the cash flow model. Company X employees have indicated that the energy yield of a solar energy project is assumed to follow a normal distribution. A specific "Probability of Exceedance" number (an explanation of this is given in Chapter 2) is used as the basic assumption for energy production in the cash flow model. To test the impact of a change in the inputs, the current cash flow model performs sensitivity analysis. For this a certain change is applied to the input of interest (e.g. +5% energy price), and its respective output is given.

This procedure is repeated for some arbitrary changes on one input at the time and provides an idea of possible outputs. The limitation of this approach is that it does not allow to generate many scenarios and the changes that can be made per scenario are limited. Additionally, there is no statistical analysis with the generated scenarios, so it is only possible to analyse the results of the sensitivity analysis by mere comparison. Finally, the arbitrary changes done in sensitivity analysis do not consider the likelihood of such change. This could leave scenarios that are unlikely to happen to not even be considered at all. **The norm for the cash flow model would be to sample different combinations of energy prices and energy yield based on their likelihood of occurrence. This would allow to generate possible CFADS, debt service, and debt coverage ratio. Then using statistical techniques, the characteristics of the distribution of possible scenarios can be obtained based on the generated sample. All these characteristics will be displayed on a dashboard that can serve as a tool for evaluating debt structures.** With this tool Team Y could evaluate a debt structure and check how likely it is to have its robustness compromised by fluctuations in energy prices an energy yield. For example, Team Y could check what is the probability that DSCR falls outside of their preferred range. If the probability for this event is high, then Team Y could consider making changes to the debt structure or not even get into the deal at all.

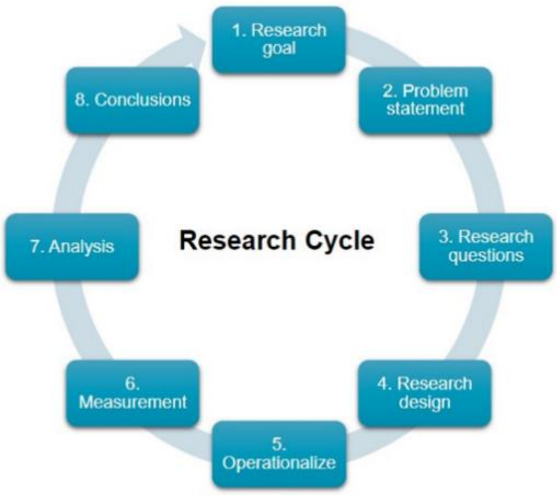
1.3 Problem Solving Approach

The methodology used to solve this problem will be the methodology for simulation by Robinson (2004) as shown in Figure 2. This consists of 4 main phases; a set of research questions has been assigned to each phase. A total of 6 main research question have been formulated for this project. After completing the research, the intended deliverables are the upgraded cash flow model and the analysed datasets. For answering each of the research questions the research cycle methodology introduced by Heerkens & met Winden (2017) in Figure 3 will be used. This methodology was chosen as the different steps are clear and make a good fit with the intended research process for upgrading the model. Below there is an overview of the research questions for each step of the methodology.



Source: *Robinson, S. (2004). Simulation: The Practice of Model Development and Use. Chichester: Wiley.*

Figure 2 – Simulation Methodology (Robinson, 2004)



Source: Heerkens, H. met Winden, A. van (2017). Solving Management Problems Systemetically Groningen: Noordhoff Uitgevers.

Figure 3 – Research Cycle (Heerkens et al., 2017)

Step 1 – Understanding the problem.

The current problem has already been explored and explained in this project plan. Before any changes are applied to the cash flow model and the way it currently works must be understood in depth. This allows to understand how the financial projections are made by the model calculated. It also ensures that the results produced by the simulation reflect the original cash flow model and are still able to provide Company X reliable and valid information for decision making. The research design for this question is exploratory as it entails to gain insights on the functioning of the current cash flow model.

1) *Knowledge problem:* How does the current cash flow model work for solar energy projects in Spain?

   Variable: Inputs (energy prices and energy yield) and outputs (CFADS, DS, DCSR).

   Action: Make an input-output mapping

The current cash flow model is very complex and has a wide array of inputs and outputs. Therefore, this research question is focused on inputs and outputs relevant to this research project. Energy yield and energy prices are the inputs to be analysed. As discussed on section 1.1 these two inputs are one of the main sources for uncertainty. Energy yield is the amount of energy produced by the project during one year by the project. Energy prices refer to the average energy price for one year. Additionally, CFADS, debt service, and debt coverage ratios are the outputs. As discussed on section 1.3 these are some of the most important figures used to build debt structures. The nature of the inputs in combination with an explanation of how they are currently outlined on the cash flow model will be discussed. For outputs an examination of how exactly they are calculated must be made. Finally, a special emphasis will be made in examining these parameters for solar energy projects in Spain to ensure validity. This will allow to create a clear input-output mapping showing an overview of how the current model works. For having a reliable understanding of the current cash flow model meetings with Company X's employees with special experience in cash flow modelling where made. Additionally, involvement in meetings with clients where the cash flow model is used will took place to understand its applications in a practical setting.

Step 2 – Conceptual modelling.

On this phase the conceptual modelling of the simulation takes place. As defined by Robison (2004) conceptual modelling consists of "abstracting a model from a real or proposed system". Therefore, this part is concerned with mathematically modelling the inputs for the cash flow model (energy prices and energy yield) such that they can be simulated. The following research questions follow a descriptive analysis as they aim to understand how the solutions proposed by academia to model a particular variable. Then it applies these methods to a research population.

2) *Knowledge problem:* What is a suitable mathematical model to simulate Spanish energy spot prices such that it represents seasonality, mean reversion, and price jumps?

   Variable: Energy spot prices.

   Research population: Dataset historical Spanish energy prices.

   Action:

1) Review existing theory and pick a model that represents seasonality, mean reversion, and price jumps.
2) Calibrate chose model to the Spanish maker using historical data.

The first step in answering this research question is to conduct a theoretical review of the main characteristics of fluctuations in energy prices. As it will be discussed on the coming chapter these are: seasonality, mean reversion, and price spikes. Then different mathematical models proposed by academia for simulating energy spot prices are reviewed. Consequently, the selected models are compared to determine which ones capture the characteristics mentioned in the theoretical review the best with a relatively low mathematical complexity. A data set of historical energy prices will be analysed to calibrate the chosen model to the Spanish market. The dataset found consists of daily Spanish spot prices since 1998. It was obtained via the Trading Economics database to which Company X provided access. This calibrated model allows to simulate daily electricity price changes and then obtain the average energy price for one year. It ensures reliability of this research design as it is a large dataset containing 6853 data points. Additionally, it ensures validity as this data is exclusively represents the phenomena tried to be capture by the model (Spanish energy spot prices). The main issue of this research design to the reliability and validity is that the past is not fully indicative of the future. Simulating based on historical does give an idea of the possible behaviour of energy prices in the future, however there are no guarantees that this will be the case.

3) *Knowledge problem:* Can the yearly energy yield of a Spanish solar energy project be assumed to follow a normal distribution? If not, what alternative can be used?

Variable: Energy Yield

Research population: Dataset historical energy yields of solar energy projects in Spain.

Action:

1. Review existing theory to check why normality is assumed, under what condition it cannot be assumed and any alternative distributions that could be used to model for energy yield.

2. Make interview with external consultant to ensure the distribution choice and its generalization is reliable and valid.

According to Company X the energy yield (amount of energy produced in a year) of a solar energy project follows a normal distribution. A literature review will be conducted to determine: What are the factors that affect the variability in energy yield? What are the limitations of modelling energy yield with a normal distribution? Are there any alternative distributions that could be used? As it will be seen in the literature review, when testing the distribution in energy yield it is recommended to have data of the specific location where the energy project is built. This is process is usually done by a technical advisor and it involves having meteorological data and other types of measurements. Additionally, this project looks to generalize a distribution in energy yield for solar energy projects in Spain (not a particular location). Therefore, conducting data analysis and test if normality holds is not a possibility for this project. As a result, an interview an energy yield technical advisor will be conducted during the research project. This will give valid arguments and the opinion of an expert on the type of distribution assumed in this project.

Step 3 – Model coding & Validation.

This step consists of coding the conceptual model and then validating that it indeed represents the real world accurately. As dictated by Company X this simulation will be coded in VBA as the current cash flow model is done in excel/VBA. With the simulated energy prices and energy yield the outputs of the model and a sample of possible scenarios can be generated with the Montecarlo method. The research design for both evaluative and descriptive. The evaluative part consists of checking whether the input-output mapping done in step 1 is correct. The descriptive part consists of finding the suitable number of iterations for simulating inputs for having precise results out of the simulation.

4) *Knowledge problem:* How to determine a suitable number of iterations for a Montecarlo Simulation?

Variable: Number of Iterations

Research population: Generated sample of simulated CFADS, DS, DSCR.

Action:

1. Literature review about number of iterations.

2. Compare standard error and computation time for different number of iterations.

As it will be discussed on Chapter 6 a Montecarlo simulation increases its accuracy with a higher number of iterations. However, this greater accuracy comes at a cost of an increased computation time. This research question involves finding a suitable number of iterations that will produce accurate results with an acceptable computation time. A literature review will be conducted to check academic information about the number of iterations for a Montecarlo simulation.

5) *Knowledge problem:* How to make sure the generated distributions for CFADS, Debt Service, and Debt Coverage Ratio are valid?

Action: Validate Input-Output mapping within Company X

The calculations for CFADS, DS, DSCR must accurately replicate the ones done in the cash flow model. Additionally, any assumptions or simplifications made must not significantly compromise the validity of the simulation. To verify these two points meetings will take place with employees of Team Y. In these meetings the Input-Output mapping made in step 1 will be reviewed to ensure validity and reliability in the calculations of CFADS, DS, DSCR. Additionally assumptions and simplifications will be discussed, and any necessary adaptations will be made such that the simulation preform up to Company X's specifications.

Step 4 – Finding solutions & Understanding

The final step of the methodology proposed by Robinson (2004) consists of finding solutions and understanding. In this phase the simulation is used to understand the modelled phenomena and it can be used as a tool to solve problems. The research design for the following question is exploratory as the simulation can then be used to gain insights on different kinds of debt structures.

6) *Knowledge problem:* How to make this simulation a tool that can help Team Y to evaluate debt structures for Spanish solar energy projects?

Action: Use statistical techniques to explore the characteristics of the distributions of simulated outputs. With these characteristics make a dashboard that can serve as a tool for Team Y.

With the use of adequate statistical techniques meaningful information of the distribution in outputs can be extracted out of the simulated samples. A dashboard will be created for visualising the characteristics of the generated distributions. This will allow to build a tool that members of Team Y can use and will help in evaluating potential Spanish solar energy projects. A presentation will be held when the simulation will be shown, and members of Team Y will receive instructions on how to use it.

These research questions were formulated to provide a solution to the core problem of section 1.1. The coming chapters aim to provide an answer to these research questions. By the end a solution to the core problem will be found via a Montecarlo simulation that generates different scenarios for energy prices and energy yield which allows to evaluate a debt structure via statistics of the resulting CFADS, DS, and DSCR.

Chapter 2- Theoretical Framework

2.1 Structure and motivation

This chapter will introduce different concepts related to modelling energy prices and energy yield that will be used in the upcoming chapters of this thesis. Out of the research questions described above there are two that present the whole basis of this research. These are:

1) *What is a suitable mathematical model to simulate Spanish energy spot prices such that it represents seasonality, mean reversion, and price jumps?*

2) *Can the yearly energy yield of a Spanish solar energy project be assumed to follow a normal distribution? If not, what alternative can be used?*

The reason being that the distribution of cashflows, debt service, and debt coverage ratio will be generated out of simulated data of energy prices and energy yield. This means that if both inputs are not correctly simulated the posterior steps for solving the core problem will not succeed. In this section background knowledge that is necessary to answer both research questions will be provided. Firstly, a description of the structure of the Spanish electricity supply chain and market will be presented. Secondly, the two main electricity markets (spot and futures) will be introduced to explain the convergence relationship between futures prices and spot prices. This relationship is used in some of the presented mathematical models for the spot price (HJM style models) discussed in this chapter. Then an overview of special attributes that electricity has will be discussed together with their effects on price fluctuations. Finally, an overview of the terminology and main mathematical approaches to model energy prices will be presented. Regarding energy yield a description of its definition and uses will be provided with focus on solar energy projects.

2.2 Modelling electricity prices

2.2.1 The Spanish electricity supply chain and market

Luna-Romera et al. (2024) present the different parties involved in the energy supply chain of the Spanish market as shown in Figure 4. Firstly, electricity needs to be produced. Electricity can be generated with different types of energy sources such as solar, nuclear, fossil fuels, etc. Then it can be transported across long distances through the high-voltage transmission grids from the system operators. After that electricity goes to a low-voltage network to be distributed between consumers. The "Nominated Electricity Market Operator" in English or OMIE is the Spanish regulating entity for energy markets. In this market the demand from distributors is matched from the supply of electricity generators by comparing the price per MWh. The matching is done by ordering all bids from generators in ascending order and from distributors in descending order. Figure 5 shows a graph of the bids with buy orders (distributors) in the blue line and sell (generators) in the green line. The point where these graphs intersect is the energy price for that hour of the day. All bids below the energy price for that hour are matched and executed with their original prices. This procedure gets repeated every hour one day before delivery. To summarize electricity gets produced, then is transported through long distances to later be distributed to the public, its price is set by comparing bids.

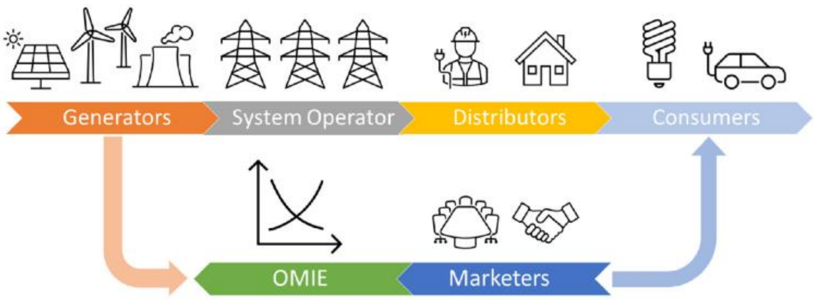


Figure 4- Spanish Energy Supply Chain & Market (Luna-Romera et al. 2024)

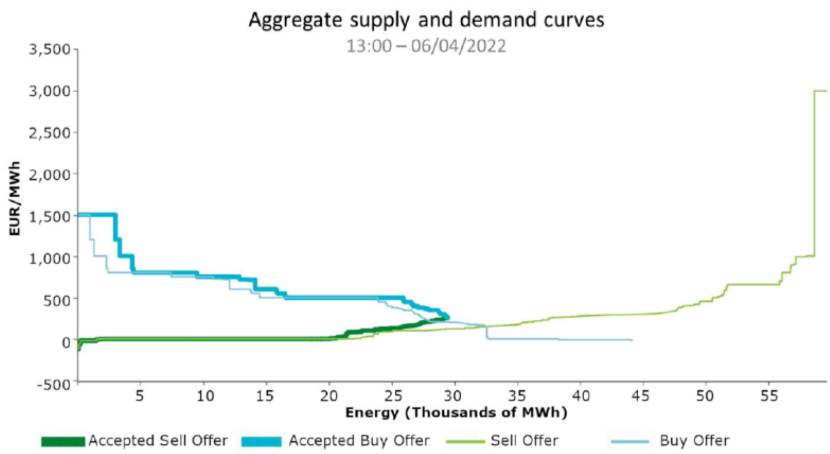


Figure 5 – Buy & Sell energy bids (Luna-Romera et al. 2024)

2.2.2 Spot market and Futures market

The market structure described in the previous section where electricity is delivered one day after is referred as the spot market for electricity. However most buying and selling of electricity is done via futures contracts. Hull (2009) defines a futures contract as "an agreement between two parties to buy or sell an asset at a certain time in the future for an established price". Futures contracts can be used for hedging, speculating, or arbitrage purposes. Aïd (2015) explains that for avoiding having an electricity futures contract for each hour of the day a certain delivery period is specified. This could be either be a month, a quarter, a year, or a day. Usually, the delivery period for contracts with longer maturity tends to be wider. Finally, an important relationship between both markets is that the settling price for futures contracts converges to the spot price at the time of maturity. To give a formal definition of this relationship consider a futures price $F(t,T)$ at time t with maturity at T. Let S_T be the spot price at the time of maturity. Equation 1 formally expresses the relationship between futures price and spot price as time t approaches maturity.

$$\lim_{t \to T} F(t, T) = S_T \tag{1}$$

2.2.3 Electricity as a commodity

Electricity has some special attributes that affect its price fluctuations. Gudkov & Ignatieva (2021) point out that the main attributes of electricity price fluctuation are price jumps, mean reversion, and seasonality. According to Aïd (2015) even though there are ways that energy can be stored efficiently such as hydroelectric reservoirs, storing electricity is generally not economical and is limited to a certain capacity. Demand for energy needs to be continuously satisfied which adds a big constraint to the rate at which electricity is produced. Additionally ten Haar (2010) describes how electricity supply is inelastic as energy plants have a limited capacity and must operate at a constant level. Having powerplants produce energy at irregular intervals adds unnecessary costs and could put the electricity network at risk. Electricity demand is also inelastic as consumers do not monitor and react to the constant fluctuation of energy prices in real time. These factors make it very challenging to precisely match electricity supply and demand, which causes the energy price to experience sudden jumps or "price spikes". Figure 6 shows daily German spot electricity prices from 2004 to 2009 (please note the authors made a typo on the plot). Price spikes are clearly visible such as the jump to €300/MWh.

Mayer et al. (2015) analyses daily electricity price data from the 1st of January 2004 to 31st of December 2009 of German markets. An analysis for the presence of these characteristics with historical data of Spanish markets will be conducted on section 4.3. Figure 7 shows a normal QQ-plot of the deseasonalized logarithmic returns. Near the mean the distribution tends to behave like a normal distribution, however, it diverts significantly near the tails. This indicates the presence of fat-tails in the distribution of returns of electricity prices, meaning that there are more extreme events that the ones modelled by the normal distribution.

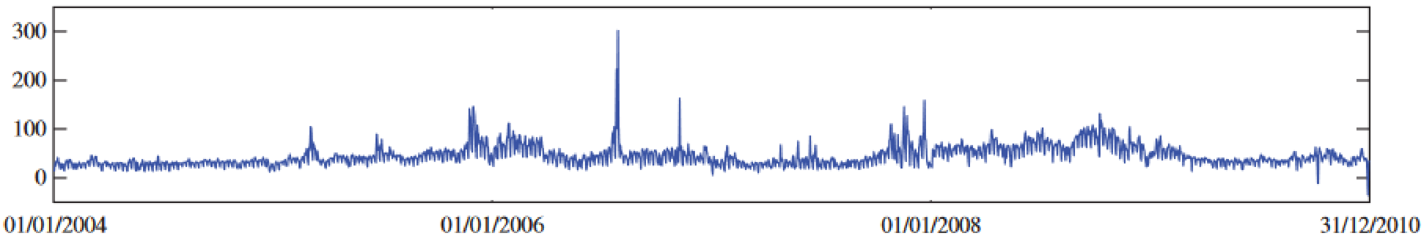


Figure 6 – German electricity spot prices (Mayer et al., 2015)

To show evidence of the impact of the spikes in electricity prices Mayer et al. (2015) run a numerical algorithm to filter the extreme events in the sample. This leaves a sample of daily electricity prices excluding the price jumps. The result is in a smaller standard deviation in the sample; however, new price spikes are formed. A normal QQ-plot of the sample is shown in Figure 8. Interestingly, Figure 8 resembles more a normal distribution than Figure 7. This comparison highlights the effects that price spikes have and that the distribution of logarithmic returns in energy prices has fat tails.

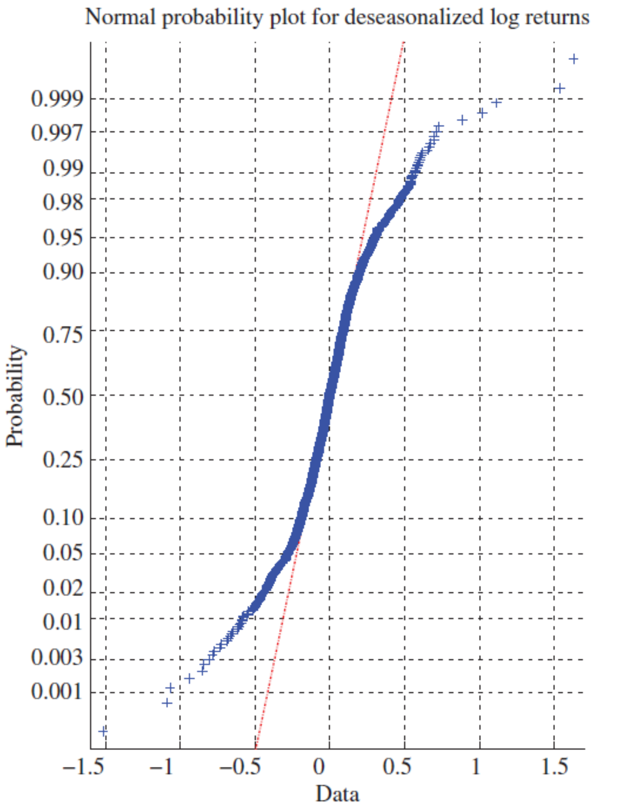


Figure 7 – QQplot log returns of electricity (Mayer et al., 2015)

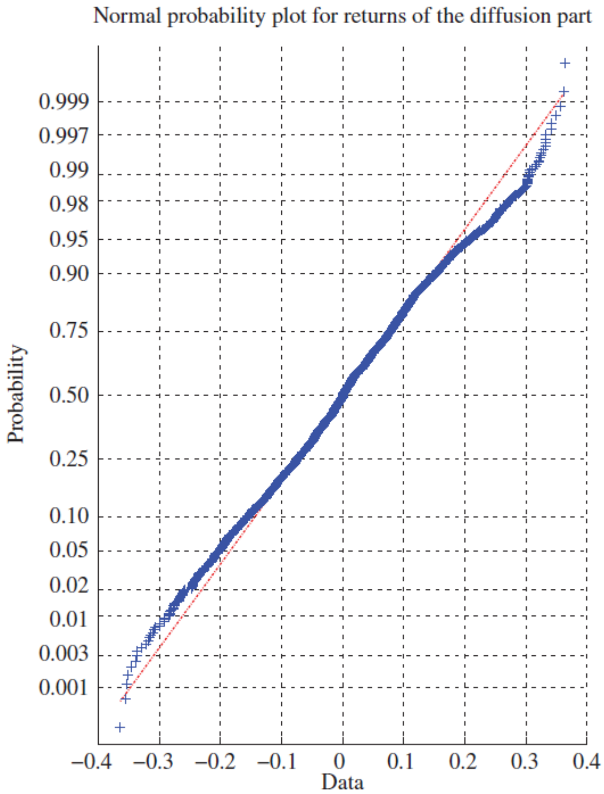


Figure 8 – QQplot excluding price jumps (Mayer et al., 2015)

Electricity exhibits seasonal patterns in futures prices, spot prices, and volatility. An analysis conducted by Størdal et al. (2023) inspects data from June 2006 and February 2021 of daily closing prices for yearly electricity futures with maturity between 1 year to 5 years from the Nordic and German markets. The study concludes that there is seasonality related to trading months. There are statistically significant higher prices in February and March and lower prices in July to October for both Nordic and German markets. These patterns can be caused by peak and off-peak demand and weather conditions. Figure 9 shows the average price for Nordic and German electricity futures contract traded on the Nasdaq and EEX. The plot on top shows futures with maturity in 1 year, the plot on the bottom show's futures with maturity in 1 to 5 years. The seasonality patters discussed earlier are clearly visible. Naturally, this could be different in the Spanish market leading to alternative patterns. However, seasonality is an effect that is present on all energy markets and is generally incorporated into pricing models.

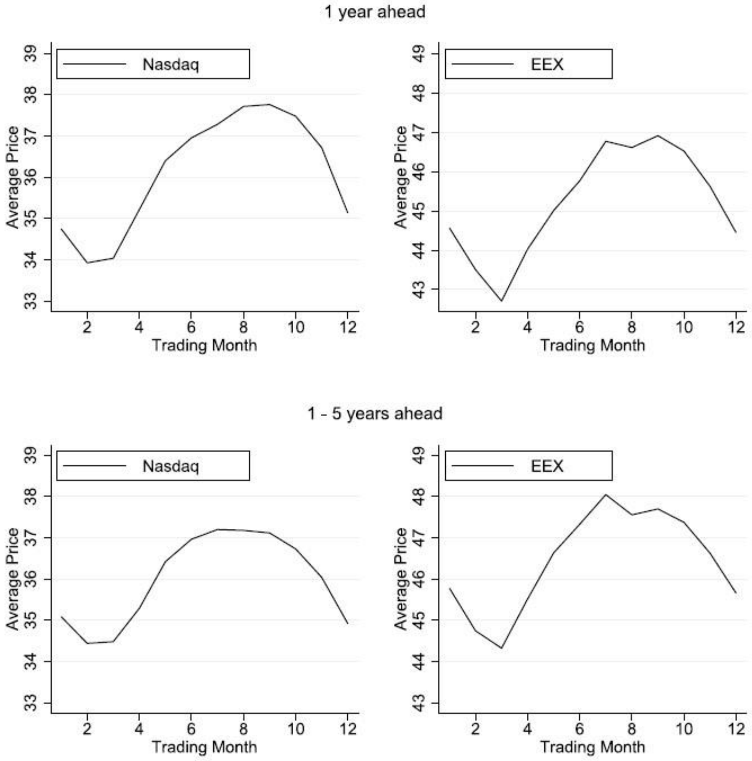


Figure 9 – Average Monthly Electricity Futures Prices (Mayer et al., 2015)

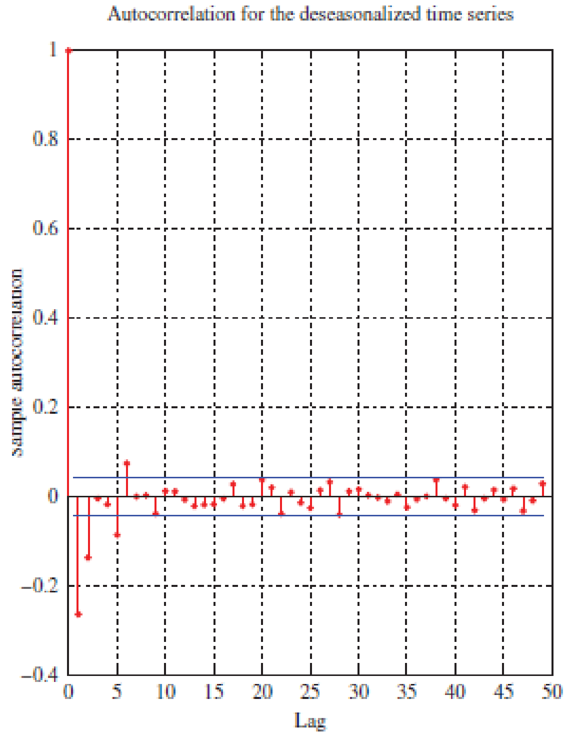


Figure 10 - ACF of deseasonalized log returns (Mayer et al., 2015)

Hull (2009) explains that commodities tend to follow a mean reversion process. This means that prices revert to a central value. Mayer et al. (2015) tests this for energy prices by analysing the time series autocorrelation function of the deseasonalized logarithmic returns plotted in Figure 10. As it can be seen there is a negative autocorrelation at lag 1 with values close to zero in posterior lags. This indicates that for an arbitrary day the returns of the next day are slightly negatively correlated, and the returns of posterior days are independent of the present day. This shows the mean reversion behaviour in energy prices.

This section highlighted three special characteristics on the behaviour of electricity prices. First, electricity prices experience jumps. Secondly, there are some periodic seasonal fluctuations over the course of one year. Finally, electricity experiences mean reversion in which prices return to a central value.

2.2.4 Pricing models

Aïd (2015) explains that there are two main approaches to model electricity spot and futures prices. The most common approach lies on modelling futures prices and using the convergence relationship to the spot price. By using this the spot price can be modelled as a futures price at its maturity date. The other approach consists of the converse, first the spot price is modelled and used to derive the futures price. This can be done either by a one factor or multi factor model. As multifactor models tend to be very complex, they will not be included on this project. A brief overview of models used for both approaches will be provided on this section.

HJM style models for futures contracts

As explained before the first approach models the spot price of electricity as a futures contract at maturity date as expressed in Equation 1. As explained by Aïd (2015) this model is based on Heath-Jarrow-Morton stochastic interest rate model. The full derivation is provided by the author, however, the main idea behind the model is presented on this section. Using the convergence relationship explained before the price of a futures at maturity $F(t,T)$ equals the price of electricity S_t Hull (2009) explains that the electricity spot price S_t price can be modelled with the generalized Wiener process of $ln\ (S_t)$ given by Equation 2.

$$d\ln(S_t) = [\theta(t) - a\ln(t)]dt + \sigma\, dz \quad (2)$$

Here a is the speed at which prices revert to the mean value. $\theta(t)$ represents seasonality and trends. Finally, σ refers to the volatility in price changes. Both a and σ are obtained from historical data and $\theta(t)$ can be obtained from futures price data.

One factor spot price model

The following models try to replicate the behaviour observed in energy prices discussed earlier: mean-reversion, seasonality, and price spikes. Aïd (2015) presents different one factor spot price models as shown in the table shown in Figure 11. These models are listed in ascending complexity.

Model	Seasonality	Process	Jumps	Change of measure
Lucia and Schwartz [127]	Deterministic	Gaussian MR		Constant
Cartea and Figueroa [61]	Deterministic	Gaussian MR	Poisson	Constant
Benth et al. [22]	Deterministic	Gaussian MR	Lévy	Time-dependant
Benth and Benth [28]	Deterministic	Non-Gaussian MR	–	Constant

Figure 11 – Factor models for electricity spot prices (Aïd, 2015)

Even though Lucia & Schwartz (2002) model is the most basic it is the foundation of more complex models. This model follows a Gaussian mean reversion process and uses a deterministic time series to represent seasonality. This is represented on the Equation 3 and 4 shown below.

$$\ln(S_t) = \theta(t) + Y_t \quad (3)$$

$$dY_t = -aY_t dt + \sigma dz \quad (4)$$

Here S_t is the daily spot price, the notations for $\theta(t)$, a, and σ are the same as presented in the HJM style model for futures contracts. Equation 4 is a generalized Wiener process for Y_T.

Even though the model presented by Lucia & Schwartz (2002) incorporates mean reversion and seasonality it fails to incorporate the price spikes observed in electricity price changes. To overcome this issue both Cartea & Figueroa (2006) and Benth et al. (2003) add a jump term to Equation 4. Benth & Šaltytė-Benth (2004) take a different approach compared to the rest by modelling energy spot prices as a non-gaussian mean reversion model. This model is too complicated for the scope of this project therefore it will not be discussed. To conclude this section, it has been discussed that fluctuations in energy prices exhibit mean reversion, seasonality, and price spikes. Different models for electricity spot prices have been discussed, some cover all these characteristics. On the following sections one of these models will be selected based on how well these three characteristics are captured.

2.3 Modelling energy yield of solar energy projects

2.3.1 Solar Energy

Böttcher (2020) explains how solar energy functions. Solar energy panels make use of the Photovoltaic effect which converts light into electricity. This is done via the emission of electrons when solar radiation gets in contact the panel, producing direct current. This can then be converted into alternative current using an inverter and later be transmitted and distributed. As one might imagine the amount of energy produced by a cell this is directly affected by the amount of sunlight received, which varies over the year. Figure 12 shows the variation in global irradiation over the different months of the year. Photovoltaics started to be used for power generation since the late 2000. Due to technological developments, photovoltaics has become the world's fastest growing renewable energy source as seen in Figure 13.

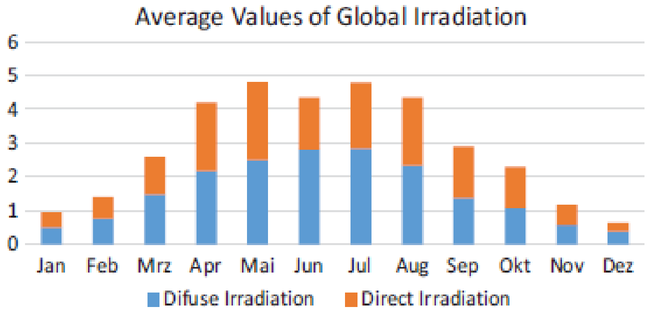


Figure 12-Global Solar Irradiation (Böttcher 2020)

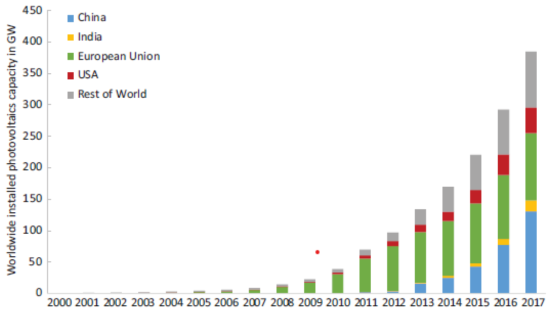


Figure 13 – Use of PVs (Böttcher 2020)

2.3.2 Energy Yield

Böttcher (2020) defines energy yield as the amount of energy produced over a certain period of time. This is usually expressed in kWh which refers to the number of kilowatts produced per hour. This is an important figure when evaluating a solar energy project as it shows how much electricity can be expected to be produced, indicating potential revenues from the project. It is also a fundamental metric for stakeholders in the project to determine the financial risk that it proposes. Energy Yield assessments consist of evaluating the projects characteristics to come up with a theoretical estimate of what will the energy yield of the project be. The energy yield assessment is usually conducted by a technical advisor. As Böttcher (2020) explains it is conducted by evaluating factors such as the influences of the construction site, technical aspects, and meteorological resources. Another important point to be analysed are the potential energy losses caused by shades, dust/snow, temperature, or reflection. Once these factors are processed an estimate for the energy yield is calculated. As previously discussed, there are many that could affect the actual energy yield, therefore there is some uncertainty related to this figure. In the case of solar energy projects, the variability in energy yield is usually modelled by a normal distribution.

2.3.3 Probability of Exceedance

As mentioned previously energy yield is an important metric for evaluating the financial risk of a solar energy project. Sengupta et al. (2017) explain that this is usually calculated by finding the "Probability of Exceedance", in short "POE" or "P". This metric refers to the probability of exceeding a certain value of energy yield. The predicted energy yield obtained from the energy yield assessment is the median of this distribution or the "P50". When evaluating a solar energy project a certain value for energy yield is assumed. The probability of exceedance allows to assume a value that will allow for a realistic assumption. For example, if a P90 value is assumed when evaluating a solar energy project then there is a 90% probability that the assumed value is either met or exceeded.

2.4 Conclusion

The structure of the Spanish electricity market was presented on this chapter. Firstly, there is the supply chain for electricity. Electricity gets produced and then transmitted at high voltages trough long distances. It then gets distributed to consumers via a low voltage network. OMIE is the main regulating entity for Spanish electricity markets. There are two different types of markets for electricity, the spot market and futures market, the price of a futures price of electricity converges to spot price at maturity date. Electricity prices have the characteristics of mean reversion, seasonality, and price spikes. Some of the mathematical models for electricity spot price include HJM style models for a futures contract at maturity date and one factor models. Energy yield is the amount of energy produced over a certain period. In the case of solar energy, the energy yield is determined by evaluating meteorological factors and potential losses. The process for determining energy yield is called energy yield assessment and is conducted by a technical advisor. After this process the advisor provides "Probability of Exceedance" numbers which are the metrics used for financial analysis.

Chapter 3- Cash flow model

This chapter covers the actions explained for phase 1 and partly of part 3 of the problem-solving methodology presented by Robinson (2004). Firstly, step 1 in which the following research question will be answered: *How does the current cash flow model work for solar energy projects in Spain?* As explained in chapter 1 an Input-Output mapping must be made. The Input-Output mapping first consists of a description of how the simulated inputs (energy prices and energy yield) are used in the cashflow model (section 3.1). Secondly, the input output mapping discusses all calculations performed by the cashflow model (section 3.2). Section 3.2.1 discusses all the simplifications and assumptions that where made such that the calculations done by the cashflow model could be adapted into the simulation. Section 3.2.2 presents all the other inputs that must be provided by the user instead of being simulated. Section 3.2.3 provides a definition of all the terms used in the calculations done by the cashflow model. Section 3.2.4 then outlines the calculations done by the cashflow model. The final step of the Input-Output mapping is explaining how to calculate CFADS, DS, and DSCR out of the cashflow statement (section 3.3). For part 3 of the problem-solving methodology the following question was answered: *How to make sure the generated distributions for CFADS, Debt Service, and Debt Coverage Ratio are valid?* This was done by reviewing the Input-Output mapping presented in this chapter with an employee of Company X to verify that the simulation made valid calculations of the outputs of the cashflow model.

3.1 Simulated Inputs

The current cash flow model has a big array of inputs. As a result, only the two inputs that are to be simulated in this project and their use will be described in detail on this section.

<u>Energy Prices</u>

The current cash flow model of Company X uses a specific price curve as an input. This price curve is a forecast by an external advisor hired by Company X, which gives a prediction for the average spot price of energy for each year in which the cash flow model makes projections. According to the external advisor, electricity markets experience short term random fluctuation and movements in the long term are driven by fundamental events. The forecast advisor constantly studies the fundamentals of energy prices out of which it calculates the predicted energy prices. The forecast advisor provides a high, middle, and low scenario for energy prices. Usually, the low scenario or an average between the middle and low scenario is used by Company X. By only using this forecast just the fundamental component is captured and random fluctuations are ignored. The mathematical model for energy prices will be used to add this random component to the cash flow model. The energy price is used to calculate revenues. Please note that there is a distinction between the types of revenue a project generates: Contracted Revenues and Floating Revenues. An energy project usually signs a Power Purchase Agreement (PPA) which is a contract that gives a party an obligation to buy a certain amount of energy at a predetermined price over a certain time. Revenues obtained from a PPA are denominated as Contracted Revenues. Clearly this is not affected by the spot price of energy as the PPA "locks" the price at which electricity is sold. Floating Revenues are the opposite, these are the sales of energy outside of the PPA and therefore are affected by the price of electricity for that given year.

<u>Energy Yield</u>

The definition of energy yield was provided on the previous section. When a project is evaluated, an external technical advisor will make meteorological measurements of the location where the project must be built. Using this the advisor will provide the following probability of exceedance numbers on the energy yield assessment: P50, P75, and P90. Usually, the P90 is used, having the P50 and P75 allows to make sensitivity analysis. Please note that the energy yield indicates the number of Megawatt hours (MWh) produced per annum for a particular asset (e.g. a solar energy plant). Another interesting observation is that there is no explicit mention of the distribution used for calculating these numbers in the energy yield assessment report.

3.2 Calculations

In this section the general assumptions that were made for simplifying and generalizing calculations where presented, followed up by a definition of all the terms used in the calculations and how are they obtained. Subsequently, an overview of the cashflow statement is given to show how the calculations within the model are made.

This cash flow model is used to evaluate renewable energy projects to which Company X finances. Its main function is to determine the profitability of the project and if it can comply with the terms in the debt structure. This model forecasts the revenues, costs, and cashflows over the lifetime of the project or debt structure. It them calculates the figures of cashflow available for debt service, debt service and debt coverage ratio. These figures are useful for Company X as they are used to determine how much debt can be made available for the project. Before the cash flow statement is introduced a brief explanation of its elements will be given. The cash flow model starts by projecting the revenues, generated from energy production. Revenues are split into two different categories. On one hand contracted revenues refers to the energy that is sold at a predetermined price via a power purchase agreement (PPA). On the other hand, floating revenues refers to energy that is sold at market spot price, which in the case of this project will be the simulated spot price via a mathematical model. Operational expenses and taxes are then subtracted to the revenues to obtain cashflows available for debt service. Using those cash flows debt repayments, interest, and fees can be covered. The result is the net cashflow of the project which can then be distributed to the shareholders via a dividend, saved in the cash reserves, or used to repay debt in advance via a cash sweep.

3.2.1 General Assumptions

The cash flow model that this simulation tries to replicate is very complex. Additionally, it always needs to be tailored to the specific characteristics of the project being evaluated. As a result, a generalization of the cash flow model needs to be done such that the simulation can be used to evaluate any type of Spanish solar energy project. For this generalization some assumptions are needed such that the cash flow model is simplified.

The assumptions are as follows:

1) Out of the total energy sold there is a constant percentage that is sold at contracted price (contracted volume rate) and a constant percentage that is sold at spot price.
2) Contracted volumes are sold at a constant predetermined price (contracted price).
3) Taxes are applied at a constant rate and are zero if there is a loss for that particular year.

4) Depreciation is applied at a constant rate over a set period. After this period, it is zero.
5) Interest payments are given by a constant rate.
6) Fees are constant overtime.
7) All net cashflows are divided at a constant rate between dividends, cash sweep, and cash reserves.
8) There is a mandatory debt service which denotes the debt payments that should be made and there is realized debt service which denotes the debt payments that were made for a particular year.
9) Cash reserves can be used to cover any shortcomings in debt service and/or negative CFADS.
10) Cash reserves, dividend payments and cash sweep cannot be negative.
11) If on a particular year the realized debt service is below mandatory debt service, the project goes into default. If a default happens the simulation continues.
12) Any shortcomings on debt service that cannot be covered by cash reserves will be accumulated for future years.
13) If CFADS is negative and cannot be covered by cash reserves, it is assumed that the difference is ignored and has no financial consequence.
14) If the total outstanding debt is zero, then debt repayments and cash sweep rate are zero.

These assumptions were taken under the supervision of a Company X employee such that the main characteristics of the cash flow model are preserved in the simulation. The assumptions allow to build a good starting concept for the simulation however they do not allow to perfectly replicate the calculations done in the cashflow model. An example is that a project might sign different PPAs at earlier years which causes energy produced to be sold at different contracted prices with different volumes. Additionally, the percentage of energy sold at merchant or contract price varies over time with exposure to merchant price usually being close to zero at earlier years and close to 100% at later years. This situation already violates assumptions 1 and 2. A point for future research would be to adapt the assumptions for a more accurate representation of the cashflow model.

3.2.2 Inputs from user

To run the cash flow model more inputs than the ones explained in section 3.1 are needed. These inputs will not be simulated, instead they will need to be provided by the user based on their respective values on the current cash flow model.

This is a list of values that are provided by the user and are inputs for the simulation. Inputs 1 to 8 together with the simulated average spot price for one year and simulated energy yield are the necessary items such that the simulation can calculate CFADS, DS, and DSCR for that year. Input 9 is necessary as initial conditions for the model for energy spot prices. Input 10 is the number of iterations that will be executed by the simulation. Inputs 9 and 10 will be explained in the coming sections.

1) Contracted volumes, tax, interest, cash sweep, cash reserves, and depreciation rate.
2) Contracted Price
3) Starting Debt Value
4) Fees
5) Asset Value
6) OPEX

7) Start and end date
8) Mandatory debt repayment for each year
9) Energy Spot Prices
10) Number of Iterations

3.2.3 Definitions

Calculating CFADS, DS, DSCR involves calculating different financial figures out of the simulated energy prices & energy yield and the inputs provided by the user. This section provides a definition to all items necessary to perform the calculations of section 3.2.4 out of which CFADS, DS, and DSCR can be calculated. These definitions are outlined such that the assumptions introduced on section 3.2.1 are met.

<u>Rates</u>

These are all the items who only have a value between zero and one.

1) v := contracted volumes rate; this is the percentage of energy yield sold at a contracted price.
2) t := tax rate; percentage of profits paid as taxes.
3) r := interest rate; percentage of debt outstanding charged as interest on debt.
4) cs := cash sweep rate; percentage of net cashflows used to repay debt in advance.
5) cr := cash reserves rate; percentage of net cashflows saved as reserves.
6) d := depreciation rate; percentage of asset value depreciated each year.

$0 \leq v, t, r, (cs+cr), d \leq 1$

<u>Financial Items</u>

These are financial items whose calculations stays constant.

1) Contracted Revenue := Simulated Energy Yield * v * Contracted Price
2) Floating Revenue := Simulated Energy Yield * (1 – v) * Simulated Average Yearly Energy Spot Price
3) EBITDA := Total Revenue – OPEX
4) Taxes := max{(EBITDA – Depreciation – Interest) * t, 0}
5) Mandatory Debt Service := Debt Repayment + Interest + Fees
6) Cash Sweep := max{(Net Cashflows * cs), 0}
7) Dividends := max{Net Cashflows * (1 – cs - cr), 0 }

<u>Conditional Items</u>

These are items whose calculation depends on certain conditions.

1) Adjust the value on cash reserves if cash was used due to a shortfall in debt service or negative CFADS.

IF CFADS ≥Mandatory Debt Service *Then*

Cash Reserves Used = 0

ELSE IF CFADS < Mandatory Debt Service *AND* Cash Reserves ≥*(*Mandatory Debt Service – CFADS) *THEN*

Cash Reserves Used = (Mandatory Debt Service – CFADS)

ELSE

Cash Reserves Used = Cash Reserves

END IF

2) Adjust the realized debt service depending on if the project produced enough cashflows.

IF CFADS $\geq$Mandatory Debt Service *THEN*

Realized Debt Service = Mandatory Debt Service

Else

Realized Debt Service = max[(CFADS + Cash Reserves Used), 0]

END IF

3) Depreciation only takes place during a fixed number of years.

IF year number $>$ d^{-1} *THEN*

Depreciation = 0

Else

Depreciation = Asset Value * d

END IF

4) Stop debt repayments and cash sweep after debt is repaid.

IF Outstanding Debt $= 0$ *THEN*

Debt repayment = 0

cs = 0

END IF

<u>Time Based Items</u>

These are items whose calculation is indexed by time.

For a year t in the lifetime of the project t $\in [1, T]$

1) Interest $_t$ = Outstanding Debt $_t$ * r
2) Outstanding debt $_t$ = Outstanding debt $_{t-1}$ - ([Realized Debt Service$_{t-1}$ – (Interest + Fees)$_{t-1}$]+ Cash Sweep $_{t-1}$)
3) Cash Reserves $_t$ = Cash Reserves$_{t-1}$ – Cash Reserves Used$_{t-1}$
4) Cash Reserves$_0$ = 0
5) Outstanding debt$_0$ = Starting Debt Value

3.2.4 Cashflow statement

Using all the terms defined on the previous section allow to build a cashflow statement for each year in which the simulation runs. CFADS, DS, and DSCR are can then be calculated out of the cashflow statement given below.

For an arbitrary year $t \in [1, T]$

Contracted Revenue	
Floating Revenue	$+$
TOTAL Revenues	
OPEX	
Taxes	$-$
CFADS	
Realized debt service	$-$
Net Cashflows	
Cash Sweep	
Cash Reserves	
Dividends	$-$
Result = 0	

3.3 Outputs

On the previous sections of this chapter the cash flow model has been introduced. These sections show which inputs will be simulated (energy prices an energy yield) and which outputs will be provided by the user. A definition for all elements used to do calculations is provided. These elements are used in the cashflow statement to model the yearly financial development of the project. This section will illustrate how the three outputs (CFADS, DS, DSCR) that this research focuses on are calculated out of the cashflow statement.

<u>Cashflow available for debt service (CFADS)</u>

This item can be clearly seen on the cashflow statement. It refers to all the cash generated on one year that is available to make debt related payments.

<u>Debt Service (DS)</u>

This item refers to the amount of cash destined to make debt related payments. It is given by the realized debt service as seen in the cashflow statement. Please note that cash sweep payments (debt repayments made in advance) do not form part of realized debt service.

<u>Debt Service Coverage Ratio (DSCR)</u>

This item indicates how well the debt payments are "covered" by current cashflows. It can be calculated by dividing CFADS by mandatory debt service. As previously mentioned, Company X usually likes to maintain this ratio between 1.2 – 1.4. Having a low debt coverage ratio indicates that the amount of cash generate by a project might fall short or close to short to make the necessary debt payments, putting the project at risk.

All the presented calculations of section 3.2.4 are performed in each year and these three outputs illustrate the following: How much cash is being generated by the project? How is debt being repaid? How well covered are the financial obligations of the project by the generated cashflows? Due to confidentiality issues an illustration of the actual outputs of the cash flow model cannot be shown. Duldinger (2023) provides a template for a cash flow model for project finance with fictional data. It is not a completely accurate representation of Company X's cash flow model and covers some areas not discussed on this section (e.g. Valuation Summary and Investment Ratios & multiples). However, it allows for an example of how a project finance model looks like and illustrates the outputs discussed on this section.

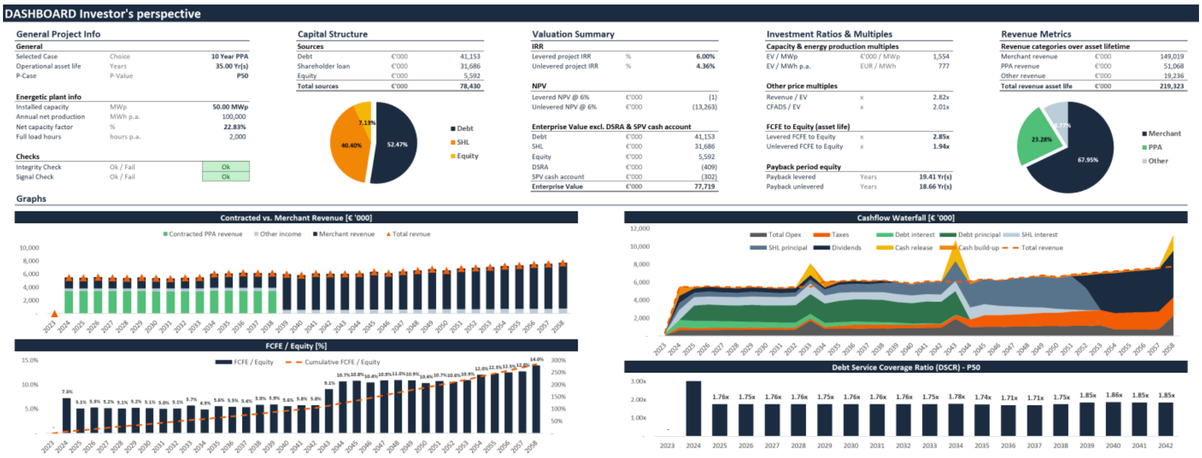


Figure 14 – Illustration outputs of a financial model for project finance (Duldinger 2023)

3.4 Conclusion

This chapter discussed the current cash flow model of Company X and how the calculations will be replicated within the simulation to obtain the yearly CFADS, DS, and DSCR. Currently Company X uses a forecast from an energy price consultant as input for the cash flow model. For energy yield a specific probability of exceedance value gets selected, usually the P90. These two inputs will get replaced for the simulated scenarios. The cash flow model also uses other inputs such as starting debt value and OPEX, these are listed on section 3.2.3 and will need to be provided by the user in the simulation. Section 3.3.3 shows how different items are defined that are used for the cashflow statement presented on section 3.3.4. Out of the cashflow statement the CFADS, DS, and DSCR can be calculated as shown in section 3.3. The content of this chapter was verified with an employee of Company X to ensure validity in the calculations of CFADS, DS, and DSCR. The employee indicated that some of the assumptions discussed in section 3.2.2 do not fully match the reality of the cashflow model. These assumptions allow to build a good starting concept whoever one of the main points of future research is to change them such that the simulation can better resemble the real cashflow model.

Chapter 4- Simulating Energy Prices

As explained in section 3.2 one of the simulated inputs is electricity spot prices. There are multiple mathematical models as shown in section 2.2.4 for simulating electricity spot prices. This section will answer the question: *What is a suitable mathematical model to simulate Spanish energy spot prices such that it represents seasonality, mean reversion, and price jumps?* After answering this question, a mathematical modelled calibrated to the Spanish market will be obtained. This section first discusses the chosen model (Cartea & Figueroa (2006)'s jump diffusion) out of the ones proposed in the theoretical framework (section 4.1). Then the Euler-Maruyama method for approximating stochastic differential equations is presented and applied to the selected model for the spot price (section 4.2). Finally, data analysis programmed in R is conducted to calibrate the model to historical Spanish energy prices (section 4.3).

4.1 Cartea & Figueroa Jump Diffusion model

As explained in the previous section there are two main approaches to model electricity spot prices. On one hand as explained by Hull (2009), the electricity spot price can be modelled by the price of a futures contract at maturity date. On the other hand Aïd (2015) introduces various factor spot price models for electricity. The criteria for selecting a model will be based on: 1) The model covers the characteristics of electricity price fluctuations discussed on the previous section (mean-reversion, seasonality, and price spikes). 2) The model is relatively easy to apply.

With this criteria the model by Lucia & Schwartz (2002) and Hull (2009) can be discarded as they fail to incorporate price spikes. Even though the models by Benth et al. (2003) and Benth & Šaltytė-Benth (2004) include mean reversion, seasonality, and price jumps they are more mathematically complex. This means that it will be difficult to apply them, therefore they are discarded as well. As a result the model by Cartea & Figueroa (2006) will be selected, which will be introduced on this section.

The notation used in the section of the theoretical framework about electricity price models will be utilized here. Cartea & Figueroa (2006) state that the electricity spot price $S_t$is given by the exponential of a seasonal and stochastic component as shown in Equation 5

$$S_t = e^{\theta(t)+Y_t} \tag{5}$$

Where $\theta(t)$ is a seasonality function and Y_t is a stochastic process given by the following stochastic differential equation as shown in Equation 6.

$$dY_t = -aY_t dt + \sigma(t) dZ_t + \ln(J)\, dq_t \tag{6}$$

Where a is the mean reversion rate, $\sigma(t)$ is time dependent price volatility, $dZ_t$is the standard wiener increment, J is the proportional random jump size, and dq_t is a Poisson process with frequency l. The Poisson process followed by dq_t is defined in Equation 7.

$$P(dq_t = x) = \begin{cases} 1 - ldt & for\ x = 0 \\ ldt & for\ x = 1 \end{cases} \tag{7}$$

The jump size J, standard wiener increment dZ_t, and occurrence of jumps dq_t are assumed to be independent. With these assumptions the proportional random jump size J follows a lognormal distribution as shown in Equation 8. Where σ_J is the standard deviation of jump sizes.

$$\ln(J) \sim N(-\frac{\sigma_J^2}{2}, \sigma_J) \tag{8}$$

From Equations 5 and 6 a stochastic differential equation for the spot price S_t can be written as Equations 9 and 10.

$$dS_t = a[\rho(t) - \ln(S_t)]dt + \sigma(t)S_t dZ_t + S_t(J-1)\,dq_t \tag{9}$$

$$\rho(t) = \frac{1}{a}\left(\frac{d\theta}{dt} + \frac{1}{2}\sigma^2(t)\right) + \theta(t) \tag{10}$$

4.2 Euler Maruyama approximation for Stochastic Differential Equations

Cartea & Figueroa (2006) do not provide an analytical solution for the SDE introduced in Equation 6. As a result, it must be numerically approximated to obtain a solution. This can be done via the Euler-Maruyama method as explained by Maruyama (1955). This method is straightforward to program and combine with the Montecarlo simulation created on this project. This approximation was developed by Japanese mathematician Gisiro Maruyama and is based on the Euler method for approximating ordinary differential equations. On this section the Euler Maruyama method will be explained and applied to Equation 6.

For this method a initial condition X_0 is needed which in the case of this project is some value for the energy spot price of the Spanish market. It can then be converted into the stochastic component of energy prices Y_0 by using Equation 5. Equation 6 states that the increment in time dt is infinitesimally small, however in the Euler-Maruyama method the time increment is approximated by a discrete time step of size Δt. For example, consider that energy prices for next year are going to be simulated. The discrete timesteps are defined by diving the period over n sub-intervals. For this example, an interval will be a day, meaning that $n = 365$. Then Δt is given by: $\Delta t = 1/n$. This implies that the rest of differentials in Equation 6 are approximated by:

$$dS_t \;\approx\; \Delta S_t \;=\; S_{t+1} - S_t$$

$$dZ_t \;\approx\; \Delta Z_t \;=\; Z_{t+1} - Z_t$$

$$dq_t \;\approx\; \Delta q_t \sim Poisson(\,l\Delta t)$$

$$where\ t\ is\ an\ arbitary\ subinterval\ and\ t\ \in [0, n)$$

Using the properties of a Wiener process presented by Hull (2009) in Appendix A it can be shown that:

$$\Delta Z_t \sim N(0, \sqrt{\Delta t})$$

This allows to approximate Equation 6 to obtain Equations 11 and 12.

$$Y_{t+1} = Y_t - \ aY_T \Delta t + \sigma(t) Y_t \Delta Z_t + \psi(\Delta q_t\ ) \tag{11}$$

$$\psi(\Delta q_t\ ) = \begin{cases} 0 & if\ \ \Delta q_t\ = 0 \\ \sum_{i=1}^{\Delta q_t} \ln{(J)}_i & if\ \Delta q_t\ > 0 \end{cases} \tag{12}$$

Equation 11 is the approximation of Equation 6 and Equation 12 is the cumulative jump size over the time interval t. Using these equations energy prices can be simulated by the following procedure. First, use the initial condition Y_0 in equation 11. This will give the log stochastic energy price component for next day, which can then be used again in Equation 11 to obtain the result in day 2. This procedure is repeated until day 365 is reached. Then using equation 5 the energy spot price can be calculated. This results for a simulated path of energy prices over a period of a year. This procedure has been coded on VBA such that possible paths can be simulated over n intervals.

The terms ΔZ_t and $ln\ (J)$ can be simulated by using excel/VBA's NORM.INV function with their respective means and standard deviations and a random number within 0 and 1. As there is no inverse Poisson function in VBA which will allow to simulate Δq_t it needs to be approximated. In here the probability of Δq_t > 2 is assumed to be negligibly small therefore it can be omitted. Using the same jump intensity as given by Cartea & Figueroa (2006) the probability of having 2 or more than jumps on the same day is already of 0.027 %. Using this assumption an inverse Poisson function can be built as shown in Equation 13.

$$let\ \pi\ be\ a\ random\ number\ between\ 0\ and\ 1$$

$$\Delta q_t = \begin{cases} 0, & \pi\ \in (0, P(\Delta q_t\ \leq 0)] \\ 1, & \pi \in (P(\Delta q_t\ \leq 0), P(\Delta q_t\ \leq 1)] \\ 2, & \pi \in (P(\Delta q_t\ \leq 1), 1) \end{cases} \tag{13}$$

4.3 Calibrating the model to the Spanish market

As it can be seen on the previous section Cartea & Figueroa (2006)'s model depends on multiple parameters which can be estimated from historical data. As this research is focused on Spanish photovoltaic projects the data will consist of historical spot prices from the Spanish market. Cartea & Figueroa (2006)'s jump diffusion model works best with a daily time step, as with monthly and yearly timesteps jumps are rarely seen. As a result, for calibrating the model daily prices from 01/01/1998 to 08/04/2024 where analysed. This dataset was obtained via the Trading Economics database which Company X provided access to. A plot of the Spanish energy prices in the dataset is shown in Figure 14. The presence of price spikes and mean reversion can be observed in the plot.

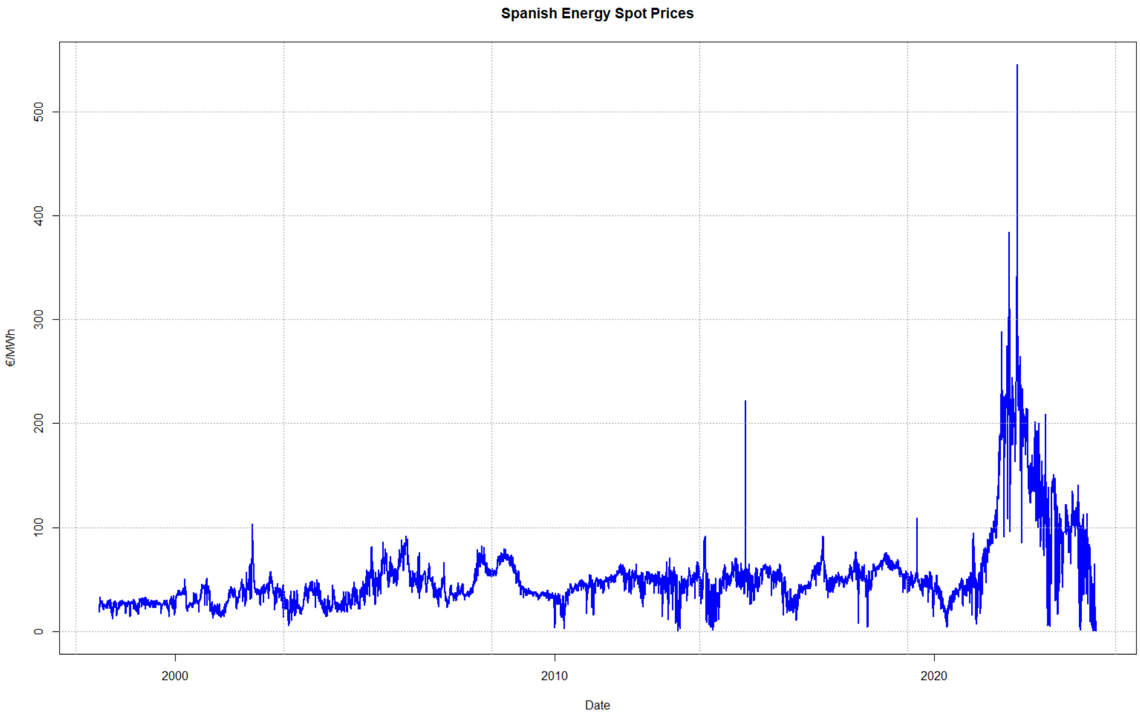


Figure 15 – Spanish energy spot prices between 01/01/1998 to 08/04/2024

After discussing with the supervisor of this thesis, the energy price forecasting consultants hired by Company X, and Company X employees is that the energy market changes over time. This means that historical data extending to many years in the past might not represent the state of the energy market today. However, limiting the analysed data to the most recent observations implies losing statistical significance. A method to deal with this issue suggested by the supervisor of this thesis is to have an exponential moving average. In here an older observation gets a lower weight when calculating statistics. This method will not be used on this thesis, but it could be a point for future research. The energy price forecasting consultants mentioned that when analysing historical data, they focus on 4 to 5 years and might exclude years with unusual conditions. Cartea & Figueroa (2006) use 5 years of data for calibrating the model.

As the jump component is an important part of this model it is interesting to observe the jump frequency and size within the dataset. Figure 15 is a plot of all the recorded jumps in log returns from 1998 to 2024. In recent years the jump frequency is somewhat similar, however the data points from 2023 and 2024 contain a high jump frequency and extremely large jump sizes. As a result, the calibration for the model will be done with data from 2018 to 2022. This allows to have a dataset of sufficient size to derive statistically significant results which focuses on the behaviour observed in recent years.

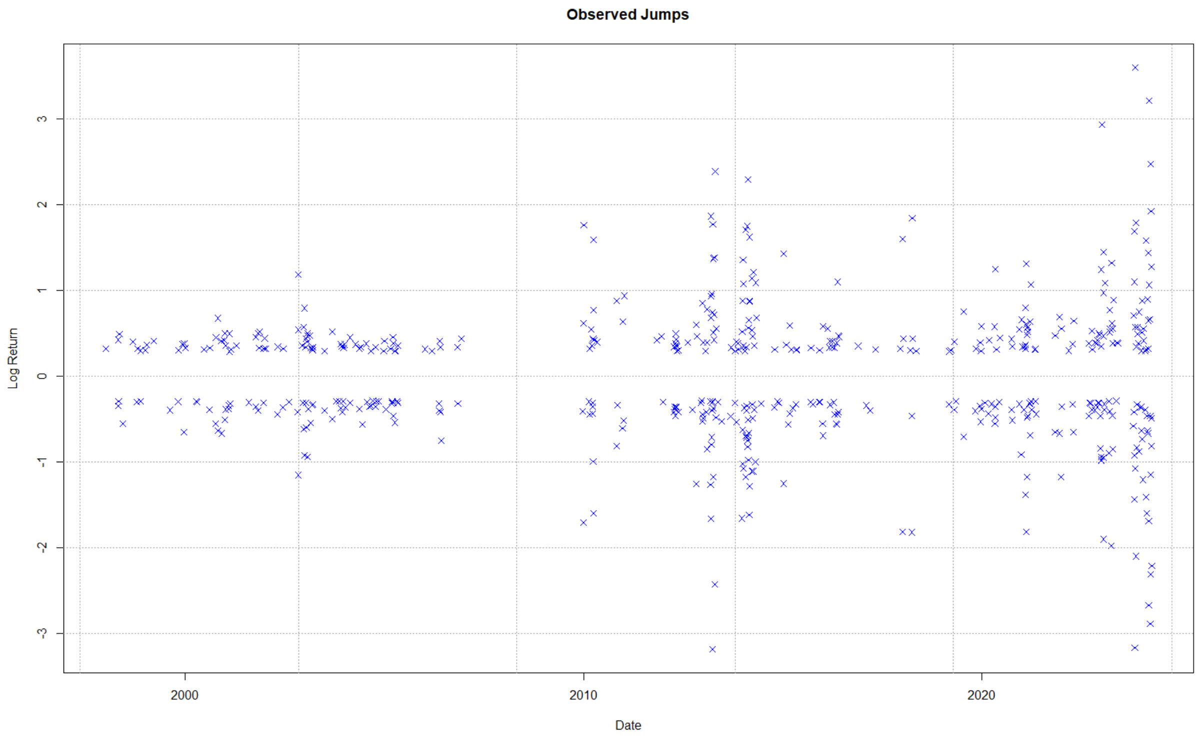


Figure 16 – Price jumps on log returns of Spanish energy spot prices between 01/01/1998 to 08/04/2024

4.3.1 Mean Reversion Rate

Cartea & Figueroa (2006) explain that the mean reversion rate a can be estimated via linear regression. Boucher et al. (2023) explain by defining ΔY_t as $\Delta Y_t = Y_{t+1} - Y_t$ the linear regression can be conducted as in Equation 14. Where $\beta_1$is the slope, β_0 the intercept, and ϵ the error.

$$\Delta Y_t = \beta_1 \Delta Y_{t-1} + \beta_0 + \epsilon \tag{14}$$

Mean reversion is present if $\beta_1 < 0$. Mayer et al. (2015) that the mean reversion rate is then obtained by $a = -\beta_1$. Please note that the log return Y_t is given by $ln(S_t \,/\, S_{t-1})$ where S_t is the electricity spot price on time t. The least squared fit of Equation 14 can be conducted to approximate its coefficients. The outputs of the regression are given in Figure 16 which shows the values for the slope $\beta_1$and intercepts β_0.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -0.001680   0.007516  -0.224    0.823
independent_variable -0.453804   0.023330 -19.452   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 17 – R output linear regression

There is a significant relationship between the independent variable ΔY_{t-1} and the dependent variable ΔY_t indicating the presence of mean reversion in the historical data. The data above suggests that the mean reversion coefficient has a value of a = 0.45. A confidence interval for the mean reversion rate can also be built in R using the results of the regression.

$$CI_{95\%} - [0.41 \leq \ a \leq 0.5\ ]$$

Hull (2009) indicates that electricity prices usually have a mean reversion rate between 0.1 and 0.2. The mean reversion rate suggested by Hull (2009) does not lie in the confidence interval for mean reversion rate for the analysed dataset. A possible reason for this as explained by Mayer et al. (2015) is that the mean reversion rate is higher when a jump occurs. As it will be seen later in this section the data set contains a lot of jumps. Cartea & Figueroa (2006) obtained a mean reversion rate of 0.24 and a jump frequency of 8.5. In this dataset the mean reversion rate is 0.45 and the jump frequency is of 20.2. Also note that the model by Hull (2009) does not include jumps, so this could be a reason why the author suggests a lower mean reversion.

4.3.2 Price Volatility

First, the definitions given by Poortena et al. (2021) of unbiased estimators for the mean and standard deviation of a sample of size n are shown below.

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

Rolling volatility assumes that the log returns are independent, normally distributed, and the standard deviation is proportional to the square root of time between observations. As it can be seen Figure 18, the log returns of energy prices in this dataset do not follow a normal distribution. Additionally, because of that the log returns experiencing mean reversion as discussed in the previous section independence cannot be assumed. Additionally research conducted by Mandelbrot (1963) and Fama (1965) shows that volatility is not constant over time. Instead, volatility tends to be grouped in clusters, where there are some periods of low volatility and other periods of high volatility. There are various methods for forecasting time dependent volatility. As explained by Hull (2009) one of the most common techniques for this is the use of Generalized Autoregressive Conditional Heteroskedasticity models (GARCH). However, using these kinds of models will add an extra layer of complexity to this project. Eydeland & Wolyniec (2003) explains that returns can be assumed to be independent and follow a normal distribution with constant volatility over short periods of time. (usually over 20-30 days). Cartea & Figueroa (2006) make use of this assumption and calculate the 30-day rolling volatility for each month of the year.

Eydeland & Wolyniec (2003) explain that rolling price volatility is given by estimating the standard deviation of the variable x_t shown in Equation 15, where r_t is the log returns and Δt is the size of the time step over the returns are calculated (e.g. 1 days, 2 days, etc).

$$x_t = \frac{r_t}{\sqrt{\Delta t}} \tag{15}$$

Let $\hat{\sigma}_m$ denote the 30-day rolling volatility and $\hat{\sigma}_d$ the standard deviation of daily log returns. The 30day rolling volatility is calculated over a time step of 30 days. It then follows from Equation 15 that $\hat{\sigma}_m$ can be calculated as shown below.

$$\hat{\sigma}_m = \sqrt{30}\,\hat{\sigma}_d$$

The 30-day rolling volatility for each month was calculated in R. The results are shown below. Hull (2009) explains that energy prices tend to have a volatility around 1 and 2. As it can be seen the results seem to be in line with this.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Volatility	1.41	1.88	1.5	1.45	0.94	0.49	0.67	0.38	0.71	0.79	0.86	1.84

Table 1 – 30-day rolling volatility

4.3.3 Seasonality

As discussed in the theoretical framework energy prices exhibit patterns of seasonality. Recall that Størdal et al. (2023) presents evidence that energy prices exhibit statistically significant price differences over different months. Additionally, periodic patters caused by different factors such as weather are observed. Therefore, the fluctuations in energy prices are divided between the deterministic seasonal component and a stochastic component as shown in Equation 5. Janczura et al. (2012) presents the three main approaches to estimate the deterministic seasonal component for energy prices. The first approach, used by Lucia & Schwartz (2002), consist on creating piecewise constant functions using dummy variables. The main drawback of this method is that it yields a non-smooth component therefore it needs additional treatment to be able to implement it on the model. The second approach, as used by Cartea & Figueroa (2006), creates a sinusoidal function using Fourier analysis. However, this comes with the drawback the periodicity in sinusoidal functions might not accurately represent the long-term evolution of prices. For solving those issues an exponentially weighted moving average together with a sinusoidal function as suggested by De Jong (2006) can be used. Finally, the third approach consists of using wavelet decomposition and smoothing. This also solves the drawbacks of the sinusoidal functions. Unfortunately using a wavelet decomposition and smoothing or the procedure of De Jong (2006) will add another layer of complexity to this project.

Using a sinusoidal function obtained by Fourier analysis as done by Cartea & Figueroa (2006) will be used in this project. This is because it allows to represent the periodic seasonal fluctuations in energy prices without adding too much complexity. However, as this project aims to make long term simulations of the evolution in energy prices it is worth mentioning that the drawback of this approach does impose a limitation to its reliability and validity. Cartea & Figueroa (2006) take the average log return for each month and then proceed to fit a Fourier Series of order 5. As suggested by the supervisor of this thesis another method to achieve this is to do the exact same procedure but to fit a spline instead of a Fourier series. This method was easier to program in R therefore a spline was fitted between the average log return for each month of the year. This was obtained by running a spline regression with functions of the splines and Ecdat packages in R. The independent variable being the day of the year and the independent variable being the seasonal component. This spline contains 5 knots, each one at the 20th, 40th, 60th, 80th, and 100th percentiles of the 365 days in the year. The plot below shows the monthly average returns on the 15th day of each month (points in red). The blue line is the fitted spline values for each one of the days in the year. Please note that in the case of a leap year it is assumed that day 366 will have the same seasonal component as day 1, therefore one more data point is added to the series in such cases. A plot of the fitted spline values to monthly average log returns is shown in Figure 17.

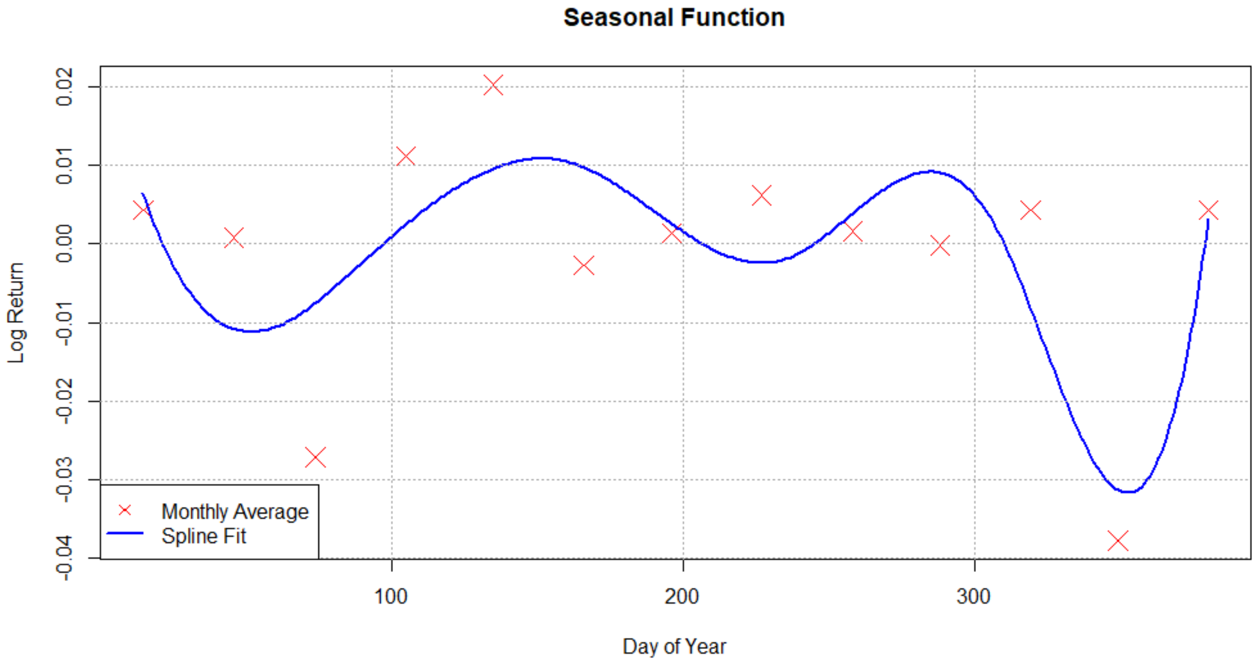


Figure 18 – Spline fit of monthly average log-returns

4.3.4 Jumps

As discussed in the section 2.2.3 the presence of jumps drops the assumption that the log returns follow a normal distribution. Figure 18 plots of the quantiles of the sample distribution of log returns

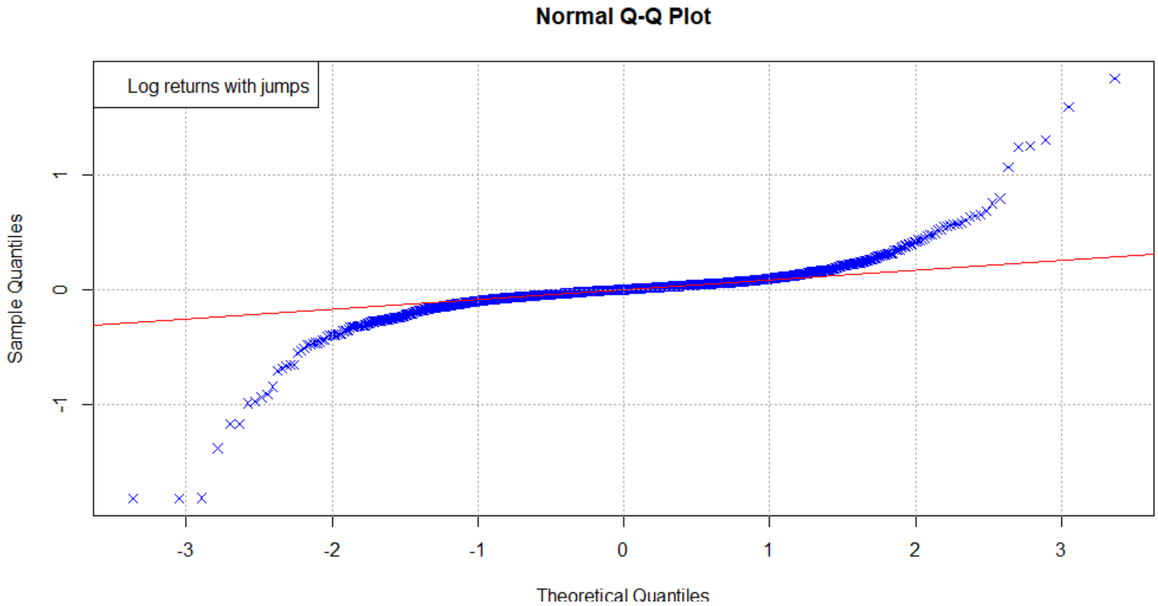


Figure 19 – Normal QQ-Plot log-returns

and the quantiles of a normal distribution. As it can be seen the deviate near the tails indicating the presence of jumps and fat-tails in the distribution.

The parameters that need to be estimated for the jump component are the yearly frequency of monthly jumps l and standard deviation of jump sizes σ_J. As explained by Cartea & Figueroa (2006) the first step to obtain the parameters is to identify all jumps on the data. This is done by the following algorithm:

Let Θ be a set

1) Deseasonalize the log returns by $R_t = r_t - \overline{r}_t$ where r_t is the log return at day t and $\overline{r}_t$ is the seasonal spline component obtained in the previous section.
2) Calculate the mean μ and standard deviation σ
3) The deseasonalized log return R_t is a jump if $|R_t| > \mu + 3\sigma$
4) Take jump out of dataset and add it to Θ.
5) Repeat steps 2, 3, 4 until there are no jumps left.

The output of this algorithm will be the set Θ which contains all jumps on the data set. σ_J can be obtained by taking the standard deviation of the elements in Θ. The frequency l can be obtained by counting the elements of Θ and dividing by the number of time periods in the analysed dataset. The plot below shows a QQplot of the log returns excluding jumps. As it can be seen it resembles more a normal distribution than in the previous case and there is no longer the presence of fat tails.

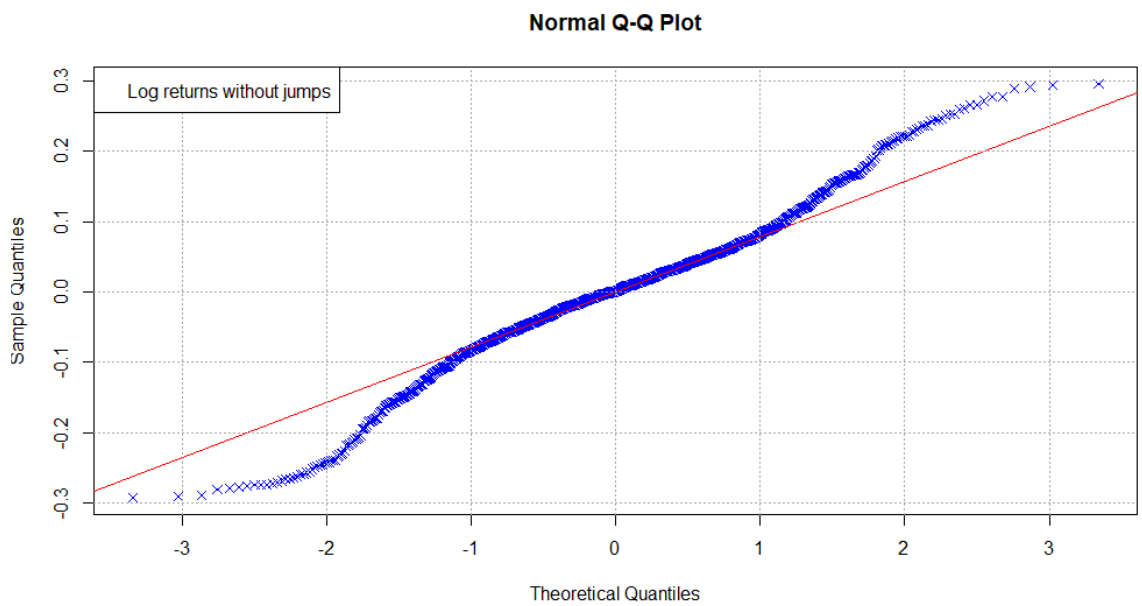


Figure 20 – Normal QQ-Plot log-returns excluding jumps

The standard deviation of jump sizes is of 0.7. Figure 20 shows a histogram of the recorded jump sizes. This is plotted together with the theoretical distribution of jump sizes given in Equation 8. As it can be seen they hold similar shapes, however the empirical distribution appears to be centred slightly closer to zero than the theoretical distribution.

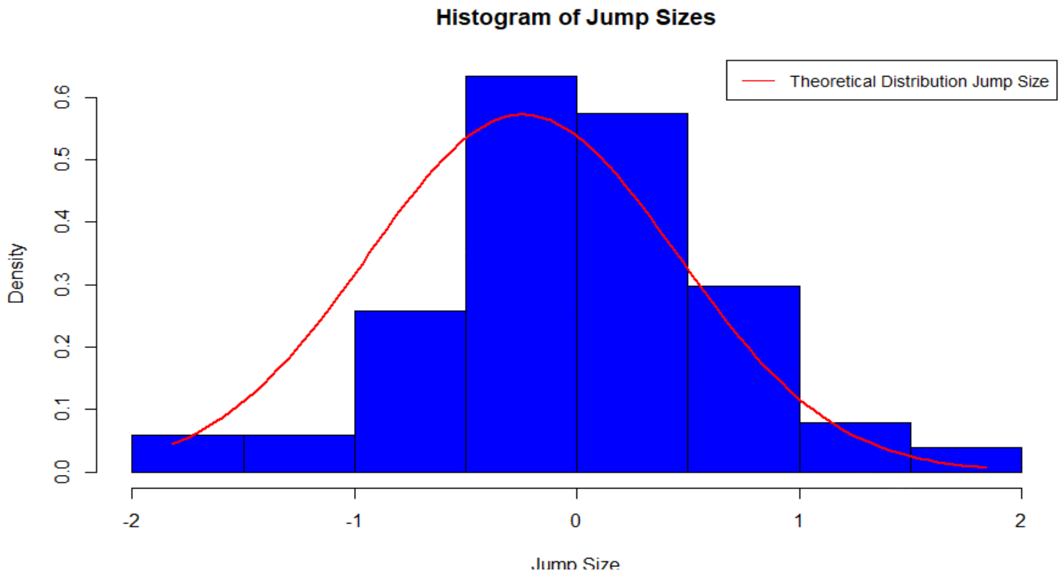


Figure 21 – Histogram of jump sizes and plot of theoretical distribution

4.4 Conclusion & Results

This chapter discussed a mathematical model used to simulate energy spot prices. Cartea & Figueroa (2006)'s jump diffusion model was selected as it represents mean reversion, seasonality, and price jumps. The SDE in the model was approximated via the Euler-Maruyama method. This model was calibrated to daily Spanish electricity spot prices from 2018 to 2022. The results from the calibration are shown in Table 2. Using these results as parameters on Equation 5, Equation 11 and Equation 12 a model for the electricity spot price calibrated to the Spanish market is obtained. Figure 21 is a plot of a simulated price path over the period of 1 year. As it can be seen the chart resembles the looks of electricity price fluctuations. Both jumps and mean reversion can clearly be observed.

Mean Reversion Rate (α)	0.45
Time dependent Volatility ($\sigma(t)$)	See Table 1
Seasonal Function ($\theta(t)$)	See Figure 17
Jump Frequency (l)	20.2
Standard Deviation Jump Size (σ_J)	0.7

Table 2 – Results calibration

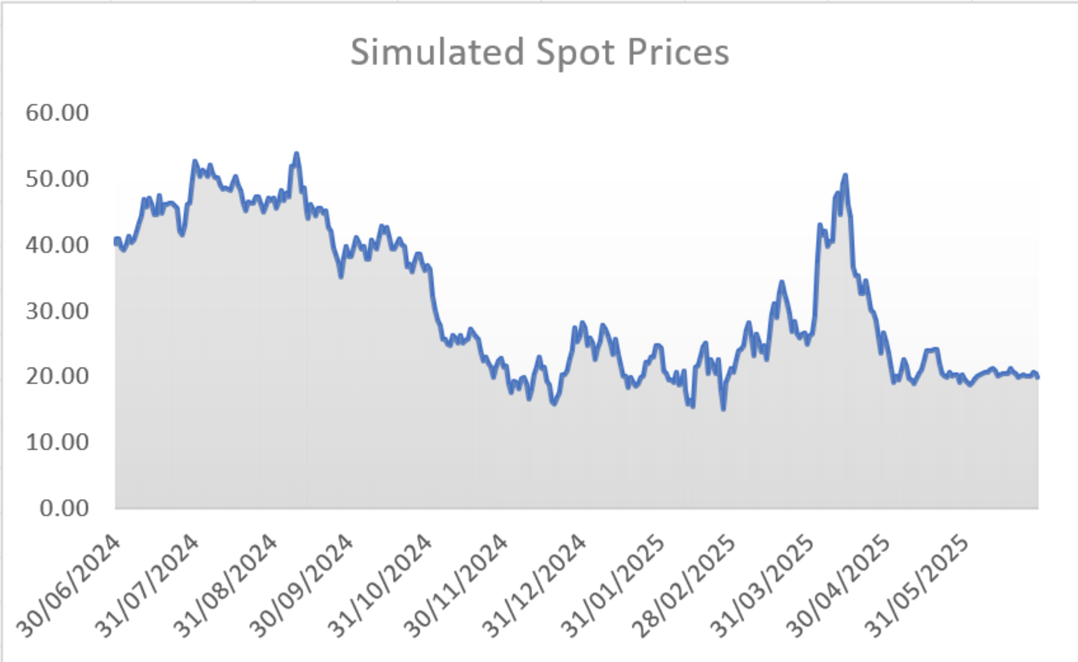


Figure 22 – Simulated spot prices with calibrated Jump Diffusion model

Chapter 5- Simulating Energy Yield

As explained in section 2.3.2 energy yield is the amount of energy produced over a certain period. This is usually expressed in terms of “Probability of exceedance”, which gives an output value such that there is certain probability of producing at least that amount of energy. The aim of this section is to answer the research question: *Can yearly energy yield be assumed to be normally distributed? If not, what alternative can be used?* By answering this question, a suitable distribution for energy yield which can then be implemented into the simulation will be found. Firstly, a description of the energy yield assessment will be made to understand the process of determining energy yield and the variables that affect its variability (section 5.1). Secondly, a discussion of the type of distribution that can be used to modelled it will be made (section 5.2).

5.1 Energy Yield Assessment

Böttcher (2020) explains the factors evaluated when making an energy yield assessment. This assessment is usually done by an external party where the technical aspects of a photovoltaic project are evaluated such that an estimate on the expected energy output of a project can be obtained. The evaluated factors are discussed below.

5.1.1 Factors Affecting Energy Production

As discussed on section 2.3 solar irradiance is the main variable affecting electricity production of a PV panel. Therefore, meteorological data is one of the main pieces of information required for this analysis. This data should be of measurements about the specific location where the project will be built. Estimating energy yield does not consist of only looking at the total energy generated as losses also need to be considered. The main cause for loses is a result of any shades that might interfere with the PV panels. Another factor is the accumulation of dust or snow on top of the panels. Any reflections of solar irradiance from the surface of the panel into its surroundings will also result on losses. The list goes on with other factors such as the temperature of the panel, degradation of the modules, etc. These depend on multiple conditions such as any surrounding buildings or vegetation, exposure to airflow, angle of the panel, etc. As a result, measuring losses is done specific to a particular project and it usually involves in-site observations with specialized equipment and evaluating the technological characteristics used in the project.

5.1.2 Estimating Energy Yield

The gathered measurements are then used to calculate an estimate for the energy yield of the project. This is done by gathering all the measured energy generation and losses and process it for a simulation study. In this simulation possible levels of solar radiation and loss scenarios are tested and used to generate results. Results are typically presented in a table as shown in Figure 22. Here all the expected gains and losses are summarized together with an uncertainty estimate. An additional measure presented on the table is the Performance Ratio (PR). This presents the resulting energy production after gains/losses as a percentage of the maximum energy production capacity of the PV panel.

Irradiation					
Step				Annual yield (kWh/m²)	Uncertainty (% (±))
Global horizontal irradiation				1142	3.5
Irradiation in module plane				1302	2.0
Electricity					
Step	Gross loss (%) *reference: PR 100 %*	Net loss (%) *reference: previous step*	PR (%)	Annual yield (kWh/kWp)	Uncertainty (% (±))
Available solar energy	-	-	-	8256	-
Nominal module efficiency	-	-	100.0	1302	-
Horizon shading	0.6	0.6	99.4	1294	0.5
Self and near shading	4.3	4.3	95.1	1239	1.0
Soiling and snow losses	1.7	1.8	93.4	1216	1.0
Reflection losses	2.5	2.7	90.9	1183	0.7
Bifacial gain	-6.0	-6.6	96.9	1262	2.0
Module quality (flash reports)	-0.9	-0.9	97.8	1273	1.0
Initial degradation	1.5	1.5	96.3	1254	1.0
Irradiation level losses	1.6	1.7	94.7	1232	0.5
Temperature losses	2.4	2.5	92.3	1202	1.0
Mismatch losses	0.4	0.4	91.9	1197	0.2
DC wiring loss	0.5	0.5	91.5	1191	0.1
Inverter losses (incl. clipping)	2.4	2.6	89.1	1160	0.7
AC wiring losses (low voltage)	0.6	0.7	88.5	1152	0.1
Medium voltage transformer losses	0.7	0.8	87.8	1143	0.1
AC wiring losses (medium voltage)	0.2	0.2	87.6	1140	0.1
Own consumption	0.2	0.2	87.4	1138	0.1
Availability (system + grid)	1.3	1.5	86.1	1121	1.0
Result			**86.1**	**1121**	**5.3**

Figure 23 – Example energy yield assessment (Böttcher 2020)

5.1.3 Results of Energy Yield Assessment

After the energy yield has been estimated different "Probability of Exceedance" (defined in theoretical framework) values can be calculated. Company X shared 2 different energy yield assessment reports about Spanish PV projects. In here the methodology for the study is outlined together with the expected production and losses for the project. Both of them follow a structure similar to the one explained by Böttcher (2020). It is a standard for the technical advisor to provide the P50, P75, and P90 numbers of the renewable energy project being evaluated. These numbers are then used as an input for Company X's cash flow model.

5.2 Distribution of Energy Yield

The probability of exceedance values are obtained by modelling the variation in energy yield according to a certain distribution. As explained in the theoretical framework this is usually a normal distribution in the case of solar energy. But is this always the case? Fernandez Peruchena et al. (2016) points out that the main variable affecting energy yield is the direct normal solar irradiance (DNI). Different statistical tests are conducted to compare the goodness of fit to historical measured data in locations. The evaluated distributions are the normal distribution and the Weibull distribution. On one hand the results of the statistical tests indicate that the hypothesis that the DNI follow a normal distribution is not rejected in all locations. On the other hand, the hypothesis that the DNI follows a Weibull distribution is rejected in all locations (by a slight value). Sengupta et al. (2017) explains that this would not always hold and there is no evidence that a particular distribution would always be the best fit. A normal, log-normal, or Weibull distribution could be good candidates. For evaluating what distribution to choose a statistical test should be done to around 10 to 20 of local meteorological measurements. Please note that as explained on the variability in energy yield does not depend only on direct solar irradiance but also on other factors such as losses. As a result, DNI might not determine completely the distribution in energy yield, however it is the main variable affecting it. Böttcher (2020) points out that the standard choice of distribution for energy yield of solar energy projects is indeed a normal distribution.

Testing normality is usually conducted by applying a statistical test to data. The theoretical review showed that for conducting such a test is it necessary to have multiple kinds of measurements (e.g. meteorological, losses, etc) of the specific location. Obtaining such measurements is out of the scope of the project therefore testing for normality based on data cannot be conducted. Out of the energy yield assessments for PV project in Spain shared by Company X one report explicitly mentioned that the energy yield best approximates to a normal distribution. The other one did not do any mention to the distribution followed by energy yield. Academic sources such as Böttcher (2020) support the assumption of normality for all solar energy projects in Spain. Additionally, an interview with an energy yield technical consultant took place to discuss normality on energy yield of PV projects. The questionnaire for the interview can be found in Appendix B, a summary of the interview with the main finding relevant for this research is given on the paragraph below.

The energy yield consultant indicated that a normal distribution is indeed generally used to model energy yield of PV systems as it is a convenient distribution to work with. The methodology used by the consultant in energy yield assessments is like the one described by Böttcher (2020). It involves finding the mean value for different objects that affect energy production and quantifying the uncertainty on each object. This allows to have a normal distribution with a specific standard deviation which can be used to calculate probability of exceedance values such as P50 or P90. From the technical advisors experience weather related factors such as DNI follow a normal distribution for PV. However other items affecting energy production might not generally follow a normal distribution, grid curtailment was given as an example. The impact that these non-normally distributed items have on the overall distribution is low therefore normality is still assumed. In case of using a generalized distribution, the advisors stated that the distribution type (normal) is the same among different locations however the mean and standard deviation changes as the variation in weather conditions is not the same in all locations. In the case of Spain assuming energy yield follows a normal distribution is a valid assumption, however the standard deviation used should represent that region within Spain. As a final remark other consulting companies might have a different methodology for energy yield assessment. However, it is highly likely that even if not stated in the report a normal distribution was used to calculate the P50 and P90 values given on the energy yield assessments.

As previously explained technical advisors provide P50, P75, and P90 numbers. Following up with the interview with the EYA consultant a suitable mean and standard deviation for each specific project to which this simulation is applied must be found. A normal distribution has a mean $\mu$and standard deviation σas parameters, which is not directly provided by technical advisors in EYA. For modelling variability in energy yield these parameters must be extracted from the data given in the energy yield assessment reports. This can be done by using the definition of probability of exceedance. As mentioned in Chapter 2 Sengupta et al. (2017) define the probability of exceedance of level X as: the level in energy yield such that the probability of exceeding it is X. For example, a P90 is the energy a project must produce to have a 90% probability of producing at least that amount. This can be expressed in terms of a standard normal distribution as shown in Equation 16.

$$P\left(\frac{P90-\mu}{\sigma} < Z\right) = 0.1 \tag{16}$$

$$where\ Z \sim N(0,1)$$

This means that using the P50 and P90 numbers provided from the energy yield assessments and Equation 16 the mean μ and the standard deviation σ can be calculated as shown below.

$$\mu = P50$$

$$\sigma = \frac{(P90 - P50)}{\alpha_{10}}$$

$$where\ \alpha_{10}\ is\ the\ 10\ percent\ quantile\ of\ a\ standard\ normal\ distribution$$

5.3 Conclusion

This chapter discussed how to model the energy yield of a solar energy project. Solar energy production mainly depends on solar irradiance, however there are multiple types of losses caused by reflections, temperature of the panel, etc. These factors are measured in meteorological and site-specific measurements. A technical advisor uses these measurements to conduct an energy yield assessment. As described by Böttcher (2020) the technical advisor quantifies the mean gains and losses in energy production and their standard deviation. After all gains and losses are mapped the overall mean and standard deviation of energy yield can be obtained. These parameters are used in a normal distribution to calculate probability of exceedance numbers such as the P50, P75, P90. An interview was conducted with a technical advisor that confirmed that the energy yield of a solar energy project is indeed normally distributed. The mean and standard deviation of this distribution are specific to the project itself. With this a normal distribution can be used to sample energy yield values. Its mean and standard deviation can be calculated using the P50 and P90 numbers provided on the energy yield assessment reports and the 10^{th} quantile form a standard normal distribution.

Chapter 6- Simulation

Now that the randomness in energy prices and energy yield can be modelled this can be combined with the calculations for CFADS, DS, DSCR showed in Chapter 3 in a simulation. This chapter deals with the phase 3 of the problem-solving methodology. Section 6.1 will describe the Montecarlo method for simulation and will provide an answer to the following question: *How to determine a suitable number of iterations for a Montecarlo Simulation?* By comparing the trade-off between computation time and precision. Section 6.2 presents the layout for the simulation and provides instructions on how to run it. Finally, section 6.3 will describe the logic of the main functions and subroutines of the simulation that were programmed using VBA.

6.1 Montecarlo Method

Hull (2009) introduces the Montecarlo method as a tool to value derivatives. In principle it consists of simulating a possible path for the value of the underlying this allows to calculate the price of the derivative for each scenario. This procedure gets repeated many times which allows to calculate the expected payoff as the mean of the sample of simulated payoffs. The expected payoff is then discounted at the risk-free rate, and this gives the price for the derivative. This method can be adjusted for Company X's cash flow model by simulating possible energy prices and energy yields for each year in which the cash flow model makes projections. This can be used to calculate CFADS, debt service, and debt coverage ratio for each year. The result is a sample of possible scenarios that the cash flow model could get for each year. Next section will discuss how can meaningful information be derived out of this sample of possible outputs.

The accuracy in the outputs of the simulation increases together with the number of iterations. As Hull (2009) explains the accuracy of a Montecarlo simulation can be evaluated via the standard error (SE). The standard error for a sample obtained by a Montecarlo simulation can be computed as shown in Equation 17, where $\hat{\sigma}$ is the sample standard deviation and n is the number of iterations.

$$\text{Standard error} = \frac{\hat{\sigma}}{\sqrt{n}} \tag{17}$$

As it can be seen from Equation 17, the standard error (SE) is inversely proportional to the square root of the number of iterations. For illustration purposes this relationship for a standard deviation of 1.69 is plotted in Figure 23. The X axis is the number of iterations and Y is the standard error. It can be observed that the higher the number of iterations the higher the accuracy. However, as the number of iterations increases the standard error reduces at a lower rate. Additionally increasing the number of iterations comes with the trade-off of increasing the computational complexity of the simulation. Hull (2009) discusses variance reduction methods that allow for an increase in accuracy without increasing the number of iterations. One of these methods is stratified sampling in which instead of sampling random values from a distribution it gets split into equally likely sub intervals. Then a sample for each subinterval is obtained. This allows to simulate a more accurate representative sample with low iterations out of regions closer to the tails of the distribution. This is a consequence that those regions can be sampled directly instead of the entire distribution where values closer to the mean are more likely to occur. These methods will not be used in this project however an implementation of a variance reduction method could be one of the main improvements and points for future research for this project.

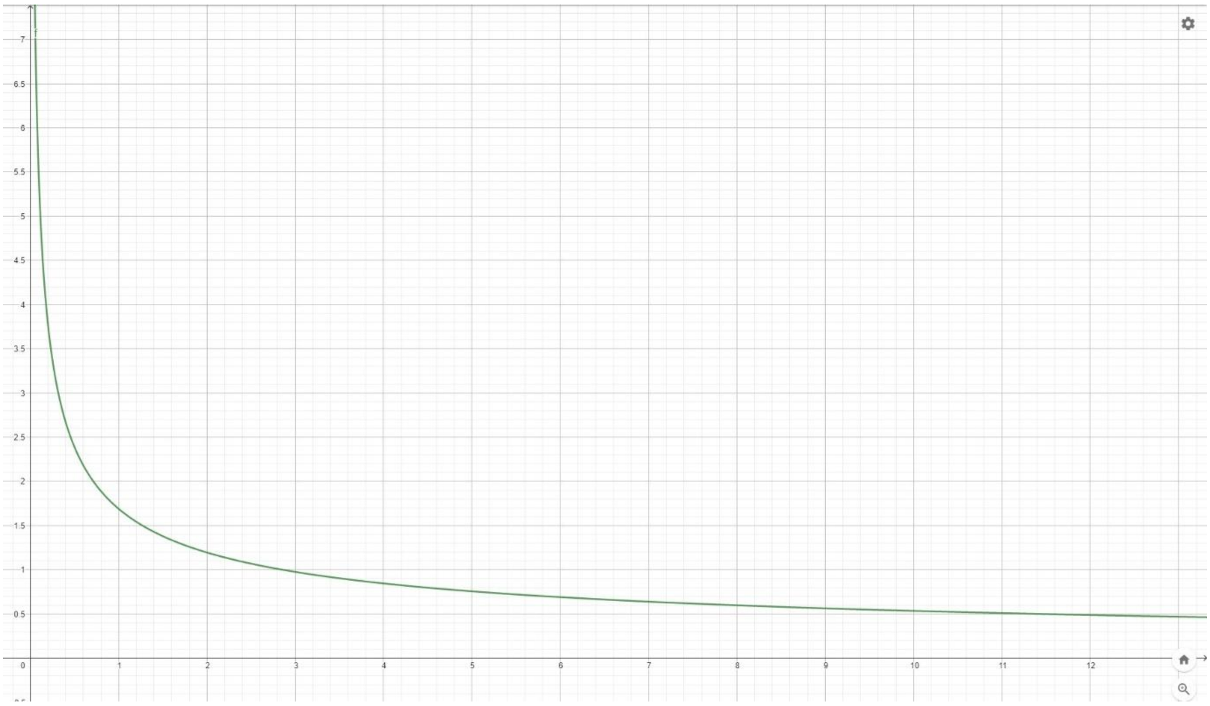


Figure 24 – Example plot of standard error

After examining suggestions by academia there is no “standard” number of iterations for a Montecarlo simulation. It really depends on the type of application as the required precision, computation time and the variance obtained by the sample might be different. Weber et al. (2011) presents a procedure to determine the number of iterations. This procedure consists of running the simulation for different numbers of iterations and recording the computation time and standard error. Then these measures are compared and the option that produces the smallest standard error with an acceptable computation time gets selected.

Poortena et al. (2021) explain that the calculation of the standard error assumes the sample is independent and identically distributed. As it will be discussed on the upcoming section there are dependencies from one year to the other. However, iterations are independent of each other. This means that the standard error can only be calculated for a sample containing the recorded values for one specific output in one particular year. To conduct the procedure presented by Weber et al. (2011) the standard error for each year for CFADS, DS, and DSCR is calculated for each year. Then it is averaged to obtain the average standard error per year for each one of the outputs of the cash flow model.

Table 3 shows the running time in seconds and average standard error for different numbers of iterations. The running time was recorded using VBA from the time the simulation initialized until every single procedure was completed. These experiments where over a time of 25 years which is the average time for which this simulation will be used. Naturally, the computation time increases with the number of years. So, a higher number of iterations can be used if the simulation is meant to be used in a lower number of years and vice versa.

# of Iterations	*Running Time (Sec)*	*SE CFADS*	*SE DS*	*SE DSCR*
5	0.13	18606.72	18606.72	$3x10^{-3}$
50	1.22	3956.21	3956.21	$6x10^{-4}$
500	16.95	1367.77	1338.39	$2x10^{-4}$
1000	46.86	962.12	923.97	$1x10^{-4}$
2500	213.7	587.84	566.43	$8.6x10^{-5}$

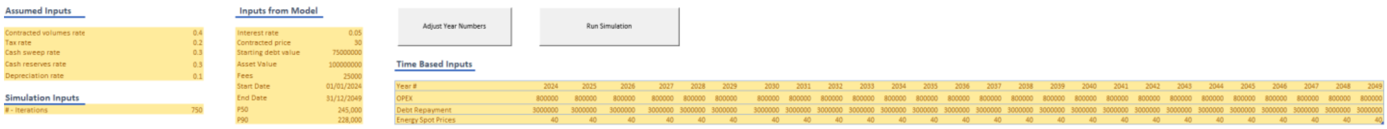
Table 3 – Comparison of computation time and accuracy between number of iterations

For running the simulation with 2500 iterations for 25 years almost 23 million different energy daily prices need to be calculated. On top of that the sampling of energy yield, calculations of the cash flow model, and statistics still need to be executed. From Table 3 even though this simulation needs to do a high number of computations the running time is relatively fast. Higher numbers of iterations than the ones seen in Table 3 were also tested, however at this point Excel started to experiment crashes. As a result, the recommended number of iterations is of 2500 as it provides sufficient accuracy, has a relatively low computation time, and does not put the simulation at risk of crashing. If Company X would like more precision, it should bear the cost of a long computation time and possible crashes unless a variance reduction method is applied. It is also recommended that when running the simulation any heavy programs or other excel workbooks are closed such that the performance of the computer is not compromised.

6.2 Layout

This simulation was built in VBA in an Excel workbook consisting of different worksheets. The first worksheet called “Manual” includes a series of short instructions on how to use this simulation. The second worksheet is called “SeedValues – Spanish Market” which includes the values of the seasonal function, rolling volatility per month and parameters for the energy spot price model. This worksheet is not intended to be modified by the user unless if it is to modify the calibration of the model. For example, if this simulation would be intended to be used in a different country (e.g. France). Then a dataset of French daily price data would need to be analysed with the R script “Calibration - Daily – 5 years”. The elements in the worksheet “SeedValues – Spanish Market” can be replaced by the outputs of the R script.

The third worksheet “Inputs” is where the user is supposed to write the inputs for the simulation on the yellow-coloured cells. Under the section “Assumed Inputs” the user needs to assume a value that would best fit the model to which this simulation is being applied. Secondly, in “Simulation Inputs” the user simply needs to write the number of iterations that the simulation will conduct. Thirdly, in “Inputs from model” the user needs to input the requested values as they are given in the original cash flow model. After this is conducted, the user will need to press the button “Adjust year numbers”, which will clear the contents on the “Time Based Inputs” section and resize the table such that it matches the period for which the simulation will run given in “Inputs from Model”. After this is done the user can enter the OPEX and debt repayment values. On the energy price section, the forecasted average price for each year must be given, it is recommended to use the same price curve as used in the cash flow model (either low curve or an average of low/medium). The reason why is being requested as an input is because an initial condition is needed to simulate energy prices. As discussed on section 3.1 energy markets experience short term random fluctuations and long-term trends from fundamental components. Therefore using exclusively Cartea & Figueroa (2006)'s model for simulating energy prices would mean that only short term fluctuations based on historical data are used to model energy prices. Excluding all the fundamental events done by the forecasting consultant. To include both aspects each year is simulated independently where the price provided by the energy consultant is used as an initial condition for each year and the short-term fluctuations are simulated with Cartea & Figueroa (2006)'s model. This allows for the simulated prices to include short term random variation and still represent the fundamental trends given by the energy price consultant.



Assumed Inputs

Contracted volumes rate	0.4
Tax rate	0.2
Cash sweep rate	0.3
Cash reserves rate	0.3
Depreciation rate	0.1

Simulation Inputs

# - Iterations	750

Inputs from Model

Interest rate	0.05
Contracted price	30
Starting debt value	75000000
Asset Value	100000000
Fees	25000
Start Date	01/01/2024
End Date	31/12/2049
P50	245,000
P90	228,000

Adjust Year Numbers

Run Simulation

Time Based Inputs

Year #	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036
OPEX	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000
Debt Repayment	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000
Energy Spot Prices	40	40	40	40	40	40	40	40	40	40	40	40	40

Year #	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049
OPEX	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000	800000
Debt Repayment	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000	3000000
Energy Spot Prices	40	40	40	40	40	40	40	40	40	40	40	40	40

Figure 25 – Inputs worksheet from simulation

After all the inputs have been filled in the button "Run Simulation" can be pressed so the simulation can start. The worksheet "Macro_Output – CI" is where the simulation will print the low, middle, and upper bounds of confidence intervals discussed on the coming chapter such that they can be plotted on the "Dashboard" worksheet. The "Dashboard" worksheet provides an overview of the results of the simulation which will be discussed on detail next chapter.

6.3 Code

The code for this simulation is too lengthy to be included as an appendix or to be described in detail. However, a description of the main functions used in it and a simplified Pseudo code of the main subroutine will be provided. This is such that the logic behind the code can be understood by the reader. One of the main priorities on this code is that it can run fast given the computational complexity of the simulation. This was done by trying to avoid doing procedures and calculations in Excel and doing them within VBA instead. Additionally automatic calculations, screen updating, and events are disabled while the simulation runs.

In the VBA code of this simulation there are five different modules. Firstly, the module "Simulation" includes the main subroutine for this simulation called "Simulate" which is executed when the button "Run Simulation is pressed". The module "Functions_simulation" includes all the functions necessary to simulate energy prices and energy yield and calculate CFADS, DS, and DSCR. The module "Functions_Satistics" includes all the functions necessary for the statistical analysis discussed in Chapter 7. The module "Outputs" includes all the functions and subroutines that use the functions in "Functions_Satistics" to produce the outputs seen on the worksheets "Macro_Output – CI" and "Dashboard". Finally, the module Resize_time includes the subroutine initialized when the button "Adjust Year Numbers" is pressed.

6.3.1 Functions

Below is a short description of the main functions of this code within the module "Functions_simulation"

Function JumpDifussion

Inputs: Mean reversion rate (a), the initial condition for that specific year (X0), jump frequency (L), jump size standard deviation (SDJ), and start and finish dates.

Output: Array containing the simulated daily energy spot prices for the selected year given initial condition X0.

Function DailyToYearly

Inputs: The intended input (daily) is the array given by function JumpDifussion as output.

Outputs: The average electricity spot price for that year.

Function EnergyYield

Inputs: Mean of energy yield distribution (mean), standard deviation of energy yield distribution (sd), number of years in the simulation (years).

Outputs: An array containing the simulated energy yields for each year during which the simulation is run.

Function CashFlow

Inputs: The array given as output by function EnergyYield (EnergyYield) and an array containing the average energy spot price for each year the simulation is ran (EnergyPrice)

Outputs: A four-dimensional array of arrays. Dimension 1 contains an array of the recorded CFADS for each year. Dimension 2 contains an array of the recorded DS for each year. Dimension 3 contains an array of the recorded DSCR for each year. Dimension 4 contains an array in which it records on what years a default occurred (0 of no default, 1 if default).

6.3.2 Main Sub Procedure

Below is a simplified pseudocode of the sub "Simulate" such that the logic behind the code of the simulation is understood by the reader. This sub is initiated when the button "Run Simulation" is pressed.

Sub Simulate()

Declare variables

Execute initial procedures

For i = 1 to number of iterations in simulation

For j = 1 to number of years in simulation

Execute initial procedure for each year

Call function JumpDifussion to obtain the simulated daily prices for year j

Call function DailyToYearly to obtain the average simulated energy price at year j

Store the average simulated energy price at year j in the j-th element of an array containing the average energy spot prices for each year.

Next j

Call function EnergyYields to obtain the energy produced from year 1 to number of years in the simulation.

Call function Cashflow to obtain the CFADS, DS, DSCR, and defaults on all years on iteration i

Store the CFADS array from the outputs of function Cashflow on the i-th entry of an array of arrays containing the recorder values of CFADS during the years in the simulation for iteration i.

Store the DS array from the outputs of function Cashflow on the i-th entry of an array of arrays containing the recorded values of DS during the years in the simulation for iteration i.

Store the DSCR array from the outputs of function Cashflow on the i-th entry of an array of arrays containing the recorder values of DSCR during the years in the simulation for iteration i.

Store the Default Count array from the outputs of function Cashflow on the i-th entry of an array of arrays containing the recorder values of Default Count during the years in the simulation for iteration i.

Next i

Execute the sub procedures on "Outputs" module

End Sub

6.4 Conclusion

This chapter combines the mathematical model for energy spot prices, the sampling distribution of energy yield, and the calculations for CFADS, DS, DSCR into a Montecarlo simulation. The simulation generates possible values for the average energy spot price and energy yield both for a period of one year. This gets repeated for each year in which the cash flow model makes projections to generate a set of simulated inputs. Using those simulated inputs, it calculates the CFADS, DS, DCSR for the given scenario. By preforming many iterations this allows to obtain a sample of the CFADS, DS, DSCR for different possible scenarios over each year where the cash flow model makes projections. By comparing the computation time and the standard error in the estimated CFADS, DS, and DSCR the recommended number of iterations of 2500. This is because 2500 iterations can be performed relatively quick, provide good precision, and do not put the simulation at risk of crashing. The layout of the simulation is presented together with a description of the main VBA functions used in the code. A simplified pseudocode of the main subroutine in the VBA code of the simulation was presented that provides the logic behind the simulation.

Chapter 7- Applications

This section will provide an answer to the last research question: *How can this simulation be made a tool that can help the team Y to evaluate Spanish solar energy projects?* As explained in chapter 6 this simulation provides a sample of possible CFADS, DS, and DSCR for each year in which projections are made. This chapter focuses on explaining how meaningful data can be extracted out of those samples to evaluate debt structures. With the use of statistics and the simulated sample the following questions will be answered:

1. What are the expected scenarios for CFADS, DS, and DSCR?

2. How to evaluate the robustness of the debt structure?

3. What is the range of possible outcomes for CFADS, DS, and DSCR?

4. What is the shape of the underlying distributions?

5. How to calculate probabilities out of data?

Section 7.1 discusses the inferential statistics used to answer the first two questions. Section 7.2 provides an answer to all remaining questions via descriptive statistics. Section 7.3 describes a dashboard that can help Team Y to visualize results and evaluate debt structures.

7.1 Inferential Statistics

7.1.1 Confidence Intervals for mean CFADS, DS, and DSCR

The current cash flow model calculates a specific value for CFADS, DS, DSCR given a set of assumed values for energy prices and energy yield. Given the uncertainty in inputs the current model does not allow to evaluate the debt structure under different scenarios. Naturally, the first question that raises when evaluating multiple scenarios is: What is the expected scenario for CFADS, DS, and DSCR? This can be answered by calculating the mean, however, given the data is a sample a confidence interval must be calculated to know the range where the actual mean of the distribution is most likely to be.

Using Appendix C and the simulated CFADS, DS, and DSCR over different iterations for a specific year the confidence interval for the mean can be calculated. By looking at this metric Team Y can compare the expected outputs versus the basic scenario used in building the debt structure on each year where the simulation makes projections. If the expected figures are slightly higher or equal to the basic scenario then the project is most likely to perform well, and the robustness of the debt structure will probably not be compromised. Additionally, if the expected figures are way above the basic scenario, it means that the debt structure is too conservative, therefore it can be altered such that debt can be repaid faster without compromising its robustness. Finally, if the figures are lower than the base case it means that the project is expected to perform worst than originally outlined in the cash flow model. In this case the robustness of the debt structure might get compromised, meaning that Company X should either adjust the debt structure or not get into the deal at all. The fact that this is being compared on a specific year allows for a more precise estimate and give Team Y an idea which years are more likely to present good/bad results.

Looking at the mean CFADS, DS, and DSCR gives a picture of what are the expected scenario for a debt structure and can be used to compare with the basic scenario. The expected value alone still does not solve the problem as distributions have more characteristics that could affect the variability in CFADS, DS, and DSCR. These characteristics will be discussed on sections 7.2.1 & 7.2.2 and can be used together with the confidence intervals for the mean to gather more information about the underlying distributions of CFADS, DS, and DSCR.

7.1.2 Confidence interval for the probability of default

As seen in Figure 1, the starting problem that a debt structure faces are that the borrower may default. This raises the question: How to evaluate the robustness of the debt structure? For Company X the robustness of a debt structure is compromised if a borrower does not pay its financial obligations. In other words, a project enters in default if its realized debt service is less than the mandatory debt service. This situation can be modelled as a Bernoulli random variable X defined below.

$$P(X = x) = \begin{cases} 1 - p & for\ x = 0 \\ p & for\ x\ = 1 \end{cases}$$

In this application $x = 1$ denotes that the project has resulted in a default for that particular year. The parameter p then denotes the probability that a project defaults in that specific year. This variable was programmed in the simulation. Using a sample of the "Defaults" over different iterations on one particular year and Appendix D the confidence interval for the probability of default can be calculated.

This interval shows what is the expected proportion of scenarios where the project goes into default. Company X should try to have the probability of default as low as possible and it should be the main indicator in the robustness of a debt structure. If the confidence intervals for the mean described in section 7.1.1 are higher than in the base case but the probability of default is high, then Company X should be cautious. This is most likely due to having a few observations with high values that push the expected value up however on average the project will not be able to meet its financial obligations, meaning that the debt structure is not robust.

7.2 Descriptive Statistics

7.2.1 5-numbers summary and boxplots

On the previous section the expected scenario for CFADS, DS, and DSCR is discussed. But what if Team Y would like to observe the range of possible scenarios that could occur? This can be done by observing the 5-numbers summary of CFADS, DS, and DSCR. Poortena et al. (2021) explain that the five numbers summary of a sample is it's 25^{th} percentile (Q_1), median (m), 75^{th} percentile (Q_3), and the upper and lower boundaries of the 1.5 x IQR rule (See Appendix E). The plot of the 5 number summary is called a box plot, and example is given in Figure 25. In the picture $X_{(1)}$ and $X_{(n)}$ refer to the first and last order statistics, however in this definition they are replaced by the boundaries of the 1.5 x IQR rule.

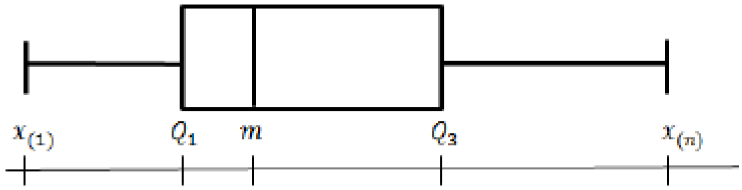


Figure 26 – Example Box Plot (Poortena et al. 2021)

By looking at the boxplots of CFADS, DS, and DSCR Team Y can see the spread in possible scenarios for an arbitrary year. The values on the box of the plot show what are the values that are most likely to be observed. The whiskers of the plot show the tails of the distribution. By looking at the boxplot Team Y can evaluate the debt structure will perform over different scenarios.

7.2.2 Skewness and Excess Kurtosis

So far, the expected scenario for each year is given as discussed on section 7.1.1 and the range of possible scenarios that could occur are shown in section 7.2.1. However, how is the shape in the underlying distribution? Does this shape in some way affect the variability in possible outcomes? To answer this questions the empirical skewness coefficient and empirical excess kurtosis will be discussed.

The skewness of a distribution shows how if it has any "shifts" and is represented by the skewness coefficient (γ_1). If the skewness coefficient is positive the distribution is skewed to the right where the mean is greater than the median. Alternatively, if the skewness coefficient is negative it is skewed to the left, where the median is greater than the mean. Finally, if it is zero the distribution is symmetric, meaning that the mean equals the median. The empirical skewness coefficient ($\hat{\gamma}_1$) can be calculated using Appendix F as shown in Equation 18.

$$\hat{\gamma}_1 = \frac{\widehat{M}_3}{\widehat{M}_2^{\;3/2}} \tag{18}$$

The skewness of the distribution can be used by Team Y to gather more information out of the confidence intervals discussed on section 7.1.1. On one hand if the underlying distribution is skewed to the right Team Y should be careful when looking at the expected scenarios as it is more likely to see values below it. On the other hand, if the underlying distribution is skewed to the left then Team Y can be more optimistic because it is more likely to see values above the mean. For example, suppose the distribution for DSCR is skewed to the left and the expected value for that for a particular year is of 1.4. This means that Team Y should reevaluate the debt structure more optimistically as it is highly likely that the observed DSCR will be over the preferred range by Company X of a DSCR between 1.2 to 1.4.

The kurtosis of a distribution (γ_2) show the length of the tails in the distribution. The longer the tails of a distribution the more outliers can occur. Poortena et al. (2021) mentions a normal distribution is usually used as reference point. Therefore, the kurtosis of a distribution is compared to the kurtosis of a normal distribution (γ_2 = 3). This can be calculated using Appendix F as shown in Equation 19. It is negative the kurtosis of the distribution is less than of a normal distribution and vice versa.

$$Empirical\ Excess\ Kurtosis = \frac{\widehat{M}_4}{\widehat{M}_2{}^2} - 3 \tag{19}$$

If the empirical excess kurtosis of the underlying distribution is positive, then Team Y should be cautious as it is more likely to observe values out of the 1.5 x IQR range discussed in section 7.2.1. As an example, suppose the empirical excess kurtosis for DS is positive. If the expected DS and values within the IQR indicate that the project will be able to pay back its debt and make its interest and fee payments, then Team Y should still be aware that it is likely to observe outliers. Even though the debt structure might appear to be robust in most scenarios these outliers could indicate a low DS meaning that the robustness of the debt structure could still be compromised.

7.2.3 Empirical CDF

Section 7.2 has so far discussed characteristics of the underlying distributions of CFADS, DS, and DSCR. However, what if Team Y would like to evaluate probabilities out of these distributions? Probabilities calculated based on the sample of simulated scenarios can be calculated via the empirical CDF. CDF stands for Cumulative Distribution Function and allows to calculate the probability that a random variable takes a realization below a certain point (t). Poortena et al. (2021) call this the empirical CDF ($\hat{F}(t)$) which is shown in Equation 20 where $\mathbb{1}$ denotes the indicator function.

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(x_i \leq t)} \tag{20}$$

An empirical CDF for CFADS, DS, and DSCR can be built with the sample of observations of each item and the value t is an input from the user. The empirical CDF serves as a tool for Team Y that could help answering multiple questions. Some of these questions and their answer using the empirical CDF are shown below as an example.

1) What is the probability that DSCR falls within the acceptable range of 1.2-1.4 on an arbitrary year?

   Answer: $\hat{F}(DSCR = 1.4) - \hat{F}(DSCR = 1.2)$

2) What is the probability CFADS are negative on an arbitrary year?

   Answer: $\hat{F}(CFADS = 0)$

3) What is the probability realized DS is below mandatory DS on an arbitrary year?

   Answer: $\hat{F}(DS = Mandatory\ Debt\ Service)$

7.3 Dashboard

The statistics discussed on this chapter need to be incorporated into the simulation and visualized on a dashboard that Team Y can use as a tool to evaluate debt structures. As discussed in Chapter 6 all the statistical concepts discussed on this chapter where coded as functions in the "Functions_Statistics" module. When the functions and sub procedures in the "Outputs" module are executed the lower bound, mean, and upper bound for each confidence interval are printed on the "Macro_Output – CI" worksheet. Then the 25-th percentile, mean, 75-th percentile, standard error skeweness and excess kurtosis for each output are printed on their respective tables on the "Dashboard" worksheet. The lower and upper bound for outliers are calculated automatically. The Empirical CDF function takes as an input any value t entered by the used on the yellow cells. All the plots on this dashboard are updated automatically after this simulation stops running. An overview of the dashboard is shown below, please note this is made with the inputs shown in Figure 24 which are fictional numbers for illustration purposes.

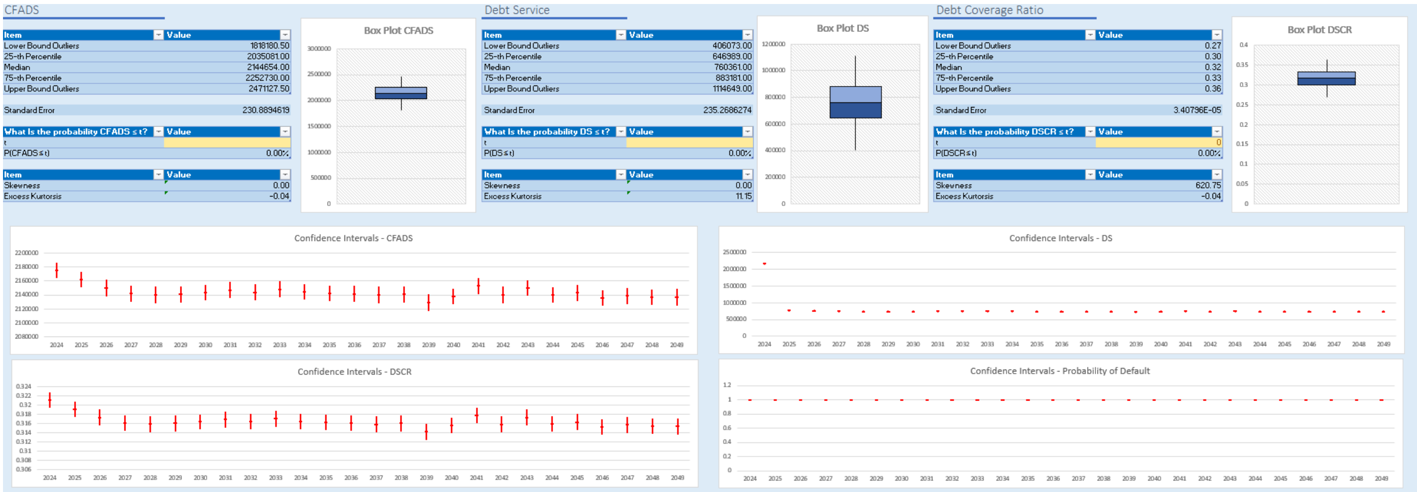


Figure 27 – Dashboard of simulation

7.4 Conclusion

This chapter provided an answer to the question: *How can this simulation be made a tool that can help the team Y to evaluate Spanish solar energy projects?* This was done by using statistical tools to analyse the samples of simulated CFADS, DS, and DSCR. Firstly, a confidence interval for the mean CFADS, DS, and DSCR was calculated such that Team Y can evaluate the expected outcome for the debt structure. A Bernoulli random variable was introduced to model if a project defaulted on a specific year. A confidence interval for the probability of default was built that serves as the main indicator of the debt structure's robustness. The 5 numbers summary for the sample of CFADS, DS, and DSCR was calculated and plotted in a boxplot. This boxplot shows Team Y the spread of the distribution given in the samples of simulated scenarios. The shape of the underlying distributions of CFADS, DS, DSCR and its impact to the variability was analysed via the empirical skewness coefficient and empirical excess kurtosis. Finally, an empirical CDF for CFADS, DS, and DSCR was created that allows Team Y to compute probabilities out of data. All these calculations are incorporated into the simulation in VBA and visualized on a dashboard that serves as a tool for Team Y to evaluate debt structures.

Chapter 8- Conclusion

8.1 Conclusion

This thesis provided an answer to multiple research questions with the objective of building a tool that allows Company X to obtain a distribution of possible outcomes for the outputs of its cash flow model. This tool solves the core problem: *The current cash flow model does not incorporate the randomness in energy prices and energy yield and its outputs are limited to one basic scenario.* As it allows to evaluate a debt structure via the CFADS, DS, and DSCR caused by possible energy price and energy yield scenarios. This was conducted according to the steps on the simulation methodology given by Robinson (2004) and the scope of the research was limited to Spanish PV projects.

The calculations done by the current cash flow model where outlined such that they can be replicated within the simulation. As the cash flow model is complex some assumptions and simplifications where done. With this the yearly CFADS, DS, and DSCR of the Spanish PV project being evaluated can be calculated. The data required for these calculations are inputs from the user based on the data in the current cash flow model (e.g. starting debt value and OPEX) and simulated average annual electricity prices and simulated energy yield for one year.

To simulate energy spot prices a mathematical model for energy spot prices was selected and calibrated to the Spanish market using data from 2018 to 2022. The Jump Diffusion model by Cartea & Figueroa (2006) was used as it represents mean reversion, seasonality, and price jumps. The SDE in the model was approximated via the Euler-Maruyama method using daily time steps. This model is used to simulate possible daily prices for one year based on an initial condition. The average of the simulated daily prices can be taken to obtain the first input required to calculate CFADS, DS, DSCR. The middle price projection given by the forecasting consultant hired by Company X is used as an initial condition for each year such that the simulated inputs show the long-term fundamental trends in energy markets studied by the consultant and short-term variability given by the Jump Diffusion model.

To simulate energy yield an appropriate distribution and its parameters where found. According to academic sources such as Böttcher (2020) and an energy yield technical consultant the distribution for energy yield for PV project follows a normal distribution. The mean and standard deviation of this distribution are specific to the project being analysed. A procedure to find the mean and standard deviation for a specific project involves using the P50 and P90 numbers given in the energy yield assessment reports and the quantiles of a standard normal distribution.

Using the simulated average yearly prices and energy yield a Montecarlo simulation was built that allows to calculate a sample of possible CFADS, DS, and DSCR. After evaluating the trade-off between accuracy and computation time the recommended number of iterations for this simulation is 2500.

Different statistics were calculated out of the sample of simulated CFADS, DS, and DSCR to obtain information about their distributions. These statistics were combined in a dashboard that serves as a tool for Team Y to evaluate debt structures. Confidence intervals for the mean CFADS, DS, and DSCR for each year provide an overview of the expected scenario. A measure of the general robustness of a debt structure is given by calculating a confidence interval for the proportion of scenarios on each year where a default occurs. The five numbers summary provides an overview of the possible range of outcomes for an arbitrary year. The empirical skewness coefficient and excess kurtosis allow to determine the shape of the underlying distributions. Finally, the empirical CDF allows to calculate probabilities to answer relevant questions to the performance of a debt structure.

8.2 Limitations and Points for future research

Even though the original research questions were answered there are multiple points of improvement and limitations on the current simulation. This section will provide a short description of each point.

1. Scope of research

This research was limited specifically to Spanish solar energy projects. Company X finances projects of multiple renewable energy sources around Europe. Therefore this simulation could be extended to other markets by calibrating Cartea & Figueroa (2006) to historical daily energy spot price data of other countries. When the data used in this project was obtained datasets for UK, Italy, France, and Germany where also downloaded. These datasets can be analysed with the same R code made in this research with minimal modifications. Besides the assumption of normality in energy yield is specific for solar energy projects. For using this simulation with different types of sources the sampling distribution used in the VBA function "EnergyYield" must be changed to a suitable alternative. For example, for wind project the inverse normal distribution could be replaced with an inverse Weibull distribution.

2. Simplification cash flow model

As explained in Chapter 3 some assumptions and simplifications were done to the cash flow model such that it could easily be replicated by the simulation. Given the short time for this research project some of these assumptions are not realistic. This is because a priority was given to deliver a first concept which might not be prefect but could be improved later. For example, this simulation assumes energy produced is sold at contracted and market price at a constant percentage. This percentage varies with energy produced being sold fully at contracted price via a PPA at earlier years and at later years fully market. Additionally, the project might have different PPA agreements that make the contract price change over the lifetime of the project instead of it being a fixed number. Making these changes does not require that much effort, however it would greatly improve the simulation. As a result, this is the main point for improvement of this research. After this thesis is submitted there are still a few days left of the internship at Company X. Therefore, a priority during this time will be to make these assumptions more accurate by adjusting the simulations to situations similar to the one described here.

3. Other stochastic inputs

This project focused on modelling the variability of only 2 inputs of the cash flow model. Company X selected average yearly energy prices and energy yield as the inputs analysed on this project. However, the cash flow model currently has more stochastic inputs that are being treated deterministically (e.g. interest rates). The principle behind this simulation could be extended and combined with such inputs with a suitable mathematical model. However, this should be done without excessively increasing the computational complexity of the simulation as it is already heavy.

4. Time horizon historical data

As discussed with the supervisor of this thesis, employees of Company X, and energy price forecasting consultants energy markets change. This means that historical data does not necessarily reflect the state of the energy market in the present and in the future. Additionally, the data used should include enough observations for statistical significance. This means that the current simulator for energy prices is limited to parameters obtained from 2018 to 2022. A future point for research would be to find different time horizon to which calibrate to model and/or use an exponential moving average to assign a higher weight to recent observations but still include past data.

4. Volatility

As it was disused on Chapter 4 the method used to model time varying volatility assumes log returns are independent and normally distributed over intervals of 30 days. This assumption does not necessarily meet the characteristics of electricity price fluctuations discussed on this research. Mean reversion implies log returns are not independent and price spikes implies log returns are not normally distributed. For having a model that better resembles these characteristics using a GARCH model instead of the 30-day rolling volatility could be a point for improvement. However, this also would increase the complexity of the model and simulation.

5. Alternative energy spot price models

This thesis used Cartea & Figueroa (2006) as it is the most simple one factor mathematical model for electricity spot price that covers mean reversion, seasonality and price jumps. However there exist more models that cover these characteristics. De Jong (2006) explains that one factor spot price models using Levy processes or Non-Gaussian Mean reversion tend to be more accurate when modelling these characteristics. Additionally, Mandelbrot & Hudson (2010) explain that price jumps tend to be grouped in clusters. If a jump occurs it is more likely to observe another jump next day. This behaviour is not shown in Cartea & Figueroa (2006)'s model. As a result, there are more models to choose from which could improve the way electricity prices are simulated, however this also could make this simulation and calibration more complex.

6. Variance reduction method

As discussed in Chapter 6 having precision in a Montecarlo simulation comes at a cost of increasing computation time. This simulation conducts the most basic form of a Montecarlo simulation in which random sampling takes place. There are different variance reduction techniques such as stratified sampling which would allow to improve precision without increasing computation time.

Appendix

A. Wiener Process

Hull (2009) defines the Wiener process as: the variable Z follows a Wiener process if the following properties are met.

Property 1

The change ΔZ during some small time period Δt is given by:

$$\Delta Z = \epsilon\sqrt{\Delta t}$$

Where ϵ follows a standard normal distribution $\phi(0,1)$.

Property 2

The values of ΔZ for any two different intervals of time Δt are independent.

B. Questionnaire Energy Yield

1) When making an energy yield assessment for a PV project do you assume a distribution, or do you find a best fitting distribution?
2) If you assume a distribution which one is it? What are the reasons behind this assumption?
3) How does your testing procedure for goodness of fit look like?
4) What are the usual distributions that energy yield for PV projects follow?
5) Suppose I want to generalize a distribution choice for energy yield for a country (e.g. Spain). Is making a generalization valid and reliable?
6) What distribution would be best suited for a generalization? Why is this the case?
7) Is a normal distribution a good "standard choice" for modelling variability in energy yield?
8) In what cases would a normal distribution not be a good choice?

C. Confidence Interval for Mean

Poortena et al. (2021) present the confidence interval for the mean of a sample of independent and identically distributed random variables. The equation for a confidence interval for the mean ($CI(\mu)$) at a level of $(1-\alpha)\%$ is shown in the equation below.

$$(1-\alpha)\% - CI(\mu) = (\bar{x} - C_{\frac{\alpha}{2}} * \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + C_{\frac{\alpha}{2}} * \frac{\hat{\sigma}}{\sqrt{n}})$$

Where $\bar{x}$ is the sample mean, $\hat{\sigma}$ is the sample standard deviation, and n is the sample size. $C_{\frac{\alpha}{2}}$ refers to the $\frac{\alpha}{2}$ quantile of a Students T distribution with $n-1$ degrees of freedom. The students T distribution approximates a normal distribution as the sample size increases. For large samples (>50) the quantile of a normal distribution can be used instead.

D. Confidence Interval for Population Proportion

Poortena et al. (2021) present the confidence interval for the population proportion of a sample of independent identically distributed Bernoulli random variables. The equation for a confidence interval for the population proportion ($CI(p)$) at level $(1-\alpha)\%$ is shown in the equation below.

$$(1-\alpha)\% - CI(p) = (\hat{p} - C_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + C_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}))$$

Where $\hat{p} = \frac{X}{n}$ with n being the sample size and X the number of successes observed in the sample. $C_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile from a normal distribution.

E. 1.5 x IQR rule

Poortena et al. (2021) define the inner quartile range (IQR) of a distribution as the difference between the 75th percentile (Q_3) and 25th percentile (Q_1). If an observation lies outside the range shown below it is considered an outlier.

$$(Q_1 - 1.5 * IQR, Q_3 + \ 1.5 * IQR)$$

F. Empirical Kth Centred Moment

Poortena et al. (2021) present the calculation of the kth centred moment $\hat{M}_k$ for sample of size n with sample mean $\overline{x}$. The calculation of $\hat{M}_k$ is shown below.

$$\widehat{M}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$$

References

Aïd, R. (2015). Electricity Markets. In R. Aïd (Ed.), *Electricity Derivatives* (pp. 5–25). Springer International Publishing. https://doi.org/10.1007/978-3-319-08395-7_2

Benth, F. E., Ekeland, L., Hauge, R., & Nielsen, B. F. (2003). A note on arbitrage-free pricing of forward contracts in energy markets. *Applied Mathematical Finance*, *10*(4), 325–336. https://doi.org/10.1080/1350486032000160777

Benth, F. E., & Šaltytė-Benth, J. (2004). The normal inverse gaussian distribution and spot price modelling in energy markets. *International Journal of Theoretical and Applied Finance*, *07*(02), 177–192. https://doi.org/10.1142/S0219024904002360

Böttcher, J. (2020). *Green Banking: Realizing Renewable Energy Projects* (1–1 online resource (XXXII, 700 p.)). De Gruyter Oldenbourg. https://doi.org/10.1515/9783110607888

Boucher, C., Jasinski, A., & Tokpavi, S. (2023). Conditional mean reversion of financial ratios and the predictability of returns. *Journal of International Money and Finance*, *137*, 102907. https://doi.org/10.1016/j.jimonfin.2023.102907

Cartea, Á., & Figueroa, M. G. (2006). *Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model with Seasonality* (SSRN Scholarly Paper 592262). https://papers.ssrn.com/abstract=592262

De Jong, C. (2006). The Nature of Power Spikes: A Regime-Switch Approach. *Studies in Nonlinear Dynamics & Econometrics*, *10*(3), 1–28.

Duldinger, L. (2023, December 2). *How to build a project finance model from scratch?* Renewables Valuation Institute. https://courses.renewablesvaluationinstitute.com/pages/academy/how-to-build-a-project-finance-model-from-scratch

Eydeland, A., & Wolyniec, K. (2003). *Energy and power risk management: New developments in modeling, pricing, and hedging*. Wiley.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *38*(1), 34–105.

Fernandez Peruchena, C. M., Ramirez, L., Silva, M., Lara, V., Bermejo, D., Gaston, M., Moreno, S., Pulgar, J., Liria, J., Macias, S., Gonzalez, R., Bernardos, A., Castillo, N., Bolinaga, B., Valenzuela, R. X., & Zarzalejo, L. (2016). A Methodology for Calculating Percentile Values of Annual Direct Normal Solar Irradiation Series. In V. Rajpaul & C. Richter (Eds.), *SOLARPACES 2015: INTERNATIONAL CONFERENCE ON CONCENTRATING SOLAR POWER AND CHEMICAL ENERGY SYSTEMS* (Vol. 1734, p. 150005). Amer Inst Physics. https://doi.org/10.1063/1.4949237

Gudkov, N., & Ignatieva, K. (2021). Electricity price modelling with stochastic volatility and jumps: An empirical investigation. *Energy Economics*, *98*, 105260. https://doi.org/10.1016/j.eneco.2021.105260

Hull, J. (2009). *Options, Futures and Other Derivatives*. Pearson/Prentice Hall.

Janczura, J., Truck, S., Weron, R., & Wol, R. C. (2012). *Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling*.

Lucia, J. J., & Schwartz, E. S. (2002). Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research*, *5*(1), 5–50. https://doi.org/10.1023/A:1013846631785

Luna-Romera, J. M., Carranza-García, M., Arcos-Vargas, Á., & Riquelme-Santos, J. C. (2024). An empirical analysis of the relationship among price, demand and CO2 emissions in the Spanish electricity market. *Heliyon*, *10*(3), e25838. https://doi.org/10.1016/j.heliyon.2024.e25838

Mandelbrot, B. (1963). New Methods in Statistical Economics. *Journal of Political Economy*, *71*(5), 421–440. https://doi.org/10.1086/258792

Mandelbrot, B. B., & Hudson, R. L. (2010). *The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. Profile Books.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti Del Circolo Matematico Di Palermo*, *4*(1), 48–90. https://doi.org/10.1007/BF02846028

Mayer, K., Schmid, T., & Weber, F. (2015). Modeling electricity spot prices: Combining mean reversion, spikes, and stochastic volatility. *The European Journal of Finance*, *21*(4), 292–315. https://doi.org/10.1080/1351847X.2012.716775

Poortena, K., Meijer, D., & Schmidt-Hieber, J. (2021). *Mathematical Statistics*. University of Twente.

Sengupta, E. M., Habte, A., Gueymard, C., Wilbert, S., Renne, D., & Stoffel, T. (2017). Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Second Edition. *Renewable Energy*.

Størdal, S., Ewald, C.-O., Lien, G., & Haugom, E. (2023). Trading time seasonality in electricity futures. *Journal of Commodity Markets*, *31*, 100291. https://doi.org/10.1016/j.jcomm.2022.100291

ter Haar, R. (2010). On modelling the electricity futures curve. *University of Twente*.

Weber, F., Schmid, T., Pietz, M., & Kaserer, C. (2011). Simulation-Based Valuation of Project Finance—Does Model Complexity Really Matter? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1594116