MODELLING PRICES IN SOLAR AND WIND LONG-TERM PPA

Oscar Bernardos Yagüe

Trabajo de investigación 21/001

Master en Banca y Finanzas Cuantitativas

Tutores: Dr. Juan Ignacio Peña

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

www.finanzascuantitativas.com

Modelling Prices in Solar and Wind Long-term PPA

Por
Óscar Bernardos Yagüe

UNIVERSIDAD COMPLUTENSE
MADRID

Máster en Banca y Finanzas Cuantitativas
Facultad de Ciencias Económicas y Empresariales

Dirigido por
Juan Ignacio Peña

Madrid, 2020–2021

Abstract

Power Purchase Agreements are a financial instrument conceived for renewable energy projects where a fixed power price is agreed upon for the long term. Their importance is increasing in recent years. Therefore, these exercise of pricing turns crucial to assure that both parties reach an agreement. That topic is the main element of this work, where an extensive investigation about the electricity markets is done, and more in detail the one of U.S. The analysis is conducted for a data-set of more than eight hundred PPA of the U.S classified by regions and technologies of generation. Specifically is adapted to the regions where information is available. The factors evaluated are power prices, different terms of the contracts, and generation profiles. Despite the lack of other factors evidenced in the results, interesting conclusions can be obtained. First, terms and quantity are studied in comparison with the prices looking for proofs if they could be considered as determinant factors for the pricing. After that, the SWAP standard pricing is computed showing a good performance with linear interpolation and extrapolation. At this point, some other factors reflect great relevance for the pricing as the time between the date of singing the contract and the start of the exchanges, the available data, or the long-term power price estimated. Last, the copula approach enables assessing the solar and wind price corrections due to correlation and volumetric risk, exhibiting negative values.

Resumen

Los acuerdos de compra de energía (PPA) son un instrumento financiero concebido para proyectos de energía renovable dónde se establece un precio fijo de la energía en acuerdos de largo plazo. Su importancia está creciendo en los útlimos años y, por lo tanto, este ejercicio de "pricing" se vuelve crucial para asegurar que ambas partes lleguen a un acuerdo. Este es el tema central de este trabajo, dónde se lleva a cabo una amplia investigación previa sobre los mercados de la electricidad y más en detalle sobre el de Estados Unidos. El análisis se hace para un conjunto de datos de más de ochocientas PPA de EEUU clasificadas por región y tecnología de generación. Además, en cada caso, se adapta el análisis a las regiones con la debida información disponible. Los factores que son evaluados son los precios de la electricidad, diferentes términos y cláusulas del contrato y así como los perfiles de generación de energía. A pesar de que los resultados evidencian la falta de otros factores, se pueden obtener conclusiones interesantes de todo ello. Primero, las variables plazo y cantidad son estudiadas en comparación con los precios en busca de pruebas de si pueden ser consideradas como variables determinantes del "pricing" o no. Después de esto, el "pricing" obtenido haciendo uso de un enfoque puramente de SWAP parece tener buenos resultados con la interpolación y extrapolación lineal. En este punto, otros factores muestran gran relevancia como son el tiempo entre la firma del contrato y el comienzo de los intercambios, la información disponible y el precio de largo plazo estimado. Por último, el uso de las cópulas permite evaluar las correciones de precio por riesgo volumétrico y riesgo de correlación tanto para los datos de PPA solares como las de viento, mostrando en ambos casos valores negativos.

Contents

List of Figures

List of Tables

Part A

Introduction

Contextualization

In the last few years, green energy has experimented a notable growth. It is becoming one of the major concerns for companies who desire to consume power produced from renewable resources. Becoming more environmentally friendly, improving the image among target customers, or avoiding pollution taxes could be considered one of the main reasons for that change of behavior.

Similarly, technologies for renewable energy generation are extending their possibilities and in constant improvement. It is difficult to identify whether the improvement in terms of costs and productivity has led to the increase in renewable power production or the other way round. However, one thing is clear: the growth differs between regions finding countries with a strong dependence on fossil fuel energy generation and others with percentages of renewable generation next to 100%. Among the existing renewable generation technologies, two can be considered the ones with the fastest development recently. These are solar photo-voltaic and wind on-shore generation.

Regarding this topic, some decades ago, a financial instrument has arisen to satisfy the renewable energy producers' needs. It is known as Power Purchase Agreement (PPA) and involves a contract between two parties that want to establish a long-term relationship of power exchanges for a predetermined price. In the same way, in hand with the growth of renewable resources, this instrument has undergone a considerable change in the last decade. However, in spite of its spread along with the countries, it is still a non-standardized contract whose terms can differ a lot from one contract to another. This fact could be explained as the particularities of each renewable project would be challenging to be satisfied with a normalized contract. That is why the study of this financial contract becomes essential. As it has many details, these can be analyzed to comprehend how they influence the PPA's price. Furthermore, with the trend already explained about renewable energies, a good understanding of PPA could be fundamental to achieve fair agreements and encourage renewable power production.

PPA use also has remarkable differences among countries. In particular, this research is focused on the EEUU region, which is the country with the highest number of GWh contracted with this instrument. That has allowed finding a database with enough numbers of contracts for this research. Moreover, as it will be seen, the power market of this country has its singularities, so the analysis is broken down into regions. What is more, only some of them have the necessary information to be analyzed, so most regions have to be discarded in parts of the analysis.

This research aims to study some of the critical factors that influence the pricing of solar and wind PPA in the United States. However, this analysis cannot be done without first investigating power markets and how they work, the different types of PPA, and the particularities that this sector involves from a financial view.

The key contributions of this work can be summarized in three aspects. The main one is the study of PPA using a database of PPA contracts from the U.S in order to analyze with detail factors that influence the pricing of PPA. Never done it before, the existent researches have focused their attention on studying individual PPA or specific elements of those instruments. Consequently, this approach tries to price market PPA using factual information and then compares the results with the market prices to see how far the model is from the market pricing. Second, since factual information is used, the analysis investigates all the possible determinants and market conditions that could influence the instrument. Last, among all factors studied, the price correction due to volumetric and correlation risk is computed for every PPA of one of the U.S. regions. And what is more, the price correction is computed on a complex intra-day model for solar data. In particular, on the financial literature, this approach has never been applied before. Thus, it is possible to see how this technique is applied to simulate an environment as realistic as possible and to compare the results of using intra-day series or daily series. In a nutshell, the work seeks to reproduce a realistic pricing to understand some of the determining components better when it comes to pricing a PPA.

The investigation is divided into four parts. The first one starts with this section and continues with a review of the financial literature related to this topic, and finishes with an explanation of the importance of this research as well as a first contact with some of the basic concepts. In the second part, the data used for the investigation is explained. First, the data about a PPA contracts database is presented and described with their main statistics. Next, analysis of the electricity markets of these two regions connecting the evidence observed with the theory. The end of this part corresponds to an explanation of the power prices data about the specific regions studied. The third part corresponds to the analysis of the mentioned factors. First of all, the effects on the prices of the term and the quantity agreed on the contracts are evaluated. Then, a model similar to a standard SWAP pricing is tested with special attention to the details. Finally, the model is extended to one which includes a price correction considering two risks present in this instrument: the correlation risk and the volumetric risk. To close the third part, the model is tested using daily and intra-day series for the solar case. Last, there is an annex with some explanations considered practical for a better comprehension of some sections.

Literature review

Off the bat, it is necessary to clarify that Power Purchase Agreements have been little investigated in the financial literature compared to many other topics. Maybe this is not totally true, but it is a field with many details still open to further investigation and increasing its importance (now and expected to continue). The obstacles in that area can be subjective, but some of them are:

- It belongs to the power commodity, which has many peculiarities which difficult its financial treatment just as available information problems.
- PPA arose some decades ago, but it is not until this decade when its use has spread all over the countries.
- The non-existence of a standardized contract makes the analysis more complex as it can be generalized for all PPA and as transparent information is very unusual.

However, in this concern, there are already some works covering these problems. For example, the data available is growing, and some investigations have collected them. For the United States, there are reports prepared by the Berkeley Lab institution. Specifically, Bolinger et al. [2019] for Solar PV energy and Wiser et al. [2020] for Wind on-shore energy. The authors provide with both reports a complete database which has been used in this investigation. For Europe, there are also reports available from specialized companies, but they mainly present aggregated data.

Then, focusing on the literature available about pricing or valuating Power Purchase Agreements could be classified into two different main approaches. The classification attends to the point of view that is taken for that objective. On the one hand, the pricing can be done from the off-taker point of view, typically using financial arguments to price the instrument as other derivatives. The reason behind it lies in the fact that the purchaser of a PPA usually has the main objective of hedging the price risk. On the other hand, if a generator does the pricing, it usually comprises a detailed estimate of all the costs that would involve the project to know the minimum price required to recover the money invested. Bearing in mind these two possibilities, each one will be deepened in regard to the literature found. See annex 1 on page 75 for a detailed review of the principal models existent in the literature.

From the off-taker point of view, the starting point could be considered the works which apply a standard SWAP pricing approach. Though most of them develop further their models, the majority of them begin at this point. For example, Peña et al. [2020], propose the standard model but also taking into account the price of renewable energy certificates. Actually, their model is expressed to valuate a PPA. However, as with any other SWAP derivative, the pricing could be done by finding the fixed price that makes the net pricing value of both parties indifferent. Another extension is done in Edge [2015], where the author proposes a stochastic process to model the spot price and a general way to take into account the credit risk in long-term PPA. Then, some works try to model the volumetric and correlation risk in order to correct the PPA price due to these elements. These are Pircalabu et al. [2017], Tranberg et al. [2020] and Kaufmann et al. [2020]. The three of them use copulas to model the dependence between production and price variables. In particular, the first two works are strictly related as the second one is an extension of the first one. Likewise, other investigations do not strictly stem from the standard SWAP approach. In this respect, Wu and Babich [2012] is focused on contingent PPA analyzing the effects of asymmetries of information in the agreements. As well, this incentive problem is taken into account in Ghiassi-Farrokhfal et al. [2021] where they suggest how to find the optimal PPA structure and terms attending to different situations. In addition, they also take into account the possibility of battery storage. In other direction Cuervo et al. [2021] present a simple model to valuate PPA, and also they apply a real options approach to determine the importance of the timing (the date to start a project). Identically, Sachs et al. [2008] do not price PPA but analyzes the quantification of the political risks from the view of the investors.

Now, from the producer's point of view, there are some researches too. These are directly linked with the LCOE. That is the levelized cost of energy of an energy project, and it has been studied in many works in the literature. The ways of computing it differs from studies, so only some works will be cited to exemplify. First, in Bruck et al. [2018] and Mendicino et al. [2019] there are presented some of the main models. In particular, the first work presents four models in ascending order of complexity. Furthermore, it has a link with PPA as the last model that they suggest includes the effect of the minimum or maximum quantities that could be agreed in PPA contracts. Meanwhile, the second-mentioned work breaks down with much detail all the possible costs that can be included in the LCOE. Many works use the NREL model for the LCOE as in Miller et al. [2017]. On Levitt et al. [2011], the authors propose a very detailed list of the cost of an offshore wind project. However, they introduce the concept of breakeven cost, differentiating it from the LCOE. As they explain, they consider the breakeven cost as the cost of equilibrium, only considering the pure costs of installation and maintenance. In this regard, there is available a software to assess all the expenses of this type of project called SAM (System advisory model). That is, for example, used in Hernandez et al. [2016] to estimate the LCOE of a geothermal project.

Apart from all the works mentioned, some works make use of PPA for other purposes. Gallardo et al. [2021] for instance, use PPA to obtain or approximate the cost of the energy. Also, it appears in studies related to the maintenance of solar or wind plants as in Lei and Sandborn [2018]. In this case, the authors use the PPA to determine when is better to schedule the maintenance of a wind farm.

As seen, a variety of models have been proposed in the literature. However, both points of view do not have a consensus yet about the model needed. This fact is expected to continue since, as already explained, PPA terms and types differ a lot. Nevertheless, despite not having a unique model, the extensions presented above do not interfere. For that reason, the extensions seen can be adapted to each case, which is the approach that defines this research.

Preliminaries

Basic concepts and types of PPA

Previously to the research, it is essential to define PPA and explain the different types depending on the classification.

First of all, a Power Purchase Agreement is a contract with two parties implied where one part wants to purchase power and the other to sell it. Usually, the parts are a utility or a company that will buy the energy and a generator of electricity who sells it. This financial instrument has been widely spread among many countries as an efficient way to cover the investment necessities of renewable projects (the benefits of this instrument are going to be explained in the following sub-section). With 20 years being the most common term, these contracts are signed for long terms. This mechanism allows managing long-term price risk assuring a price to the consumer for a quantity of energy at a predetermined frequency. On the other side, the price eliminates part of the uncertainty about the project's performance for the seller. Some classifications could be proposed to distinguish between types of PPA attending to different characteristics.

If the structure is the characteristic taken into account, there are three different main types:

a). On-site PPA: The generator is established next to the company that has entered the PPA. In that way, there is a direct supply (called direct or behind the meter PPA too).

b). Off-site PPA: Founded as sleeved or physical PPA too. It is referred to PPA where the purchase of energy is at the meter point, and the customer receives the power through the grid.

c). Virtual PPA: It is called virtual, synthetic or financial PPA. In this case, the energy is not purchased physically, and the transaction is purely financial. In this structure, the buyer and the seller do not need to be in the same grid region compared to the physical PPA.

For a more detailed description of these types, see Mendicino et al. [2019]. In some cases, it is important to say that certificates of green energy are exchanged (depending on the region, these are called RECs, GOs or I-RECs). When the PPA is associated with the exchange of the correspondent RECs, it is called a bundled PPA.

According to the type of buyer, there are utility PPA if the contract is between the energy generator and a utility (seller in the wholesale market). It means that the buyer does not use the electricity purchased. Instead, the utility sells it to its clients. On the other hand, corporate PPA is when the exchange is made directly between a company and the energy generator. Currently, utility PPA are predominant in the market, but corporate PPA have also experienced massive growth in the last few years.

Another classification can be used. It is possible to differentiate based on the payment structure. The PPA can have a fixed or variable price (following an escalating rule, for example). As well, this agreement can be for the entire production of the supplier, for a percentage of the production or for a fixed amount (each day, month, year). A PPA is "Baseload" if there is established a fixed price for an hourly contract volume. This volume can be identical for each hour and can be established for different frequencies: for a given year, for a given month. Likewise, the volume can change between different hours. On the contrary, in PPA contracts "Pay as produced", there is a fixed price for all the produced energy or some percentage. So, as it is easy to imagine, the PPA price will depend on the price structure, and for that reason, this characteristic turns to be very important.

Importance of PPA from a Economic approach

First of all, the importance of renewable energy can be pointed out. The efforts to reduce the contamination and the impact of climate change are continuously growing. More and more agents become concerned about this topic, from families to big companies or countries. That can be confirmed by looking at the following graph (figure 1), representing the last 15 years of development in renewable technologies.

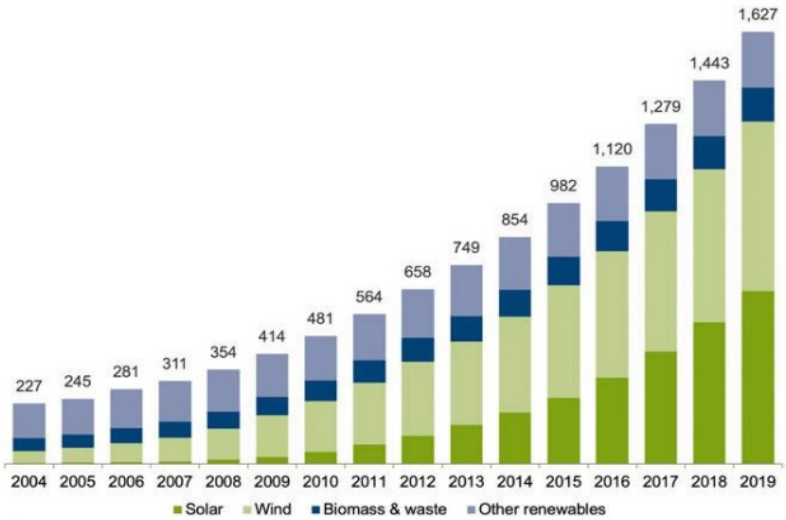

Figure 1: Global Capacity in Renewable Power (Unit of measure: GWh). Source: UNEP, Frankfurt School-UNEP, BloombergNEF. *Large hydro-electric projects of more than 50MW are not included

It is easy to say that attending to the graph, the energy generated by renewable sources is increasing. Moreover, the most remarkable growth comes from wind and solar technologies that have started representing a low percentage of the capacity but turning to represent now more than the 75% of the renewable energy produced. Those technologies have benefited from governments and public policies. Many countries have encouraged the development of these projects at the beginnings. Nevertheless, this behaviour becomes more testimonial once the region has developed a more mature renewable energies market. In this situation, in a more mature market than the last decade, different instruments are evolving between the participants of these markets to satisfy new needs. Renewable energy projects have different inversion structures than traditional sources. Because of that and other reasons, new financial instruments are appearing to respond to these new challenges and attracting inversion for renewable projects

In figure 2 it is possible to see some of the different instruments that are being used in order to carry out new renewable projects (from the producer's view) and to meet its sustainable goals (from the buyer's view). The data is about the United States, and for a better understanding of the different existent mechanisms, it should be checked the following reference Ajadi et al. [2020].

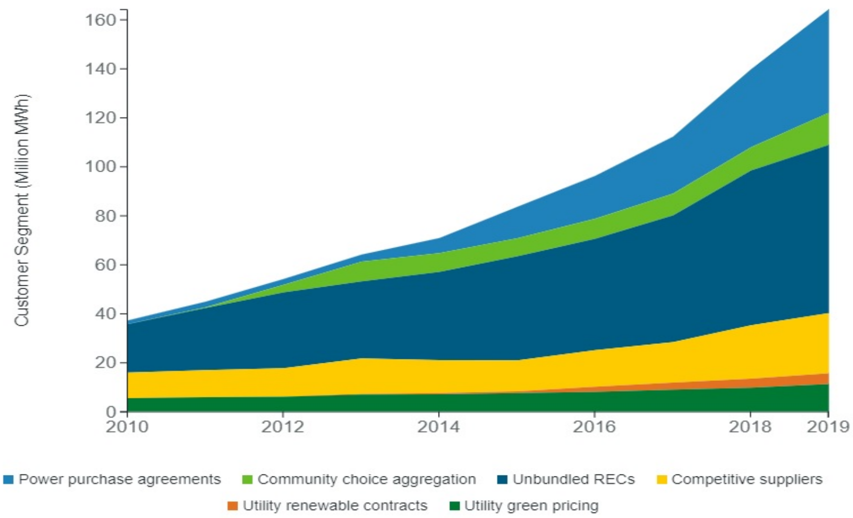

Figure 2: Sales of Voluntary Green Power. Source: National Renewable Energy Laboratory.

In figure [1] 3 it is visible that the use of these contracts by companies has grown with giant steps in the last years. In particular, in the last four years, it has been especially fast. The region with an earlier start has been the one denominated as AMER, with two breaks in the trend but with a positive tendency. The situation in the region "EMEA" shows a later growth but with a notable jump in the last year (opposite with AMER, whose volume has descended the last year compared with the previous one). The "APAC" zone has had a similar increase in the volume contracted by the companies but with separate paths in the last two years, maintaining the volumes in lower levels.

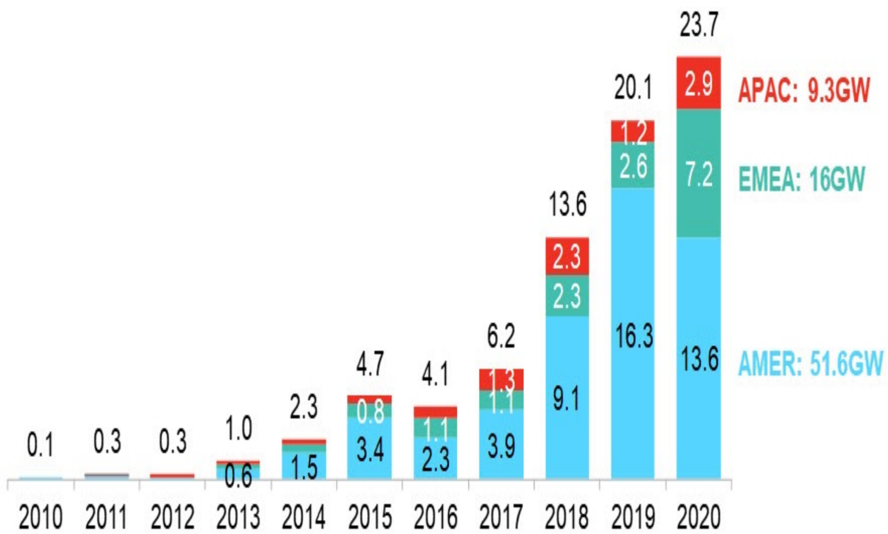

Figure 3: Global Corporate PPA Volumes. Source: BloombergNEF.

This graph allows seeing the increasing importance of PPA between companies in the last years. Besides, the positive trend is shared in all the zones reflecting growing importance worldwide. The cumulative value could be interpreted as a proxy of the power bought that year by companies through this type

[1] The unit of measure is in gigawatts per hour and is referred to as the DC Capacity contracted (total wattage of all the project). This data does not include onsite PPA (direct PPA) and other exceptions as Australian sleeved PPA or pre-market reform Mexico PPA. The quantities are not definitive because more information could become available, and specifically, the APAC value is an estimate. AMER includes North, Central and South America. EMEA includes Europe, Middle East and Africa. Furthermore, APAC includes the Pacific part of Asia. The values at the end of the columns represent the cumulative of each zone.

of financial instrument (to be more accurate, as the PPA are not signed for the perpetuity, this proxy should take into account that the contracts have different terms, but this is not the objective of this epigraph so as the graph only shows the last eleven years, it does not seem a critical assumption to think that PPA with terms lower than ten years are not so common)

Between the corporate buyers, there are multi-national companies such as Amazon, Google, and Facebook that want to reduce their carbon emission while reducing their power market risk assuring a fixed price. Also, it is used between industrial or other companies with high electric necessities. Virtual PPA have played a fundamental role in this growth, given that contracts between companies of different countries or power grid regions are possible. That incentives the competition and usually reduces the purchase price, becoming more attractive for the buyers.

Importance of PPA from a Financial Risk approach

Attending to pure financial arguments, Power Purchase Agreements constitute a very attractive instrument for both parties if the terms and structure are adequate for the individual needs of both parties. From the point of view of the generator of energy it gives him the following main benefits:

- It assures a price for the energy during the long-term that in some cases can cover all the expected life of the project. As with other commodities, it is usual that the producer turns to future markets to lock the price when he knows with some certainty the quantity that is going to have available to sell. Indeed, the seller objectives do not include speculating with the commodity price, so he will prefer to erase the price uncertainties. Furthermore, renewable energy projects usually can estimate their expected production for the whole active period of the project.
- Fixing a price for the exchanges allows the generator to plan its cash-flows and estimate its cost structure finer.
- Linked with the previous reason, PPA improve the so-called bankability of these projects. In other words, banks would see better a project endorsed by a PPA. That turns into lower interests and more favourable financial conditions too.

However, despite the benefits, there are as well some cons. With PPA, some risks are covered or partially covered, but some others arise for the producer. These are the risks associated with the uncertainty on the production volume and the negative correlation between prices and production. Both of them will be seen in detail in the analysis of this research.

Attending to the point of view of the off-taker, the key benefits are:

- As for the seller, the buyer avoids the market price risk in the sense that the amount that will have to pay would be known with a PPA. Nonetheless, as for every hedging strategy, future scenarios where the buyer would have had better results without the PPA could be possible (if the market price goes in the other direction, so, if the price decreases above the agreed PPA price). Despite this, electricity markets are very volatile. As in the other case, buyers whose intentions are related to using the energy for their business or basic needs do not want to be exposed to these risks.
- In some sense, it could be interpreted that it allows obtaining more competitive prices (however, this will not always be true as it will depend on the market structure and the negotiation power).
- It also helps the plans for the companies as they will be able to know their expenses related to this element. Indeed, this could be crucial for companies with extensive use of power.

Part B

Data

1 Power Purchase Agreements Data

1.1 Source

This research needs information about power purchase agreement contracts. Nonetheless, there is little public data available as the contract is not negotiated in a financial market, but it is usually an agreement between two parties. Besides, as explained before, this contract appeared not so many decades ago and is not until the last decade when its use has experimented a significant increase. That is why it is not so well-known yet. Currently, companies whose work is related to those instruments collect data and offer advice to valuate or price these derivatives. After a deep search, a database has been found. Berkeley Lab, the national laboratory of the U.S energy department, has two data sets, updated frequently, with information about solar PPA and wind PPA contracts. It contents the PPA execution date of the contract, the term in years, the capacity (quantity of MWh agreed to exchange), the levelized price, and the region where the contract is going to take place. The sample available for that research comprises 842 contracts where all this information is specified. It is fundamental to point out that all the PPA collected in the database are bundled PPA. That type of PPA always includes the selling of RECs simultaneously as the contract is signed. So, this aspect is essential to keep in mind as it will influence the pricing.

Then, as this investigation seeks to propose a model for pricing these instruments, a transformation is done. In this way, the actual prices agreed in the contract are obtained from the levelized prices, as these are expressed to be compared between them, but this research aims to do the realistic pricing. This procedure involves obtaining the price that makes both parties indifferent to entering into the contract or not (at the date of signing the contract). Previously, the data needs to be transformed as the prices available are the levelized prices for 2019. A detailed description of this process is included in annex 2 on page 78.

1.2 Descriptive Analysis of the data

For a better understanding of the available data, a brief descriptive analysis is provided. First of all, the main statistics of each series are computed. These statistics are the sample size, mean, standard deviation, median, mode, skewness, kurtosis, and maximum and minimum prices. Furthermore, this is obtained for the prices, the terms, and the quantities. The results are presented in tables below. Table 1 shows the descriptive statistics of the prices by type of energies where the differences between solar and wind data are notable. The sample has 353 solar PV PPA contracts and 489 wind on-shore PPA contracts. The mean or the median values of solar prices are more than 50 % higher than wind prices. Also, the variability seems to be much more significant in solar prices. The mode shows a relevant success since the more common price is higher on the wind data. The skewness and kurtosis reflect that the distributions of the prices are not similar to a Normal distribution. Regarding extreme values, solar prices have higher values in the distribution's right tail. In particular, it could remind of a log-normal distribution. Indeed, using the Jarque-Bera test, the null hypothesis of log-normality cannot be rejected in the wind and general samples[2], as seen in the Jarque-Bera statistics shown in table 1.

[2] The critical value is compared with the statistic obtained for each case, and the null will be rejected if the critical value is lower than the statistic. For a confidence level of 5% the crtical value is 5.99.

	Sample size	Mean	Std dev	Median	Mode	Skew	Kurt	Max	Min	Jarque-Bera Log-normality
Solar	353	125.56	76.03	105.97	92.82	1.41	5.35	494.6	21.99	12.8
Wind	489	78.35	41.28	69.79	131.02	1.05	3.77	207.7	17.27	5.5
All	842	98.14	62.86	82.60	92.82	1.80	7.66	494.6	17.27	1.9

Table 1: Price descriptive statistics of the PPA contracts database. Dollars per MWh

After that, in tables 2 and 3 the descriptive statistics of the term and capacity variable can be seen. In general terms, the contract's duration is very similar in both technologies of generation. Meanwhile, the quantities agreed for the exchange of power are higher on wind projects (except for one solar PPA with an agreement to exchange 690 MWh being the highest observation). That is reasonable as large-scale wind projects are more common than solar projects.

	Mean	Standard deviation	Median	Mode	Skewness	Kurtosis	Maximum	Minimum
Solar	22.20	3.93	20	20	-0.21	4.49	34	3
Wind	20.10	4.27	20	20	-0.32	4.23	34.20	5
All	20.98	4.26	20	20	-0.31	4.25	34.20	3

Table 2: Term descriptive statistics of the PPA contracts database. Years

	Mean	Standard deviation	Median	Mode	Skewness	Kurtosis	Maximum	Minimum
Solar	66.30	78.81	40	20	3.16	18.59	690	3.60
Wind	99.61	74.29	80	50	1.05	3.85	400	1.98
All	85.64	77.93	59.90	20	1.86	9.26	690	1.98

Table 3: Capacity descriptive statistics of the PPA contracts database. Megawatt hour

After that analysis, a very similar one has been replicated but, in this case, differentiating between regions. That has made clear the need to separate the models by region. Tables 4 and 5 present the same statistics for each region PPA prices. Generally, the differences are noticeable between regions in both technologies. Indeed, the differences observed in prices between solar and wind projects can be better explained due to the different percentages that each region represents in the available contracts than due to the difference between the generation technologies. The solar data have relevant samples for the CAISO and West-non-ISO regions overall. However, the wind data have relevant samples for the SPP, MISO, and West-non-ISO regions. Only attending regions with a sufficient number of contracts, there are great differences between the CAISO region and the West-non-ISO solar median price (and also concerning Southeast-non-ISO). The same is true for the wind data where SPP reflects fairly low median prices, then MISO, West-non-ISO, PJM, and CAISO have higher median prices (in that order).

This fact is consistent from a financial view because these regions represent different power markets, as explained before. So, the electricity prices in each market are expected to differ between different regions due to different demand and supply forces operating in each market. A power purchase agreement is directly related to the electricity, so, for coherent pricing, the agreement should consider the level of the prices in the particular region where the contract is being signed (spot prices and futures prices as well).

	Sample size	Mean	Standard deviation	Median	Mode	Skewness	Kurtosis	Maximum	Minimum
CAISO	123	139.93	74.18	119.09	92.82	1.15	3.86	365.55	34.15
West (non-ISO)	100	104.45	66.05	78.92	32.89	1.05	3.51	346.63	21.99
MISO	12	92.32	27.80	83.85	48.74	0.38	2.10	142.05	48.74
SPP	7	103.82	46.87	87.43	70.22	1.49	3.83	201.76	70.22
ERCOT	10	112.17	115.98	65	31	1.64	4.05	381.31	31
PJM	15	118.22	99.10	73.63	57.35	1.77	4.68	356.10	43.74
NYISO	3	342.99	131.92	279.67	254.67	0.68	1.50	494.63	254.67
ISONE	9	127.25	30.85	141.48	77.96	-0.73	2.15	161.82	77.96
Southeast (non-ISO)	46	102.82	68.76	96.99	52.48	2.44	9.85	396.14	33.64
Hawaii	28	179.81	46.60	174.80	124.69	1.14	3.74	307.18	124.69

Table 4: Price regions descriptive statistics of the Solar PPA contracts database. Dollars per MWh

	Sample size	Mean	Standar deviation	Median	Mode	Skewness	Kurtosis	Maximum	Minimum
CAISO	45	132.49	52.13	143.89	196.89	0.01	1.48	207.79	54.35
West (non-ISO)	96	87.06	37.06	89.57	131.02	0.22	2.41	171.75	17.27
MISO	121	76.2	32.49	72.4	67.46	1.04	4.43	190.13	26.9
SPP	130	48.34	19.81	43.45	32.29	0.96	3.44	108.33	18.31
ERCOT	19	53.45	18.48	44.77	34.23	1.08	3.43	102.16	34.23
PJM	53	95.83	33.02	100.67	129.59	0.67	3.44	204.21	44.1
NYISO	1	121.22	0	121.22	121.22	-	-	121.22	121.22
ISONE	15	109.98	48.20	81.27	65.8	0.780	1.92	199.37	65.80
Southeast (non-ISO)	9	69.11	28.30	68.24	33.50	0.25	1.89	110.76	33.50

Table 5: Price regions descriptive statistics of the Wind PPA contracts database. Dollars per MWh

Apart from this descriptive analysis, it could also be very revealing the following two figures. In particular, in figure 4 the long-term annual average solar resource is presented and, in figure 5 the wind speed. This information can help explain why some regions have more PPA contracts than other regions (not being the unique factor, obviously). As can be expected, regions will attract more projects the higher their production expectations are. That could explain to some extent the number of wind PPA contracts available on the SPP or MISO regions or even the West-non-ISO region. The same argument for the solar PPA contracts where CAISO and West-non-ISO seem to be the regions with the highest irradiance values coinciding with the number of solar PPA available for that regions.

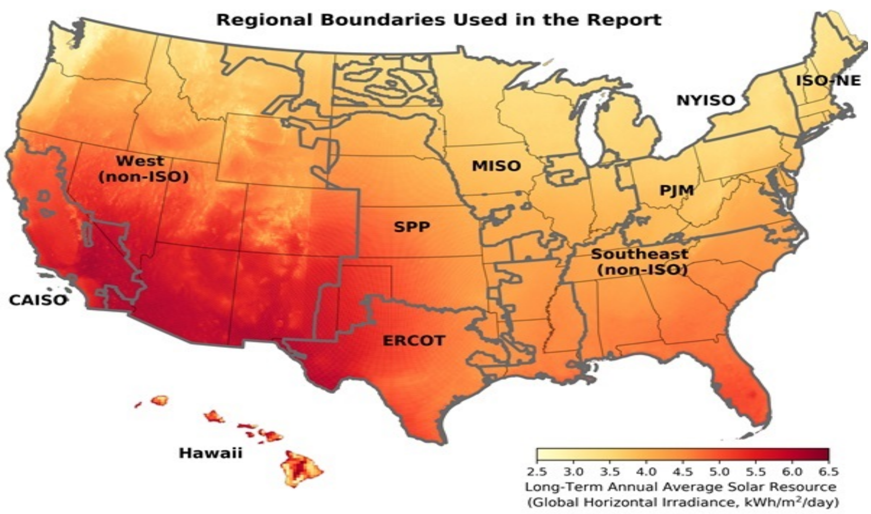

Figure 4: U.S radiation map. Source: National Renewable Energy Laboratory (NREL)

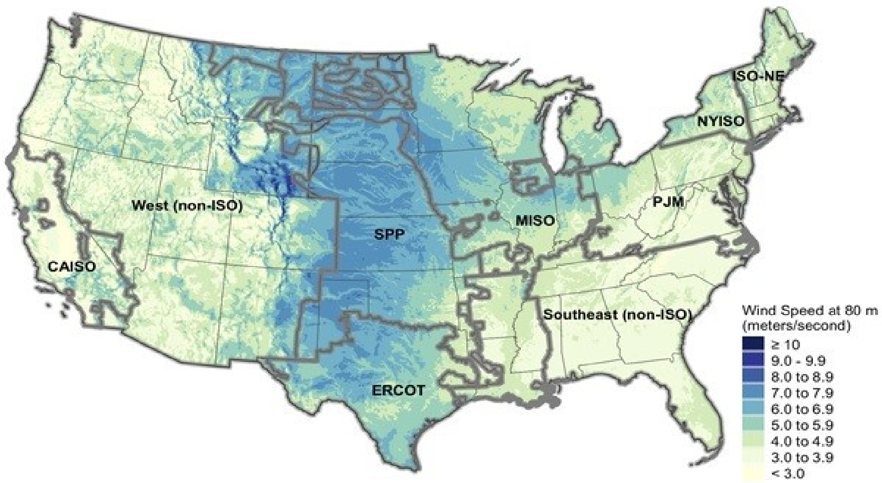

Figure 5: U.S wind speed map. Sources: AWS Truepower, National Renewable Energy Laboratory (NREL)

In conclusion, attending to the analysis carried out, it seems clear that a different treatment should be applied to wind and solar contracts as well as for different electricity regions of the United States. Furthermore, not all regions have sufficient contracts neither the minimum required information to develop the analysis. So, only two regions are considered to propose and test models.

In that direction, the first region considered is the CAISO region with the biggest amount of solar PPA contracts usable and rich information on electricity prices (spot and future prices especially). The second region considered is the West-non-ISO region. Its power prices are very similar in characteristics to CAISO, and the region constitutes a representative part (almost one quarter) of all the PPA contracts available.

2 Electricity U.S. Markets

2.1 Background of Electricity U.S. Market

The generation and supply of electricity in EEUU are governed by a complex framework of laws at the national, federal, and local levels. There are three zones of interconnection, eight regions in charge of the accuracy of the supply and the wholesale market, and various organized markets competing among them.
Three electricity interconnections that are almost independent between them (with limited energy exchanges) divide the United States at the first level. Those are:

- Western interconnect
- Eastern interconnect
- ERCOT or Texas Interconnect (that includes the majority of Texas)

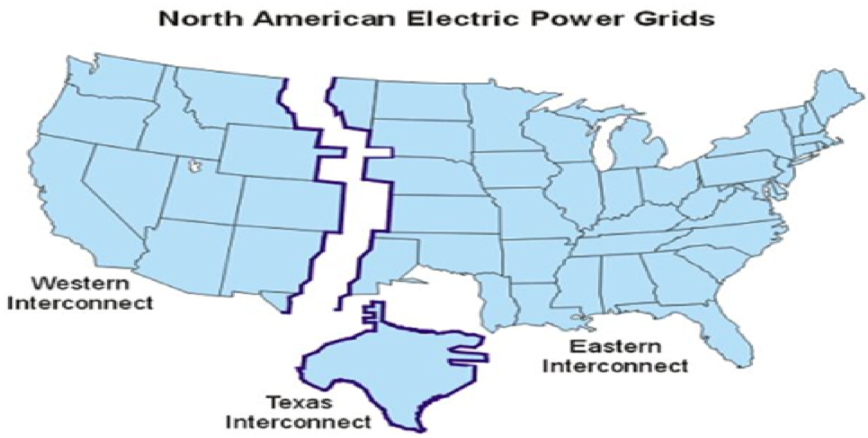

Figure 6: Power grid interconnections of EEUU. Source: United States Environmental Protection Agency

Going deeper, the North American Electric Reliability Corporation has eight regional entities to ensure that the power grid works appropriately.

Figure 7: Wholesale Electricity Power Markets

In this research, the northwest and southwest regions are treated as one unique region. Moreover, in the case of solar technology, the region of Hawaii is also included.

The FERC (Federal Energy Regulatory Commission) regulates the wholesale electricity markets (except the ERCOT market). Two types form these markets: On the one hand, there are restructured markets that compete with each other (whose transport management is done by independent system operators, ISOs, or regional transmission organizations, RTOs) where independent power producers and non-utility generators trade power. Those markets are usually more attractive for renewable producers. On the other hand, there are bilateral regulated markets where the power companies are vertically integrated to control the transport, generation, and distribution. Those second types of markets are the Southeast, Southwest, and Northwest regions of the map.

The structure of the market determines the renewable energy options, as said above. The explanation is that traditionally regulated markets are bounded to green power products offered by their utility (products that include renewable energy certificates, RECs). Meanwhile, customers in competitive electricity markets can shop for various electricity service providers, competitive products, or green marketing products.

Besides, the most important for this study is that the market structure also significantly influences the opportunities to enter into a power purchase agreement contract. For a physical or direct PPA, the electricity consumer must be in a competitive retail market and the project in a competitive wholesale market with connections with the consumer's ISO. However, for a financial PPA, an electricity consumer can be anywhere in the U.S., and the project must be in a competitive wholesale market.

2.2 Structure of energy generation

Currently, energy is a commodity that must be produced and consumed simultaneously (maybe in the future, if storage systems continue improving their characteristics, this affirmation could lose importance). This fact implies that the resources (energy generation plants) need careful monitoring in each electricity grid region. In the short term equaling the demand and supply of power by choosing the plants that should be actively generating energy and in the medium/long term forecasting the expected needs of energy and promoting new plants for the future when needed. This resource monitoring in the short term is done trying to have in operation the plants which lead to the minimum cost simultaneously assuring that the energy needs are covered. The different types of energy generators that operate in a region are crucial to price formation. With the introduction of renewable energy generation, some of these requirements have become more important, and others have arisen. One of the causes of these changes is the non-dispatchability of wind or PV-solar (currently the primary renewable sources). The non-dispatchability is a typical feature of both technologies and refers to sources of electricity that cannot deliver the energy on demand. In particular, this energy generation is dependent on the weather conditions and, more specifically, the wind's speed and solar irradiance. That is highly problematic in every market where these technologies are being integrated because it challenges the existing procedures to meet the region's electricity needs. In response to it, there is a vast literature about how to face this problem on integrating efficiently non-dispatchable resources as on Perera et al. [2017]. Another influencing feature is the ramping of the different plants in each region. Ramp rate is referred to the rate, expressed in megawatts per minute, that a generator can change its output. Applied to energy generators, it is related to the costs and ability to shut down or start production. It is relevant since it causes inefficiencies because of the inter-temporal constraints of some technologies whose production cannot be stopped or started immediately. This characteristic is not so problematic if the energy production of the existing and active generators is regular. However, with the inclusion of renewable energy, as the production of these plants can be very irregular, it implies that the ramping of each available plant becomes very relevant to avoid inefficient scenarios.

To better understand electricity prices, it is necessary to comprehend how they are formed following the so-called merit-order effect. This effect means that power prices are the marginal cost of the last needed electricity generation technology. In particular, renewable energies very often have very low marginal costs. Hence, including renewable generation in this process shifts the supply curve to the right (concerning their priority dispatch), resulting in lower energy prices. Another characteristic of electricity is its highly inelastic demand in the short run due to consumer difficulties changing their consumption patterns of this commodity. This economic argument has been investigated so far in many countries and applying diverse methods. In Würzburg et al. [2013] the authors make an exhaustive review of the existing researches about the price effect of renewable generation. One of the main conclusions is that despite the differences observed between countries and the different methods applied, it could be accepted that including renewable resources in a market region results in a decrease in electricity prices. However, as renewable production is irregular and intermittent (due to its dependence on weather conditions), that is not the unique effect observed in the article. Indeed, renewable production could challenge electricity markets, increasing its uncertainty and turning more complex to match the supply and demand. For example, for the case of Germany, Kyritsis et al. [2017] present an empirical study about this topic focused on wind and solar generation. The results for that sample showed that solar power had reduced prices and its volatility. On the other hand, wind power had also reduced prices (at a greater level) but had increased the volatility by introducing electricity price spikes. In this way, it could be remarkable the importance of having a deep knowledge about the energy generators in the region to be studied. Likewise, the arguments explained before are related to the main objective of this paper. PPA are a financial instrument likely to promote renewable projects. As it can be viewed, this development must go hand in hand with a correct comprehension of the phenomenons explained above.

In the descriptive analysis of the database PPA prices, the regions with a more significant sample and the required power prices information are CAISO and West-non-ISO. Furthermore, the results showed that different wind and solar PPA treatment is needed, but it also seems the same for the different regions. In accordance with this, it has been decided to do the modelling of the PPA prices with these two regions, expecting that as more contracts are available, the better for the model's reliability. For that reason and the arguments explained above, before developing the models, it has been considered necessary to do a brief research about the main characteristic of these two regions and their power markets with particular attention to the generation structure. For that objective, it is essential to define the capacity ratio previously. This ratio is between the energy produced by a generator and its capacity at its maximum performance. In this case, it will be used to talk about the capacity ratio of each type of generation. For example, nuclear generation usually has a ratio of nearly 90% or more, and some fossil fuels as natural gas or coal have a ratio around 50%. Meanwhile, wind generation has a lower ratio of around 40 % and solar generation even lower with a 25-30%. For this reason, it is crucial to differentiate between generation and capacity when talking about types of energy generation.

2.3.1 CAISO

The California Independent System Operator is a non-profit ISO covering most California (80%) and a small part of Nevada. It grants access to electricity in the region with almost 26,000 circuit miles of transmission lines and coordinates the energy resources from different term perspectives. As can be noticed in the map of figure 7 is the unique ISO of the Western U.S.

For an adequate comprehension of the market prices, it is fundamental to know how the market operates and the composition attending to the different energy generation technologies available in the region. In figure 8 it is the percentage that represents each type of generator of the maximum on-peak available capacity in the summer of 2020. According to this, the availability of energy generation was based on a significant percentage of natural gas power. Gas is a fossil fuel and typically with higher costs than renewable energies. That implies that the price structure is going to depend a lot on the price of this underlying. Also, hydropower represents a significant 16% of the maximum capacity available. In the third position, solar technology can provide up to 9% of the total energy consumed on the maximum on-peak.

Likewise, it is helpful to see the percentage of energy generated by each type of technology because the capacity available with each technology is not usually the same as the percentage of use of each one. In figure 9 this data is presented for the year 2019. On the left graph, the percentages have changed, and gas is not as important as in the full capacity graph. That means that on-demand peaks of energy prices will depend more on gas prices than in everyday situations. It is also noticeable that renewals (including solar, wind, biomass, and others) have a percentage higher than hydro-power[3] as well as the significant importance of the imports of energy. Along the months, it is visible a changing profile with fewer percentages of gas energy generation approximately during the spring. The increase of energy production in renewals and hydroelectric technologies could explain that. So, price expectations are lower in those months due to the higher importance of technologies with fewer production costs.

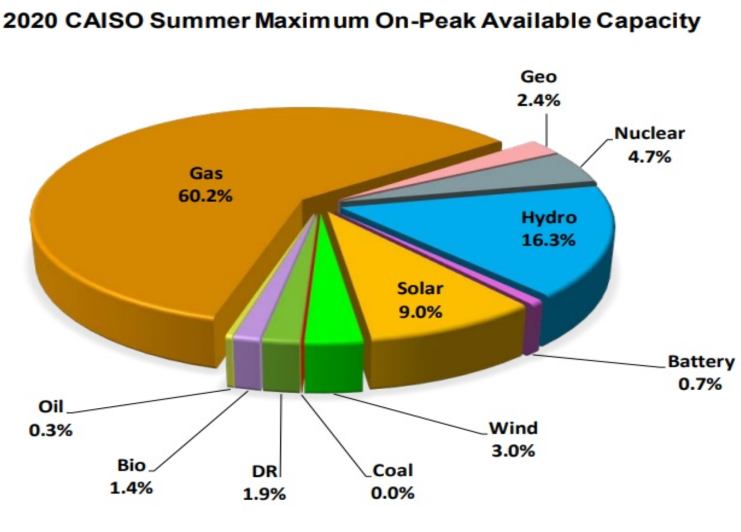

Figure 8: Percentage of generation by each type of fuel on the Maximum Available Capacity. CAISO. *Source: Report of CAISOs website*

This changing production structure occurs during the day too. In figure 10 this is shown with the daily generation profile of each technology. It has to be remembered that solar production has a relevant impact on the total production, and attending to this last graph, the influence of this generation would

[3]As it represents a relevant percentage, it would be helpful to know the percentage of hydroelectric generation due to conventional damns or to pumped storage plants. Hydroelectric generation is considered dispatchable. More specifically, when the plant is a pumped-storage one, the possibility to pump the water (on off-peak electricity demand periods) allows to generate power and storage it. These characteristics can be advantageous to mix in markets where renewable energy is being integrated as it permits to match the demand because it is dispatchable and allows energy storage

be more decisive in the sun hours. In this interval, gas, imports, and hydroelectric (at a fewer level) experiment significant reductions. The hours of maximum energy demand are usually on the interval between 16 - 20. Thanks to the amount of solar energy, this market probably does not show its higher prices in this interval. Instead, maximum prices will probably be around 20-22 in concordance with the decrease of solar production and increase of gas and imports. This can be confirmed in figure 11, where the interval hours with higher prices of 2019 coincided with the hours where the solar production was decreasing or almost zero between 18-22 interval hours. The same occurs for the lower prices that matched the peak solar production hours.

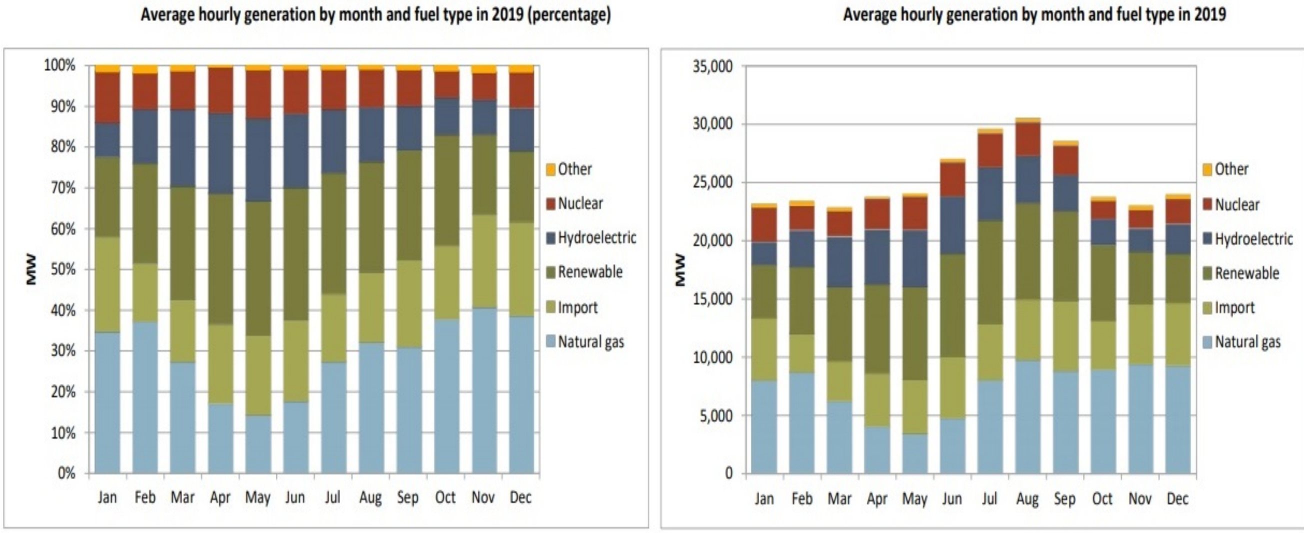

Figure 9: Percentage produced each month by types of fuel on 2019. CAISO.
Source: Report of CAISOs website

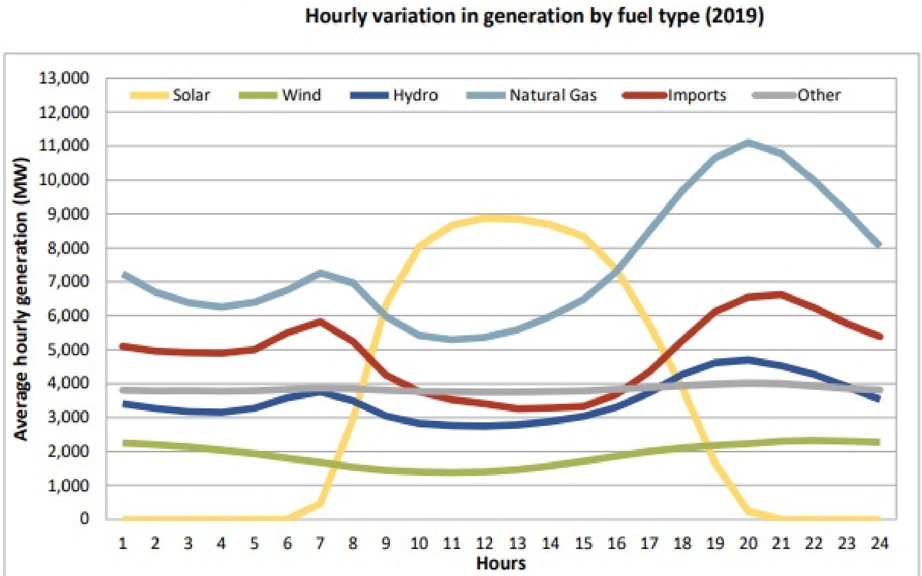

Figure 10: Percentage produced each hour by types of fuel on 2019. CAISO.
Source: Report of CAISOs website

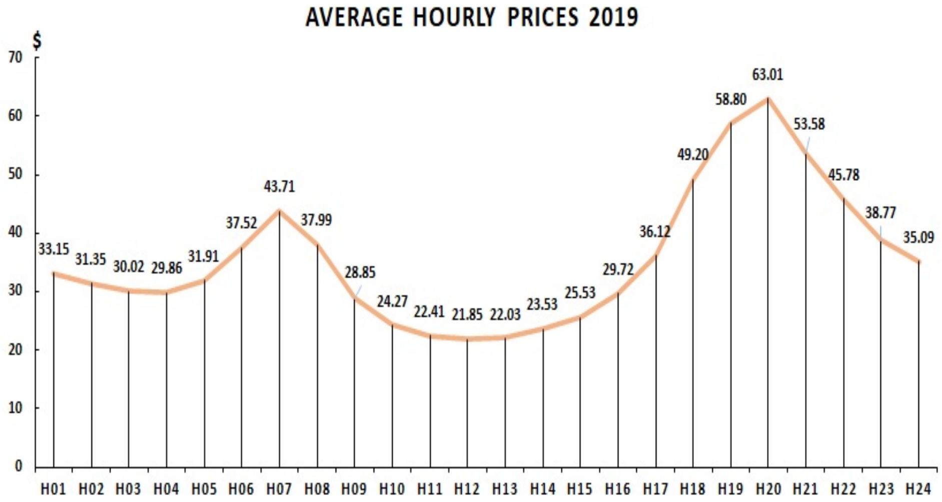

Figure 11: Average hourly prices 2019. CAISO.*Source: Data from Reuters, own elaborated graph*

2.3.2 West-non-ISO

The West-non-ISO covers the Southwest and Northwest regions. It is the part of the west of the U.S without an ISO operator. Both regions are examined as a unique region in this research as the PPA database considers both regions as the West-non-ISO region. Nevertheless, it includes the Northwest Power Pool (NWPP), the Rocky Mountain Power Area (RMPA), and the Southwest Power Area (where are listed Arizona, New Mexico, and Southern Nevada). All of them belong to the Western Electricity Coordinating Council (WECC). A common characteristic of these areas is that they contain many balancing authorities (B.A.s) with different competencies (dispatching generation, obtaining power, assuring the electricity grid reliably, among other things).

The NWPP comprises seven states and one small portion of California, covering 1.2 million square miles. It has up to 20 Balancing Authorities. The Southwest electric market encompasses the three states already mentioned. The Rocky Mountain Power area is the smallest one, responsible for the electricity grid in parts of Utah, Wyoming, and Idaho.

Now, for a better comprehension of this market, a brief analysis of the capacity and net generation by fuel types is presented. Attending to figure 12 natural gas is the central capacity resource in the region. However, the North-West region has a higher percentage of hydroelectric generation. Indeed, this type of generation has a significant percentage in all regions. Nuclear is only available (in significant quantities) in the South-West region. Coal is a fossil fuel used in the three regions too. Last, wind and solar generation differ from each sub-region, but they similarly have a significant percentage.

However, focusing on figure 13 renewable generation with solar or wind technologies had residual importance in the period shown in the figure but increasing with the years. Furthermore, as seen, coal has greater use than natural gas, or the difference is not as big as with the maximum capacity.

So, the market has a similar composition as the CAISO market except for the coal that has significant importance in the generation. In contrast, in the CAISO grid, the generation is residual. Besides, attending to the graphs at the top of figure 13, base-load fuel types predominate in the region, so the prices are expected to be more regular than if non-dispatchable generators represent a higher percentage. Though hydroelectric generation represents a significant part of the net generation, as stated before, it would be useful to know the production originated by conventional damns and by pumped storage plants. Knowing that would determine the percentage of energy generation that can be considered purely base-load type (but as with the CAISO market, this information is unavailable).

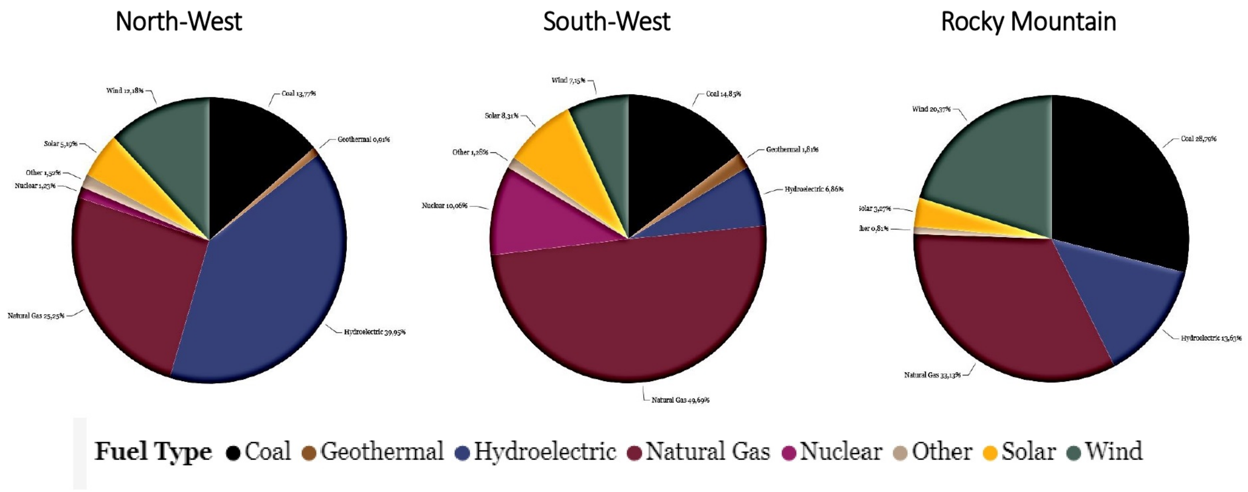

Figure 12: Capacity by types of fuel in the three regions of West-non-ISO (2021 information). *Source: WECC website*

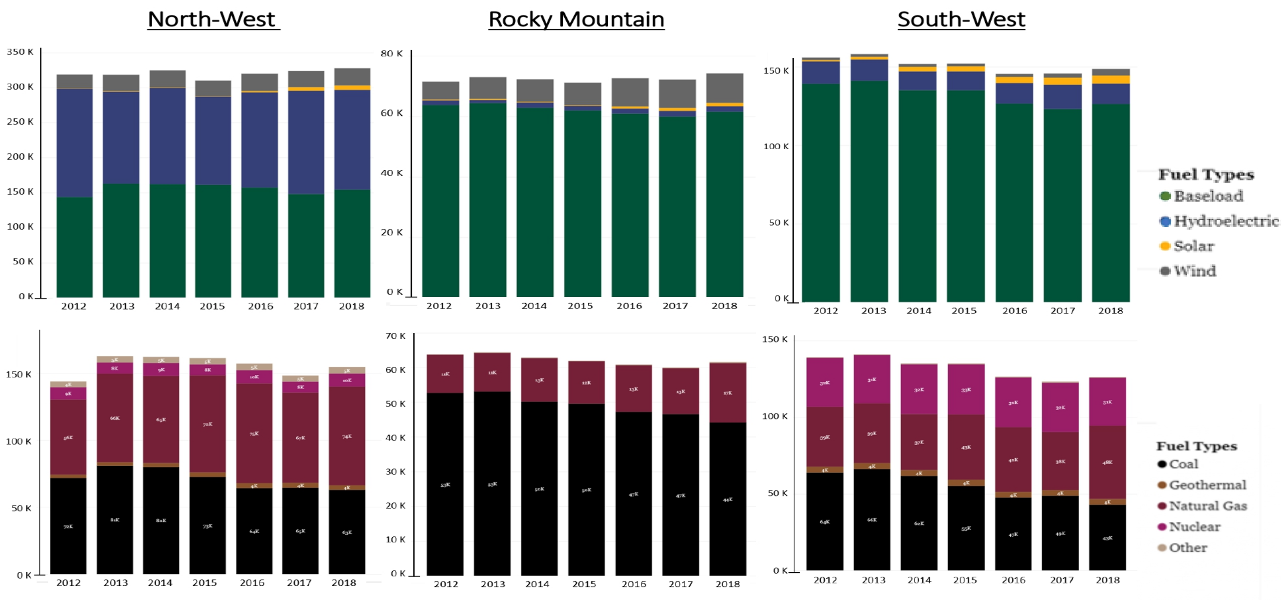

Figure 13: Net generation by types of fuel in the three regions of West-non-ISO (2019 information). *Source: WECC website*

3 Power prices data

A brief description of the data used related to power prices (both spot and forwards) corresponds to this section. In particular, two markets have been analyzed: the CAISO market and the West-non ISO market.

First, the CAISO market available data is explained. This region is divided into three main sub-regions called SP-15, NP-15, and ZP-26. Among these three regions, there is only available data about SP-15 and NP-15. So, these two regions have to be used as a proxy of the CAISO's prices. To unify the series, the weighted average has been computed to have the price of all the region. Nevertheless, SP-15 seems to be predominant between these two regions, so the individual SP-15 prices series has been considered a possible proxy too. Indeed, data is broken down into regions for both spot prices and future prices, so this explanation is handy for both cases.

Second, the West-non-ISO region available data has different particularities. As seen in section 2.3.2, this region covers a vast territory of the west of the U.S. Additionally, being a non-ISO region could make it more complex to find transparent and precise data. Despite these problems, there is data available of future prices of the Mid-C hub, and spot series also from this hub and some others as Palo-Verde or Mead hub. Consequently, taking into account the different options, the best one seems to use for the future price series and the spot series the Mid-C hub data. Although there are many hubs in this region, and the Mid-C hub only represents one of them, it seems more reasonable to use as a proxy only this data since for future prices is the unique hub that information has been found. This hub is from the North-West power region specifically, but the spot series seemed to be very similar to the other ones regarding the trend.

The following paragraphs refer to both regions but divided for the spot series first and the future series second. The spot series used are available on the data source "reuters" where the explained information is organized. In particular, the series from the ICE day ahead prices provided on this data source are used because the data starts some years before the original Reuters series. In fact, for the West-non-ISO region, the original Reuters series for the day ahead prices are not available.

Second, the future prices, also provided by Reuters, are described with more precision, as they have other characteristics fundamental to be explained and necessary to understand the first model later. In particular, the future contracts available data is from the Tullet and Prebon broker. More precisely, there are available prices for energy purchases in the next three ends of the months, in the next six ends of the quarters, and in the next eight ends of the years. However, only the quotes of the year contracts are used.

At each date, there is the spot price of the last day quoted and the future price for the end of the next eight years. So, for example, in the first of January 2019, the database has the spot price of this day and the future yearly prices with maturity on 31/12/19, 31/12/20, 31/12/21, 31/12/22, 31/12/23, 31/12/24, 31/12/25, and 31/12/26. Then, the series are ordered so that at each date all the row has the future prices ordered from the nearest maturity contract to the farthest one. So, a curve of prices can be formed with all this information covering the actual price and the next eight prices at the end of the years to purchase 1 MWh of power between the next day when the contract expires and the hole next year.

In figures 14 and 15 the histograms of the spot series used are presented. The CAISO prices reflect a very similar distribution comparing the SP-15 data and the SP-15 and NP-15 mixed data. Likewise, the West-non-ISO prices have also a very similar distribution with the main difference in the left tail, which has more negative values. To sum up, the three distributions have heavy right tails and a remarkable asymmetry, with the mode being the lowest value, after the median, and last the mean.

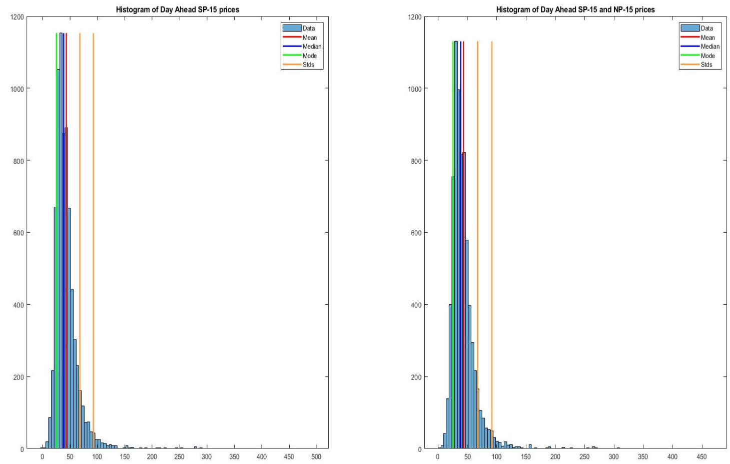

Figure 14: Histogram of the price series used for the CAISO market

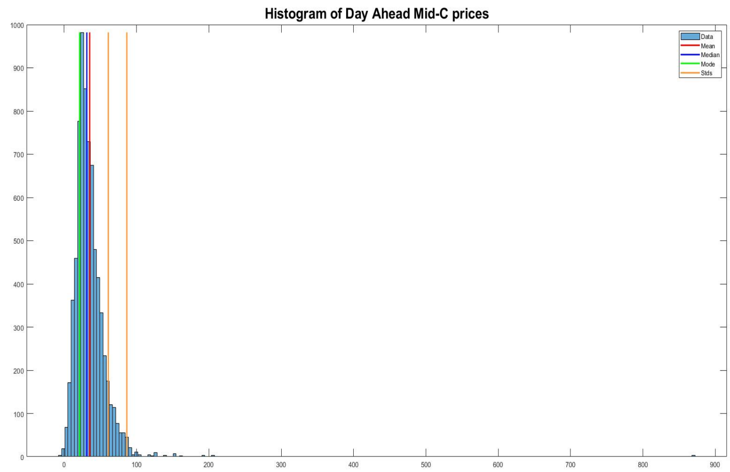

Figure 15: Histogram of the price series used for the West-non-ISO market

Part C

Analysis

4 Quantity and term analysis

This chapter presents a brief analysis of the relationships between the PPA prices and other variables to see if these variables can be considered determinant factors of the PPA pricing. For that purpose, to evaluate the relationship between variables, a graphical and correlation analysis is done. The correlation analysis includes the following three existing correlation measures:

- Pearson correlation
- Kendall correlation
- Spearman correlation

Moreover, to complete the analysis, a formal test is going to be computed to see if the correlation coefficient (Pearson's correlation) is different from zero or not. For that purpose, the t-test is used. In that way, it could be evaluated if a conclusion about the population can be drawn from the available sample or not. That test is useful if the variable to be considered as the response is not clear. Formally, the null hypothesis for the t-test of the population correlation coefficient will be:

$$H_0 : \rho = 0 \;\; and \;\; H_1 : \rho \neq 0 \tag{1}$$

The t-statistic is:

$$t^* = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

Then, the p-value is calculated assuming that the statistic follows a Student-t distribution with n-2 degrees of freedom. After that, the hypothesis is rejected or not if the value is lower than the significance level of $\alpha < 0.01$ (**) or not.

4.1 Relationship between PPA prices and quantity.

The first part of the analysis is about the connection between the price and the quantity. The underlying argument in this relationship could be that the quantity may be related to the price as could be interpreted that the larger the project is, the lower cost the project has. Nevertheless, arguments in favour of possible dis-economies of scale for especially large projects could also be considered as Bolinger et al. [2019] explain in their empirical analysis of PV Utility-scale solar projects. The authors hypothesize higher costs to face (administrative, regulatory, or some others) or the longer time needed to construct these projects.

4.1.1 Whole Database

As it could be expected, the data shows a negative relation and more or less similar between solar or wind data. This relationship is relatively low, confirmed by the scatter plot where the relationship does not seem clear. However, the t-test rejects the null hypothesis. It means that sufficient statistical evidence exists to conclude a linear relationship between the variable price and quantity in the database.

Correlation \Sample	Solar Database	Wind Database
Pearson	-0.1534 (0.0039)**	-0.1877 (0.000)**
Kendall	-0.2520	-0.1668
Spearman	-0.3603	-0.2485

Table 6: Correlation Measures between PPA prices and quantity. All regions (p-value of the t-test)

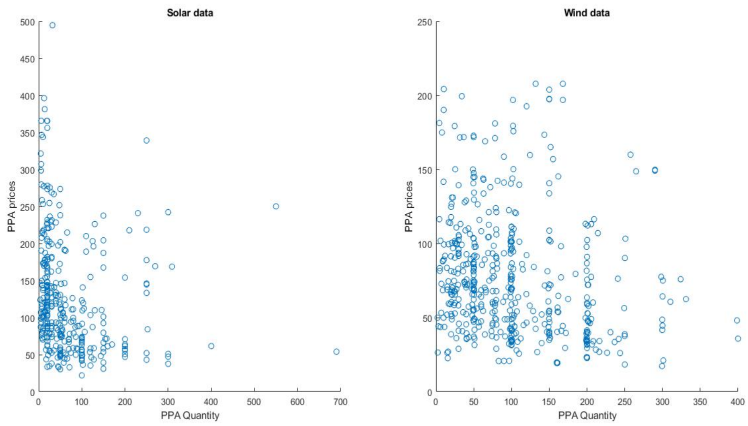

Figure 16: Scatter plot between PPA prices and PPA quantities

Identically, the analysis can also be done by separating the sample into four different size groups. In that way, it is possible to see if there are differences between small, medium, large, or very large projects (lower than 20 MWh, between 20-50 MWh, between 50-100 MWh, and higher than 100 MWh projects, respectively).

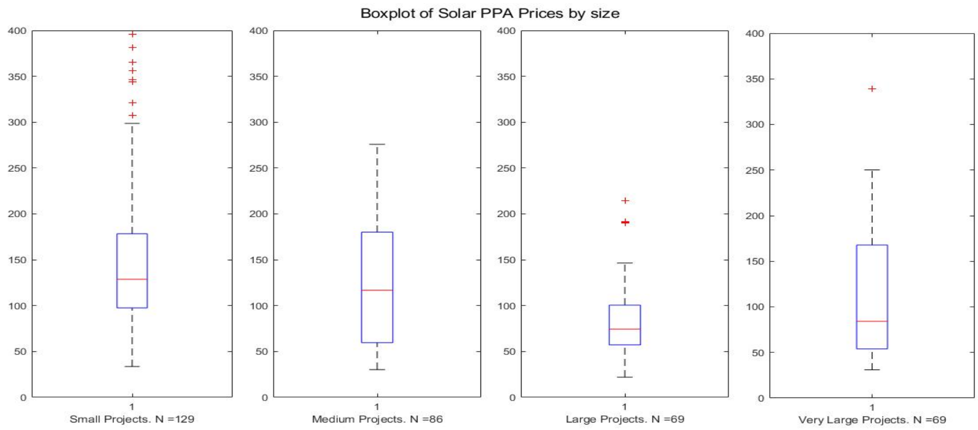

Figure 17: Prices Box-Plots of different Solar projects sizes

Figures 17 and 18 show that Solar and Wind prices of the database reflect a decreasing trend. However, very large solar projects are more dispersed and with a slightly higher median price.

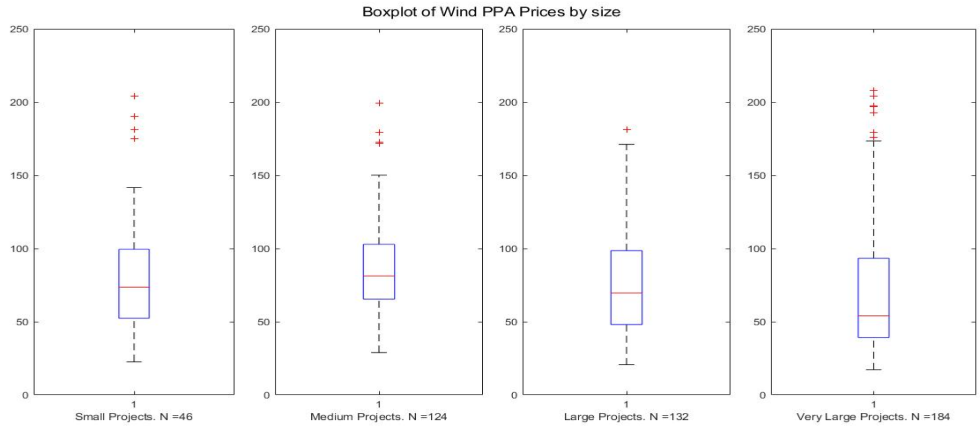

Figure 18: Prices Box-Plots of different Wind projects sizes

As seen in section 1.2, the database is very heterogeneous among all the regions. As the model proposed is tested on CAISO and West-non-ISO data, the analysis is now focused only on these regions.

4.1.2 CAISO Region

When only CAISO data is contemplated, the results change significantly. On the one hand, solar data shows a non significant relationship. In addition, the t-test does not show enough statistical evidence to reject the null hypothesis. On the other hand, surprisingly, wind data reflect a positive relationship. Indeed, the sample shows a strong relation confirmed with the t-test that rejects the hypothesis of the in-existence of a linear relationship between wind PPA prices and quantities.

Correlation \Sample	Solar CAISO	Wind CAISO
Pearson	0.1309 (0.1491)	0.6282 (0.000)**
Kendall	-0.0083	0.4365
Spearman	-0.0135	0.6055

Table 7: Correlation measures between PPA prices and quantity. CAISO region. (p-value of the t-test)

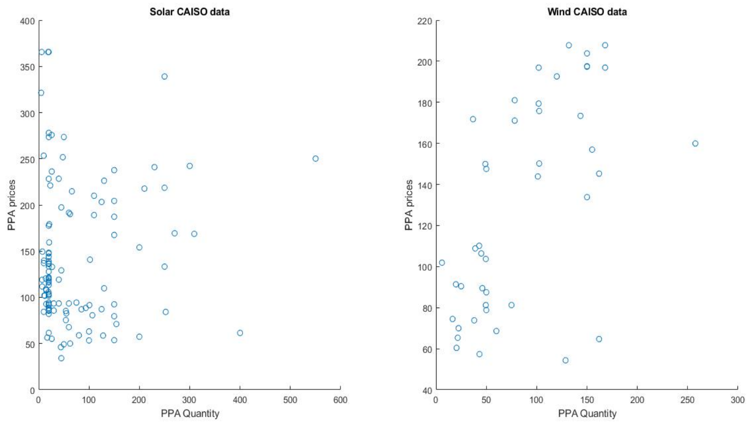

Figure 19: Scatter plot between PPA prices and PPA quantities. CAISO Region

Same as before, the Box-Plots are presented below to see the trends by project size but only considering the CAISO data. Figure 20 reflects an increasing median price for the case of solar very-large projects compared with the previous size. Moreover, for the wind case, figure 21 reflects the facts already seen. When increasing the size of the project, the PPA prices seem to increase too.

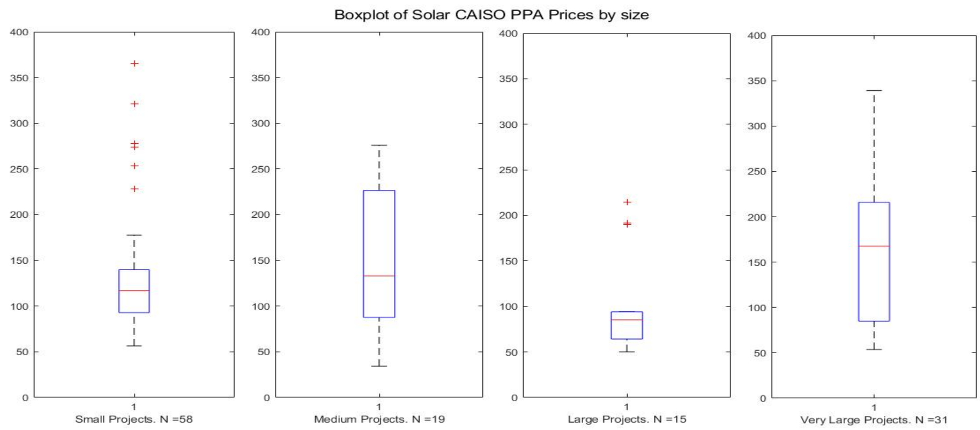

Figure 20: Prices Box-Plots of different CAISO Solar projects sizes

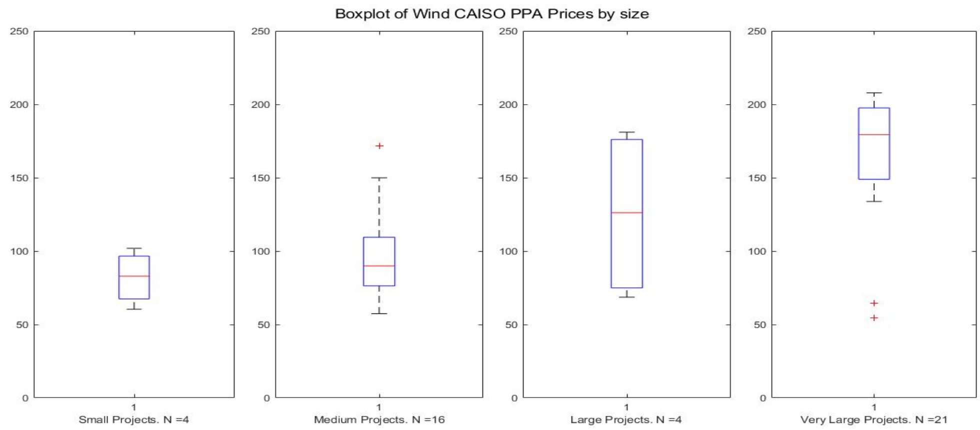

Figure 21: Prices Box-Plots of different CAISO Wind projects sizes

4.1.3 West-non-ISO Region

Now, the West-non-ISO region is examined. This time, solar and wind data reflect a negative correlation. Indeed, it is stronger than the correlations found for the full sample. The t-tests also confirm the existence of a linear relationship. In solar at the significance level of 1%, and in wind it is non significant at the level 1% but at 5% yes.

Correlation \Sample	Solar West-non-ISO	Wind West-non-ISO
Pearson	-0.3204 (0.0012)**	-0.2225 (0.0293)
Kendall	-0.3851	-0.1592
Spearman	-0.5468	-0.2502

Table 8: Correlation measures between PPA prices and quantity. West-non-ISO region. (p-value of the t-test)

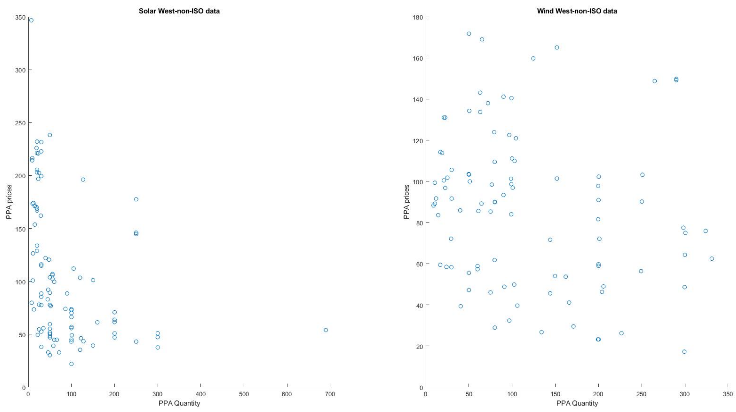

Figure 22: Scatter plot between PPA prices and PPA quantities. West-non-ISO Region

The Box-Plots are presented as in the other cases. Figure 23 reflects a decreasing median price along with the different sizes. In the wind case, figure 24 shows a decreasing trend but less evident. Furthermore, the dispersion seems to decrease by the size in the solar case while it seems to increase for the wind data.

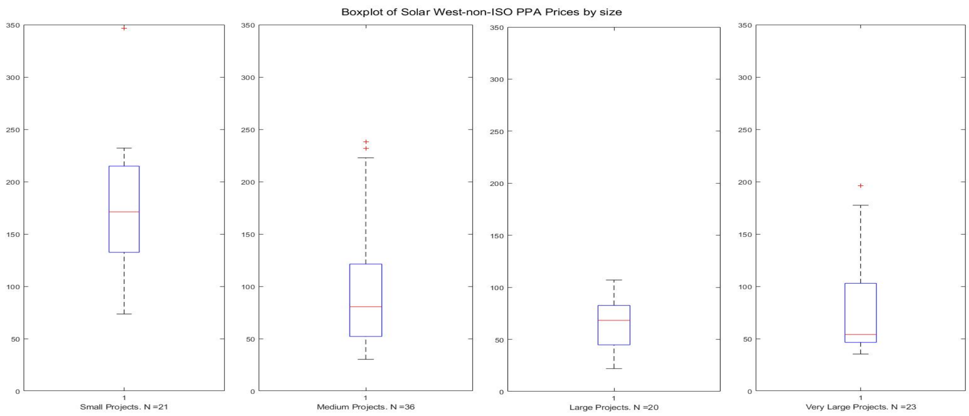

Figure 23: Prices Box-Plots of different West-non-ISO Solar projects sizes

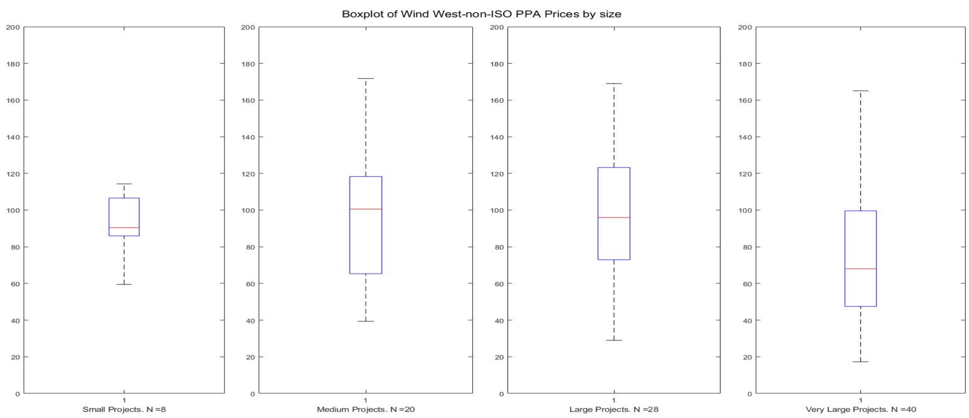

Figure 24: Prices Box-Plots of different West-non-ISO Wind projects sizes

In general terms, different relationships have been found. The two specific regions present significantly different results. That could confirm that the analysis should be done separated by regions and technologies of generation as the data seems to differ attending to the evidence.

4.2 Relationship between PPA prices and term.

This second part of the analysis can be justified as some approaches consider that some prices have implicit a term-premium in their values. This framework has been intensely studied in the literature of interest rate markets. So, considering that this argument is also possible in this market, the relation between these variables is analyzed.

4.2.1 Whole Database

Positive but not so high values can be seen for all the regions. Nevertheless, the results cannot be considered valid for all the population if a significance level of 0.01 is determined.

Correlation \Sample	Solar Database	Wind Database
Pearson	0.1330 (0.0124)	0.1018 (0.0244)
Kendall	0.1207	0.0425
Spearman	0.1614	0.0546

Table 9: Correlation Measures between PPA prices and term. All regions (p-value of the t-test)

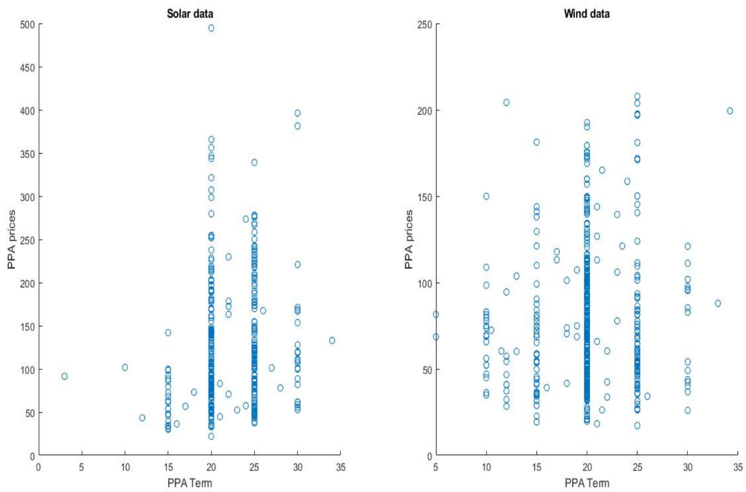

Figure 25: Scatter plot between PPA prices and PPA terms

4.2.2 CAISO Region

Again, when only the CAISO region contracts are considered, the situation changes. Now, the relationship seems to be positive in solar data as well as in wind data. Furthermore, focusing on the graphs, the direct relationship could be expected before seeing the correlation coefficients. Finally, this idea is strengthened as the t-test allows concluding with sufficient statistical evidence that a linear relationship exists in solar and wind CAISO data.

Correlation \Sample	Solar CAISO	Wind CAISO
Pearson	0.2761 (0.002)**	0.6732 (0.000)**
Kendall	0.2989	0.5558
Spearman	0.3894	0.6823

Table 10: Correlation Measures between PPA prices and term. CAISO region. (p-value of the t-test)

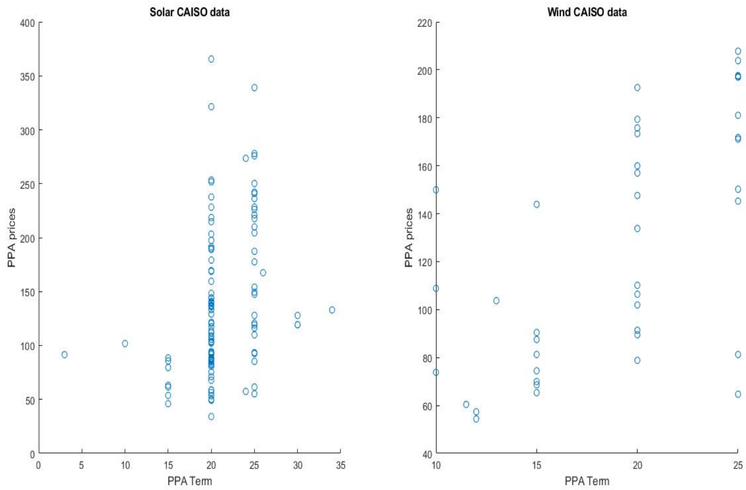

Figure 26: Scatter plot between PPA prices and PPA terms. CAISO Region

4.2.3 West-non-ISO Region

Another time, the West-non-ISO region differs a lot from the CAISO region. Both correlations do not reject the null hypothesis of the inexistence of a linear relationship. And what is more, the wind sample reflects a negative correlation (though it is very weak).

Correlation \Sample	Solar West-non-ISO	Wind West-non-ISO
Pearson	0.1274 (0.2066)	-0.0546 (0.5969)
Kendall	0.1255	-0.0842
Spearman	0.1619	-0.0921

Table 11: Correlation Measures between PPA prices and term. West-non-ISO region. (p-value of the t-test)

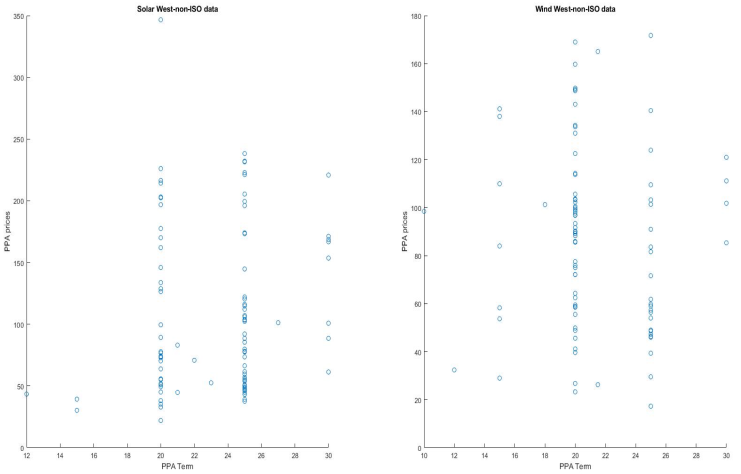

Figure 27: Scatter plot between PPA prices and PPA terms. West-non-ISO Region

In conclusion, on the one hand, the correlation between price and size reflect significant negative relationships for the full sample and the West-non-ISO solar data. Nevertheless, CAISO does not have a significant correlation for solar data, but for wind data, a positive relationship is confirmed by the t-test. In addition, wind West-non-ISO data seems to be non-significant. So, the prices and size depend on the sample considered, and despite expectations of negative or non-significant values, the CAISO sample does not follow that assumption. On the other hand, between prices and terms, less significant relationships are observed. Specifically, only the CAISO region shows a significant positive correlation for both technologies (in fact, high values). Meanwhile, the terms appear to not indicate a relationship in the West-non-ISO case and in the whole sample.

5 First model: Standard SWAP pricing

5.1 Methodology

The first model to be tested is based on the traditional approach for swap valuation but with some modifications. Generally, in a swap, a fixed leg has to pay flat or pre-agreed prices, and there is also a float leg in which the payments will depend on the price of the predetermined underlying. Specifically, a Power Purchase Agreement is a regular daily/monthly/yearly exchange of a predetermined quantity of power for a fixed price (or a price based on some rule established before starting the contract). The float leg of the PPA is the part that must deliver the energy, and the fixed leg is the part that receives and pays the power. In this way, there are cash-flows of the fixed and float leg. In particular, the net present value of each leg (supposing an exchange of 1 Mwh) is given by:

$$NPV^{fixed} = \sum_{i=1}^{n} P_{PPA_i} \cdot e^{-r \cdot (t_i - t_0)} \ ; \ NPV^{float} = \sum_{i=1}^{n} P(t_i) \cdot e^{-r \cdot (t_i - t_0)} \tag{3}$$

Where P_{PPA_i} is the price agreed for the PPA for the exchange number "i". In particular, in this research, the price agreed is taken as it remains at the same value in all the exchanges, so the sub-index i from P_{PPA} could be erased.

Furthermore, t_i is the date associated with the cash-flow number "i". $(t_i - t_0)$ are the years between the execution date and each concrete cash flow (obtained dividing by 365 the number of days in the interval). It is essential to remember the period between the agreement and the start of the trade-offs (t_1 - t_0).

Likewise, "n" is the number of exchanges that the PPA involves. It is obtained by multiplying the term of the contract with the frequency of the exchanges. For example, if the contract is for monthly deliveries of energy, the multiplier will be 12, but if there are exchanges each day, it will be 365.

Last, $P(t_i)$ is the electricity quote in the market where the PPA has been signed on the date t_i. It represents the variable and unknown price of electricity. As the pricing is done when the PPA contract is signed, these prices will not be certainly known. For that reason, those should be estimated. Taking expectations and under the risk-neutral measure, it can be substituted by the forward's electricity prices available at that moment. In particular, following the notation used in the literature, as for example in Benth and Koekebakker [2008], a future contract of power is expressed in that way: $\mathbb{E}_{t_0}^{\mathbb{Q}}[\sum_{i=1}^{n} P(t_i)] = \sum_{i=1}^{n} F(t_0, t_i, t_{i+1})$. As electricity is a commodity that cannot be storable, these contracts involve that future prices are not for a specified moment. Instead, the price includes an interval after the contract's maturity (the interval can be hours, days, months). For that reason, future contracts on energy markets can be considered swaps as the exchange of power is agreed upon for an interval of time. That is why in the expression below, there are three dates. It represents the day of observation of the future price of a contract with delivery over the last two dates. Additionally, attending to non-arbitrage conditions, a SWAP with delivery over $[t_1, t_n]$ and N - 1 swaps with deliveries over the disjoint periods $[t_i, t_{i+1}]$ with i = 1,2,..., N-1 and the union of all the periods equals $[t_1, t_n]$ must have the same value. Expressed formally: $\frac{1}{t_n - t_1} \sum_{i=1}^{n-1} F(t_0, t_i, t_{i+1}) = F(t_0, t_1, t_n)$. Now, it is easy to see that float leg could be expressed as the price of the forward $F(t_0, t_1, t_n)$. In particular, it is assumed that the forward involves exchanges of power with the same frequency as the PPA. This means that if the PPA price is for the exchange of the quantity agreed each day of the period, the forward needed, also will need to be for daily exchanges. In that way, the discounting factors can be erased, and it is easy to the that the final equation is:

$$P_{PPA} = F(t_0, t_1, t_n) \tag{4}$$

So, the best method to calculate the PPA price would be if a future contract for the exchange of power during the interval of t_1 and t_n existed. Nevertheless, this is not going to happen in most of the electricity markets. That is why, as this data is not available, an approximation may be used with the existing future contracts. In this regard, interpolation and extrapolation can be used to estimate the prices of contracts with lower intervals. In figure 28, a scheme exemplifies one of the possible procedures used with the existing contracts on CAISO's market (that coincides with the one used with the West-non-ISO market, as the future prices available have the same characteristics in both markets).

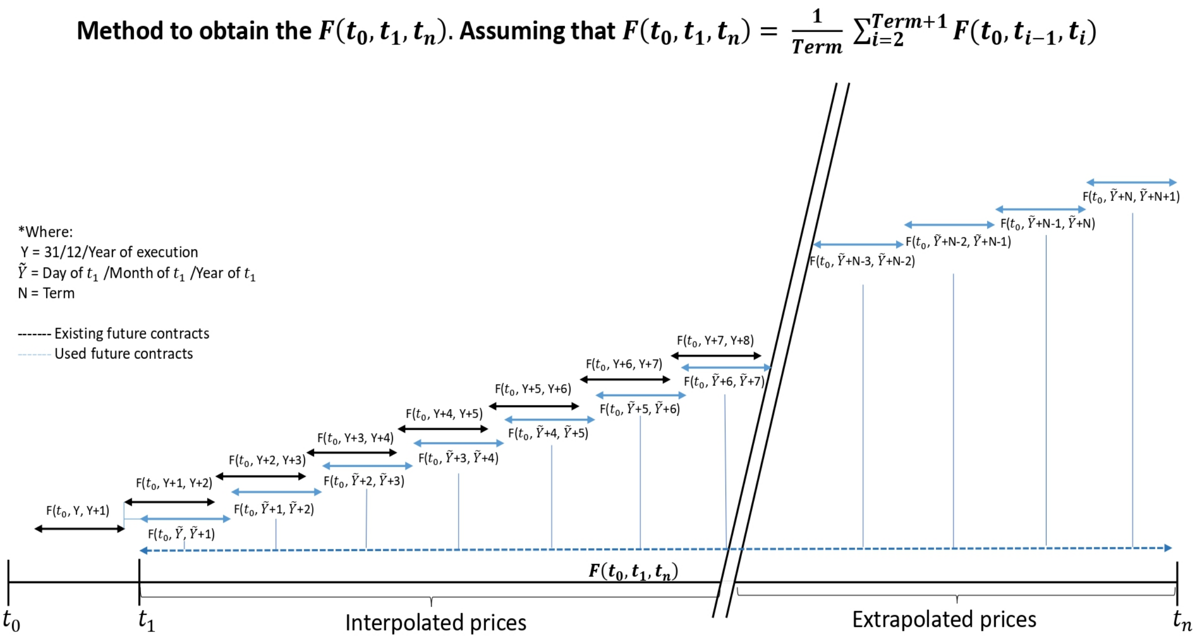

Figure 28: Estimation of $F(t_0, t_1, t_n)$

The black lines represent the existing contracts in that example. Meanwhile, the blue lines represent the concatenated contracts for one year whose prices are interpolated or extrapolated. The extrapolation method becomes relevant as the PPA terms are usually longer than 15 years, so more than half of the future yearly contracts will be extrapolated. For that purpose, different extrapolations and interpolations have been tested. First of all, between the different interpolations tried, the output differs insignificantly. Secondly, to extrapolate, the possibilities considered are mainly two. It turns out to be crucial to determine a long-term future price. Otherwise, the procedure can result in "exploding" extrapolations with extreme future prices (really high or low prices outputs).

The options considered [4] are the price of the last maturity future contract available or the average of

[4] More complex options exist to estimate the long-term future price. For example, to model historical data with an

the last eight/seven years [5] electricity spot prices in the market analyzed. In that way, the procedure has become an interpolation where there are eight future yearly contracts and a long-term price at each time. Different methods have also been tried showing that "spline", "makima", or others more complex have shown strange behaviors. At the same time, the linear interpolation outputs seem to be the more reasonable one [6]. Then, comparing the two long-term price possibilities considered, the output is very similar. On the basis of the argument behind each one, the first one assumes that the best estimate of the long-term price with the available information is the last future price available, and the second one relies on the fact that history is the best estimator of the future. With this in mind, the average last eight/seven years' spot price could be less biased by temporary external factors than the other estimate. Even so, both methods are applied and present to see different behaviors of the model variants.

To sum up, to price each PPA, it is needed the forward's electricity curve [7], and the contract information. In this case, the information needed is the term and the date of execution of the contract. As it can be observed, the capacity is not used.

5.2 Results

Model 1 proposed before is tested now with the markets already analyzed in detail. In this way, nearly half of the PPA contracts available in the database will be evaluated with this model.

5.2.1 CAISO results

First of all, there is an explanation of the pricing model variations considered. Actually, the models are pretty the same, but they differ in two main things:

- Price series: The SP-15 price series is used in the first two cases and the mixed SP-15 and NP-15 (weighted by their volumes) in the other two.
- Extrapolation method: As explained, the methods contemplated are the latest final maturity future price available or the long-term price as the average of the eight last years. The long-term price corresponds to the second and fourth models.

The CAISO region has the biggest sample of solar PPA contracts in the database and a large enough sample of wind PPA contracts. Using the contracts information and applying model 1, the following results have been obtained. First, in figure 29 the solar PPA are priced with model 1 but using the different variations explained. Then, the output is shown in the figure, where an apparent underestimation can be remarked. The same occurs for the wind PPA, whose results are in figure 30.

ARMA specification. Nevertheless, it does not seem to involve a significant improvement compared with the requirements implied (it will be needed to model the last spot prices for each PPA execution date). The objective of this research is not to model electricity prices at all. For that reason, simpler estimators are considered

[5] The choice of the last eight/seven years is justified as is the largest sample available previous to the first PPA execution dates.

[6] The "interp1" function of the software Matlab has been used for that aim. The software also has a cubic interpolation, but the results are very similar to those of the linear interpolation.

[7] In particular, as explained before, the future and spot electricity price series are two possible options: only using the SP-15 data as a proxy for the electricity prices of this region, or a weighted series between SP-15 and NP-15 data.

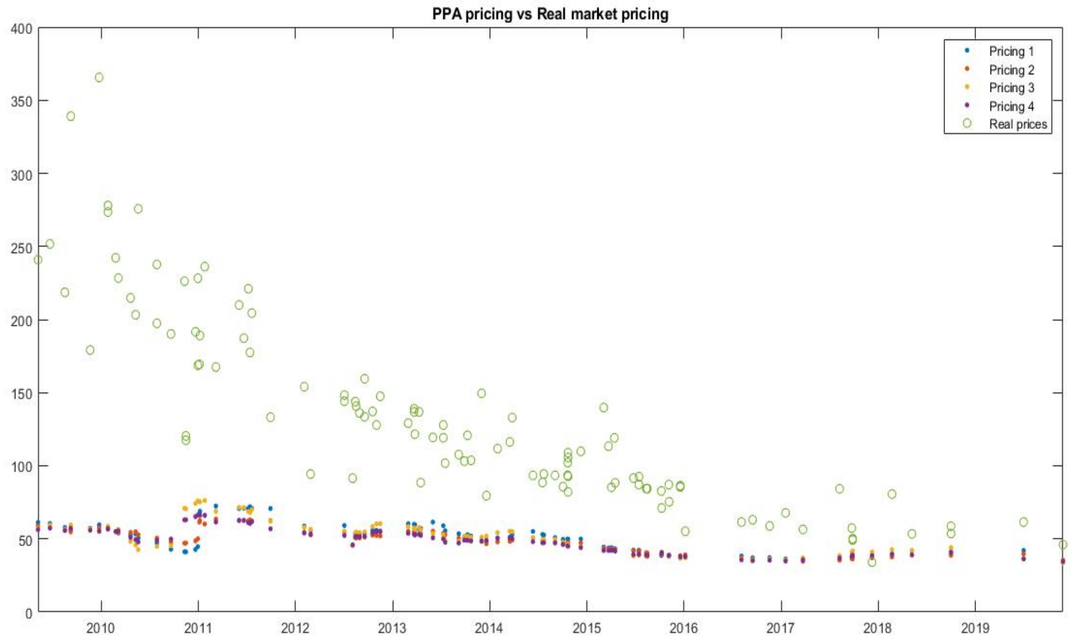

Figure 29: Market PPA pricing compared with the pricing using model 1. Solar CAISO data.

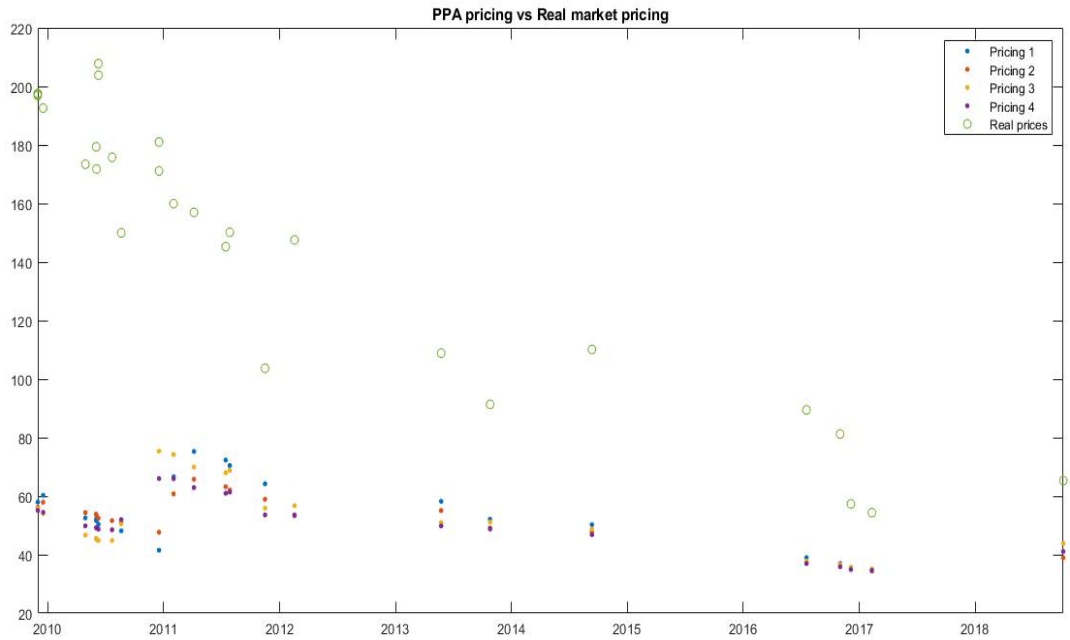

Figure 30: Market PPA pricing compared with the pricing using model 1. Wind CAISO data.

After the results, an analysis of these results is done. For that aim, different measures of forecast errors are presented. The explanation of them is in annex 3 on page 81.

Error measure \Model	Pricing 1	Pricing 2	Pricing 3	Pricing 4
Mean Squared Error Root	108.04	109.41	107.31	109.09
Mean Absolute Error	85.32	87.22	84.92	87.12
BIC	504.53	504.07	503.40	503.40
AIC	501.75	501.29	500.62	500.62
Correlation (Market vs Model)	0.5681	0.7105	0.5703	0.6930

Table 12: Measures of the errors on the solar CAISO data

As seen in table 12 the MSER and MAE are better using SP-15 data prices only and the same if the extrapolation used is the second one. However, attending to the information criterion statistics, the second extrapolation reflects better or equal results. The SP-15 price data has worse results than mixed prices data in the BIC and AIC cases. Concerning the correlation coefficient, whose justification is explained in the annex 3, the second extrapolation method captures better the trend of the pricing in the market. Meanwhile, between the CAISO or SP-15 data, the results are not the same clear.

Error measure \Model	Pricing 1	Pricing 2	Pricing 3	Pricing 4
Mean Squared Error Root	109.19	109.58	109.69	110.09
Mean Absolute Error	99.49	100.63	99.89	101.07
BIC	121.94	120.80	122.16	121.00
AIC	120.51	119.36	120.73	119.57
Correlation (Market vs Model)	0.3945	0.6396	0.3519	0.5820

Table 13: Measures of the errors on the wind CAISO data

In table 13 the SP-15 data gets lower MSER and MAEs. Furthermore, the first extrapolation method returns higher values too. As in the previous case, with the BIC and AIC, the second extrapolation would be preferred. Nevertheless, in this case, SP-15 also seems better attending to AIC and BIC, opposite with the solar case. The correlations are lower than in solar data and have higher values if the extrapolation method uses the historical average. In addition, in this case, the SP-15 data detects better the movements attending to the results.

Despite this brief analysis, the main conclusions are that the models have significant deviations from the market prices. In addition, the performance of each model does not seem to be very different among them (the differences are minimal in all the measures).

Loss Function \DM p-value	Pricing 1	Pricing 2	Pricing 3	Pricing 4
First Loss Function	0.0000	0.0000	0.0000	0.0000
Second Loss Function	0.0081	0.0082	0.0097	0.0088
Third Loss Function	0.0000	0.0000	0.0000	0.0000

Table 14: Solar Diebold Mariano test p-values (CAISO)

Loss Function \DM p-value	Pricing 1	Pricing 2	Pricing 3	Pricing 4
First Loss Function	0.0000	0.0000	0.0000	0.0000
Second Loss Function	0.0019	0.0014	0.0013	0.0011
Third Loss Function	0.0000	0.0000	0.0000	0.0000

Table 15: Wind Diebold Mariano test p-values (CAISO)

Looking at tables 14 and 15 the expectations are confirmed by the DM test. As the output shows, there is enough evidence to reject the null hypothesis in all the cases. This result would mean that the losses caused by the forecast errors are statistically different from zero in the four models and both wind and solar data (and the loss functions do not matter a lot as the three of them conclude the same results).

Lastly, the forecast errors are studied by years. The objective of this is to explore the changes over time of the model estimates versus the market prices. Furthermore, the relationship between that errors and the size and term is computed too. Specifically, the forecast errors computed are the ones of the pricing 4, as it uses the mixed prices of the two available hubs information and the average of the last eight years for the long-term price. So, the results are presented in tables 16 and 17.

The evidence reflects an important underestimation in the first years of the sample. In particular, solar data has high values at the beginning and these have become moderate in comparison since 2013. Attending to the correlations, the size does not have a significant linear relationship except in the year 2011 where a high negative correlation is found. Meanwhile, the duration of the PPA contract seems to be significant more often, finding that the full sample has a positive significant relationship as well as 2010 and 2014 when strong relationships are observed. Secondly, attending to wind data, only a few years allow computing the correlations. Nonetheless, the whole sample reflects significant positive relationships for both variables. Besides, in 2010 exists a very high correlation between the size and the forecast errors and in 2009, term and forecast errors have a significant relationship almost perfectly positive.

Compared with the results of the previous chapter, the correlations are very similar to the ones found for the values of the market prices contrasted with the two variables. Overall, the wind data still shows strong positive relationships and the solar data for the case of the term. That fact could help to find other influential factors in the pricing since the results show the persistence of the relationships previously detected.

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	87.05	234.7	159.0	126.9	82.9	66.5	53.3	51.8	23.6	19.6	21.6	18.4
2	0.0849 (0.35)	-0.301 (0.46)	-0.049 (0.85)	-0.742 (.01)**	0.2859 (0.34)	-0.5 (0.04)	-0.069 (0.77)	-0.383 (0.11)	0.5447 (0.45)	0.6249 (0.13)	-0.448 (0.55)	-
3	0.2711 (.00)**	-0.012 (0.97)	0.6723 (.00)**	0.5785 (0.08)	0.5538 (0.05)	0.533 (0.03)	0.7153 (.00)**	0.1742 (0.49)	-0.971 (0.03)	-0.012 (0.97)	0.4373 (0.56)	-

Table 16: Analysis of the Solar forecast errors by years in CAISO. 1: Average forecast error.
2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	101.1	141.5	131.8	82.2	94.0	50.8	63.3	-	40.1	19.8	24.2
2	0.5758 (0.000)**	0.4511 (0.37)	0.8151 (0.002)**	0.7476 (0.14)	-	-	-	-	0.6844 (0.52)	-	-
2	0.6587 (0.000)**	0.9836 (0.000)**	0.5624 (0.07)	0.7378 (0.15)	-	-	-	-	0.8136 (0.39)	-	-

Table 17: Analysis of the Wind forecast errors by years in CAISO. 1: Average forecast error.
2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

5.2.2 West-non-ISO results

In this case, as explained before, only one price series is used. So, there are two pricing outputs: one extrapolating with the latest final maturity future price available, the "Pricing 1", and the other with the long-term price calculated as the average of the seven last years, the "Pricing 2" (the sample is a little shorter than with CAISO data, so, to be able to price PPA of 2009 the average should be done with the last seven years instead of eight).

This market has more or less the same amount of wind and solar PPA contracts (around one hundred each). Nevertheless, many wind contracts are from dates earlier than 2009. So, as 2009 is the first date where future yearly prices are available, the model is only tested in approximately half of the wind contracts (this also occurs in CAISO data, but the percentage of discarded PPA contracts is smaller). When model 1 is applied, the results obtained are shown in figures 31 and 32 . Three series are represented on the graph: the real market pricing, the pricing with the last future price available extrapolation, and the pricing extrapolating with the historical average price. Similar to the previous market results, there is an underestimation in both types of PPA. This repeated fact reveals the existence of other factors which influence the pricing not being taken into account.

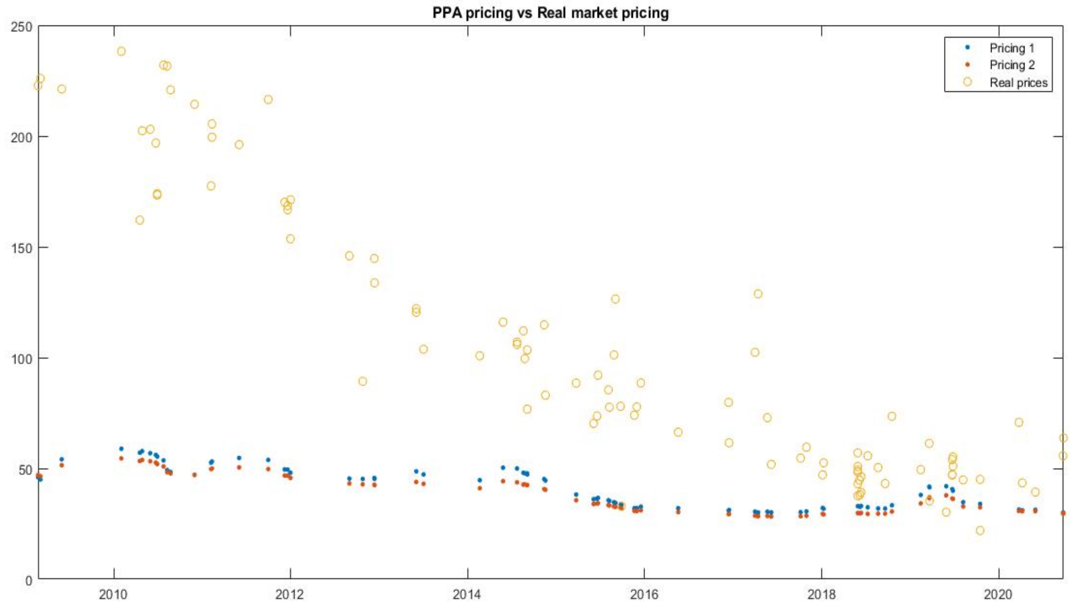

Figure 31: Market PPA pricing compared with the pricing using model 1. Solar West-non-ISO data.

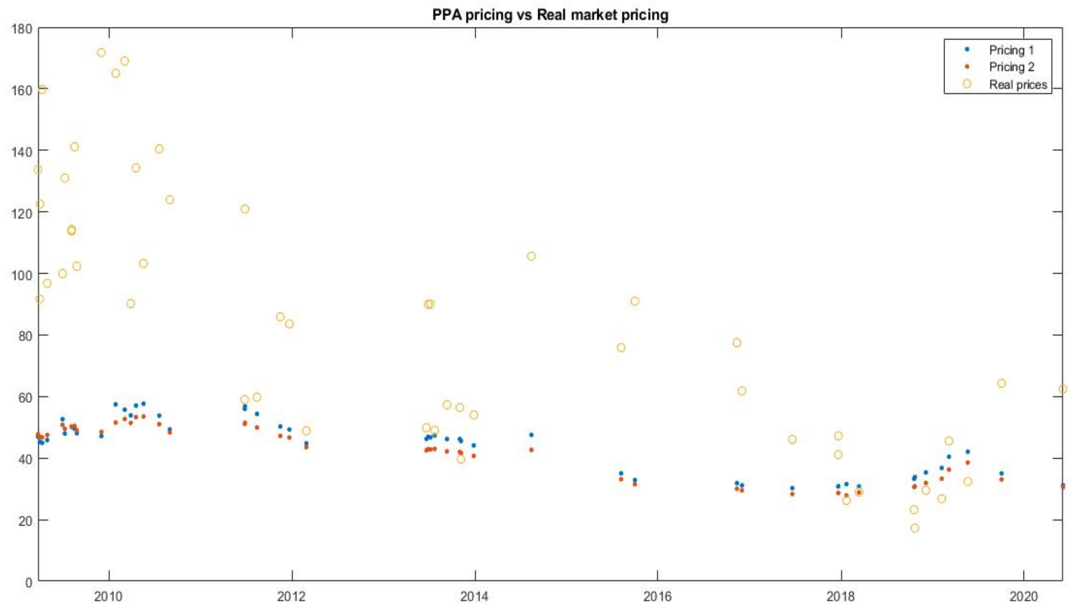

Figure 32: Market PPA pricing compared with the pricing using model 1. Wind West-non-ISO data.

To quantify the errors, the same procedure as with the CAISO market is done. The different measures of the forecast errors proposed are presented in order to be able to evaluate the output better.

Error measure \Model	Pricing 1	Pricing 2
Mean Squared Error Root	82.05	84.07
Mean Absolute Error	61.99	64.55
BIC	400.86	400.65
AIC	398.27	398.06
Correlation (Market vs Model)	0.8124	0.8718

Table 18: Measures of the errors on the solar West-non-ISO data

Table 18 shows that the MSER and MAE are higher when the extrapolation with the average of the last seven years is used. However, the information criteria reflect lower values for the mentioned extrapolation method. As it can be seen, the correlations in this market are stronger, with values very near to 1. Indeed, the second extrapolation method reflects a coefficient of almost 0.9.

Error measure \Model	Pricing 1	Pricing 2
Mean Squared Error Root	54.56	55.37
Mean Absolute Error	42.63	43.60
BIC	203.45	201.93
AIC	201.44	199.92
Correlation (Market vs Model)	0.6459	0.7468

Table 19: Measures of the errors on the wind West-non-ISO data

Similarly, table 19 manifests results in the same direction as with solar data. The errors measured with the MSER and MAE are shorter if the extrapolation with the latest future price available is applied, and the opposite occurs with the information criteria values. The correlation coefficients for wind data are lower but still high, compared to the other market. Moreover, better results are found for the second extrapolation method.

Loss Function \DM p-value	Pricing 1	Pricing 2
First Loss Function	0.0000	0.0000
Second Loss Function	0.0001	0.0001
Third Loss Function	0.0000	0.0000

Table 20: Solar Diebold Mariano test p-values (West-non-ISO)

Loss Function \DM p-value	Pricing 1	Pricing 2
First Loss Function	0.0000	0.0000
Second Loss Function	0.0000	0.0000
Third Loss Function	0.0000	0.0000

Table 21: Wind Diebold Mariano test p-values (West-non-ISO)

Lastly, tables 20 and 21 show that the null hypothesis of the losses due to forecast errors being not statistically different from zero is rejected in all the cases firmly.

Consequently, the conclusion does not differ a lot from the conclusion of CAISO data. The model has significant errors in both cases, and the differences between the extrapolation methods are not deep enough to confirm the best performance of one of them. The most relevant difference seems to be in the correlation, as with the information available, the trend of the prices agreed for the PPA has been captured better for the West-non-ISO market.

Identically to the CAISO analysis, the study of the forecast errors by years is presented below in tables 22 and 23. In this case, as previously seen in the graph, the solar forecast errors are lower but the trend is similar. However, wind data differs in the trend as in 2012 the forecast average error is very low, then it experiments an increase of two years and another decrease even with a negative value in 2018.

Solar forecast errors seem to have a significant negative relationship with the size when the full sample is used. In contrast, the term shows a significant correlation neither in the full sample nor in specific years. For the wind case, the results are almost equal to the ones of the previous chapter. The size reflects a negative significant correlation in the full sample but in specific years not. And the duration of the contract has an exception as in 2009 a positive significant relationship is detected. So, in the West-non-ISO case, the results also are close to the ones of the previous chapter except that the wind data now shows a significant relationship at the confidence level of 1% (and in the previous chapter was at the confidence level of 5%).

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	64.14	175.0	153.0	138.8	96.0	71.8	59.5	45.7	39.5	46.5	18.9	9.8	24.3
2	-0.323 (.00)**	0.049 (0.96)	0.401 (0.22)	-0.187 (0.66)	0.081 (0.88)	-0.716 (0.49)	0.038 (0.92)	-0.091 (0.76)	-0.942 (0.22)	-0.112 (0.81)	-0.415 (0.10)	0.294 (0.35)	0.606 (0.28)
3	0.184 (0.07)	-0.7934 (0.42)	0.365 (0.27)	-0.358 (0.38)	0.677 (0.14)	0	0.610 (0.06)	0.38 (0.18)	0	-0.304 (0.50)	0.289 (0.26)	0.628 (0.03)	0.905 (0.03)

Table 22: Analysis of the Solar forecast errors by years in West-non-ISO. 1: Average forecast error. 2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	41.9	73.5	80.6	32.6	5.3	18.5	62.9	51.1	39.9	16.2	-5.4	7.0
2	-0.3763 (.00)**	0.1753 (0.53)	-0.6948 (0.08)	-0.6062 (0.28)	-	-0.3867 (0.34)	-	-	-	-0.9966 (0.05)	-0.8048 (0.05)	0.9468 (0.05)
3	-0.0204 (0.18)	0.2557 (.00)**	-0.1934 (0.42)	0.4319 (0.55)	-	-0.3663 (0.37)	-	-	-	0.9924 (0.07)	-0.5073 (0.30)	0.4918 (0.51)

Table 23: Analysis of the Wind forecast errors by years in West-non-ISO. 1: Average forecast error. 2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

In general, the correlations suggest that PPA with higher terms or larger size have higher forecast errors. That could be related to the fact that more specific terms could be defined in this type of contracts turning more difficult to price a PPA with a standard model.

6 Second model: Volumetric and correlation risk price correction in CAISO

6.1 Context

Not only exists price risk in power but also volumetric risk and correlation risk are crucial factors to manage in renewable energy projects. Volumetric risk is related to the uncertainty on production. Meanwhile, the correlation risk concerns the relationship between power generation and prices. Specifically, it is when higher production volumes coincide with low prices. Indeed, if the importance of renewable production in a market increase, this event is more likely to occur as this type of generation is expected to lower prices. So, a negative relationship can be expected.
In particular, renewable energies have irregular productions dependent mainly on weather conditions. Furthermore, as stated before in section 2.2, these technologies are non-dispatchable. That is why markets where renewable power generation is increasing its importance are experiencing structural changes due to the complexity to plan production. These alterations have lead to changes in prices too. All this, coupled with the fact that renewable energies are increasing their weight in many electricity markets, makes that risk modelling fundamental.

Regarding PPA, these risks are very important since price risk is hedged with this instrument, but these others not. For that reason, a correction of the price can be calculated referred to this element. The correction will usually result in a lower fixed price agreed for the exchanges. The justification lies in the fact that the buyer of energy will require a better price for the energy since it is exposed to uncertainty in the already mentioned events. Nevertheless, this argument is based on a negative correlation, but the price correction would be added to the forward price if the correlation was positive. All this assuming that PPA have a payment structure known as "pay-as-produced" previously defined in the preliminaries.

That topic has been covered in different researches. Their objectives have mainly been to study the relationship between price and renewable production in different markets and propose hedging strategies or instruments. Namely, this research follows an approach with ideas from Pircalabu et al. [2017] and Kaufmann et al. [2020]. In both cases, the authors model the price and generation series in different ways. After that, using a copula approach, they fit the standardized residuals of the marginal series to different types of copulas to replicate the relationship between both variables. Similar to the first investigation, Tranberg et al. [2020] propose a different way to realize the procedure made in Pircalabu et al. [2017]. Moreover, the copula approach can be applied with three variables as Pircalabu and Jung [2017] developed with a vine-C copula.

This research will apply this methodology to price the mentioned risk correction that each PPA should incorporate in its pricing. For that purpose, there is data available of solar and wind generation between 10/04/2018 and 31/03/2021. Altogether the sample has 1081 observations (five days are not available from the source). For a more rigorous correction, each PPA price correction should be calculated with the generation data of the dates before the day of pricing. Thus, if the pricing is done at the PPA execution date, the correction at this time could only be done with historical data. Nevertheless, this fact will be omitted justified by two arguments. First, there is no data available before 2018 of power generation in CAISO. Second, even if this information is available, the procedure seems to be very large computationally as the method will require to be replicated for each PPA with a different sample. So, keeping in mind the limitations of these assumptions, the procedure can be interpreted as the existing way to estimate in this market the volumetric and correlation risks but losing some accuracy.

As already mentioned, the analysis will be done for both solar and wind generation technologies. In the literature, the most common technology to apply this procedure is wind. The reason usually comes from the more irregular generation profile compared with the solar generation profile. The irregularity makes more likely to find a negative correlation between generation and prices. Nevertheless, in this research, the approach is followed for both technologies. That is because solar generation has the highest importance in renewable generation and since the number of contracts is larger for solar PPA. In addition, if the results reflect an absence or small values for the price correction, that would also be very relevant.

6.2 Methodology

A detailed formalization of the procedure followed is going to be presented. The first part of the method, the modelling of the series, is based on the approach followed by Kaufmann et al. [2020]. Specifically, section 6.2.2 closely follows the steps done in the cited article. Then, a copula approach is used to fit the modelled data. After that, the model proposed to price the volumetric and correlation risk correction by Pircalabu et al. [2017] is used with some adjustments. In particular, the equations of section 6.2.6 stem from this mentioned article. Furthermore, the goodness of fit measures for marginal models as well as for copulas appear in both works (and in related literature). Likewise, there are some other measures to extend the ones of these two investigations

6.2.1 Previous processing of the data

In figures 33, 34 and 35 it can be observed the series that are going to be used to model the solar or wind volumetric and correlation risks. Moreover, in figure 36 the marginal distributions and joint distribution of solar generation and prices are presented. Similarly, in figure 37 wind generation and price distributions are shown.

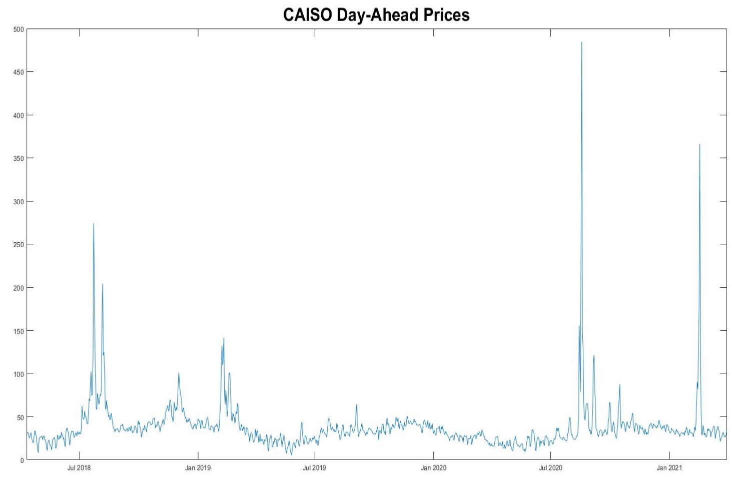

Figure 33: Price series sample used for the volumetric and correlation risk modelling

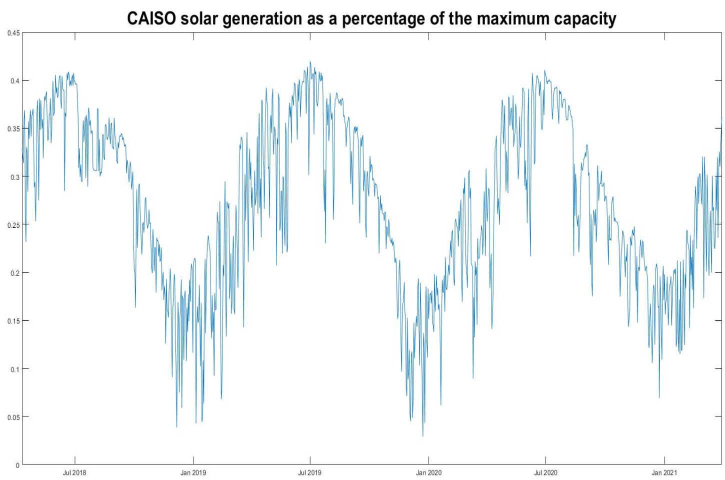

Figure 34: Solar generation expressed as the infeed factor

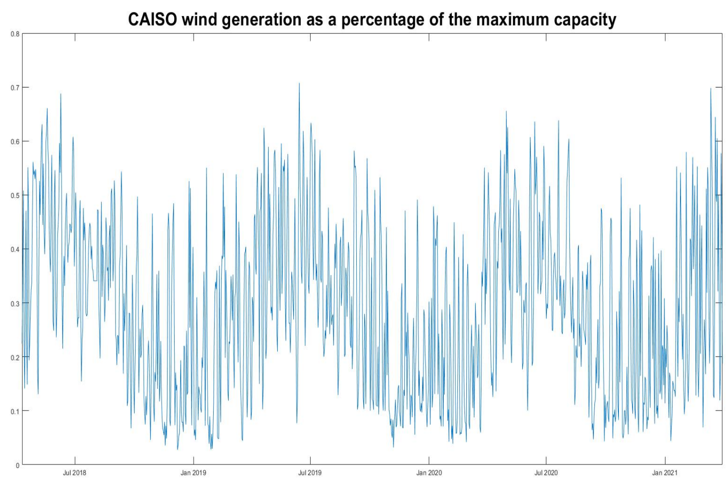

Figure 35: Wind generation expressed as the infeed factor

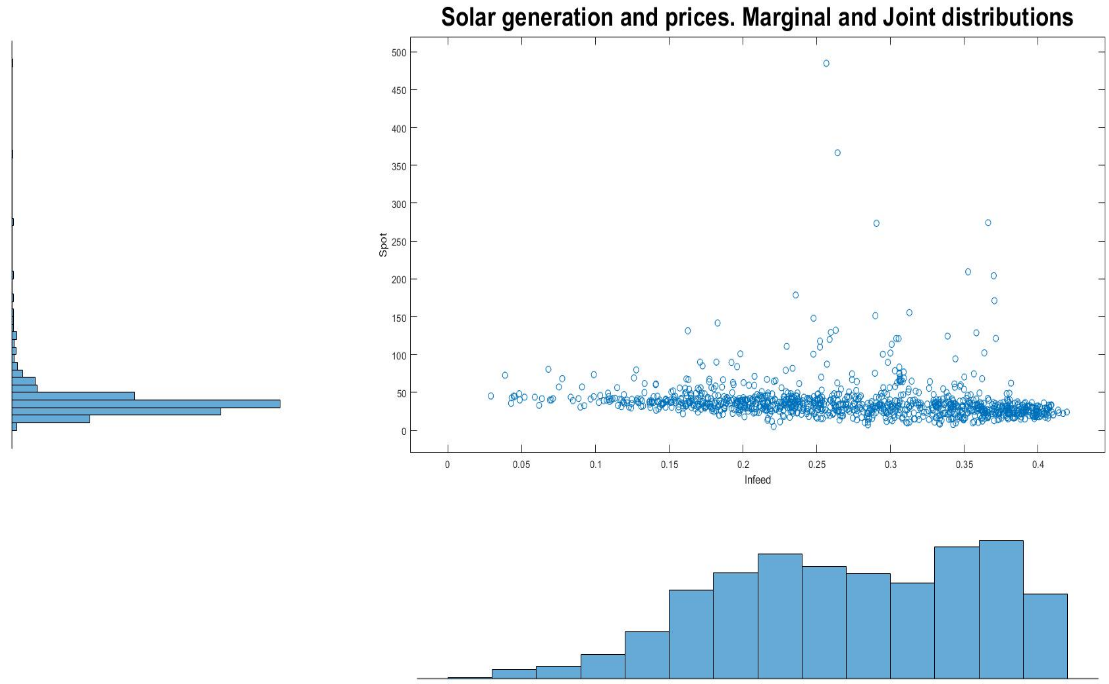

Figure 36: Scatter-plot of solar generation and prices and their marginal distributions

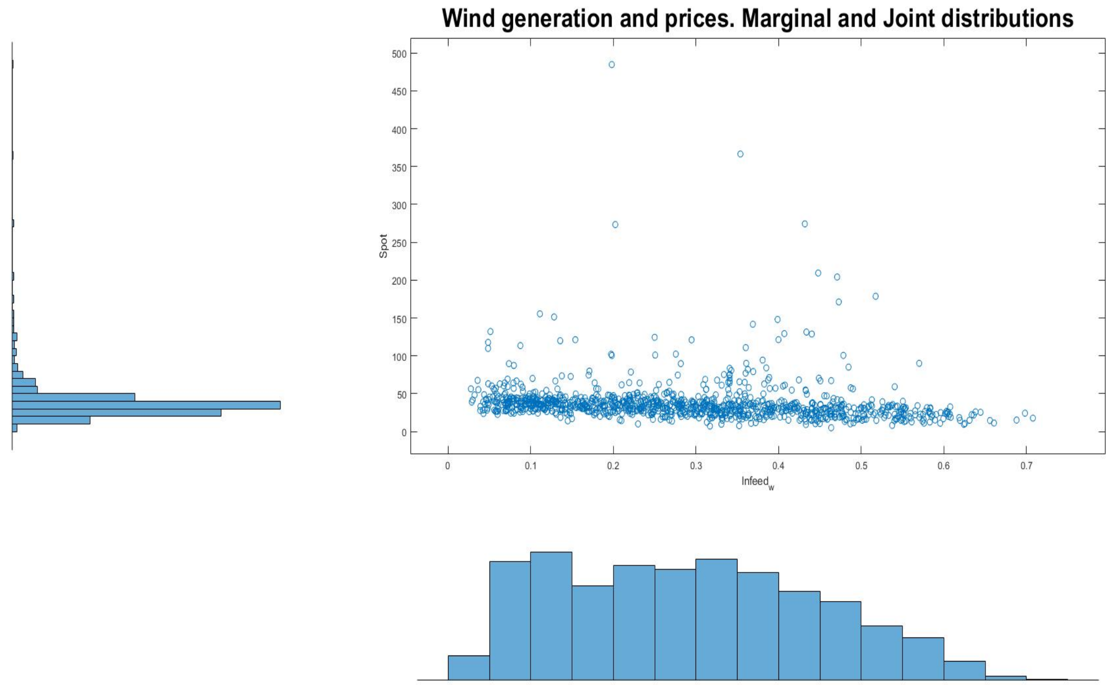

Figure 37: Scatter-plot of wind generation and prices and their marginal distributions

Now, different transformations and models applied are explained in detail. First of all, the generation series is transformed in two stages. First, it is expressed as the value of the total installed capacity, commonly known in the literature as the infeed factor. Second, as the domain of the infeed factor is between 0 and 1, a logistic transformation is applied to obtain values along the entire real line. Formally, the logistic transformation is:

$$F(x) = ln\left(\frac{x}{1-x}\right) \tag{5}$$

In Kaufmann et al. [2020] the authors apply an outlier treatment before modelling the marginal price series. However, this research differs from this approach, as the prices observed are considered values caused in the market. Thus, it only seems necessary to treat any value if exogenous factors interfere with that value. So, the point of view is that the PPA valuation will be biased if this treatment is done, and the objective will be to find suitable distributions which fit well with extreme observations.

6.2.2 Marginal models

After that, the different resources used for the marginal models are presented. The aim is to account for seasonality and other factors in the data separating these effects from the series. For the case of the prices, denominating P_t the day-ahead price, the proposed model has short and long-term seasonal components and a stochastic component:

$$P_t = STSC_t + LTSC_t + X_t \tag{6}$$

Where each component tries to capture different patterns. The LTSC is a sinusoidal function for yearly periods to capture the existent pattern along the year (the different levels of the prices each period of the year). The use of sinusoidal functions has already been used in similar analyses showing a good performance. In particular, the parameters are the following ones:

$$LTSC_t = A_0 + A_1 \cdot sin\left[2\pi(\frac{t}{365.25})\right] + A_2 \cdot cos\left[2\pi(\frac{t}{365.25})\right] \tag{7}$$

*Where A_1 and A_2 determine the phase and amplitude of the function and A_0 the mean.

Secondly, STSC is intended for capturing daily patterns as differences in prices can be expected depending on the day of the week. For that objective, dummy variables of each day of the week are defined:

$$STSC_t = d_1 \cdot \mathbb{1}_{\{Day(P_t)=Monday\}} + ... + d_7 \cdot \mathbb{1}_{\{Day(P_t)=Sunday\}} \tag{8}$$

*Where $d_1, d_2, ..., d_7$ will be the coefficients assigned with each dummy.

Lastly, to estimate the marginal distribution, the stochastic element is a model for the mean and variance (ARMA-GARCH). The model chosen would be the one that returns an error series independent and identically distributed $n_t \sim Distribution(0, \sigma^2)$ (It is usually the Normal distribution). Formally this is:

$$X_t = k + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{9}$$

$$\epsilon_t = n_t \cdot \sigma_t \tag{10}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 \tag{11}$$

Also, to consider possible asymmetries, a leverage term is used. In particular, is the specification of a GJR-GARCH model considering only one lag for the leverage parameter:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \gamma \mathbb{1}_{\{\epsilon_{t-1}<0\}} \tag{12}$$

With the following restrictions for the parameters:

- The autoregresive, ϕ_i, and moving average, θ_i, parameters should satisfy the stationarity and invertibility conditions.
- Likewise, to ensure stationarity, the GARCH parameters should satisfy: $\omega > 0$, α_i and $\beta_i \geq 0$ and $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j + \frac{\gamma}{2} < 1$

If the model is well-specified, the variable n_t will be white noise for all t. Moreover, transformed using the cumulative distribution function of the fitted distribution, it should return a uniform distribution.

For the infeed transformed factor, the model estimated is the same except for the short-term component that is not used since different generation patterns between the days of the week are not expected.

6.2.3 Goodness of fit models

For this part of the procedure, different evaluations are used to choose the best possible model. First, to chose the lag order of the ARMA and GARCH parameters, the BIC or AIC can be utilized. Second, to test the goodness of fit, the simple and partial autocorrelation functions should reflect no significant values. Specifically, this can be tested with the Ljung-Box test of serial independence. The statistic is the following one:

$$Q = N(N+2) \sum_{k=1}^{h} \frac{\hat{\rho_k^2}}{N-K} \tag{13}$$

*Where the null hypothesis is that the data is independently distributed. Furthermore, N is the sample size, $\hat{\rho_k^2}$ the sample autocorrelation at lag k and h the last lag for which the test is calculated.

Third, the Kolmogorov-Smirnov and Cramer-von-Misses tests are computed too. These are used to see the goodness of fit of the residuals from the estimated distributions. In other words, the standardized residuals need to be fitted to a specified distribution. Additionally, these two tests compare if two series can be considered having the same distribution. That is why the empirical distribution of the data is compared with a specified distribution using the mentioned tests. The null hypothesis is $H_0 : D = D_0$, where D is the sample distribution and D_0 the known distribution. In a formal way, the statistics are:

$$KS_i = \underset{t}{\text{Max}} |U_t - \hat{U}_t| \tag{14}$$

$$CM_i = \sum_{t=1}^{T} (U_t - \hat{U}_t)^2 \tag{15}$$

*where U_t is the fitted parametric distribution and $\hat{U}_t$ the empirical CDF.

Moreover, after modelling the marginal models, two other verification can be done. With the chosen distribution F, using its cumulative distribution function, the probabilities of the series can be obtained ($F(y_t)$). These values could be interpreted as uniform numbers and are very useful to see if all the procedure is well-specified. First, the resulting series can be compared to the standard uniform distribution with the already presented KS and CvS tests. Second, Berkowitz [2001] proposed a more robust test where the null hypothesis is that the probabilities are a sequence of random numbers. For that aim, the uniform numbers are transformed into normal numbers with the inverse of the CDF of the normal standard distribution since tests for the uniform distribution are not as straightforward as tests on the normal distribution. So, being $Z_t = \Phi^{-1}(F(y_t))$ the null hypothesis would be now that Z_t is identically and independently distributed as a N(0,1) against the alternative of Z_t being identically and independently distributed as a $N(\mu, \sigma^2)$. Then, the test is based on the likelihood ratio statistic:

$$-2ln(LR) = -ln\left(\frac{L_0}{L_1}\right) \sim \chi_2^2 \tag{16}$$

That can be expressed as:

$$-2ln(LR) = \sum_{t=1}^{N} Z_t^2 - \sum_{t=1}^{N}\left(\frac{Z_t - \hat{\mu}}{\hat{\sigma}}\right)^2 - N \cdot ln(\hat{\sigma}^2) \tag{17}$$

This last verification can also be shown with the graph of the probability integral transformation, where it could be checked if all the intervals have similar frequencies.

6.2.4 Copula approach for modelling the dependence

The copula model is now applied. These instruments allow modelling the relationship between variables when the relationship is certainly complex due to non-linear dependences or asymmetric relationships. Some previous definitions are presented below:

A multivariate copula is a distribution function defined on the $[0,1]^d$ plane with marginal uniform distributions. For the particular case of this analysis, the bivariate copula has the following properties:

It acts on the range of the values taken by both distribution functions:

$$C : [0,1] \times [0,1] \to [0,1]$$

The marginal uniform distributions should satisfy:

$$C(u_1, 1) = u_1; C(1, u_2) = u_2$$

And also two conditions from every bivariate distribution function:

$$C(u_1, 0) = 0; C(0, u_2) = 0$$

$$for\ \ all\ \ u_1, u_2, v_1, v_2\ \ in\ \ [0,1]\ \ with\ \ u_1 \leq v_1, u_2 \leq v_2 :\ \ C(v_1, v_2) - C(u_1, v_2) \geq C(v_1, u_2) - C(u_1, u_2)$$

*Non-decreasing in every component

In addition, with the assumption of differentiability of the marginal distributions, the copula can be expressed as:

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))$$

Nevertheless, this analysis works with conditional distributions. So, an extension of Sklar's theorem is applied. Having a joint conditional distribution function $F(\Theta|\mathcal{F}_{t-1})$ and the correspondent marginal distribution functions $F_1(\Theta|\mathcal{F}_{t-1})$ and $F_2(\Theta|\mathcal{F}_{t-1})$. The conditional copula is defined in the following way:

$$F((x_1, x_2|\mathcal{F}_{t-1}) = C(F_1(x_1|\mathcal{F}_{t-1}), F_2(x_2|\mathcal{F}_{t-1})|\mathcal{F}_{t-1})$$

And if the marginal distributions are continuous, the copula is unique: $U_t|\mathcal{F}_{t-1} \sim C(\Theta|\mathcal{F}_{t-1})$

After these previous statements, a brief description of the copulas to be used will come in the following paragraphs.

Copulas can be divided into two big groups. The first group involves the ones based on elliptical distributions. The most common elliptical or implicit copulas are the Gaussian and Student's t distribution. Their formal expressions are:

$$Gaussian = C_\rho(u_1, u_2) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{18}$$

*Where ρ is the correlation, Σ the correlation matrix, and Φ the cumulative distribution function of a Normal standard

$$Student \;\; t = C_{\nu,\Sigma}(u_1, u_2) = t_{\nu,\Sigma}\Sigma(t^{-1}(u_1), t^{-1}(u_2)) \tag{19}$$

*Where Σ is the correlation matrix, t the cumulative distribution function of the student t with ν degrees of freedom.

The second group involves the Archimedean copulas, which use the generator functions explicitly defined (so, these are explicit copulas). Two of the best-known ones are the Clayton and Gumbel copulas.

$$Gumbel = C_\theta(u_1, u_2) = exp\left(-[(-ln(u_1)^\theta) + (-ln(u_2)^\theta)]\right)^{\frac{1}{\theta}} \tag{20}$$

$$Clayton = C_\alpha(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-\frac{1}{\alpha}} \tag{21}$$

To fit these copulas to the data the maximum log-likelihood method is going to be used. In other words, the objective is to find the values of the parameters that return the higher value of the logarithm of the likelihood function. Formally, defining α_i as the vector of parameters of the marginal distributions, θ as the parameters of the copula distribution and x_i the vector or variables:

$$lnL(\alpha, \theta, x_1, x_2) = \sum_{t=1}^{T} lnc(F_1(x_1; \alpha_1), F_2(x_2; \alpha_2); \theta) + \sum_{t=1}^{T} lnf_1(x_{1t}; \alpha_1) + \sum_{t=1}^{T} lnf_2(x_{2t}; \alpha_2) \tag{22}$$

From this equation, it is easy to see that an estimation in two stages can be done. First, the parameters of the marginal density can be calculated and then the parameters of the copula using the already estimated marginal parameters $\hat{\alpha_i}$:

$$\operatorname*{Max}_{\alpha_i} \sum_{t=1}^{T} lnf_i(x_{it}; \alpha_i) \quad for \quad i = 1, 2. \tag{23}$$

$$\operatorname*{Max}_{\theta} = \sum_{t=1}^{T} lnc(F_1(x_1; \hat{\alpha_1}), F_2(x_2; \hat{\alpha_2}); \theta) \tag{24}$$

Actually, equation 23 is not true as the marginal models are estimated as explained above. However, it is helpful to show that the estimation can be divided in two stages where the marginal models are first estimated individually. Second, the copula model is estimated with the maximum log-likelihood method.

Now, the likelihood functions of the different copulas mentioned are presented below:

Gaussian copula:

$$c(u_1, u_2; \rho) = \frac{1}{\sqrt{1-\rho^2}} exp\left(-\frac{\rho^2\xi_1^2 - 2\rho\xi_1\xi_2 + \rho^2\xi_2^2}{2(1-\rho^2)}\right) \tag{25}$$

Student t copula:

$$c(u_1, u_2; \rho) = K \frac{1}{\sqrt{1-\rho^2}} \left[1 + \frac{\xi_1^2 - 2\rho\xi_1\xi_2 + \xi_2^2}{\nu(1-\rho^2)}\right]^{-\frac{\nu+2}{2}} [(1+\nu^{-1}\xi_1^2)(1+\nu^{-1}\xi_2^2)]^{\frac{\nu+1}{2}} \tag{26}$$

Where K is:

$$K = \Gamma\left(\frac{\nu}{2}\right)^{n-1} \Gamma\left(\frac{\nu+1}{2}\right)^{-n} \Gamma\left(\frac{\nu+n}{2}\right) \tag{27}$$

Gumbel copula:

$$\begin{aligned} c(u_1, u_2; \theta) = ([(-ln(u_1)^\theta) + (-ln(u_2)^\theta)]^{\frac{1}{\theta}} + \theta - 1)([(-ln(u_1)^\theta) + (-ln(u_2)^\theta)]^{\frac{1}{\theta}})^{1-2\theta} \\ exp(-[(-ln(u_1)^\theta) + (-ln(u_2)^\theta)]^{\frac{1}{\theta}})(u_1u_2)^{-1}(-ln(u_1))^{\theta-1}(-ln(u_2))^{\theta-1} \end{aligned} \tag{28}$$

*In this case and the next one, the uniform numbers are obtained with the CDF of the marginal distribution.

Clayton copula:

$$c(u_1, u_2) = (\alpha+1)\left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-2-\frac{1}{\alpha}} u_1^{-\alpha-1}u_2^{-\alpha-1} \tag{29}$$

6.2.5 Goodness of fit copulas

At this point, to choose the best copula fitting, some of the mentioned goodness of fit measures can be used. The ones used in this case are the log-likelihood of each copula, the information criteria (BIC and AIC), and the Kolmogorov-Smirnov and Cramer-von-Mises tests. The last ones, in that case, are used to compare the estimated copula with the empirical results, so they are formally defined as follows:

$$KS_C = \underset{t}{\mathrm{Max}}|C(U_t; \hat{\Theta}_T) - \hat{C}_T(U_t)| \tag{30}$$

$$CvM_C = \sum_{t=1}^{T}(C(U_t; \hat{\Theta}_T) - \hat{C}_T(U_t))^2 \tag{31}$$

*Where the empirical copula is computed with the following formula:

$$\hat{C}_T(u) \equiv \frac{1}{T}\sum_{t=1}^{T}\prod_{i=1}^{n}\mathbb{1}_{\{\hat{U}_{it}=u_i\}} \tag{32}$$

6.2.6 Quantification of volumetric and correlation risk

After all this process to price the risk correction, the pricing model is presented. In particular, this new model is the same presented in Pircalabu et al. [2017] but changing the sign of "c" as it is considered better in this way. The reason is that with this change, the price correction will have a sign that reflects the impact on the forward price and the PPA price. In other words, if c is negative, it means that the price correction is subtracted from the forward price decreasing the PPA price and the other way round. In addition, the equation is very similar to the one of model 1 but adding the volumetric and correlation risk correction on the price. First, defining the profit obtained with a PPA contract as:

$$\sum_{t=t_1}^{t_n} Q_t(S_t - (F+c)) \tag{33}$$

Here, Q_t and S_t represent the power production and price at time t (It is necessary to assume that $Q_t = \mathbb{E}_{t-1}[Q_t]$, i.e., assuming no balancing risk). F denotes the futures price as explained in a previous section, i.e, the forward price at t_0 for the delivery period from t_1 to t_n ($F = F(t_0, t_1, t_n)$). Last, c denotes the compensation due to the uncertainty due to power generation and the negative correlation between prices and production interpreted as the volumetric and correlation risks.

Then, to obtain the value of c, the discounted conditional expectation of the payoff from the previous equation is equaled to zero:

$$\mathbb{E}_{t_0}^{\mathbb{Q}}\left[\sum_{t=t_1}^{t_n} Q_t(S_t - (F+c))\right] = 0 \tag{34}$$

*Assuming an interest rate of 0 to simplify (considering it a non-strong assumption).

And isolating c:

$$c = \frac{\mathbb{E}_{t_0}^{\mathbb{Q}}\left[\sum_{t=t_1}^{t_n} Q_t S_t\right]}{\mathbb{E}_{t_0}^{\mathbb{Q}}\left[\sum_{t=t_1}^{t_n} Q_t\right]} - F \tag{35}$$

Therefore, simulating the chosen copula, the value can be obtained by substituting the values of the prices and production in the above equation. In particular, in reference to the risks, the denominator is the correlation risk (it reflects the relationship between renewable generation and prices), and the numerator is the volumetric risk.

At this point, a key remark is necessary. Being the model fitted to historical data means that the price correction calculated above would correspond to the objective measure $\mathbb{P}$. Nevertheless, the expectations are taken under the risk-neutral measure $\mathbb{Q}$ for what is necessary a measure for the market price of risk. In this regard, the particularity of electricity assets and their non-storability implies that a hedging strategy cannot be used in the usual way. The market price of risk of the prices becomes very difficult to obtain. Furthermore, that measure does not even exist for production. The solution widely used in the literature is to set an assumption that the market price of risk is zero, and thus, the risk-neutral measure and the real-measure are the same.

6.3 Results

The methodology widely explained above is applied to the series. First of all, the previous statistics of the series are presented in table 24. The Pearson correlation coefficients have been tested with the t-test, confirming that solar production and wind production have negative covariation with the prices. In both cases, the null hypothesis is rejected with p-values of 0.0000015 and 0.0000011, respectively, for solar and wind series.

	Prices	Solar Generation	Wind Generation
Mean	37.8235	0.2719	0.2895
Standard Deviation	28.4790	0.0871	0.1555
Pearson Correlation		-0.1457	-0.1472
Kendal Correlation		-0.2827	-0.2827
Spearman Correlation		-0.4294	-0.4294

Table 24: Main statistics of the CAISO spot prices and solar/wind infeed generation.

The coefficients and details of the model can be seen in tables 25, 26 and 27. A brief description of the specific treatments to the series. The price series is fitted with linear regression with the long-term and short-term seasonal components. On the one hand, all the sinusoidal parameters are statistically relevant, and on the other hand, the dummies with the days of the weeks, not all of them are significant. Fridays and Saturdays reflect a negative value and relatively low p-values. Mondays and Wednesdays have a coefficient around 2.5-3 and p-values higher than 0.15 but lower than 0.25. Tuesday seems to be significant with a coefficient of 4. Lastly, the less significant dummies are the Thursdays and Sundays, having coefficients next to zero and p-values of 0.87 and 0.48, respectively.

At this point, the ARMA-GARCH model is fitted to the residuals of the previous adjustment. Different combinations of AR, MA lags are proved as well as different GARCH or GJR-GARCH models. In this case, the model that seems to result in a better fit is an ARMA(1,2) (only with the second term of the MA). To correct the heteroscedasticity, a GJR-GARCH(1,1) seems to fit well. Finally, the standardized residuals are suited to a generalized Student-t distribution with a mu parameter of -0.0016, a sigma parameter of 0.1807, and a nu parameter of 4.05. Attending to the Kolmogorov-Smirnov and Cramer-von-Mises tests, both series can be considered as the same one. Furthermore, transforming the standardized residuals into uniform numbers using the cumulative distribution function of the mentioned student t, these tests are computed another time but for the uniform distribution this time. The results still confirm the proper fitting. Last, the LR statistic mentioned above is computed, showing that the null hypothesis of the PIT (probability integral transform) series being a uniform distribution cannot be rejected.

Next, for the marginal model of the solar infeed data, a logistic transformation is computed as explained above. After that, a long-term component is fitted with linear regression. Similar to the case for the prices, all the parameters of the sinusoidal appear to be significant. The ARMA-GARCH model chosen in this instance is an AR(1) to model the mean (without constant) and a GJR-GARCH(1,1) to model the variance. Thanks to that, there is no sign of autocorrelation and heteroscedasticity on the standardized residuals. These are fitted to a Stable distribution since the Student t fit results in a nu lower than 4, meaning that the third and fourth moments of the distribution are not defined. The Stable distribution has four parameters: α ($\in (0, 2]$) called the stability parameter that determines the rate at which the tails of the distribution decline, β ($\in [-1, 1]$) the skewness parameter, γ the scale parameter determining the dispersion of the probability density function and δ the location parameter defining the peak of the distribution. It is a more general distribution, and for example, it includes the Gaussian distribution when the alpha parameter is equal to 2. In particular, for this data the parameters are $\alpha = 1.4802$, $\beta = -0.4631$, $\gamma = 2.5198$ and $\delta = 0.9836$. As seen in the test computed, the fitting can be considered to be good as any of them can reject the null hypothesis of a bad fitting

Last and similarly to the previous treatment, the wind infeed data have a logistic transformation. After that, the long-term component is adjusted, revealing statistically significant values. The ARMA-GARCH chosen comprises the first and second moving average terms for the mean model. And for the variance model, the second term of a GARCH and the second and third of an ARCH. Finally, the standardized residuals are fitted to a generalized Student-t distribution with a mu parameter of 0.0211, a sigma parameter of 1.6017, and a nu parameter of 9.10, reflecting good values of the goodness of fit tests. Indeed, these standardized residuals seem to be the ones with the best fitting.

As well, for the three marginal models, the PITs histograms are presented in figure 38. As already seen in the table's results, the wind generation series has the better fit. However, the three series do not reject that each probability integral transform can be considered a uniform series.

Prices marginal model			
Model / Distribution	Parameter	Coefficient	P-value
Long-term Component	A_0	37.73	0.000
	A_1	-9.17	0.000
	A_2	3.95	0.0008
Short-term Component	Monday	3.0604	0.1659
	Tuesday	4.0291	0.0682
	Wednesday	2.5815	0.2425
	Thursday	0.3566	0.87167
	Friday	-3.3116	0.1338
	Saturday	-5.2127	0.0187
	Sunday	-1.5673	0.4821
ARMA(1,[2])	k	-0.6580	0.000
	ϕ_1	0.8602	0.000
	θ_1	-0.1018	0.0008
		Ljung-Box (6)	0.000
		Ljung-Box (10)	0.000
GJR-GARCH(1,1)	ω	6.2849	0.000
	β	0.37186	0.000
	α	1.0	0.000
	γ	-0.9674	0.000
		Ljung-Box (6)	0.0622
		Ljung-Box (10)	0.0628
Generalized Student-t	μ	-0.0016	
	σ	0.1807	
	ν	4.0457	
		K-S test	0.1674
		CvM test	0.1204
Uniform numbers		K-S test	0.1983
		CvM test	0.1215
LR statistic (critical value = 5.9915)		Statistic	0.0166

Table 25: Prices marginal model parameters and associated coefficients.

Solar generation marginal model			
Model / Distribution	Parameter	Coefficient	P-value
Long-term Component	A_0	-1.043	0.000
	A_1	0.0274	0.0304
	A_2	-0.5826	0.000
AR(1)	ϕ_1	0.75718	0.000
		Ljung-Box (6)	0.1408
		Ljung-Box (10)	0.2522
GJR-GARCH(1,0)	ω	0.0028	0.000
	α_1	0.7610	0.000
	γ_1	0.3940	0.000
		Ljung-Box (6)	0.8055
		Ljung-Box (10)	0.9368
Stable Distribution	α	1.4802	
	β	-0.4631	
	γ	2.5198	
	δ	0.9836	
		K-S test	0.0761
		CvM test	0.2039
Uniform numbers		K-S test	0.1053
		CvM test	0.2244
LR statistic (critical value = 5.9915)		Statistic	0.5845

Table 26: Solar generation marginal model parameters and associated coefficients.

Wind generation marginal model			
Model / Distribution	Parameter	Coefficient	P-value
Long-term Component	A_0	-1.0567	0.000
	A_1	0.2574	0.000
	A_2	-0.6239	0.000
MA(1,2)	θ_1	0.7122	0.000
	θ_2	0.2054	0.000
		Ljung-Box (6)	0.1777
		Ljung-Box (10)	0.2967
GARCH([2],[2,3])	ω	0.0066	0.0670
	β_2	0.8663	0.000
	α_2	0.0487	0.01485
	α_3	0.0694	0.0003
		Ljung-Box (6)	0.0000
		Ljung-Box (10)	0.0000
Generalized Student-t	μ	0.0211	
	σ	1.6017	
	ν	9.1063	
		K-S test	0.6155
		CvM test	0.6329
Uniform numbers		K-S test	0.6458
		CvM test	0.6320
LR statistic (critical value = 5.9915)		Statistic	0.0116

Table 27: Wind generation marginal model parameters and associated coefficients.

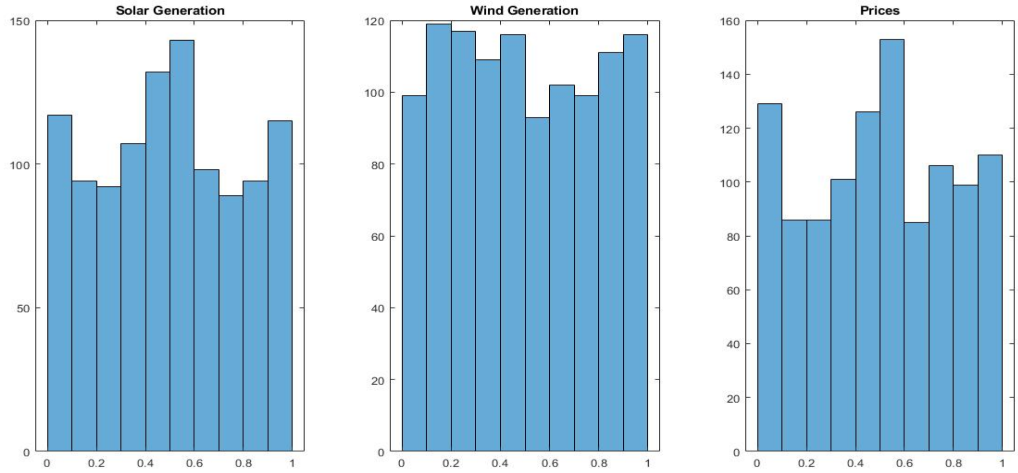

Figure 38: Probability integral transformation histograms

After the solar modelling, the price and solar-generation series reflect a Pearson correlation of -0.0167 (with a p-value of the t-test of 0.55, showing statistical evidence of the lack of a linear relationship of both transformed variables). Likewise, 0.0143 and 0.0207 values for the Kendall and Spearman correlation, respectively. This result shows that in the solar case, when the model is deseasonalized and fitted to an ARMA-GARCH, the residuals have an insignificant correlation. For the case of wind generation, the negative correlation has increased to -0.2188 for the Pearson correlation coefficient (confirming the existence of a linear relationship with the t-test that returns a p-value next to zero). The Kendall and Spearman correlation has not increased, but their values are still negative, being -0.1502 for the Kendall correlation and -0.2235 for the Spearman correlation. So, that confirms the fact about the wind series reflecting a negative correlation after adjustments and the solar not (or lower).

Having the marginal models fitted to a specific distribution allows transforming both series of standardized residuals into uniform numbers using the cumulative distribution function of the specific distribution chosen. Then, different copulas are calibrated to the data in order to model the dependence between both variables. Bearing in mind the goodness of fit measures, the best copula is chosen. As shown in table 28[8], the best results for the solar data correspond to the Student-t copula with $\rho = 0.0148$ and 21.9275 degrees of freedom. Wind data, attending to table 29[9], seems to fit better with the Gaussian copula (considering that the Student-t converges to the Gaussian as the degrees of freedom are high and the Archimedean copulas converge to the independence copula). Thus, these two copulas are used for the simulations of both dependent relationships. For that purpose, the procedure is backward to the one conducted for the marginal models.

Copula	Parameter(s)	logL	AIC	BIC	K-S Test Statistic (critical value = 0.0414)	CvM Test Statistic (critical value = 0.220)
Gaussian	$\rho = 0.0125$	0.0799	1.8401	6.8258	0.0337	0.0882
Student-t	$\rho = 0.0148$ $\nu = 21.9275$	0.9282	0.1436	5.1292	0.0328	0.0829
Gumbel	$\theta = 1.0140$	0.3110	1.3779	6.3636	0.0321	0.0825
Clayton	$\alpha = 0.000$	0.0000	-	-	0.0814	0.2514

Table 28: Copula solar generation vs price estimates.

[8] The t-stat is computed returning a p-value of 0.6814 for the Gaussian ρ and 0.6278, meaning that the correlation coefficient cannot be considered to be different from zero.

[9] In this case the t-stat returns values of 0.000 for both Gaussian and Student-t correlation which means that a linear relationship can be considered to exist

Copula	Parameter(s)	logL	AIC	BIC	K-S Test Statistic (critical value = 0.0414)	CvM Test Statistic (critical value = 0.220)
Gaussian	ρ = -0.2232	27.61	-53.23	-48.24	0.0198	0.0362
Student-t	ρ = -0.2232 $\nu = 52.43$	27.69	-53.59	-48.60	0.0195	0.0349
Gumbel	$\theta = 1$	0.000	-	-	0.0336	0.4555
Clayton	$\alpha = 0.000$	0.000	-	-	0.0924	4.3456

Table 29: Copula wind generation vs price estimates.

Explained with detail, first, extractions simulated from the copula chosen are obtained. Specifically, the number of extractions needed is equal to the number of days each PPA has between the first day where the exchange of power starts and the last day plus the timeout time. Later, these uniform simulations are converted with the inverse of the distributions fitted to the standardized residuals of the marginal models. In this way, the result is a series of simulated standardized residuals later introduced in the models applied before. Namely, the series are re-transformed into prices and solar-generation values using the ARMA-GARCH of each case and the models for the short and long-term seasonal components. Following this procedure, the resulting series can be interpreted as a simulation/forecast of both series imposing the dependence on the data with the copula approach.

The last step is to quantify the risk correction using both series obtained. Here, to do more accurate pricing, the specific characteristics of each PPA contract will be considered. That means three main things:

- The execution date of the PPA is the date when the historical prices are observed in order to give them as an input to the re-transforming procedure. Specifically, the parameter A_0 of the marginal model for the prices is substituted with the average of the previous two years. In this way, each PPA is particularized with their specific characteristics. In the generation series, this cannot be applied as there is no information available. So, the initialize values will be the last values of the solar-generation series (after the LTSC model is applied), acting as this sample ends the previous date of the PPA execution date. That is not true, but there is no other option better to apply. As the generation-solar data is first standardized between 0 and 1, the production values of out-of-sample observations are not expected to differ significantly from the available sample.
  Likewise, the execution date is considered to determine at what point of the season will start each particular case. The sinusoidal model will start on the specific day of the year and the week-day model on the specific day of the week when the contract is signed.
- The term determines when the exchanges will end. Knowing the end date, the days between the PPA execution date and the end date are the number of extractions needed for each case.
- As well, the $F(t_0, t_1, t_n)$ estimates of each PPA are needed to compute the formula shown in the methodology.

With all this information, the values of the volumetric-risk and correlation-risk price correction are obtained and presented on figures 39 and 40.

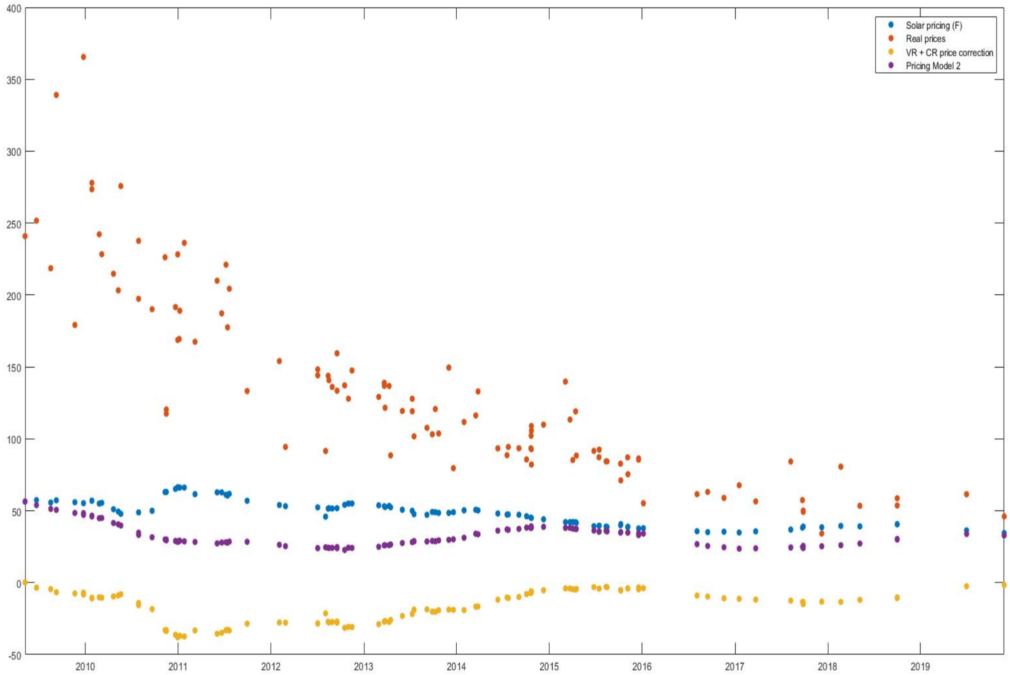

Figure 39: Actual prices, volumetric and correlation risk correction, and model 1 and 2 pricing

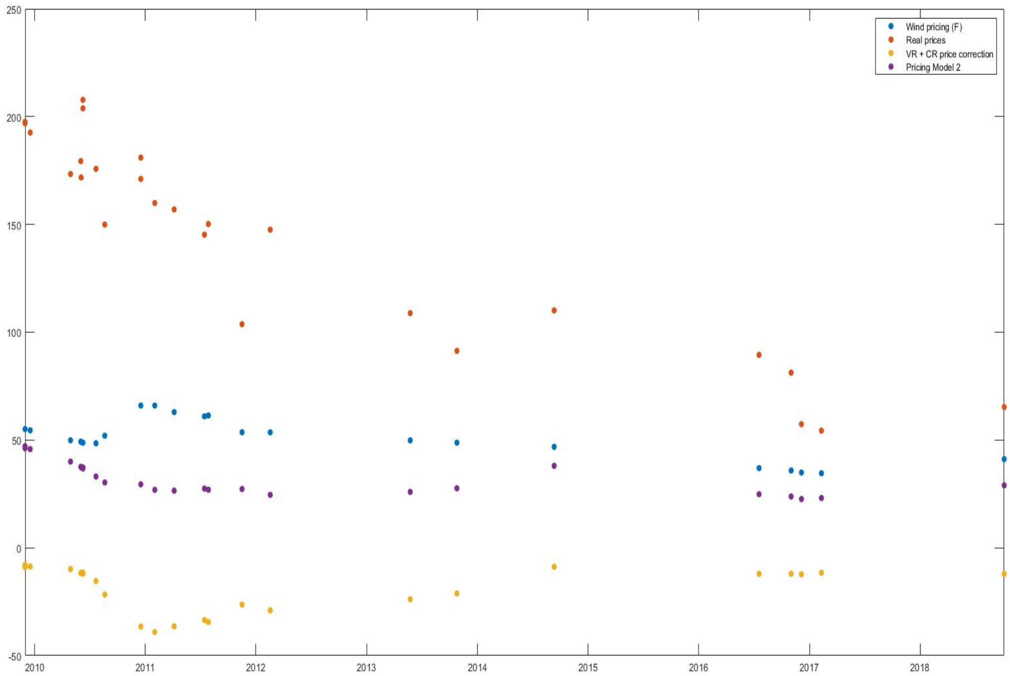

Figure 40: Actual prices, volumetric and correlation risk correction, and pricing models 1 and 2

Off the bat, it is necessary to make explicit that the forecasts of $F(t_0, t_1, t_n)$ used in both cases correspond to the pricing 4.

The results show that the pricing obtained on the first model influences the risk price correction in some sense. This fact makes sense, as the forward pricing appears in the formula. Furthermore, the model 2 pricing seems to fit the actual prices worse and underestimates the majority of the contracts. The results for wind data are similar to solar data. In both cases, the price correction is lowest between the years 2011 and 2012. In recent years, this correction has decreased his importance, with values next to zero.

As in model one, now the forecast error measures of this models are presented in order to have a deeper comprehension of the results.

As expected, table 30 shows that the forecast errors have increased compared with the pricings of model 1. In particular, both MSER and MAE are higher, as well as the BIC and AIC statistics. The correlation is much lower to the one of the pricing 4 in model 1. The DM statistics continue to reject the null hypothesis (shown in table 31).

Error measure \Model	Pricing Model 2
Mean Squared Error Root	123.36
Mean Absolute Error	102.95
BIC	507.32
AIC	504.54
Correlation	0.4922

Table 30: Measures of the errors on the solar data with the pricing model 2

Loss Function \DM p-value	Pricing Model 2
First Loss Function	0.0000
Second Loss Function	0.0026
Third Loss Function	0.0000

Table 31: Solar Diebold Mariano test p-values of pricing model 2

Now, in table 32 the worse performance of the extension can be confirmed. The trend has a similar behavior but in general, the forecast errors are higher. Regarding the correlations, the results are very similar to the ones of model 1. The significant correlations are the same except for the year 2011 that in this case, the size has not a significant correlation with the forecast errors under the level of confidence of 1%.

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	102.95	240.3	178.9	161.2	110.9	89.0	62.4	55.8	31.9	32.5	33.2	20.4
2	0.0849 (0.36)	-0.317 (0.44)	-0.034 (0.89)	-0.724 (0.02)	0.2623 (0.38)	-0.482 (0.05)	-0.090 (0.71)	-0.373 (0.12)	0.6051 (0.39)	0.6510 (0.11)	-0.481 (0.52)	-
3	0.2783 (.00)**	-0.0306 (0.94)	0.6666 (.00)**	0.5503 (0.010)	0.6268 (0.02)	0.5079 (0.04)	0.7433 (.00)**	0.1857 (0.46)	-0.912 (0.09)	0.0022 (0.99)	0.4410 (0.56)	-

Table 32: Analysis of the Solar forecast errors (model 2) by years in CAISO. 1: Average forecast error. 2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

Tables 33 and 34 shows the forecast error measures and DM p values for the case of wind data. The error measures show that the model has lower MSER and MAE too. However, the BIC and AIC statistics are lower or equal to the ones of model 1. Furthermore, in this case, the correlation reflects a higher value than the pricing of models 1. Compared with the solar case, it could be interpreted as a better performance model for wind data than for solar data. Going deeper, solar data will probably capture better the solar volume and correlation risk if the analysis is done with intra-day data as the profile generation of solar power is more characteristic. Similarly, the DM test in this case, returns rejections in the three cases.

Error measure \Model	Pricing Model 2
Mean Squared Error Root	126.29
Mean Absolute Error	118.95
BIC	120.15
AIC	118.71
Correlation	0.7390

Table 33: Measures of the errors on the wind data with the pricing model 2

Loss Function \DM p-value	Pricing Model 2
First Loss Function	0.0000
Second Loss Function	0.0001
Third Loss Function	0.0002

Table 34: Wind Diebold Mariano test p-values of pricing model 2

Table 35 shows as in the solar case the worse performance. As well the results are close to the ones of model 1, with strong relationships for both variables. Therefore, the results are still in favor of the possible influence of the variables in the PPA pricing. Including this extension of model 2, has not changed at all the correlations, and that strengthens the idea. In particular, in CAISO the evidence observed reflects a positive relationship between the forecast errors and the terms or the size, and for the solar data only with the term.

	All	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	118.9	150.0	149.1	116.2	123.0	73.3	72.1	-	52.3	31.3	36.3
2	0.6057 (.00)**	0.6145 (0.19)	0.7493 (.00)**	0.7805 (0.12)	-	-	-	-	0.6821 (0.52)	-	-
3	0.6610 (.00)**	0.9550 (.00)**	0.7236 (0.01)**	0.7079 (0.18)	-	-	-	-	0.8117 (0.40)		

Table 35: Analysis of the Wind forecast errors (model 2) by years in CAISO. 1: Average forecast error. 2: Correlation between forecast error and size. 3: Correlation between forecast error and term.

7 Extension of model 2: Intra-day solar volumetric and correlation risk price correction

7.1 Context and methodology

To finish the study, an extension of the previous model is computed. This extension seeks to capture better the risk price correction due to correlation and volumetric risk. In particular, the modification is to use intra-day data for the solar case rather than daily data. As already seen, solar generation has a very strong pattern along the different hours of the day. Likewise, the prices of power differ at each hour observed. That is why the model is extended for the solar data since it seems very relevant to see if intra-day information improves the results of the model. Attending to the arguments explained it seems crucial to account for the intra-day data as when solar generation is at its maximum the prices could be at its minimum, which would mean a higher price correction.

The sample used is the same as the other one, but this time there are 24 series of prices as well as of production. Similarly, the methodology followed is the same as in the previous chapter for each hour. So, in this case, there are going to be 24 marginal models of prices, 24 models of generation, and 24 copulas to capture the dependence of each pair of variables. The steps are the same and the test to evaluate the quality of the adjustments too. The unique difference is that this time the procedure is repeated 24 times (actually, 15 times as some of the series are not needed to be modeled as will be explained below). As this approach has never been emulated in the financial literature it appears to be very interesting and original. Furthermore, having computed before the price correction with daily data will allow compare the results and appreciate the utility of this extension.

The extension expressed formally is:

$$c = \frac{\mathbb{E}_{t_0}^{\mathbb{Q}}\left[\sum_{t=t_1}^{t_n}\sum_{h=1}^{24} Q_{t,h}S_{t,h}\right]}{\mathbb{E}_{t_0}^{\mathbb{Q}}\left[\sum_{t=t_1}^{t_n}\sum_{h=1}^{24} Q_{t,h}\right]} - F \tag{36}$$

In particular, the differences with the previous model are mainly two. First, the long-term component of some of the series includes a fourth parameter (A_3). The parameter is employed to account for periods where the production is zero. As this is difficult for the previous model to be captured, a dummy with ones in the days of the year when each time frame reflects the almost nonexistence of production (caused in non-sun hours). Attending to the data, the time frame between 06-07, 17-18, 18-19, and 19-20 are the unique ones where there are some periods of the year without regular sun hours[10]. Second, as in the previous case, the value of A_0 in the prices is changed with the value of the last two years previous to the execution date. The sample of the prices of the 24 time-frames is smaller than the other one. Specifically it starts in 2009. That is why 2 years are used instead of 8, to being able to price the majority of the

[10]The time frame 20-21 has a very low average generation, so, the difference between the normal values and the ones in periods of almost no production seem to not differ.

PPA. Nevertheless, the PPA with a execution date before april 2011 cannot use the average. So, in these cases the value is not modified.

Another important aspect is that in spite of having 24 series, only 15 are used. The argument is linked with the previous one. In the hours 00 to 06 and 21 to 24, there is no period of the year when the radiation is significant. Indeed, this is confirmed doubly. First, looking at the generation data, the generation in these nine time-frames have residual values or zero values. Additionally, the average generation of these time frames is lower than 0.001, verifying that this series would not affect significantly the result of the equation 36. Second, that fact can be proved with meteorological information about the hours of sunrise and sunset at each different period of the year. So, it does not seem a strong assumption to forget these nine time-frames as the justification is consistent with the evidence and the implications do not make the model significantly less robust.

7.2 Results

Thus, with all that previous clarifications, the procedure is computed and the results are below. First, the marginal models fitted parameters and goodness of fit measures are in the Annex 4 in tables 38 and 39. Similarly, the copulas fitted are in table 40. The Gumbel and Clayton copula results are not shown as in the majority of the cases they tend to the independence copula, and as in any case they reflect the best goodness of fit measures. As seen in the table, the Student-t copula has the best performance in all the cases.

Now, in figure 41 the results can be seen. The first thing that is evidenced, is that the PPA between 2009 and 2011 are influenced by the different procedure applied (due to the smaller sample for the 24 time-frames as explained above). Furthermore, despite the more demanding model, the price correction has not changed significantly. However, the price correction reflects positive low values in some of the periods.

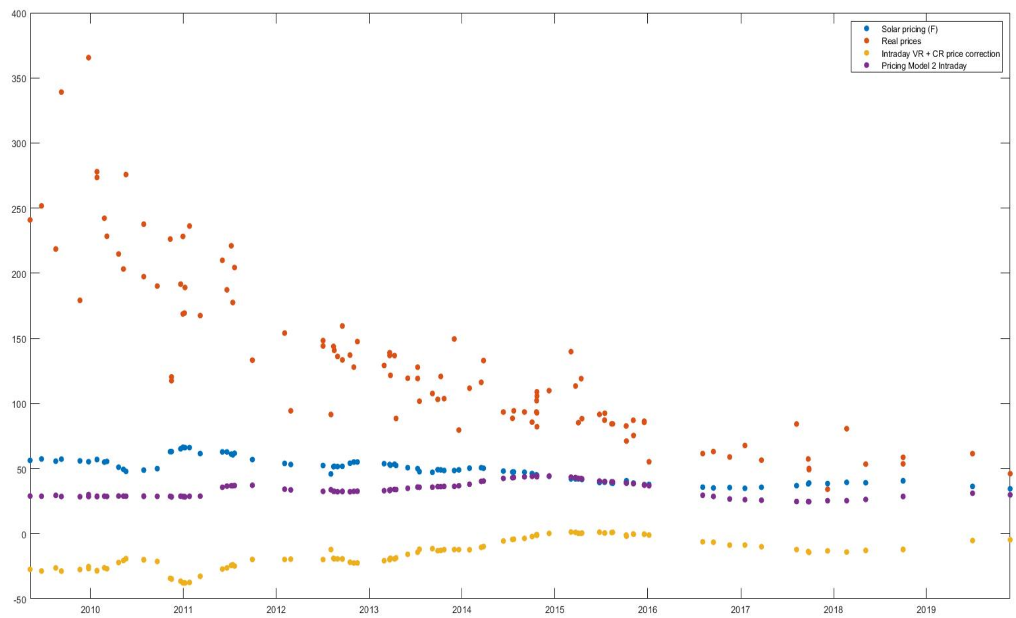

Figure 41: Actual prices, volumetric and correlation intra-day risk correction, and model 1 and extension of model 2 pricing

To see the differences between the model with diary prices and intra-day prices, figure 42 is presented as well. It can be observed that after the first two years, where the average of the historical prices is not used, the model presents better results in some sense. The price corrections are lower and in the last years of the sample very similar between them.

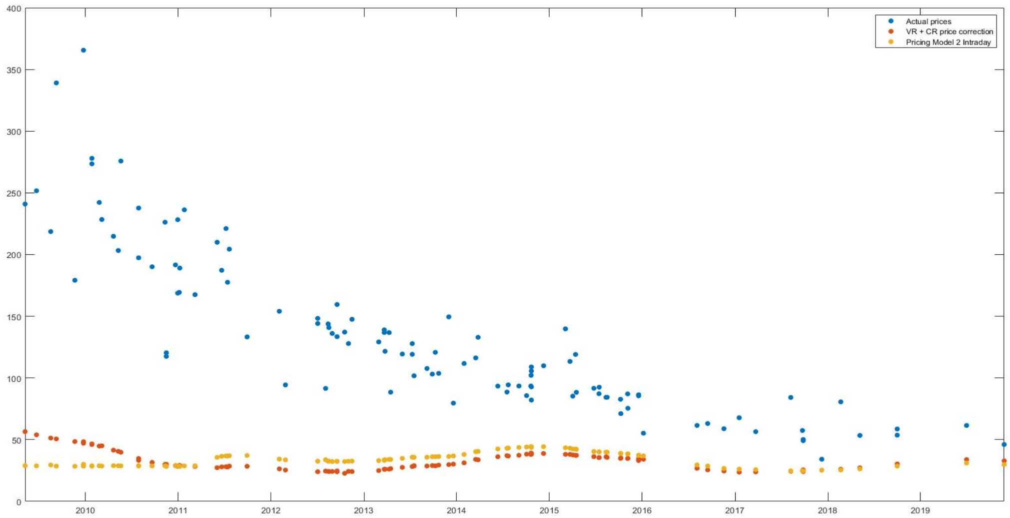

Figure 42: Actual prices and model 2 with daily and intra-day series

As well as in all the previous models, the measures of the forecast errors are presented. In table 36 the results are slightly worse than in the daily model. The MSER, MAE and the information criteria reflect better values with the daily model. Nonetheless, the results seem to be better if only the sample between 20011-2021 is considered. Besides, the most relevant fact is that the correlation has changed a lot. However, the change seems to be justified by the success observed in the first two years of the PPA sample. Indeed, if the correlation is computed with only the sample after april 2011, the correlation is positive. Table 37 indicates that with the second loss function de DM test is more nearer to not reject the null hypothesis. Nonetheless, the main conclusion is a slightly worse or equal performance of the model with intra-day series due to the lack of longer price series, whose availability could allow comparing more accurately the two models.

Error measure \Model	Pricing Model 2
Mean Squared Error Root	125.37
Mean Absolute Error	101.56
BIC	516.67
AIC	513.88
Correlation	-0.3326

Table 36: Measures of the errors on the solar data with the pricing intra-day model 2

Loss Function \DM p-value	Pricing Model 2
First Loss Function	0.0000
Second Loss Function	0.0157
Third Loss Function	0.0000

Table 37: Solar Diebold Mariano test p-values of pricing intra-day model 2

Part D

Conclusion

From the literature review, it has been clearly seen that most of the works analyze PPA from a theoretical approach or individual PPA. So, the innovation of this work is to present a PPA study applying those analyses for a group of PPA contracts. This approach has allowed evaluating the factors that influence the pricing and exhibiting other ones not analyzed yet. Thanks to the research, some of the main aspects have been evaluated. For example, a copula approach to model the dependence between prices and production has been applied as in other literature articles but in this case for more than 150 PPA contracts. What is more, that approach only had been tested in wind data, and in this research, the results are available about both technologies, wind and solar.

This work has investigated Power Purchase Agreements. First, presenting the information needed to understand well how PPA work and what factors have a significant role in their functioning. As can be seen before, PPA varies in terms and structures, offering flexible possibilities to ease the agreements between two agents. Its utility has increased along the last decade, helped by the characteristics seen, making this instrument rather appealing for renewable energy projects. However, information is not so transparent yet, which makes more complex and limited the possible analysis. Despite that fact, data has been found and described for the United States. In particular, analyzing the data, the evidence showed that different treatment of the electricity regions of the U.S would be more appropriate. In the same way, the available data has information about solar and wind PPA, which also have exhibited that a separate analysis seems more adequate. Likewise, the importance of the generation mix and the structure of the market of each region has been highlighted. Additionally, the different particularities of power have been explained and connected with implications for the study.

The analysis has focused on the pricing of this instrument with particular attention to the variables with stronger effects. First of all, with the analysis of the PPA terms and PPA capacities agreed upon, the conclusions are not straightforward. The results confirm that different treatment for regions is better as the relationships depend on the sample. Then, model 1 has been proposed resulting in underestimations of the PPA prices. That could demonstrate that PPA are not the same as a usual SWAP since the comparisons evidence other relevant factors influencing the pricing. Last, the model has been extended to account for other risks that arise with the signing of a PPA. In this sense, the process of quantification of the volumetric and correlation risk could be seen for both CAISO solar and wind PPA. The results reflected a negative price correction to compensate the buyer of the energy for that risk. Moreover, the underestimation is still very strong. Actually, the performance is worse than model 1. That fact can be surprising as the second model could be considered more complete. Nevertheless, the argument that explains the negative correction is clear and lies in the negative correlation between prices and production. So, despite the worse performance, that correction is necessary for fair pricing. Additionally, it would be useful to know if this price correction have been applied in the market for all the PPA along with the sample. Maybe, at the beginning, with a less significant renewable power penetration in the market, the price correction due to these risks was not considered in the pricing. Additionally, based on the evidence, it was considered more adequate to compute the price correction with intra-day series for the solar case. Therefore, the second model extension uses the 24 hourly series of generation and prices in order to price the volumetric and correlation risk correction more accurately. The results do not have an easy interpretation. Some difficulties have arisen because of the data that do not allow to price the correction along with the whole sample, turning more complex the comparison. Even so, the model has changed confirming that the approach used influences the pricing in solar data, although it is not easy to determine, based on the evidence, which model has a better performance over the other. Despite that fact, the procedure has enabled to see how to employ each approach and opens a line of research for future investigations.

The outcome of the analysis has made clear an underestimation with the factors taken into account. In greater detail, the differences between the estimates and the market prices are not homogeneous along with the full sample but with very similar behaviours in solar data and wind data. In general, the farther is the PPA contract evaluated, the higher differences are observed. On the one hand, this could be interpreted as a factor or various factors related to the dates or changes as time progresses. However, on the other hand, the similarities between solar and wind data indicate that the missing factors are related to both technologies. Indeed, after 2012-2013, the estimations differences can be considered moderate compared to 2009-2011, where the estimations are no more than half or a quarter of the market prices. Furthermore, in the most recent dates, better performance is noticeable. For example, in solar CAISO, the differences are minor, and the trend is very well captured (when the market prices rise a little, the estimates also and vice versa). Similarly, West-non-ISO captures much better the trend in recent dates, and even some overestimations are observed.

Nevertheless, apart from that, analyzing the PPA contracts database, PPA with the same characteristics and very similar dates have relevant differences in market prices. That turns more difficult the exercise of pricing since the market could be working in a non-efficient way. Likewise, in all the research, it is assumed that the database and all the information are accurate but could be false or better information could be used. In fact, as already seen, this field has very poor information yet, so, as the information available would increase and become more accurate, the analysis in that direction would have better performances.

Subsequent works could develop further this analysis as it still underestimates the pricing observed on the market. As seen, the underestimations are not uniform along with the sample, so the successive extensions should look for elements that could fix this. The approach should be to search the missing components that could help to explain the differences between the models and the market prices. Apart from the studied variables size and term, other ideas are exposed for future researches.

One factor that could be very interesting to consider would be the Renewable Energy Certificates (RECs) as the PPA database only includes bundled PPA. Likewise, hypothesizing about other factors, one could be the credit risk, which may explain the greater underestimations as the farther is the PPA (maybe because the instruments used to eliminate the credit risk have changed along with the sample). The first contact with this topic has been presented in Edge [2015]. Another factor could be negotiation power since the prices are from the market, and maybe these are not formed in a completely competitive market at the beginning of the sample. In the same way, the analysis could be improved with the use of more information. For instance, in the second model, the lack of a larger production series has determined to assume the same sample and copula fitting for all the PPA contracts, which could be improved. Other factors to be considered could be those related to political measures or legal issues. In other words, different measures or laws may have existed in specific periods along with the sample. For example, a political measure to encourage the use of renewable energy could increase the PPA prices as these become more attractive. Also, legal issues can put obstacles to developing renewable energy projects, implying that energy generators would expect a higher price in the PPA to compensate for these issues. Last, from the point of view of the expenses, the technologies have improved over the years, meaning a remarkable decrease in the costs and increasing the efficiency of solar PV and wind plants. Similarly, some of the steps could be improved too. For example, for the extrapolation method a more sophisticated one can be considered as in Leoni et al. [2018], where the authors employ a multivariate constrained robust M-regression method to refine the forward curve in electricity markets. Moreover, the method is based on arbitrage arguments.

What is clear is the necessity of fair pricing in Power Purchase Agreements, and the more investigations in that field, the more spread could experiment these financial instrument. Furthermore, that will imply the production of more clean energy that, as already seen, is a process that presents some difficulties and should be considered for correct development.

Bibliography

Tayo Ajadi, Victoria Cuming, Rohan Boyle, David Strahan, Matthias Kimmel, Michael Logan, Angus McCrone, et al. Global trends in renewable energy investment 2020. 2020.

Hirotugu Akaike. Prediction and entropy. In *Selected Papers of Hirotugu Akaike*, pages 387–410. Springer, 1985.

Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

Jeremy Berkowitz. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474, 2001.

Mark Bolinger, Joachim Seel, and Dana Robson. Utility-scale solar: Empirical trends in project technology, cost, performance, and ppa pricing in the united states–2019 edition. 2019.

Maira Bruck, Peter Sandborn, and Navid Goudarzi. A levelized cost of energy (lcoe) model for wind farms that include power purchase agreements (ppas). *Renewable Energy*, 122:131–139, 2018.

Peter Christoffersen, Kris Jacobs, et al. *The importance of the loss function in option pricing.* Citeseer, 2001.

Felipe Isaza Cuervo, Carlos Andrés Arredondo-Orozco, and Gabriela Carolina Marenco-Maldonado. Photovoltaic power purchase agreement valuation under real options approach. *Renewable Energy Focus*, 36:96–107, 2021.

Francis X Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1, 2015.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

Paul Edge. An approximation of counterparty credit risk in long term power purchase agreements (ppas). In *2015 12th International Conference on the European Energy Market (EEM)*, pages 1–4. IEEE, 2015.

Felipe Ignacio Gallardo, Andrea Monforti Ferrario, Mario Lamagna, Enrico Bocci, Davide Astiaso Garcia, and Tomas E Baeza-Jeria. A techno-economic analysis of solar hydrogen production by electrolysis in the north of chile and the case of exportation from atacama desert to japan. *international journal of hydrogen energy*, 46(26):13709–13728, 2021.

Yashar Ghiassi-Farrokhfal, Wolfgang Ketter, and John Collins. Making green power purchase agreements more predictable and reliable for companies. *Decision Support Systems*, 144:113514, 2021.

Kevin Hernandez, Christopher Richard, and Jay Nathwani. Estimating project lcoe–an analysis of geothermal ppa data. In *Proceedings of 41st Workshop on Geothermal Reservoir Engineering*, pages 1–8, 2016.

Johannes Kaufmann, Philipp Artur Kienscherf, and Wolfgang Ketter. Modeling and managing joint price and volumetric risk for volatile electricity portfolios. *Energies*, 13(14):3578, 2020.

Evangelos Kyritsis, Jonas Andersson, and Apostolos Serletis. Electricity prices, large-scale renewable integration, and policy implications. *Energy Policy*, 101:550–560, 2017.

Xin Lei and Peter A Sandborn. Maintenance scheduling based on remaining useful life predictions for wind farms managed using power purchase agreements. *Renewable Energy*, 116:188–198, 2018.

Peter Leoni, Pieter Segaert, Sven Serneels, and Tim Verdonck. Multivariate constrained robust m-regression for shaping forward curves in electricity markets. *Journal of Futures Markets*, 38(11): 1391–1406, 2018.

Andrew C Levitt, Willett Kempton, Aaron P Smith, Walt Musial, and Jeremy Firestone. Pricing offshore wind power. *Energy Policy*, 39(10):6408–6421, 2011.

Luca Mendicino, Daniele Menniti, Anna Pinnarelli, and Nicola Sorrentino. Corporate power purchase agreement: Formulation of the related levelized cost of energy and its application to a real life case study. *Applied Energy*, 253:113577, 2019.

Lindsay Miller, Rupp Carriveau, Scott Harper, and Someshwar Singh. Evaluating the link between lcoe and ppa elements and structure for wind energy. *Energy strategy reviews*, 16:33–42, 2017.

Juan Ignacio Peña, Rosa Rodríguez, and Silvia Mayoral. *Productos financieros para la transición energética*. Funcas, 2020.

Amarasinghage Tharindu Dasun Perera, Vahid M Nik, Dasaraden Mauree, and Jean-Louis Scartezzini. Electrical hubs: An effective way to integrate non-dispatchable renewable energy sources with minimum impact to the grid. *Applied Energy*, 190:232–248, 2017.

Anca Pircalabu and Jesper Jung. A mixed c-vine copula model for hedging price and volumetric risk in wind power trading. *Quantitative Finance*, 17(10):1583–1600, 2017.

Anca Pircalabu, Thomas Hvolby, Jesper Jung, and Esben Høg. Joint price and volumetric risk in wind power trading: A copula approach. *Energy Economics*, 62:139–154, 2017.

Tillmann Sachs, Alberto Dalla Rosa, and Robert LK Tiong. Case study on quantifying the impact of political risks on demand and pricing in a power project. *The Journal of Structured Finance*, 14(2): 77–84, 2008.

Gideon Schwarz et al. Estimating the dimension of a model. *Annals of statistics*, 6(2):461–464, 1978.

Bo Tranberg, Rasmus Thrane Hansen, and Leopoldo Catania. Managing volumetric risk of long-term power purchase agreements. *Energy Economics*, 85:104567, 2020.

Ryan H Wiser, Mark Bolinger, Ben Hoen, Dev Millstein, Joseph Rand, Galen L Barbose, Naïm R Darghouth, Will Gorman, Seongeun Jeong, Andrew D Mills, et al. Wind technology data and trends: Land-based focus, 2020 update. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2020.

Owen Q Wu and Volodymyr Babich. Unit-contingent power purchase agreement and asymmetric information about plant outage. *Manufacturing & Service Operations Management*, 14(2):245–261, 2012.

Klaas Würzburg, Xavier Labandeira, and Pedro Linares. Renewable generation and electricity prices: Taking stock and new evidence for germany and austria. *Energy Economics*, 40:S159–S171, 2013.

Part E

Appendix

Annex 1: Review of the literature models

After a deep review of the existent models related to power purchase agreements, the most important ones are presented in the following pages:

Article	Model	Description
Edge [2015]	$PPA = \sum_{t=s}^{s+l} \frac{(P(t)-P_c)}{(1+i)^t}$	Model equation which the authors start the article. P_c is the price agreed for the exchanges and the spot price is P(t). It is similar to a standard valuation of a SWAP.
Edge [2015]	$PPA_y(t) = \frac{(P_y(t \vee s)-\overline{P})a_{l,i+\alpha}+(\overline{P}-P_c)a_{l,i}}{(1+i)^s}$	Introducing in the previous equation the price process (an Ornsetin-Uhlenbeck mean reverting process) and using the fact that a geometric series of payments can be expressed as an annuity $a_{l,i}$. $\overline{P}$ is the long term average market price (the level where the prices are supposed to revert)
Tranberg et al. [2020], Pircalabu et al. [2017]	$\mathbb{E}_{t0}^{\mathbb{Q}}\left[\sum_{t=T_1}^{T_2} Q_t(S_t-(F-c))\right]$	Where Q_t is the quantity of power produced, S_t is the spot price of the electricity, F the forward price at each instant T_i and c the compensation due to the price-production correlation. The exchange of power starts at T_1 and ends at T_2. This equation is also very similar to the standard one but on the one hand it does not discount the cash-flows and has a new term to reflect a volumetric and correlation risk correction.
Kaufmann et al. [2020]	The authors propose a very detailed method to model Day Ahead power prices (P_t^{DA}) decomposing the variable price in three components: a short term component, a long term and a stochastic component each one with its model equations (Page 6 of the article). After that, the equation to obtain the price of the PPA is: $P_{t_0}^{fixed} = \frac{\mathbb{E}_{t0}^{\mathbb{Q}}\left[\sum_{t=h_0}^{h_0+23} Q_t P_t^{DA}\right]}{\mathbb{E}_{t0}^{\mathbb{Q}}[Q_t]}$	h_0 accounts for each hour of the day. $\mathbb{Q}$ is the risk-neutral measure that under the rational expectation hypothesis it can be set equal to the physical measure $\mathbb{P}$ that takes into account the uncertainty of using historical data.
Cuervo et al. [2021]	$PPA_t = Ce_{t-1}(1+g_{Ce})(1-PPA_{DF})$	C_e is the retailing price of electricity, g_{Ce} the growth rate in the cost of electricity calculated with historical data and PPA_{DF} is the PPA discounted factor. The authors also apply a Real Option Analisys to see the value of the timing of the project (to see when is better to develop the project). Actually, it is not a pricing of a PPA.

Article	Model	Description
Ghiassi-Farrokhfal et al. [2021]	There is no model in itself, but from pages 3 to 7 the authors present as a method to calculate the fair price of a PPA.	The fair price is considered the one that improves the situation of both parties, and it is usually a range of prices not a unique price.
Peña et al. [2020]	$PPA_t = \sum_{i=1}^{m} e^{-r(T_i - t)}(F(t, T_i) + REC_i - K)$	$F(t, T_i)$ is the forward electricity price in T_i observed at time t. REC_i is the renewable energy certificates prices at time i. And K is the fixed price agreed in the PPA. Therefore, this is a equation of valuation of a PPA (some previous equations are pricing models of the PPA) but it could be used for pricing too. As shown, this equation takes into account the certificates of renewable energy. However, to use them it would be needed to have historical REC prices.

In addition to these pure PPA valuation models, it can also be reviewed the literature models developed to calculate the LCOE. The reason is justified since the PPA price can be calculated as the one that covers the cost of producing the energy plus a margin for the generator. That is why the following table presents some models for LCOE valuation too.

Article	Model	Description
Bruck et al. [2018]	$LCOE = \sum_{i=1}^{n} \frac{TLCC}{\frac{E_i}{(1+WACC)^i}}$	Where TLCC is the total life cycle cost of the wind plant. WACC is referred to the average weighted cost of the capital and E_i is the energy generated each year.
Bruck et al. [2018], Mendicino et al. [2019]	$LCOE = \frac{\sum_{i=0}^{n} \frac{CPE_i}{(1+WACC)^i}}{\sum_{i=0}^{n} \frac{E_i}{(1+WACC)^i}}$	CPE_i is the cost of production of the energy in the year i. This is only a model expanding the previous equation. It is the model used in the software SAM, System Advisory Model, from NREL.
Bruck et al. [2018], Mendicino et al. [2019]	$LCOE = \frac{\sum_{i=0}^{n} \frac{I_i + OM_i + F_i - PTC_i - D_i - T_i + R_i}{(1+WACC)^i}}{\sum_{i=1}^{n} \frac{E_i}{(1+WACC)^i}}$	The authors make reference of another model of the literature more complete. In this one is taken into account the cost of fuel of each year (F_i), the depreciation (D_i), the credit taxes (PTC_i), the general taxes (T_i) and some royalties (R_i). Furthermore OM_i are the operational costs and I_i the cost of the inversion.
Bruck et al. [2018]	$LCOE = \frac{\sum_{i=0}^{n} \frac{I_i + OM_i + F_i - TC_i - Pen_i}{(1+WACC)^i}}{\sum_{i=1}^{n} \frac{E_i}{(1+WACC)^i}}$	In TC_i some of the taxes are encompassed and Pen_i is a variable where the possible maximum or minimum limits of the energy delivered can be included. That is the model that the authors propose to reflect the possible limits that a PPA contract can have.
Mendicino et al. [2019]	$LCOE = \frac{\sum_{i=1}^{n} \frac{I_0 + M_i + F_i}{(1+r)^i}}{\sum_{i=1}^{n} \frac{E_i}{(1+r)^i}}$	Similar notation as the previous models. In this case I_0 are the initial cost and M_i the operational costs.

Article	Model	Description
Mendicino et al. [2019]	$LCOE_{vcppa} = \frac{\sum_{i=1}^{n} \frac{I+GT_i+GM_i+GE_{X i}+OM_i+EMC_i+CI(p)_i+CN(p)_i+CM(p)_i}{(1+r)^i}}{\sum_{i=1}^{n} \frac{E_i}{(1+r)^i}}$	The authors propose this very complete model. They use the figure of a intermediary to explain all the cost that should be accounted for the LCOE. Between them there are some cost of guarantee due to different reasons, costs of maintenance, costs of disequilibrium, costs of congestion, cost of dis-adjustments etc...
Miller et al. [2017]	$LCOE = \frac{(FRC \cdot ICC)AOE}{AEP_{net}}$	FRC is the flat rate of collection , ICC is the cost of the installed capital, AOE the annual costs of operation and AEP_{net} the annual net energy production. This equation is the one that NREL (the National Renewable Energy Laboratory) uses.
Hernandez et al. [2016]	$Levelized - Revenue/Cost = \frac{\sum_{n=1}^{N} \frac{P_n \cdot Q_n}{(1+d)^n}}{\sum_{n=1}^{N} \frac{Q_n}{(1+d)^d}}$	It is not an equation to obtain the LCOE. Q_n is the energy in KWh per hour, P_n the price per KWh per hour (both in each year n) and d is the annual rate of discount. It is a useful link between PPA and LCOE because the authors obtain the LCOE from PPA prices.

Annex 2: Processing of the PPA database

The following paragraphs explain the prices available in the database and the procedure applied to obtain them. On the available database, the PPA prices are expressed as levelized prices. That means two main things: the values are calculated considering the level of the prices and expressed as the actual value of the expected cash flows. Specifically, the prices are in constant 2019 dollars as an average of the expected discounted cash flows.

To formalize this, first, some variables or concepts are defined:

- P_{lev}: Levelized price. As defined above, is the price in constant 2019 dollars and as an average of the expected discounted cash flows.
- P_{PPA}: PPA price agreed in the contract. It is supposed as a fixed price for the exchanges of electricity of one MWh each month during all the contract length.
- DF_i: Discount factor for a logarithmic interest rate of i years (i does not need to be an integer because it also includes months)
- $DEF_{i,2019}$: Deflator applied to each i year to convert the nominal prices into constant (real) 2019 prices. In this case, the same deflator for all the months of each year is assumed.
- $GDP - DEF_i$: GDP Deflator value of the year i.
- CF_i: Discounted and real cash-flow of the period i
- T: Duration or term of the PPA contract.
- t_i: time from the initial date of signing the contract to the i date
- r_{real}: Real interest rate.
- N: Number of cash-flows in the PPA contract.

After defining the primary variables, it is now necessary to explain how the levelized prices are obtained. For better comprehension figure 43 represents a scheme of the transformation done. The starting point is the monthly series of the nominal prices (assuming that the PPA of the database are monthly base-load). These series are in current dollars, so a conversion to constant dollars is done using the GDP deflator (historical or projected). Then, these cash flows are discounted with the real interest, and the average of the resulting series is obtained. Applying this method results in a price that permits comparisons between different PPA of different periods or countries (applying each country deflator and each country real interest rate).

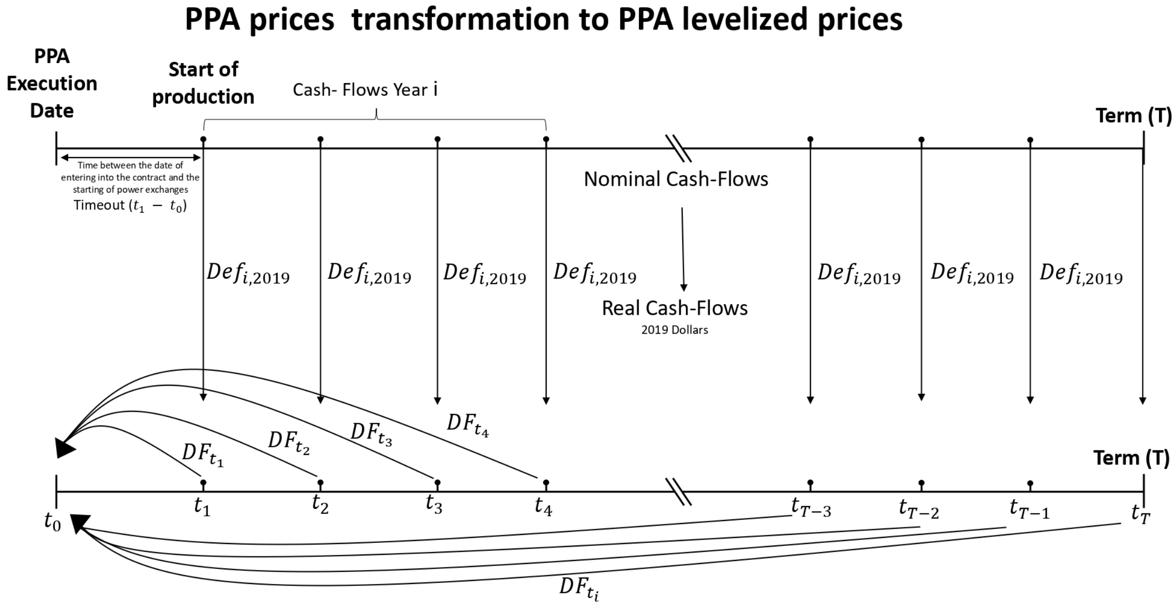

Figure 43: Transformation done in the database to obtain the levelized prices

With this in mind, it is easy to show the reverse process (obtaining the PPA price of the contract from the levelized 2019 prices database). In the database, it is only available the Execution Date and not the Commercial date. Therefore, it is needed to make some assumptions about the average length between these two dates. Attending to the scheme, this is going to be called as "timeout". For that purpose, brief research on this topic has been done. The conclusion is to use an average difference between t_0 and t_1 of 18 months[11].

As it is easy to show, from the levelized price can be computed the total sum of the constant (2019 dollars) and discounted prices of all the contract term. This is because the value is used to obtain the average of this cash flows: $P_{lev} = \frac{1}{N}\sum_{n=1}^{N} CashFlows_n$. So multiplying N and the levelized price results in the total sum. So, with the database available, the exact cash-flows of each PPA are not possible to be obtained. However, with the total sum mentioned and the total equivalent sum using the P_{PPA} (explained below), it can be computed the desired price.

To compute the total sum of the cash flows starting from P_{PPA}, a similar process can be done. First, P_{PPA} will be the nominal price and cash flow for each period.

$$NominalCF = (NominalCF_1 = P_{PPA}, ..., NominalCF_N = P_{PPA})$$

The correction of prices is then applied to these nominal cash flows, converting them into 2019 real dollar prices.

$$DEF_{i,2019} = \frac{GDP - DEF_{2019}}{GDP - DEF_i} \quad , for \quad i = 1, ..., N$$

$$RealCF_i = DEF_{i,2019} \cdot NominalCF_i \quad , for \quad i = 1, ..., N$$

Similarly, the present value of this cash flows is computed.

[11] Based on the PPA tracker of the PexaPark website. In particular, the average of the United States PPA contracts (available with that source) where the Execution and Commercial date were specified

$$DF_i = \frac{1}{e^{r_{real} \cdot t_i}} \quad , for \quad i = 1, ..., N$$

$$CF_i = DF_i \cdot RealCF_i \quad , for \quad i = 1, ..., N$$

Then, summing all these cash flows, the equivalent total sum of the cash-flows that should be equal to the other can be obtained. So that, the P^*_{PPA} is obtained from this equivalence. Expressed mathematically, the P^*_{PPA} value is the one that equals the right part of the equation below with the left part.

$$P_{lev} \cdot N = \sum_{i=1}^{N} CF_i$$

Expanding the expression and isolating P^*_{PPA}:

$$\frac{P_{lev} \cdot N}{\sum_{i=1}^{N} \frac{1}{e^{r_{real} \cdot t_i}} \cdot \frac{GDP-DEF_{2019}}{GDP-DEF_i}} = \frac{P_{lev} \cdot N}{\sum_{i=1}^{N} DF_i \cdot DEF_{i,2019}} = P^*_{PPA}$$

That is the procedure applied to every PPA to obtain its PPA price agreed in the contract.

Annex 3: Measurement of forecast errors

To evaluate the forecasts obtained with the different models there are going to be used the following measures:

- Mean squared error root:

$$MSER = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{37}$$

- Mean absolute error:

$$MAE = \frac{1}{N}|\hat{y}_i - y_i| \tag{38}$$

*Where y_i is the actual price on the database, $\hat{y}_i$ the forecasted price with each model and N the number of observations. The variable to forecast does not oscillate around a non-zero average, so it is preferable to use these measures and not their percentage versions.

Likewise, information criteria statistics are also computed to complement the testing of the models. The statistics used are:

- The Akaike information criterion from Akaike [1985]:

$$AIC = 2k - 2 \cdot ln(L) \tag{39}$$

  * Where k are the degrees of freedom (number of parameters to be estimated) and L the maximum value of the likelihood function of the model.

- The Schwarz information criterion from Schwarz et al. [1978] (also known as the Bayesian information criterion).

$$SIC/BIC = k \cdot ln(N) - 2 \cdot ln(L) \tag{40}$$

  * Where N is the length of the sample used and the other variables follow the same notation as in the AIC.

  That can be simplified under the assumption that the forecasted errors are identically, independently, and normally distributed (and other conditions).

$$SIC/BIC = k \cdot ln(N) + N \cdot ln(\hat{\sigma}_{\epsilon}^2) \tag{41}$$

  *Where $\hat{\sigma}_{\epsilon}^2$ is the variance of the forecasted errors. This assumption is also made to compute the AIC.

These statistics can be interpreted as a measure of the relative quality of a model for the data available. They are used as criterion model selection choosing the lower ones. Both are closely related since the two statistics use the maximum likelihood as a criterion of the goodness of fit. Nevertheless, as can be seen, the BIC statistic penalizes free parameters more than the AIC as long as N is greater than e^2, which is expected in most cases. Consequently, the inclusion of the two of them allows a better evaluation.

Additionally, the correlation between the market pricing series and the pricing model output is presented. That would help to see better which model predicts more accurate prices. Or from another point of view, which model is more likely to be the one used in the market. The model with higher correlation could be interpreted as the one which does the pricing in a direction more similar to the approach followed by the market. For example, if the pricing is being done using the prices of the main node of the electricity region, the model that replicates this will experiment changes in the same directions.

Furthermore, for a better comparison, it is possible to use the test proposed by the authors Diebold and Mariano [2002] originally intended for comparing different forecast outputs. Related to this, one of the authors, Diebold, has recently written a new paper where analyses the extended use of this test (see Diebold [2015]). The author emphasizes an abuse of the test when the assumption required to compute the test is relaxed. From his perspective, the test is being used to compare models when it was first created to compare forecasts. He argues that if a model presents better results with a particular sample, it does not mean that the result can be extrapolated considering this model to better performance in all possible scenarios. Moreover, he enumerates other measures to test models created more specifically for this purpose. These are some of the already presented below. Because of the arguments explained above, the DM test will be used to test the null hypothesis that the losses due to forecasts error of each model are statistically different from zero and that there is no sign of auto-correlation or heteroscedasticity.

Formally, following a similar notation as in the literature:
The purpose is to forecast the pricing of an M number of PPA contracts. Then, a loss function is needed to evaluate the performance of each model. The ones to be used are the following:

First, quantifying the losses as the original errors:

$$L_1(e_{model,i}) = y_i - [E_{model}(y_i)] \tag{42}$$

Second, quantifying the losses as the quadratic errors to avoid positive values compensate negative ones:

$$L_2(e_{model,i}) = (y_i - [E_{model}(y_i)])^2 \tag{43}$$

Last, quantifying the losses as the costs that may involve these forecast errors:

$$L_3(e_{model,i}) = (y_i - [E_{model}(y_i)]) * Term_i * Quantity_i \tag{44}$$

*where i = 1, 2, ..., M.

The last loss function is proposed as it seems better specified for this particular research. As Christoffersen et al. [2001] argue, the choice of the loss-functions used to test the forecast of a model should not be underestimated. The authors encourage choosing a function in line with the purpose of the model. So, this final loss function evaluates a model's performance better as it has a deep relation with the objective of the pricing. It penalizes more the deviations from the real values of contracts with higher terms or with higher quantities agreed for the exchanges. The loss function weights the forecast errors giving more importance to the ones where the real money losses will be more significant.

After that, in a different way as the usual procedure used in the Diebold-Mariano test, an extension is applied. The statistic can be calculated by computing a linear regression of the differential (applying the loss function to the forecast error) on a constant but under a robust estimation where heteroscedasticity and autocorrelation robust standard errors (HAC) are used. For that purpose, there is available a function in Matlab called *hac*. The function returns the robust estimation, the vector of corrected coefficient standard errors, and the coefficient estimates. So, the robust estimator of the t-stat is obtained by dividing the coefficient by the robust standard error. After that, the p-value is easily obtained assuming a t-student distribution with N - 1 degrees of freedom (being N the number of observations) and taking into account that the test is of two tails.

The null hypothesis is that the losses caused by the forecast errors of each model are statistically not different from zero. On the other way, the alternative hypothesis (if the null is rejected) is that the losses are different from zero.

Annex 4: Intra-day marginal models and fitted copulas

Model	A_0	A_1	A_2	A_3	ARMA	LB	GARCH	LB	Distribution	KS	CvM	KS_2	CvM_2	LR
D1H07	-4.54**	0.53**	-1.76**	-5.23**	MA(1,2)	0.0000	GARCH(1,1)	0.9648	Stable(1.77,-0.03,0.82,-0.11)	0.07	0.07	0.07	0.07	0.32
D1H08	-1.51**	0.24**	-0.96**	-	AR(1,3,7)	0.0763	GJR-GARCH	0.9438	Stable(1.44,-0.67,1.25,0.55)	0.10	0.26	0.11	0.26	0.52
D1H09	-0.01	0.10**	-0.63**	-	AR(1,3)	0.5664	GARCH(1,1)	0.9628	Stable(1.58,-0.78,1.27,0.54)	0.41	0.49	0.46	0.51	0.33
D1H10	0.65**	0.01	-0.61**	-	AR(1,3)	0.9744	GARCH(1,1)	0.7789	Stable(1.60,-0.83,1.26,0.58)	0.19	0.16	0.18	0.16	0.59
D1H11	0.95**	-0.01	-0.74**	-	AR(1,3)	0.8654	GJR-GARCH	0.5935	Stable(1.70,-0.98,1.33,0.50)	0.49	0.43	0.48	0.45	0.21
D1H12	1.09**	-0.04	-0.87**	-	AR(1,3)	0.8363	GJR-GARCH	0.6609	Stable(1.73,-0.96,1.39,0.43)	0.69	0.69	0.74	0.71	0.23
D1H13	1.11**	-0.06**	-0.93**	-	AR(1,3)	0.9151	GARCH(1,1)	0.9069	Stable(1.73,-0.94,1.37,0.44)	0.17	0.45	0.19	0.47	0.25
D1H14	1.07**	-0.07**	-0.94**	-	AR(1,3)	0.9409	GARCH(1,1)	0.8911	Stable(1.74,-0.87,1.34,0.39)	0.33	0.37	0.37	0.39	0.23
D1H15	0.94**	-0.06*	-0.98**	-	AR(1,3)	0.5268	GARCH(1,1)	0.6754	Stable(1.76,-0.80,1.29,0.33)	0.23	0.14	0.25	0.16	0.22
D1H16	0.55**	0.01	-1.21**	-	AR(1,3)	0.1110	GJR-GARCH	0.9084	Stable(1.73,-0.85,1.28,0.36)	0.20	0.14	0.26	0.15	0.25
D1H17	-0.38**	0.24**	-2.11**	-	ARMA(1,[2])	0.0000	GARCH(1,1)	0.4604	Stable(1.56,-0.58,0.91,0.21)	0.48	0.38	0.61	0.42	0.39
D1H18	-1.29**	0.11*	-2.41**	-4.98**	AR(1,3)	0.0132	GARCH(1,1)	0.9998	Stable(1.69,-0.40,0.58,0.14)	0.12	0.19	0.12	0.19	0.01
D1H19	-2.89**	0.50**	-2.76**	-3.39**	AR(1,3)	0.0000	GARCH(1,1)	0.9989	Stable(1.38,-0.22,0.96,-0.07)	0.13	0.09	0.17	0.08	0.79
D1H20	-5.58**	0.59**	-2.94**	-2.09**	ARMA(1,1)	0.0000	GJR-GARCH	0.9851	Stable(1.25,-0.08,0.68,-0.29)	0.17	0.21	0.18	0.20	0.61
D1H21	-7.37**	0.74**	-0.39**	-	AR(1)	0.0000	GJR-GARCH	0.2567	Stable(1.58,0.26,0.41,0.04)	0.26	0.29	0.47	0.29	0.73

Table 38: Generation marginal models coefficients and goodness of fit measures

*The statistic significance of the parameters is denoted with ** when the parameter is significant at a 1% level of confidence and with * at a 5%. Additionally, in all the marginal models the coefficients of the parameters for the mean and variance models are not shown due to the lack of space, but in all the cases the parameters included are significant at a 1% level.

** The GJR-GARCH off the series 16, does not have an ARCH parameter.

Model	D1H07	D1H08	D1H09	D1H10	D1H11	D1H12	D1H13	D1H14	D1H15	D1H16	D1H17	D1H18	D1H19	D1H20	D1H21
A_0	42.4**	36.5**	27.6**	23.7**	22.6**	22.6**	23.6**	25.5**	29.1**	34.5**	42.6**	58.5**	74.1**	74.6**	58.18**
A_1	-0.51	-2.43**	-3.47**	-5.02**	-6.89**	-9.06**	-11.5**	-13.6**	-16.3**	-18.0**	-20.1**	-25.0**	-32.7**	-20.7**	-5.1
A_2	14.69**	15.23**	11.92**	8.28**	5.16**	2.26**	-1.01	-2.92**	-4.11*	-1.51	3.42	10.38**	-6.3	-19.0**	-5.11*
L	2.3	3.2*	2.4*	2.6**	3.0**	3.2**	3.9**	4.2*	3.4	6.7	7.8	8.7	8.1	9.2	7.1
M	3.8	4.1**	3.0**	3.1**	2.9**	3.1**	2.9	3.1	7.7*	6.9	9.6*	10.1	9.9	11.9	8.7
X	6.4**	4.8**	3.1**	3.2**	3.0**	2.7*	2.4	2.3	1.5	0.6	0.2	3.3	5.0	5.2	3.3
J	2.3	2.0	1.8	1.6	1.7	1.9	1.5	1.4	0.7	-0.3	-0.3	-1.0	-0.6	-2.2	-3.2
V	-4.7*	-4.7**	-3.1**	-2.7**	-2.9**	-3.3**	-3.5*	-3.7*	-4.5	-5.0	-5.9	-6.0	-7.4	-9.0	-6.0
S	-7.7**	-7.4**	-5.6**	-6.0**	-5.9**	-6.1**	-6.3**	-6.1**	-6.7*	-6.9	-9.2*	-11.7	-11.3	-11.2	-7.4
D	-2.2	-2.0	-1.6	-1.7	-1.7	-1.5	-0.8	-1.1	-1.9	-1.8	-2.0	-3.1	-3.5	-3.6	-2.3
ARMA	ARMA([1,7],[1])	ARMA(1,[7])	ARMA(1,[7])	ARMA([1,7],[7])	ARMA([1,7],[7])	ARMA([1,7],[7])	AR(1,7)	AR(1,7)	ARMA([1],[7])	ARMA([1,7],[2])	AR(1,7)	ARMA([1,7],[1])	AR(1,7)	ARMA([1],[4,7])	AR(1,7)
LB	0.0000	0.0000	0.0001	0.0019	0.0175	0.0651	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH	GJR-GARCH
LB	0.0635	0.0000	0.0012	0.0004	0.0009	0.0000	0.0002	0.0003	0.0502	0.2493	0.6394	0.0446	0.4196	0.2660	0.0550
Distribution	Stable (1.57,-0.02, 0.10,-0.01)	Stable (1.60,0.08, 0.10,-0.01)	Stable (1.67,0.01, 0.11,-0.01)	Stable (1.73,0.00, 0.11,-0.01)	Stable (1.63,-0.01, 0.10,-0.01)	Student-t (-0.01,0.12,4.04)	Stable (1.63,-0.06, 0.09,-0.01)	Stable (1.62,-0.08, 0.09,0.00)	Stable (1.61,-0.13, 0.07,0.00)	Stable (1.60,-0.19, 0.07,0.00)	Stable (1.75,0.04, 0.06,-0.01)	Stable (1.64,0.01, 0.05,0.00)	Stable (1.57,0.14, 0.05,-0.01)	Stable (1.58,0.13, 0.05,-0.01)	Stable (1.59,-0.02, 0.08,-0.01)
KS	0.18	0.07	0.23	0.65	0.35	0.22	0.20	0.09	0.18	0.09	0.05	0.11	0.19	0.11	0.18
CvM	0.13	0.08	0.30	0.45	0.25	0.37	0.16	0.07	0.13	0.09	0.15	0.12	0.13	0.11	0.11
KS_2	0.20	0.07	0.28	0.81	0.47	0.27	0.29	0.14	0.21	0.13	0.06	0.16	0.20	0.15	0.18
CvM_2	0.14	0.08	0.34	0.52	0.26	0.37	0.17	0.08	0.14	0.10	0.15	0.13	0.14	0.12	0.12
LR	0.94	0.94	0.57	0.53	0.70	0.03	0.79	1.02	0.64	0.81	0.01	0.58	0.73	0.84	1.23

Table 39: Price marginal models coefficients and goodness of fit measures

	Student-t Copula							Gaussian Copula					
	ρ	ν	LogL	AIC	BIC	KS-test (statistic)	CvM-test (statistic)	ρ	LogL	AIC	BIC	KS-test (statistic)	CvM-test (statistic)
07	0.0250	204	20.78	-39.56	-34.59	0.0229	0.0556	0.0212	0.24	1.50	6.47	0.0227	0.0583
08	-0.0725	67.4	2.66	-3.32	1.65	0.0231	0.0612	-0.0713	2.55	-3.09	1.87	0.0234	0.0633
09	-0.1437	27.2	10.96	-19.92	-14.95	0.0232	0.0341	-0.1418	10.32	-18.64	-13.67	0.0241	0.0383
10	-0.1329	13.7	10.15	-18.30	-13.33	0.0234	0.0478	-0.1281	8.26	-14.51	-9.54	0.0250	0.0587
11	-0.0866	15.0	5.70	-9.38	-4.42	0.0240	0.0421	-0.0817	3.44	-4.87	0.10	0.0244	0.0416
12	-0.0516	24.0	2.04	-2.08	2.88	0.0233	0.0342	-0.0481	1.19	-0.37	4.59	0.0240	0.0395
13	-0.0364	13.8	2.87	-3.75	1.22	0.0212	0.0349	-0.0349	0.63	0.75	5.72	0.0218	0.0415
14	-0.0394	20.8	1.76	-1.53	3.44	0.0236	0.0697	-0.0381	0.75	0.51	5.48	0.0238	0.0758
15	-0.0461	17.8	2.42	-2.84	2.13	0.0278	0.0783	-0.0340	0.95	0.10	5.08	0.0283	0.0876
16	-0.0864	44.3	4.02	-6.05	-1.08	0.0249	0.0877	-0.0857	3.76	-5.52	-0.55	0.0254	0.0911
17	-0.5860	29.6	2.28	-2.55	2.42	0.0219	0.0355	-0.0613	1.92	-1.83	3.13	0.0255	0.0382
18	-0.0771	19.4	4.14	-6.28	-1.31	0.0266	0.0317	-0.0725	2.83	-3.66	1.31	0.0255	0.0349
19	-0.0081	20.0	0.86	0.26	5.23	0.0228	0.0850	-0.0074	0.03	1.95	6.92	0.0237	0.0899
20	0.0093	12.5	2.11	-2.22	2.74	0.0236	0.0577	0.0041	0.01	1.98	6.95	0.0242	0.0624
21	-0.0587	204.2	56.44	-110.9	-105.9	0.0289	0.1022	-0.0596	1.76	-1.52	3.44	0.0290	0.1026

Table 40: Fitted copula parameters and goodness of fit measures of the intra-day solar model.