UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede amministrativa: Universitá degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

CORSO DI DOTTORATO DI RICERCA IN SCIENZE MATEMATICHE

CURRICOLO: MATEMATICA

CICLO XXXIII

Modeling and Forecasting Electricity Market Variables Using Functional Data Analysis

Coordinatore: Ch.mo Prof. Martino Bardi

Supervisore: Ch.mo Prof. Tiziano Vargiolu

Dottorando: Mariia Soloviova

Abstract

In this thesis we use a relatively new modeling technique based on functional data analysis for demand and price prediction. The basic novelty of our problem is that we are going to predict not just a value at some point, but a whole function of the price depending on the cumulative offered quantity. As far as we know, non-parametric mesh-free interpolation techniques were never considered for the problem of modeling the daily supply and demand curves. The main goal of this thesis is to model and forecast the whole supply and demand curves and the variables related to electricity markets, such as prices and demand. We will show that the forecasting of the whole curves gives deep insight into the electricity market and allows to improve the accuracy of forecasting.

Chapter 1 provides a brief overview of previous research on short term forecast. Short term forecast proved to be a very challenging task due to some specific features. In the literature, different methods have been discussed. Functional data analysis is extensively used in other fields of science, but it has been not much explored in the electricity market setting.

In **Chapter 2** the mathematical preliminaries regarding the infinite dimensional stochastic processes relevant for this thesis are provided. Mainly, we follow the monograph by Bosq, which introduces functional linear time series.

Chapter 3 describes radial basis function interpolation techniques. The first task in our thesis is to elaborate an appropriate algorithm to present the information about electricity prices and demands, in particular to approximate a monotone piecewise constant function. This problem is similar to another one already studied in numerical analysis, in particular in the context of approximation theory with meshless methods. The use of radial basis functions have

attracted increasing attention in recent years as an elegant scheme for high-dimensional scattered data approximation, an accepted method for machine learning, one of the foundations of mesh-free methods and so on.

In **Chapter 4** we present a parsimonious way for representing supply and demand curves, using a mesh-free method based on radial basis functions. Using the tools of functional data analysis, we are able to approximate the original curves with far less parameters than the original ones. Namely, in order to approximate piecewise constant monotone functions, we are using linear combinations of integrals of Gaussian functions.

We also test this new approach with the aim of forecasting supply and demand curves and finding the intersection of the predicted curves in order to obtain the market clearing price. In assessing the goodness of our method, we compare it with models with similar complexity in terms of dependence of the past, but only based on the clearing price. Our forecasting errors are smaller compared with these univariate models. In particular, our analysis show that our multivariate approach leads to better results than the univariate one in terms of different error measures.

In **Chapter 5** we consider supply and demand curves as stochastic processes with values in a functional space. In order to deal with the huge amount of bid data, we study linear transformations of multivariate stochastic processes. It is a known fact that a linear transformation of a vector ARMA process is again an ARMA process. However, in general, there are transformations of a finite order AR(p) process that do not admit a finite order AR representation, but just a mixed ARMA representation. We obtained a characterization result regarding the conditions that guarantees that a linear transformation of a vector AR process is again an AR process both in finite and in infinite dimension, and we apply these results to the model of Ziel and Steinert from [75].

Riassunto

In questa tesi utilizziamo una tecnica di modellizzazione relativamente nuova basata sull'analisi dei dati funzionali per la previsione della domanda e dei prezzi. La novità fondamentale del nostro problema è che prediremo non solo un valore in un determinato punto, ma un'intera funzione del prezzo che dipende dalla quantità cumulativa offerta. Per quanto ne sappiamo, le tecniche di interpolazione senza mesh non parametriche non sono mai state prese in considerazione per il problema della modellizzazione delle curve di domanda e offerta giornaliere. L'obiettivo principale di questa tesi è modellizzare e prevedere tutte le curve di domanda e offerta e le variabili relative ai mercati elettrici, come i prezzi e la domanda. Dimostreremo che la previsione di tutte le curve fornisce una visione approfondita del mercato elettrico e consente di migliorare l'accuratezza delle previsioni.

Il **Capitolo 1** fornisce una breve panoramica delle ricerche precedenti sulle previsioni a breve termine. Le previsioni a breve termine si sono rivelate un'attività molto impegnativa a causa di alcune caratteristiche specifiche. In articoli del settore sono stati discussi diversi metodi. L'analisi dei dati funzionali è ampiamente utilizzata in altri settori disciplinari, ma è stata poco esplorata nel contesto del mercato elettrico.

IL **Capitolo 2** presenta i preliminari matematici riguardanti i processi stocastici a dimensione infinita rilevanti per questa tesi. Principalmente, seguiamo la monografia di Bosq, che introduce serie storiche lineari funzionali.

Il **Capitolo 3** descrive le tecniche di interpolazione delle funzioni radiali di base. Il primo compito per la nostra tesi è quello di creare un algoritmo appropriato per presentare le informazioni sui prezzi e le richieste dell'elettricità, in particolare per approssimare una funzione monotona costante a tratti. Questo problema è simile ad un altro già studiato in analisi numerica, in particolare nell'ambito della teoria dell'approssimazione con metodi meshless. Negli ultimi anni l'uso delle funzioni radiali di base ha attirato una crescente attenzione

come metodo elegante per l'approssimazione di dati sparsi ad alta dimensione, un metodo accettato per machine learning, uno dei fondamenti dei metodi meshless etc.

Nel **Capitolo 4** presentiamo un metodo parsimonioso per rappresentare le curve di domanda e offerta, usando un metodo meshless basato su funzioni radiali di base. Utilizzando gli strumenti di analisi dei dati funzionali, siamo in grado di approssimare le curve originali con molti meno parametri di quelli iniziali. Per approssimare funzioni monotone costanti a tratti, stiamo usando combinazioni lineari di integrali di funzioni gaussiane.

Inoltre, testiamo questo nuovo approccio con l'obiettivo di prevedere le curve di domanda e offerta e trovare l'intersezione delle curve previste per ottenere il prezzo di equilibrio di mercato. Nel valutare l'efficacia del nostro metodo, lo confrontiamo con modelli con complessità simile in termini di dipendenza dal passato, ma basati solo sul prezzo di equilibrio di mercato. I nostri errori di previsione sono minori rispetto a questi modelli univariati. In particolare, la nostra analisi mostra che il nostro approccio multivariato porta a risultati migliori rispetto a quello univariato in termini di diverse misure di errore.

Nel **Capitolo 5** consideriamo le curve di domanda e offerta come processi stocastici con valori in uno spazio funzionale. Per gestire l'enorme quantità di dati di offerta, abbiamo studiato trasformazioni lineari di processi stocastici multivariati. È noto che una trasformazione lineare di un processo ARMA vettoriale è di nuovo un processo ARMA. Tuttavia, in generale, ci sono trasformazioni di un processo AR(p) di ordine finito che non ammettono una rappresentazione AR di ordine finito, ma solo una rappresentazione ARMA mista. Abbiamo ottenuto un risultato di caratterizzazione relativo alle condizioni che garantiscono che una trasformazione lineare di un processo AR vettoriale sia ancora un processo AR sia di dimensione finita che di dimensione infinita, e applichiamo questi risultati al modello di Ziel e Steinert da [75].

Acknowledgements

First of all, I would like to express my deep gratitude to my scientific advisor Prof. Tiziano Vargiolu for his help during the Ph.D. program. Without his constant support and encouragement this work would not have been done.

The material appearing in this Thesis has been presented at various conferences and workshops. I would like to thank the organizers and the conference participants at Energy Finance Christmas Workshop (2018) in Bolzano; Quantitative Finance Workshop (2019) in Zurich; Energy Finance Italia Workshop in Milan (2019), the Freiburg-Wien-Zurich Workshop (2019) in Padova.

I am also thankful for helpful suggestions and discussions to Enrico Edoli, Marco Gallana, and Emma Perracchione. The authors wish to thank also Florian Ziel, Carlo Lucheroni, Stefano Marmi, Sergei Kulakov, Enrico Moretto for their comments and suggestions.

I am very grateful to my first research supervisor, Dr. V. M. Kadets at Karazin Kharkiv National University, Ukraine, for introducing me to the beauty of theoretical mathematics and encouraging me to take the new path that brought me to Italy.

I also acknowledge the Fondazione Cassa di Risparmio di Padova e Rovigo (CARIPARO) which has funded my Ph.D. and gave me the possibility to study and collaborate with talented youth and excellent tutors of the Department of Mathematics of Padova University during these three years.

Finally, to my family, thank you so much for being always there for me.

Contents

Notation

Throughout the dissertation, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively, the symbol $\mathbb{N}$ denotes the set of natural numbers, $\mathbb{Z}$ denotes the set of integers.

H denote a real separable Hilbert space with its norm $\|\cdot\|$ and its scalar product $\langle\cdot,\cdot\rangle$.

$C(K)$ – the space of continuous real-valued functions on K, with the norm

$$\|f\| = \max_{t\in K}|f(t)|.$$

$\mathscr{B}(\Omega)$ – the Borel σ-algebra on a topological space Ω.

$L(X,Y)$ – the space of continuous linear operators from X to Y with the norm

$$\|T\| = \max_{x\in X}\frac{\|Tx\|_Y}{\|x\|_X}.$$

$L(X)$ – the space of continuous linear operators from X to X.

T^* – adjoint operator of T.

$(\Omega, \mathscr{A}, P)$ – a probability space

$\mathbf{1}_A\colon X \to \{0,1\}$ – the indicator function of a subset, which for a given subset A of X, has value 1 at points of A and 0 at points of X/A

Chapter 1

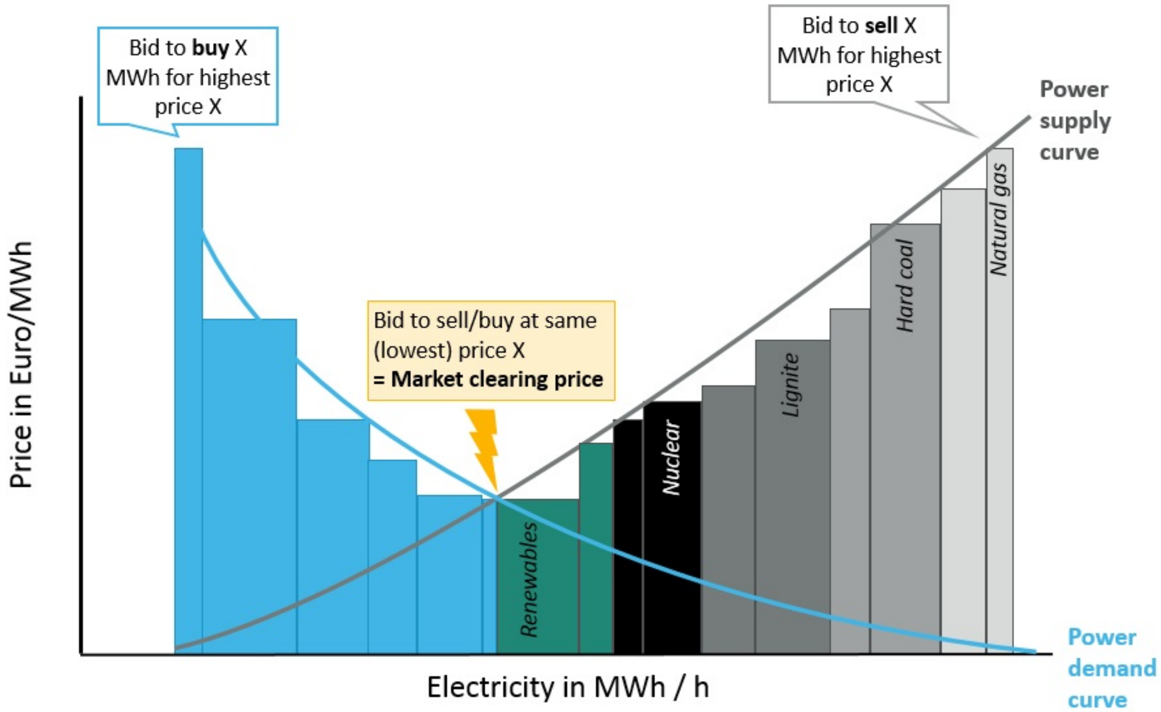
Introduction

1.1 Supply and demand curves

In microeconomics, supply and demand is an economic model of price determination in a market. An equilibrium is defined to be the price-quantity pair where the quantity demanded is equal to the quantity supplied. It is represented by the intersection of the demand and supply curves.

Before liberalization of the electric sector, when the market was highly regulated and controlled by state owned companies, the electric utilities were mainly interested in efficient forecasting of electric load as the variation in the electricity prices was minimal and changes in prices were considered after regular time intervals. However, because generation is actually a competitive market with upward-sloping supply curves, it does not need to be regulated as part of a "rate case", as is the case for distribution and transmission. So, in most of the country, in the 1980s and 1990s, the generation part of the system was sold off or spun off into separate companies. These companies are called "merchant generators" or "unregulated generators", because they are selling at marginal cost into a competitive marketplace, bidding against other firms. They are not natural monopolies that need to be controlled by public utility commissions. There is lots of evidence that market systems for generation deliver lower costs and better service, but many areas are comfortable keeping generation under the control of utility commissions. They are trading off lower costs and potential innovations for stability and less price volatility.

Figure 1.1: Supply and demand curves in electricity market



1.2 Price formation process

Consider, for example, the Italian electricity market (IPEX). IPEX consists of different markets, including a day-ahead market. The day-ahead market is managed by Gestore del Mercato Elettrico (GME) where prices and demand are determined the day before the delivery by means of hourly concurrent auctions. For each delivery day the market session starts at 8 a.m. of the ninth day before the day of physical delivery and closes at mid-day (12 a.m.) of the day before delivery.

The producers submit **offers** where they specify the quantities and the minimum price at which they are willing to sell. The demanders submit **bids** where they specify the quantities and the maximum price at which they are willing to buy. These bids and offers typically consist of a set of energy blocks and their corresponding prices with other relevant information for every hour of the next day and they are submitted through an online web portal called "market participant interface" that is also used to manage and display invoicing data and payables/receivables resulting from transactions that are already concluded in the previous days. They are then aggregated by an independent system operator (ISO) in order to construct the supply and demand curves. Only one agent is responsible for this task and his role is very important for many reasons including reliability, independence, non-discrimination, unbundling and efficiency. ISO ensures reliability of power grid by coordinating short term operations, independence by not allowing any entity to control the criteria or operating procedures and non-discriminatory access for all market participants without distinction as to customer identity or affiliation. Services unbundling for utilization by the market participants and efficient operating procedures and pricing of services are also responsibility of ISO.

Since bidders are expected to buy electricity at lower prices and sell at higher prices, corresponding quantities in the hourly bid must be a non-increasing sequences. In a competitive market, each generator enters bids for how much of its output power it wants to sell at what price. That is, each generator gives an individual supply curve to the system operator. The system operator is a quasi-governmental non-profit firm that is responsible for collecting all of the bids, arranging them in ascending order of price, and then figuring out which power plants shall be turned on, and when.

When we add together each individual supply curve, we are left with an aggregate supply curve that is called a "generation stack" - literally, all of the generators are "stacked up" in ascending order of marginal cost, and only the lowest cost ones necessary to meet expected demand will be turned on the next day. Although an hourly bid consists of a discrete set of quantity price pairs, it is in fact a **monotone increasing piecewise constant function**. This is done on a "day-ahead" basis, where generators enter their bids for tomorrow and, after computer runs, are told if and when they will be expected to turn on the next day.

In electricity markets, the demand side is called the "load". The load is simply the sum of all demands for electricity in a market at any given time. Load changes continuously as people turn devices on and off, as temperature changes, as the natural light comes and goes, and so on. This pattern of changing load is called a "load shape". We can have daily load shapes, weekly ones, and annual ones. The demand curve for electricity was classically represented as a **vertical line**, i.e. a perfectly inelastic demand curve. However, in recent work [40] a study of wholesale demand elasticities were conducted.

Once the offers and bids are received by the ISO, supply and demand curves are established by summing up individual supply and demand schedules. In the case of demand, the first step is to replace "zero prices" bids by the market maximum price (for IPEX, the market maximum price is 3000 Euro) without changing the corresponding quantities. After this replacement, the bids are sorted from the highest to the lowest with respect to prices. The corresponding value of the quantities is obtained by cumulating each single demand bid. For supply curve, in contrast, the offers are sorted from the lowest to the highest with respect to prices and the corresponding value of the quantities is obtained by cumulating each single supply offer. The market equilibrium is the point where both curves intersect each other and the price balances supply and demand schedules (see Figure 1.2). This point determines the market clearing price and the quantity. Accepted offers and bids are those that fall to the left of the intersection of the two curves and all of them are exchanged at the resulting price.

However, at GME the equilibrium price is different from the market clearing price as the latter accounts for other transactions, e.g. transmission capacity limits between zones, electricity imports from other countries etc. All demand bids and supply offers pertaining to both, pumping unit and consuming units, belonging to foreign virtual zones, that are accepted, are valued at the marginal clearing price of the zone to which they belong. The accepted demand bids pertaining to consuming units belonging to Italian geographical zones are valued at the "Prezzo Unico Nazionale" (national single price, PUN); this price is equal to the average of the prices of geographical zones, weighted for the

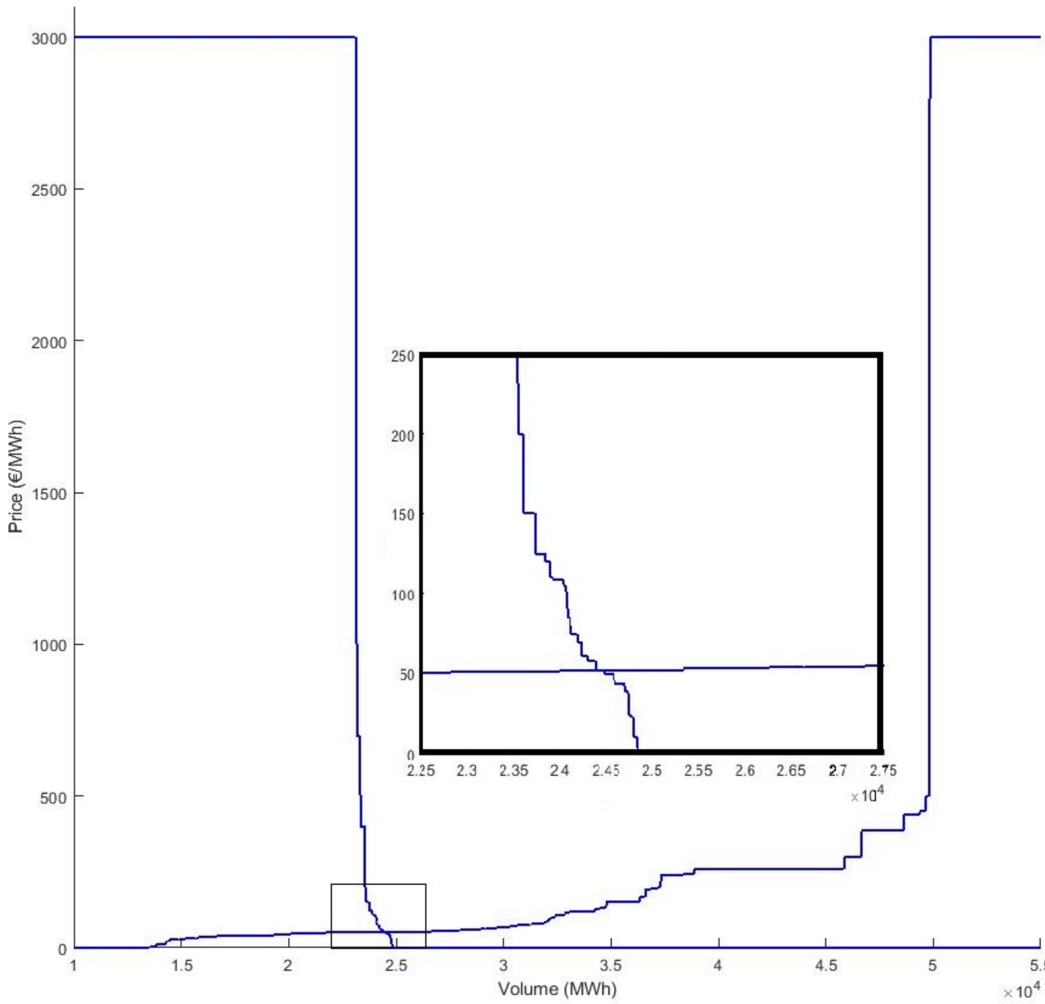


Figure 1.2: The market equilibrium point

quantities purchased in those zones (more information on the GME website www.mercatoelettrico.org). The results (market clearing prices and quantities for each hour for the following day) of the day-ahead market (MGP) are made available within 12.55 p.m. of the day before that of delivery.

1.3 Literature review

In the beginning of the 2000s the amount of papers focused on electricity price forecasting started to increase dramatically. A great variety of methods and models occurred during last twenty years. Weron [74] made an overview of the existing literature on electricity price forecasting. Electricity price models in literature can be broadly classified under the following classes:

1. Multi-agent models, which simulate the operation of a system of heterogeneous agents (generating units, companies) interacting with each other, and build the price process by matching the demand and supply in the market.

2. Fundamental models, which describe the price dynamics by modeling the impacts of important physical and economic factors on the price of electricity. These models manifest electricity price dynamics by incorporating and modeling impact of all physical factors and economic factors. These models are believed to be better suited for medium-term electricity price forecasting compared to short term electricity price modeling and forecasting.

3. Quantitative models, which characterize the statistical properties of electricity prices over time, with the ultimate objective of derivatives evaluation and risk management. These models have their practical application in valuation of derivatives and for risk management motive and purpose.

4. Statistical approach. These techniques are direct applications of methods inspired by electrical load forecasting or time series econometric models. The effectiveness, efficiency and appropriate usefulness of adopting technical analysis approach is often questioned in financial markets, however, the same techniques stand better chance in power markets irrespective of the time period considered. Statistical models are attractive because some physical interpretation may be attached to their components, thus allowing engineers and system operators to understand their behavior.

5. Artificial Intelligence techniques. In these techniques, spot electricity prices are modeled by adopting neural networks, expert systems, support vector machines, fuzzy logic etc which are non-parametric tools having the advantage of being flexible and capable of handling complexity and most

importantly non-linearity. Being non-intuitive and often performing below par has been their biggest drawback.

Forecasting models for electricity prices also can be classified on the base of the time frame for which prediction of electricity price needs to be done as follows:

1. Forecasting of electricity prices for long-term (more than 1 year). The prime objective is for analyzing and planning long term investment and political decisions.

2. Forecasting of electricity prices for medium-term (3 months to 1 year). These classes of models are normally favored for balance-sheet calculations, derivatives pricing and also risk management. The focus is on distributions of future electricity prices for medium term rather than exact point forecasts.

3. Forecasting of electricity prices for short-term price (up to 3 months). Power market participants belonging to auction-type spot markets are particularly interested with forecasting of electricity prices for short-term where they should participants communicate their bids quoting the price for buying/selling along with quantities. Statistical models and artificial intelligence based approaches are useful for short-term electricity price forecasting purpose.

The multi-agent (or equilibrium) models, and hybrid models which, given the particular characteristics of electricity, explain price formation based on state variables that are mainly associated to supply and demand. For example, Pirrong and Jermakyan (1999) [54] and Pirrong and Jermakyan (2000) [?] proposed to model the equilibrium price as a function of two state variables, electricity demand and the futures price of the marginal fuel. Moreover, the authors considered that electricity prices should be an increasing and convex function of demand.

Bessembinder and Lemmon (2002) [6] adopted an equilibrium perspective and explicitly modeled the economic determinants of the forward market. In their model, producers face marginal production costs that may increase steeply with output and aggregate demand is exogenous and stochastic. They showed that the forward premium, defined as the forward minus the expected spot price, is positively (resp. negatively) related to the skewness (resp. variance) of the spot price.

Longstaff and Wang (2004) [43] focused on the question of how electricity forward prices are related to expected spot prices. Their goal was to provide an empirical analysis of the theoretical predictions presented in Bessembinder and Lemmon (2002) [6]. They found a significant forward premium in the PJM market which they consider as being the result of ”the rationality and risk aversion of economic agents participating in the market“. They pointed out that ”total demand approaching or exceeding the physical limits of power generation“ is an important economic risk (related also to quantity risk) and ”the risk of price spikes as demand approaches system capacity is an extreme type of risk which may have important implications for the relation between spot and forward prices“. Therefore in those situations where the demand level is near the maximum capacity of the system, the behavior of electricity prices can be quite abrupt, since electricity must be generated by plants with higher marginal costs (convexity of the supply function). Barlow (2002) [4] proposed a non-linear Ornstein-Uhlenbeck process for the description of observed electricity prices.

In 2007 A.Cartea and P.Villaplana [18] proposed a model for the electricity spot price as a function of demand and generation capacity . They derived analytical expressions to price forward contracts and to calculated the forward premium. They applied their model to the PJM, England and Wales, and Nord Pool markets. They assumed that both volatility of capacity and the market price of capacity risk are constant and found that, depending on the market and period under study, it could either exert an upward or downward pressure on forward prices. Most models have in common that they focus on the price itself or related time series. In such a way these models does not take into account the underlying mechanic which determines the price process – the intersection between the part of the electricity supply and demand. Some of the recent approaches try to to analyse the real offered volumes for selling and purchasing electricity. This commonly leads to a problem of a large amount of data and, therefore, high complexity.

Eichler, Sollie, Tuerk in 2012 [25] investigated a new approach that exploits information available in the supply and demand curves for the German day-ahead market. They proposed the idea that the form of the supply and demand curves or, more precisely, the spread between supply and demand, reflects the risk of extreme price fluctuations. They utilize the curves to model a scaled supply and demand spread using an autoregressive time series model in order to construct a flexible model adapted to changing market conditions. Furthermore, Aneiros, Vilar, Cao, San Roque in 2013 [2] dealt with the prediction of residual demand curve in elecricity spot market using two functional models. They tested this method as a tool for optimizing bidding strategies for the Spanish day-ahead market.

In 2016 Shah [65] also considered the idea of modeling the daily supply and demand curves, predicting them and finding the intersection of the predicted curves in order to find the predicted market clearing price and volume. He used the functional approach, namely, B-spline approximation, to convert the resulted piece-wise constant curves into smooth functions.

In 2016 Ziel and Steinert described and showed a new methods for the day-ahead electricity market of Germany and Austria [75]. Instead of directly modeling the electricity price, they modeled and utilized its true source: the sale and purchase curves of the electricity exchange. They analyzed the hourly day-ahead electricity price auction data of Germany and Austria provided by the EPEX Spot from 01.10.2012 to 19.04.2015, using a subtle data processing technique as well as dimension reduction and lasso-based estimation methods. Their model consists of three parts:

1. Construction of price classes in order to overcome the massive amount of data.

2. Forecasting for each price class by using time series model.

3. Reconstruction of supply and demand curves and computation of market clearing price.

We describe the model of Ziel and Steinert in more details in Chapter 5.

1.4 Our approach to price prediction

Short term forecast proved to be very challenging task due to these specific features. Figure 1.3 and 1.4 demonstrate changing of electricity equilibrium price and quantity during one week. The hourly load forecasting of the next 24 up to 48 hours ahead or more is needed to support basic operational planning functions, such as spinning reserve management and energy exchanges, as well as network analysis functions related to system security, such as contingency analysis. Functional data analysis is extensively used in other fields of science, but it has been little explored in the electricity market setting.

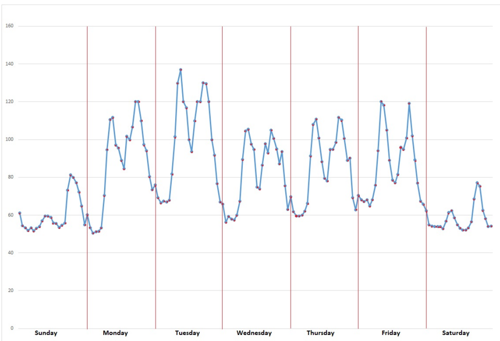


Figure 1.3: Electricity equilibrium prices during a week

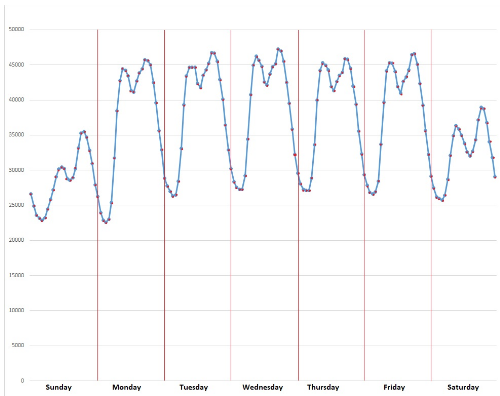


Figure 1.4: Electricity equilibrium quantities during a week

In this thesis we are going to use a relatively new modeling technique based on functional data analysis for demand and price prediction. The basic novelty of our problem is that we are going to predict not just a value at some point, but a whole function of the price depending on cumulative offered quantity. The first task for this purpose is to make an appropriate algorithm to present the information about electricity prices and demands, in particular to approximate a monotone piecewise constant function. This problem is similar to another one already studied in numerical analysis, in particular in the context of approximation theory with meshless methods, namely, approximation by radial basis functions. As far as we know, non-parametric mesh-free interpolation techniques were never considered for the problem of modeling the daily supply and demand curves. The use of radial basis functions have attracted increasing attention in recent years as an elegant scheme for high-dimensional scattered data approximation, an accepted method for machine learning, one of the foundations of meshfree methods and so on. We will show that the forecasting of the whole curves gives deep insight into the electricity market.

After presenting the original supply and demand curves from the Italian day-ahead electricity market with far less parameters than the original ones we will show that there is no direct relationship between the number of offer and bid layers and the hour of the day, the day of the week, and the time of the year. We also will test this new approach with the aim of forecasting supply and demand curves and finding the intersection of the predicted curves in order to obtain the market clearing price. In assess the goodness of our method, we will compare it with models with similar complexity in terms of dependence of the past, but only based on the clearing price.

In order to deal with the huge amount of bid data, we will study linear transformations of multivariate stochastic processes. It is known fact that a linear transformation of a vector ARMA process is again an ARMA process. Instead, a linear transformation of a finite order AR(p) process does not admit in general a finite order AR representation, but just a mixed ARMA representation. We will obtain a characterization result regarding the conditions that guarantee that a linear transformation of a vector AR process is again an AR process both in finite and in infinite dimension. We will then apply them to the model of Ziel and Steinert from [75].

Chapter 2

Mathematical preliminaries for stochastic modelling in large dimension

In this chapter we provide the mathematical preliminaries regarding the stochastic calculus relevant for this thesis. Mainly, we follow the monograph by Bosq [13], which introduces functional linear time series

Let us give a simple example where infinite-dimensional modeling is a useful tool for applications. If one observes temperature in continuous time during N days, and wants to predict its evolution during the $(N+1)$ day, then $(X_n), n \in \mathbb{N}$ is a sequence of random variables with values in a suitable function space, say $C([0, 24])$.

Another example of modeling in large dimensions is the following: consider an economic variable associated with individuals. At instant n, the variable associated with the individual i is $X_{n,i}$. In order to study the global evolution of that variable for a large number of individuals, and during a long time, it is convenient to set

$$X_n = (X_{n,i}, i \geqslant 1), \quad n \in \mathbb{Z},$$

which defines a process $X = (X_n, n \in \mathbb{Z})$ with values in some sequence space F.

2.1 Stochastic processes and random variables in functional spaces

We need to recall the definitions of the main types of vector-valued integrals [22, 23]. The Bochner integral is a straightforward generalization of the Lebesgue integral to Banach space valued functions.

Let B be a Banach space and $(\Omega, \mathscr{A}, \mu)$ be a measurable space. A function $f : \Omega \to B$ is called *simple*, if it is of the form $f = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}$ with $x_n \in B$, and $A_i \in \mathscr{A}$ forming a partition of Ω. A function $f : \Omega \to B$ is said to be *measurable* if $f^{-1}(U) \in \mathscr{A}$ for every Borel subset $U \subset X$; f is said to be *scalarly measurable* if the composition of f with every linear functional is a measurable scalar function; f is said to be *strongly (or Bochner) measurable* if there is a sequence of simple functions converging to f a.e..

For an arbitrary Banach space B we have the following characterization [23, Theorem 3.3]: a random variable $\xi : (\Omega, \mathscr{A}, P) \to B$ is Bochner integrable if and only if ξ is Bochner measurable and $\mathbb{E}\|\xi\| < \infty$. Notice that for separable B strong measurability, scalar measurability and measurability are equivalent, but in non-separable case this equivalence no longer takes place.

Let $(\Omega, \mathscr{A}, P)$ be a probability space. Let H be a real separable Hilbert space with its norm $\|\cdot\|$ and its scalar product $\langle\cdot,\cdot\rangle$, $L(H)$ denote the space of continuous linear operators from H to H, and $\mathscr{B}$ be the Borel σ-algebra generated by the norm topology on the space H.

A mapping $X : \Omega \to H$ is said to be a *random variable taking values in a Hilbert space* H if $X^{-1}(B) \in \mathscr{A}$ for every $B \in \mathscr{B}$. Define

$$P_X(B) = P(X^{-1}(B), B \in \mathscr{B}).$$

P_X is a probability measure on the measurable space$(H, \mathscr{B})$ generated by the random variable X.

We consider the space $L_H^2 := L_H^2(\Omega, \mathscr{A}, P)$ of random variables X, defined on the probability space $(\Omega, \mathscr{A}, P)$, with values in H, and such that $\mathbb{E}\|X\|^2 < \infty$. If $\mathbb{E}\|X\|^2 < \infty$, then the mathematical expectation $\mathbb{E}X$ exists as an element of H (e.g. as a Bochner integral $\int_\Omega X(w)dP(w)$). The mean $\mathbb{E}X$ is the unique element of H such that

$$\langle \mathbb{E}X, h\rangle = \mathbb{E}\langle X, h\rangle \text{ for all } h \in H. \tag{2.1.1}$$

We now list some important properties of the expectation [10].

Proposition 2.1.1. *1. The space L_H^2 of equivalence classes of integrable H-random variables X (with respect to the equivalence relation $X = Y$ a.s.), defined on the probability space $(\Omega, \mathscr{A}, P)$, with values in H, and such that $\mathbb{E}\|X\|^2 < \infty$ is a Hilbert space with scalar product*

$$\langle X, Y\rangle_{L_H^2} = \mathbb{E}\langle X, Y\rangle. \tag{2.1.2}$$

2. $\mathbb{E}$ defines a continuous linear operator from L_H^2 to H, which satisfies the contractive property

$$\|\mathbb{E}X\| \leqslant \mathbb{E}\|X\| \tag{2.1.3}$$

2. Let H_1 and H_2 be two separable Hilbert spaces and let T be a continuous linear operators from H_1 to H_2. If $X \in L_{H_1}^2$, then $T(X) \in L_{H_2}^2$ and

$$\mathbb{E}T(X) = T(\mathbb{E}X) \tag{2.1.4}$$

*4. **Dominated convergence:** If $X_n \to X$ a.s. in H and $\|X_n\| \leqslant Y$ a.s., where $n \geqslant 1$ and Y is an integrable real random variable, then $X_n \in L_H^2, n \geqslant 1$, $X \in L_H^2$ and*

$$\mathbb{E}\|X - X_n\| \to 0. \tag{2.1.5}$$

If X and Y are in L_H^2, the cross-covariance operator of X and Y, which is an infinite dimensional analogous to the covariance matrix, is defined as

$$C_{X,Y}(h) = \mathbb{E}[\langle X - \mathbb{E}X, h\rangle(Y - \mathbb{E}Y)] : H \to H. \tag{2.1.6}$$

The covariance operator $C_{X,X}$ of X is denoted by C_X. The covariance operator C_X is positive symmetric operator, i.e. $\langle C_X x, x\rangle \geqslant 0$ and $\langle C_X x, y\rangle = \langle x, C_X y\rangle$ for all $x, y \in H$.

We can indicate a characterization of covariance operators.

Theorem 2.1.2. *[10, Theorem 1.7] An operator $C : H \to H$ is a covariance operator if and only if it is symmetric, positive, and nuclear (i. e. a compact operator with finite trace).*

Moreover, the following properties holds: if v_i, $i \geqslant 1$, denotes an orthonormal basis of H consisting of eigenvectors, $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant 0$ are corresponding eigenvalues, then C has decomposition:

$$C(h) = \sum_{i=1}^{\infty} \lambda_i \langle h, v_i \rangle v_i, \quad h \in H.$$

$$\sum_{i=1}^{\infty} \lambda_i = \mathbb{E}\|X\|^2 \quad \textit{and} \quad \sum_{i=1}^{\infty} \lambda_i^2 = \sqrt{\sum_{i=1}^{\infty} (\mathbb{E}\langle X, v_i \rangle^2)^2}$$

Remark 2.1.3. Recall that any bounded linear operator in a separable Hilbert space can be viewed as an infinite matrix. Fix an orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ for a separable Hilbert space H. Let $X = \sum_{i=1}^{\infty} x_i e_i, Y = \sum_{i=1}^{\infty} y_i e_i$ and $h = \sum_{i=1}^{\infty} h_i e_i$. We can define

$$YX^T := \begin{pmatrix} y_1x_1 & y_1x_2 & y_1x_3 & \cdots \\ y_2x_1 & y_2x_2 & y_2x_3 & \cdots \\ y_3x_1 & y_3x_2 & y_3x_3 & \cdots \\ \vdots & \vdots & \ddots & \end{pmatrix}.$$

Notice that, if $\mathbb{E}X = \mathbb{E}Y = 0$, then the matrix $\mathbb{E}(YX^T)$ represents the operator C_{XY}. Indeed,

$$\begin{aligned} C_{X,Y}(h) &= \mathbb{E}[\langle X, h \rangle Y] = \mathbb{E}\left[\sum_{i=1}^{\infty} x_i h_i \sum_{j=1}^{\infty} y_j e_j\right] \\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \mathbb{E}(x_i y_j) h_i e_j = \mathbb{E}(YX^T)h. \end{aligned}$$

2.2 Linearly closed subspaces

The problem of linear approximation of a nonobserved random variable X by a linear function of observed random variables $(X_i, i \in I)$ has a simple and well known statement in a finite-dimensional setting.

If $X \in L^2(\Omega, \mathscr{A}, P)$ and $X_i \in L^2(\Omega, \mathscr{A}, P), i \in I$ are zero-mean, the best linear approximation of X is its orthogonal projection over the smallest closed subspace of $L_2(\Omega, \mathscr{A}, P)$ containing $(X_i, i \in I)$. This subspace is the closure of

$$\operatorname{span}\{X_i, i \in I\} = \left\{ \sum_{i \in J} a_i X_i : J \subset I - \text{ finite}, i \in J, a_i \in \mathbb{R} \right\}.$$

If the variables are in $X \in L^2_{\mathbb{R}^d}(\Omega, \mathscr{A}, P)$, the usual procedure is to consider the closed subspace generated by the components of the observed random vectors and then to project each component of the nonobserved random vector.

More generally, in an infinite-dimensional Hilbert space it is convenient to project over a rich enough subspace of $L^2_H(\Omega, \mathscr{A}, P)$. In this context, we introduce the notion of linearly closed subspace (LCS) (or hermetically closed subspace) in $L^2_H(\Omega, \mathscr{A}, P)$.

Definition 2.2.1. $\mathscr{G}$ is said to be a linearly closed subspace of $L^2_H(\Omega, \mathscr{A}, P)$ if

1. $\mathscr{G}$ is a closed subspace of $L^2_H(\Omega, \mathscr{A}, P)$.
2. If $X \in \mathscr{G}$, then $\ell(X) \in \mathscr{G}$ for any $\ell \in L(H)$.

For any random variable $X \in L^2_H$ we can consider the linearly closed subspace generated by X:

$$\mathscr{G}_X = \overline{\operatorname{span}\{\ell(X) : \ell \in L(H)\}}.$$

Note that, in general, $\mathscr{G}_X$ is infinite-dimensional and that the elements of $\mathscr{G}_X$ are not necessary of the form $\ell(X), \ell \in L(H)$.

Example 2.2.2. Let $H = \ell_2$, $\Omega = [0,1]$, $X(w) = we_1 : \Omega \to \ell_2$.

$$\begin{aligned} \mathscr{G}_X &= \overline{\operatorname{span}\{\ell(X) : \ell \in L(H)\}} = \overline{\operatorname{span}\{w \cdot \ell(e_1) : \ell \in L(H)\}} \\ &= \overline{\{w \cdot h : h \in H\}} \subset L^2_H. \end{aligned}$$

So, $\mathscr{G}_X$ is infinite dimensional and $\mathscr{G}_X \neq L^2_H$ $(Y(w) \equiv e_1 \notin \mathscr{G}_X)$.

By $\Pi^{\mathscr{G}_X} = \Pi^X$ we denote the orthogonal projector onto the subspace $\mathscr{G}_X$. Now it is of our interest to give conditions that yields existence of $l \in L(H)$ such that $\Pi^X(Y) = l(X)$ a.s..

Theorem 2.2.3. *[12, p.233] Let X, Y be zero-mean H-valued random variables in L_H^2. Then the following conditions are equivalent:*

(i.) The cross-covariance operator $C_{X,Y}$ is dominated by the covariance operator C_X, i.e. there exists $\alpha \geqslant 0$ such that $\|C_{X,Y}(h)\| \leqslant \alpha \|C_X(h)\|$ for all $h \in H$.

(ii.) There exists $l \in L(H)$ such that $\Pi^X(Y) = l(X)$ (a.s.).

2.3 Stationary processes in Hilbert spaces

Definition 2.3.1. An H-valued process $X = (X_n, n \in \mathbb{Z})$ is said to be *strictly stationary* if the joint probability distribution of X does not change when shifted in time, i.e.

$$P_X = P_{\tau^m(X)}, \quad \forall\, m \in \mathbb{Z},$$

where $\tau^m(X) = (X_{n+m}, n \in \mathbb{Z})$.

Definition 2.3.2. An H-valued process $X = (X_n, n \in \mathbb{Z})$ is said to be *(weakly) stationary* if

1. $\mathbb{E}\|X_n\|^2 < \infty$ and $\mathbb{E}X_n$ do not depend on n;
2. $C_{X_{n+j}, X_{m+j}} = C_{X_n, X_m}$ for any $n, m, j \in \mathbb{Z}$.

Example 2.3.3 (Discrete Ornstein-Uhlenbeck equation in the infinite-dimensional case.)**.** Let $X = (X_n, n \in \mathbb{N})$ be a random ℓ_2-valued vector. Consider the dynamics

$$X_n = AX_{n-1} + W_n,$$

where W_n are independent identically distributed ℓ_2-valued random vectors such that $W_n \sim \mathcal{N}(0, S)$. Suppose that A is invertible with $\|A\| < \infty$. We want to find a covariance operator C_0 such that, if $X_0 \sim \mathcal{N}(0, C_0)$, then (X_n) is weakly stationary.

Evidently, $\mathbb{E}X_n = 0$ for all $n \in \mathbb{N}$. So, we need to guarantee that the covariance matrix is invariant under time shift. We will use the representation

$C_{XY} = \mathbb{E}(YX^T)$. Suppose that $m = n + k, k \geqslant 0$ and calculate $C_{X_m X_n}$.

$$
\begin{aligned}
C_{X_m X_n} &= \mathbb{E}(X_n X_m^T) \\
&= \mathbb{E}\left[(W_n + AW_{n-1} + \ldots + A^n X_0)(W_m + AW_{m-1} + \ldots + A^m X_0)^T)\right] \\
&= \mathbb{E}[(W_n W_n^T + AW_{n-1}W_{n-1}^T A^T + \ldots \\
&\qquad\qquad + A^{n-1} W_1 W_1^T ((A^{n-1})^T) + A^n X_0 X_0^T (A^n)^T)(A^k)^T] \\
&= \left(S + ASA^T + \ldots + A^{n-1} S (A^{n-1})^T + A^n C_0 (A^n)^T\right)(A^k)^T
\end{aligned}
$$

Denote $F(n) = S + ASA^T + \ldots + A^{n-1}S(A^{n-1})^T + A^n C_0 (A^n)^T$. We want to find sufficient condition such that $C_{X_m X_n}$ depends only on k. This is equivalent to the condition $F(n-1) = F(n)$, so, we obtain the equation for C_0:

$$
C_0 = S + AC_0A^T. \tag{2.3.1}
$$

This is the so-called discrete time Lyapunov equation. The solution can be expressed as an infinite sum

$$
C_0 = \sum_{k=0}^{\infty} A^k S (A^T)^k. \tag{2.3.2}
$$

The operator C_0 is defined correctly if $\|A\| < 1$. So, we showed that (X_n) is weakly stationary, if A is an invertible operator with norm less than 1 and C_0 satisfies (2.3.2). Notice that in the one-dimensional case ($A, S, C_0 \in \mathbb{R}$) this means that $C_0 = \sum_{k=1}^{\infty} A^2 S = S/(1 - A^2)$.

Definition 2.3.4. An H-valued process $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is said to be a H-*white noise* if

1. $0 < \mathbb{E}\|\varepsilon_n\|^2 = \sigma^2 < \infty$;
2. $\mathbb{E}\varepsilon_n = 0$,
3. $C_{\varepsilon_n} := C_\varepsilon \neq 0$ do not depend on $n \in \mathbb{Z}$;
4. ε_n are pairwise orthogonal in the strong sense

$$
\mathbb{E}\left(\langle \varepsilon_n, x\rangle \langle \varepsilon_m, y\rangle\right) = 0 \quad \forall\, x, y \in H, n \neq m.
$$

ε_n is called a H *strong white noise* if it satisfies 1)-3) and

4') ε_n is a sequence of i.i.d. H-random variables.

An strong white noise is a white noise and the converse fails. It holds if ε_n is Gaussian. Let us now give examples of Hilbertian white noises.

Example 2.3.5. Consider $H = L^2([0,1], \mathscr{B}([0,1]), \mu)$, where μ denotes the Lebesgue measure. Let W_s be a bilateral Wiener process (i.e $W_s = W_s^{(1)}\mathbf{1}_{\mathbb{R}_+}(s) + W_{-s}^{(2)}\mathbf{1}_{\mathbb{R}_-}(s)$, and $W_s^{(1)}, W_s^{(2)}$ are two independent standard Wiener processes). Fix $h \in H, h \neq 0$ and set

$$\varepsilon_n(t) = \int_n^{n+t} h(n+t-s)dW_s,\ t \in [0,1]\ n \in \mathbb{Z}.$$

Then $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is a strong white noise, since increments of W are independent stationary.

Definition 2.3.6. Let $X = (X_n, n \in \mathbb{Z})$ be H-valued weakly stationary process and let M_n be the linearly closed subspace generated by $(X_s, s \leqslant n)$, i.e. $M_n = \overline{\text{span}\{\ell(X_s) : \ell \in L(H), s \leqslant n\}}$. X is called *regular* process if, for the process

$$\varepsilon_n = X_n - \Pi^{M_{n-1}}(X_n),$$

it holds that $\sigma^2 := E\|\varepsilon_n\|^2 > 0$.

In this case $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is an H-white noise. Moreover $\varepsilon_n \in M_n$ and ε_n is strongly orthogonal to M_{n-1}, i.e. $C_{\varepsilon_n,\xi} = 0$ for any $\xi \in M_{n-1}$. (ε_n) is called the *innovation* process of X.

Definition 2.3.7. An H-valued weakly stationary and regular process $X = (X_n, n \in \mathbb{Z})$ is a *linear* process (LPH) if for all $n \in \mathbb{Z}$

$$X_n = \Pi^{I_n}(X_n),$$

where I_n is the linearly closed subspace generated by $(\varepsilon_s, s \leqslant n)$.

So, every LPH $X = (X_n, n \in \mathbb{Z})$ can be written in the form

$$X_n = \varepsilon_n + \sum_{k=1}^{\infty} \Pi^{\varepsilon_{n-k}}(X_n), n \in \mathbb{Z}$$

and $\Pi^{\varepsilon_{n-k}}(X_n)$ only depends on $C_{\varepsilon_0}, C_{\varepsilon_0,X_j}$ and X_n. However, linear processes which depend only on a finite number of parameters are more tractable than general linear processes from a statistical point of view.

2.3.1 Gaussian random variables in Hilbert spaces

In this subsection we recall the basic definitions and some classical properties of Gaussian random variables. Historically, the study of Gaussian random vectors and processes may indeed be considered as one of the fundamental topics of the theory since it inspired many other parts of the field in results and techniques of investigation.

A real valued random variable X in $L^2(\Omega, \mathscr{A}, P)$ is said to be Gaussian if its characteristic function is given by

$$\varphi_X(t) = e^{(i\mu t - \sigma^2 t^2/2)},$$

where $\mu = \mathbb{E}X, \sigma^2 = Var[X]$.

A random vector $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$ is Gaussian if for all real numbers $\alpha_1, \ldots, \alpha_d$, a linear combination $\sum_{k=1}^d \alpha_k X_k$ is a real valued Gaussian random variable. An equivalent definition is the following: a random vector X is Gaussian in $\mathbb{R}^d$ if there is a d-vector μ and a symmetric, positive semidefinite $d \times d$ matrix S, such that the characteristic function of X is

$$\varphi_X(t) = e^{\left(it^T\mu - \frac{1}{2}t^T St\right)}.$$

Recall that the characteristic functional φ_X of the random variable X taking values in a Hilbert space H is given by

$$\varphi_X(y) = \int_H e^{i\langle x,y\rangle} P_X(dx) = \int_\Omega e^{i\langle X(w),y\rangle} P(dw) = \mathbb{E}\left[e^{i\langle X,y\rangle}\right], \quad y \in H.$$

It is known that $\varphi_X(\cdot) : H \to \mathbb{C}$ is continuous in the norm topology, and satisfies the properties:

1. $\varphi_X(0) = 1$;
2. $|\varphi_X(y)| \leqslant 1$, $y \in H$;
3. $\varphi_X(y) = \overline{\varphi_X(-y)}$, $y \in H$;
4. If X and Y are independent random variables with values in H, then $\varphi_{X+Y}(y) = \varphi_X(y)\varphi_Y(y)$, $y \in H$;

The proof of these results can be found in [32].

Definition 2.3.8. A random variable X on a Hilbert space H is said to be *Gaussian* if its characteristic functional $\varphi_X(y)$ is of the form

$$\varphi_X(y) = e^{\left(i\langle\mu,y\rangle - \frac{1}{2}\langle Cy,y\rangle\right)},$$

where $\mu \in H$ and $C : H \to H$ is semi-definite positive Hermitian operator with finite trace (that is, for some orthonormal basis $\{e_i\}_{i=1}^{\infty}$ of H, the sum $\sum_{i=1}^{\infty}\langle Ce_i, e_i\rangle < \infty$).

It can be shown that μ is the mean and the operator C is the covariance operator for the Gaussian random variable X. The multivariate Gaussian distribution of a infinite dimensional random variable X can be written with the notation $X \sim \mathcal{N}(\mu, C)$.

The next theorem gathers some important properties of H-valued Gaussian random variables.

Theorem 2.3.9. *[32, p. 141]*

(i) Suppose that X and Y are two H-valued independent random variables, $X \sim \mathcal{N}(\mu_X, C_X)$, $Y \sim \mathcal{N}(\mu_Y, C_Y)$. Then $(X+Y) \sim \mathcal{N}(\mu_X + \mu_Y, C_X + C_Y)$.

Conversely, if $Z = X+Y$ is H-valued Gaussian random variable, and X, Y are independent, then X and Y have to be Gaussian random variables.

(ii) If $X \sim \mathcal{N}(\mu_X, C_X)$, then X can be represented as

$$X = \mu_X + \sum_{i=1}^{\infty} \psi_i e_i,$$

where $\{e_i\}_{i=1}^{\infty}$ is an orthonormal basis on H, $\{\psi_i\}_{i=1}^{\infty}$ are independent zero-mean Gaussian random variables with $Var(\psi_i) = \sigma_i^2$ and $\{\sigma_i^2\}_{i=1}^{\infty}$ are the eigenvalues of C_X. Furthermore the infinite series is convergent (strongly) with probability 1.

(iii) If $X \sim \mathcal{N}(\mu_X, C_X)$, and $A \in L(H)$ is a bounded linear operator from

H to H, then the random variable $Y = AX$ is also Gaussian and $Y \sim \mathcal{N}(A\mu_X, AC_X A^*)$.

2.3.2 The Wold decomposition

The classical Wold decomposition theorem states that any covariance stationary process can be decomposed into two mutually uncorrelated component processes, one a linear combination of lags of a white noise process and the other a process, future values of which can be predicted exactly by some linear function of past observations. The Wold theorem plays a central role in time series analysis. It implies that the dynamic of any purely nondeterministic covariance-stationary process can be arbitrarily well approximated by an ARMA process. So, one reason for the popularity of the ARMA models derives from Wold's Theorem. On the other hand, the Wold decomposition of a stationary process is analogous to the Lebesgue decomposition of the spectral measure into its absolutely continuous and singular parts.

We are using the Wold decomposition theorem for vector-valued processes in the proof of Lemma 2.4.1.

The Wold decomposition – real valued processes

The Wold representation theorem says that every weakly stationary process can be written as the sum of two processes, one deterministic and one stochastic.

Let $\{x_t,\ t \in \mathbb{Z}\}$ be a real valued weakly stationary process and define

$$M_n = \overline{\mathrm{span}}\{x_t,\ t \leqslant n\}.$$

Definition 2.3.10. The process $\{x_t\}$ is said to be deterministic if and only if the one-step squared error

$$\sigma^2 = \mathbb{E}|x_{n+1} - \Pi^{M_n} x_{n+1}|^2$$

equals to 0. In other words, the values $x_{n+j}, j \geqslant 1$ are perfectly predictable in terms of elements of M_n.

It is important to note that deterministic does not mean that x_t is non-random.

Example 2.3.11. Let $\{x_t,\, t \in \mathbb{Z}\}$ be a stochastic process defined by

$$x_t = A\cos(t) + B\sin(t)$$

where A and B are independent standard normal random variables. This process is deterministic. In fact it is possible to show that $x_t = \frac{\sin(2)}{\sin(1)} x_{t-1} - x_{t-2}$.

Proposition 2.3.12. *Any zero-mean weakly stationary process $\{x_t\}$ with $\sigma^2 > 0$ can be expressed as*

$$x_t = \sum_{i=0}^{\infty} \psi_i z_{t-i} + \mu_t$$

where

(i) $\psi_0 = 1$ *and* $\sum_{i=0}^{\infty} \psi_i^2 < \infty$,

(ii) $\{z_i\} \backsim WN(0, \sigma^2)$,

(iii) $z_t \in M_t$ *for each* $t \in \mathbb{Z}$,

(iv) $\mathbb{E}(z_t \mu_s) = 0$ *for all* $t, s \in \mathbb{Z}$,

(v) $\mu_t \in \bigcap_{n \in \mathbb{Z}} M_n$ *for each* $t \in \mathbb{Z}$,

(vi) μ_t *is deterministic.*

The usefulness of the Wold Theorem is that it allows the dynamic evolution of a variable x_t to be approximated by a linear model. If the innovations ε_t are independent, then the linear model is the only possible representation relating the observed value of x_t to its past evolution. However, when ε_t is merely an uncorrelated but not independent sequence, then the linear model exists but it is not the only representation of the dynamic dependence of the series. In this latter case, it is possible that the linear model may not be very useful, and there would be a nonlinear model relating the observed value of x_t to its past evolution. However, in practical time series analysis, it is often the case that only linear predictors are considered, partly on the grounds of simplicity, in which case the Wold decomposition is directly relevant.

The Wold decomposition – vector valued processes

Proposition 2.3.13. *Any zero-mean stationary vector process $X = (X_n, n \in \mathbb{Z})$ admits the following representation:*

$$X_n = \sum_{i=1}^{\infty} C_i \varepsilon_{n-i} + \mu_n$$

where

(i) $C_0 = I$ *and* $\sum_{i=0}^{\infty} \|C_i\|^2 < \infty$,

(ii) ε_i *is white noise*

$C(L)\varepsilon_n$ *is the stochastic component with* $C(L) = \sum_{i=0}^{\infty} C_i L^i, C_0 = I$ *and* μ_n *the purely deterministic component.*

If $\mu_n = 0$ the process is said regular.

The result is very powerful since holds for any covariance stationary process. However the theorem does not implies that (2) is the true representation of the process. For instance the process could be stationary but non-linear or non-invertible.

The Wold decomposition – H-valued processes

For the sake of clarity, first, we present the concept of The Wold decomposition of H-valued process for linear process based on the paper [47].

Definition 2.3.14. $X = (X_n, n \in \mathbb{N})$ be an H-valued linear process. Then the representation

$$X_n = \mu + \sum_{j=0}^{\infty} a_j(\varepsilon_{n-j}), \tag{2.3.3}$$

where $\mu = \mathbb{E}X \in H$, $(a_k)_{k\in\mathbb{N}}$ is a sequence of elements from $L(H)$, $a_0 = I$ and $\varepsilon_n, n \in N$ is a sequence of i.i.d. centered random variables in H, is called *the Wold Decomposition* of X.

In the work [47] we can find the following invertibility property:

Theorem 2.3.15. *Let $X = (X_n, n \in \mathbb{N})$ be an H-valued linear process defined by* (2.3.3)*, and*

$$1 - \sum_{j=1}^{\infty} z^j \|a_j\| \neq o \text{ for any } |z| < 1.$$

Then $X = (X_n, n \in \mathbb{N})$ is invertible, i.e.

$$X_n = \varepsilon_n + \sum_{j=1}^{\infty} \rho_j (X_n - j), \tag{2.3.4}$$

where $\rho_j \in H$ and $\sum_{j=1}^{\infty} \|\rho_j\| < \infty$.

Let X be a weakly stationary process and let M_n be the linearly closed subspace generated by $(X_s, s \leqslant n)$, i.e. $M_n = \overline{\text{span}\{\ell(X_s) : \ell \in L(H), s \leqslant n\}}$. X is called a *regular* process if, for the process

$$\varepsilon_n = X_n - \Pi^{M_{n-1}}(X_n),$$

it holds that $\sigma^2 := E\|\varepsilon_n\|^2 > 0$.

In this case (ε_n) is an H-white noise. Moreover $\varepsilon_n \in M_n$ and ε_n is strongly orthogonal to M_{n-1}, i.e. $C_{\varepsilon_n,\xi} = 0$ for any $\xi \in M_{n-1}$. (ε_n) is called the *innovation* process of X.

Now, if J_n is the linearly closed subspace generated by $(\varepsilon_s, s \leqslant n)$, the *Wold docomposition* of X is defined by

$$X_n = \Pi^{J_n}(X_n) + \Pi^{J_n^{\perp}}(X_n) := Y_n + Z_n, \quad n \in \mathbb{Z}.$$

This definition remains essentially the same as in Equation (2.3.3), but the operators $a_j, j \in \mathbb{N}$ may then be unbounded; this finally generalizes the notion. Properties of this decomposition are similar to those in the real case. In particular, one has ε_s is strongly orthogonal to Z_n for any $s, n \in \mathbb{Z}$, i.e. $C_{\varepsilon_s,\xi} = 0$ for any $\xi \in Z_n$. and $Z_n \in \bigcap_{j=0}^{\infty} M_{n-j}, n \in Z$.

2.3.3 Moving average processes in Hilbert spaces

Definition 2.3.16. A *moving average process of order* q in H (MAH(q)) is a linear process $X = (X_n, n \in \mathbb{Z})$ such that $\mathbb{E}\|\Pi^{\varepsilon_{n-q}}(X_n)\| > 0$ for all $n \in \mathbb{Z}$ and

$$\Pi^{M_{n-1}}(X_n) = \Pi^{J_{n-1,q}}(X_n), \ n \in \mathbb{Z},$$

where $J_{n-1,q}$ is the linearly closed subspace generated by $(\varepsilon_{n-1}, \ldots, \varepsilon_{n-q})$.

Example 2.3.17 (Truncated Ornstein-Uhlenbeck process)**.** Let $H = L^2\left([0,1], \mathscr{B}([0,1]), \mu + \delta_{(1)}\right)$, where μ is the Lebesgue measure on $[0,1]$ and $\delta_{(1)}$ denotes the Dirac measure centered on point 1. We choose a version of an orthonormal basis $\{e_j\}_{j=0}^{\infty}$ such that $e_0 = \mathbf{1}_{\{1\}}$ and $e_j(1) = 0, j \geqslant 1$.

Consider the real continuous time process

$$\xi_t = \int_{\lfloor t-1 \rfloor}^{t} e^{s-t} dW_s, \;\; t \in \mathbb{R},$$

where W_s is a bilateral Wiener process (i.e $W_s = W_s^{(1)} \mathbf{1}_{\mathbb{R}_+}(s) + W_{-s}^{(2)} \mathbf{1}_{\mathbb{R}_-}(s)$, and $W_s^{(1)}, W_s^{(2)}$ are two independent standard Wiener processes), and $\lfloor t-1 \rfloor$ is the biggest integer $\leqslant t-1$. $(\xi_t, t \in \mathbb{R})$ is a fixed continuous version of the stochastic integral.

Let us set

$$Y_n(x) = \xi_{n+x}, \;\; x \in [0,1], n \in \mathbb{Z}.$$

Then we can identify Y_n with an H-valued random variable by putting

$$Y_n(\cdot) = Y_n(1)e_0(\cdot) + \sum_{j=1}^{\infty} \left[\int_0^1 Y_n(s)e_j(s)ds\right] e_j(\cdot).$$

We claim that (Y_n) is MAH(1) process. Indeed, let us define the operator $\ell \in L(H)$:

$$(\ell(f))(x) = f(1)e^{-x}, \;\; f \in H.$$

If $0 \leqslant x < 1$ we can write

$$Y_n = \int_{\lfloor n+x-1 \rfloor}^{n+x} e^{s-n-x} dW_s = e^{-x} \int_{n-1}^{n} e^{s-n} dW_s + \int_{n}^{n+x} e^{s-n-x} dW_s.$$

Then (Y_n) has decomposition

$$Y_n = \ell(\varepsilon_{n-1}) + \varepsilon_n, \;\; n \in \mathbb{Z}, \tag{2.3.5}$$

where (ε_n) is defined as follow:

$$\varepsilon_n(x) = \int_n^{n+x} e^{s-n-x} dW_s, \;\; x \in [0,1), \;\; \varepsilon_n(1) = Y_n(1) - e^{-1} Y_{n-1}(1).$$

From (2.3.5) we have that $\Pi^{M_{n-1}}(Y_n) = \Pi^{\varepsilon_{n-1}}(Y_n) = \ell(\varepsilon_{n-1})$. Obviously, $\mathbb{E}\|\Pi^{\varepsilon_{n-1}}(Y_n)\| = \mathbb{E}\|\ell(\varepsilon_{n-1})\| = \mathbb{E}\|e^{-x} \int_{n-1}^{n} e^{s-n} dW_s\| > 0$.

2.3.4 Autoregressive processes in Hilbert spaces

Definition 2.3.18. Let $X = (X_n, n \in \mathbb{Z})$ be a H-valued weakly stationary process, M_n be the linearly closed subspace generated by $(X_s, s \leqslant n)$, and M_n^p be the linearly closed subspace generated by $(X_s, n-p \leqslant s \leqslant n)$. X is called *autoregressive Hilbertian process of order* p (ARH(p)) if,

$$\Pi^{M_{n-1}}(X_n) = \Pi^{M_{n-1}^p}(X_n),$$

and, if $p > 1$,

$$\mathbb{E}\left\|\Pi^{M_{n-1}^p}(X_n) - \Pi^{M_{n-1}^{p-1}}(X_n)\right\| > 0.$$

Remark 2.3.19. One may characterize an ARH(1) by a relation of the form

$$X_n = \lambda_n(X_{n-1}) + \varepsilon_n, n \in \mathbb{Z}, \tag{2.3.6}$$

where λ_n are measurable mappings from H to H, and $(\varepsilon_n, n \in \mathbb{Z})$ is a *H-white noise* .

If X is strictly stationary, it is possible to choose $\lambda_n = \lambda$ not depending on n. If also we have that $C_{X_{n-1}X_n}$ is dominated by $C_{X_{n-1}}$, then Theorem 2.2.3 yields existence of $\rho \in L(H)$ such that

$$X_n = \varepsilon_n + \rho(X_{n-1}). \tag{2.3.7}$$

In this case we will say that ρ is the *autocorrelation operator* of X.

The next theorem shows the existence of $X = (X_n, n \in \mathbb{Z})$ satisfying (2.3.7) with a given white noise (ε_n) and $\rho \in L(H)$. First, we need to prove a simple but somewhat surprising lemma.

Lemma 2.3.20. *Let $\rho \in L(H)$. The following conditions are equivalent:*

(i). $\sum_{j=0}^{\infty} \|\rho^j\| \leqslant \infty$;

(ii). $\exists\, j_0 \in \mathbb{N}$ *such that* $\|\rho^j\| < 1$ *for all* $j \geqslant j_0$;

(iii). $\exists\, j_0 \in \mathbb{N}$ *such that* $\|\rho^{j_0}\| < 1$;

(iv). $\exists\, j_0 \in \mathbb{N}, a > 0$ *and* $b \in (0,1)$ *such that* $\|\rho^j\| \leqslant ab^j$ *for all* $j \geqslant j_0$.

Proof. Obviously, (iv) entails (i):

$$\sum_{j=j_0}^{\infty}\|\rho^j\| \leqslant \sum_{j=j_0}^{\infty} ab^j \leqslant \infty, \text{ so } \sum_{j=0}^{\infty}\|\rho^j\| \leqslant \infty.$$

The implications $(i) \Rightarrow (ii) \Rightarrow (iii)$ are also trivial.

Now we are going to prove the most substantive part of the lemma that from (iii) follows (iv). We have $0 < \|\rho^{j_0}\| < 1$ and suppose that $j > j_0$. Then we can write

$$j = j_0 q + r,$$

where $q \geqslant 1$ and $r \in [0, j_0 - 1]$ are integers. Using inequality $\|sv\| \leqslant \|s\|\|v\|$ for any $s, v \in L(H)$, we obtain

$$\|\rho^j\| = \|\rho^{j_0 q + r}\| \leqslant \|\rho^{j_0}\|^q \|\rho^r\|.$$

Notice that $q = \frac{j}{j_0} - \frac{r}{j_0} > \frac{j}{j_0} - 1$. As $0 < \|\rho^{j_0}\| < 1$ we can estimate

$$\|\rho^j\| \leqslant \|\rho^{j_0}\|^{\frac{j}{j_0}-1}\|\rho^r\|. \tag{2.3.8}$$

Let us choose

$$a = \frac{\max\{\|\rho^r\| : 0 \leqslant r \leqslant j_{0-1}\}}{\|\rho^{j_0}\|} \text{ and } b = \|\rho^{j_0}\|^{\frac{1}{j_0}}.$$

Then $a > 0$, $b \in (0,1)$ and from (2.3.8) it follows that

$$\|\rho^j\| \leqslant ab^j \text{ for all } j \geqslant j_0.$$

□

Remark 2.3.21. Observe that $(i) - (iv)$ does not imply $\|\rho\| < 1$, contrarily to the one-dimensional case. The simplest example to see that in two dimensional case could be

$$\rho = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

A less trivial example can be found in the Hilbert space $H = L^2\left([0,1], \mathscr{B}([0,1]), \mu + \delta_{(1)}\right)$ with the operator $(\rho(f))(x) = f(1)e^{-\frac{1}{2}x}$, $f \in H$.

Theorem 2.3.22. *[11, p. 245] Let (ε_n) be a white-noise (see Definition 2.3.4), $\rho \in L(H)$, and there is $m \in \mathbb{N}$ such that $\|\rho^m\| < 1$. Then (2.3.7) has a unique stationary solution given by*

$$X_n = \sum_{m=1}^{\infty} \rho^m(\varepsilon_{n-m}), \ n \in \mathbb{Z},$$

where the series converges almost surely and in L_H^2. Moreover (ε_n) is the innovation of (X_n).

Example 2.3.23 (Ornstein-Uhlenbeck process)**.** Consider again $H = L^2\left([0,1], \mathscr{B}([0,1]), \mu + \delta_{(1)}\right)$. Let $\xi = (\xi_t, t \in \mathbb{R})$ be a measurable version of the Ornstein-Uhlenbeck process:

$$\xi_t = \int_{-\infty}^{t} e^{s-t} dW_s, \ t \in \mathbb{R},$$

where W_s is a bilateral Wiener process.

Let us set

$$Y_n(x) = \xi_{n+x}, \ x \in [0,1], n \in \mathbb{Z}.$$

We claim that (Y_n) is ARH(1) process. Indeed, let us define the operator $\rho \in L(H)$:

$$(\rho(f))(x) = f(1)e^{-x}, \ f \in H,$$

and define the H-white noise

$$\varepsilon_n(x) = \int_{n}^{n+x} e^{s-n-x} dW_s, \ x \in [0,1), \ \varepsilon_n(1) = Y_n(1) - e^{-1}Y_{n-1}(1).$$

Also we have

$$Y_n(x) = \int_{-\infty}^{n+x} e^{s-n-x} dW_s = e^{-x} \int_{-\infty}^{(n-1)+1} e^{s-n} dW_s + \int_{n}^{n+x} e^{s-n-x} dW_s.$$

Therefore, (Y_n) has decomposition $Y_n = \rho(Y_{n-1}) + \varepsilon_n$. Notice that

$$\|\rho\|^2 = \int_0^1 e^{-2x} d(\mu + \delta_{(1)})(x) = \frac{1-e^{-2}}{2} + e^{-2} = \frac{1+e^{-2}}{2} < 1.$$

So, the assumption of Theorem 2.3.22 holds, and (Y_n) is ARH(1) with innovation (ε_n) and autocorrelation operator ρ.

Example 2.3.24 (Cartea-Villaplana). Let $H = L^2([0,M],\mathscr{B}([0,M]),\mu)$, where μ is the Lebesgue measure on $[0,M]$, M is maximum electricity price, for example $M = 3000$ Euro. Let $C = (C_n, n = 0,1,2,\ldots)$ be a random valued variable which represents capacity. Consider the dynamics

$$C_n(x) = -2bX_n - \lg x,\ \ x \in (0,M], b > 0$$

where X_n is the solution of a discrete Ornstein-Uhlenbeck equation:

$$X_{n+1} = \lambda X_n + W_{n+1}, \quad n = 0,1,\ldots,$$

where $\lambda \in (0,1), X_0 \sim N\left(0,\frac{1}{1-\lambda^2}\right)$, $W_n \sim N(0,1) - i.i.d.$ - H-white noise. So, we can write

$$C_n = -2b\left(\lambda^n X_0 + \sum_{i=1}^{n}\lambda^{n-i}W_i\right) - \lg x.$$

We claim that (C_n) is ARH(1) process. First, we should verify that (C_n) is weakly stationary.

$$\begin{aligned}\|C_n\|_H^2 &= \int_0^M \left(-2b\left(\lambda^n X_0 + \sum_{i=1}^{n}\lambda^{n-i}W_i\right) - \lg x\right)^2 dx \\ &\leqslant \int_0^M \lg^2 x dx + 4b\left(\lambda^n X_0 + \sum_{i=1}^{n}\lambda^{n-i}W_i\right)\int_0^M \lg x dx \\ &+ \left(-2b\left(\lambda^n X_0 + \sum_{i=1}^{n}\lambda^{n-i}W_i\right)\right)^2\int_0^M dx.\end{aligned}$$

Therefore, $\mathbb{E}\|C_n\|_H^2 < \infty$. We now compute

$$\begin{aligned}&C_{C_nC_{n+k}}(h) = \mathbb{E}\left[\langle C_n - \mathbb{E}C_n, h\rangle (C_{n+k} - \mathbb{E}C_{n+k})\right] \\ &= \mathbb{E}\left[\langle -2bX_n - \lg x - (2b\lambda^n\mathbb{E}X_0 - \lg x), h\rangle (-2bX_{n+k} - \lg x - (2b\lambda^{n+k}\mathbb{E}X_0 - \lg x))\right] \\ &= -4b^2\mathbb{E}\left[\langle X_n - \mathbb{E}X_n, h\rangle (X_{n+k} - \mathbb{E}X_{n+k}))\right] = -4b^2 C_{X_nX_{n+k}}(h).\end{aligned}$$

Using the reasoning from Example 2.3.3 we can conclude that the covariance matrix is invariant under time shifts.

We can write

$$C_n = \lambda C_{n-1} - (1-\lambda)\lg x - 2bW_n,$$

so, equation (2.3.6) holds, which means that (C_n) is ARH(1) process.

Similarly, let $D = (D_n, n = 0, 1, 2, \dots)$ be a random valued variable which represents the evolution electricity demand. Consider the dynamics

$$D_n(x) = -2aY_n + \lg x, \; x \in (0, M], a > 0$$

where Y_n is the solution of a discrete Ornstein-Uhlenbeck equation:

$$Y_{n+1} = \lambda Y_n + V_{n+1}, \quad n = 0, 1, \dots,$$

where $\lambda \in (0, 1), Y_0 \sim N\left(0, \frac{1}{1-\lambda^2}\right)$, $V_n \sim N(0, 1) - i.i.d.$ - H-white noise. Obviously, (D_n) is also ARH(1) process. Finally, the wholesale power prices P_n can be found as the intersection of the capacity and demand:

$$-2bX_n - \lg x = -2aY_n + \lg x$$

Therefore,

$$P_n = e^{aY_n - bX_n},$$

which is the model of price proposed by Cartea and Villaplana.

2.4 Linear transformation of stochastic processes

In the paper [44] it is proved that a linear transformation of a process possessing an $MA(q)$ representation gives a process that also has a finite order MA representation with order not greater than q. The more general fact that a linear transformation of a vector ARMA process is again an ARMA process is also proved. These results are of importance because many temporal as well as contemporaneous aggregation procedures can be represented as linear transformations.

Proposition 2.4.1. *[44, Lemma 1] Let $X = (X_n, n \in Z)$ be an m_2-dimensional $MA(q)$ process, and $T = [t_{ij}]_{i,j} \neq 0$ be a real $m_1 \times m_2$ matrix. Then $(Y_n) = (T(X_n))$ is an m_1-dimensional $MA(q^*)$ process, where $q^* \leqslant q$.*

Proof. If $X_t = (x_{1n}, x_{2n} \dots, x_{m_2 n})'$ is $MA(q)$ process, then it can be written as follows

$$X_n = U_n + M_1 U_{n-1} + \dots + M_q U_{n-q}, \tag{2.4.1}$$

where $U_n = (u_{1n}, u_{2n} \dots, u_{nn})'_n$ is an m_2-dimensional white noise process, (i.e. $\mathbb{E}U_n = 0$, $\mathbb{E}(U_n U'_{n+k}) = 0$ if $k \neq 0$ and $\mathbb{E}(U_n U'_n) = \Sigma_u$) and $M_k = [\mu^k_{i,j}]_{i,j=1,\dots,m_2}$ are $(m_2 \times m_2)$ matrices.

Let us denote the back-shift operator B (i.e $B^k U_n = U_{n-k}$) and

$$M(B) = \left[\sum_{k=0}^{q} \mu^k_{i,j} B^k \right]_{i,j=1,\dots,m_2}.$$

Then we can rewrite (2.4.1) as

$$X_n = M(B) U_n. \tag{2.4.2}$$

For $Y_n = (T(X_n)) = (y_{1nt}, y_{2n} \dots, y_{m_2 t})'$ we define a Hilbert space H to be the closure of

$$\text{span}\{y_{i,t} : i = 1 \dots, m_1, n \in \mathbb{Z}\}$$

with inner product $\langle x, y \rangle = \mathbb{E}(xy)$, where $x, y \in H$. Consider also closed subspaces of H:

$$H_n = \overline{\text{span}\{y_{i,s} : i = 1 \dots, m_1, s \leqslant n\}}.$$

Let us denote by M_n the orthogonal complement of H_{n-1} in H_n, so $H_n = H_{n-1} \bigoplus M_n$.

By the Wold Decomposition Theorem,

$$Y_n = \sum_{k=0}^{\infty} \Phi_k V_{n-k}, \quad \Phi_0 = I_{m_1},$$

where v_{in} is the projection of y_{in} on M_n, and thus $V_n = (v_{1,n}, \dots, v_{m_1 n})'$ is white noise with variance-covariace matrix Σ_v say, and

$$\Phi_k = \mathbb{E}(Y_t V'_{t-n}) \Sigma_v^{-1}, \quad n > 0$$

where Σ_v^{-1} is the generalized inverse of Σ_v. Since $\mathbb{E}(Y_n Y'_{n-k}) = 0$ for $k > q$, y_n is orthogonal to H_{n-k} for $k > q$, and hence $\mathbb{E}(Y_n V'_{n-k}) = 0$ for $k > q$. Consequently, $\Phi_k = 0$ for $k > q$, and thus we have a representation of Y_n as an $MA(q^*)$, where q^* is less than q if $\mathbb{E}(Y_n V'_{n-k}) = 0$ for $n > q^*$, otherwise $q^* = q$. $\square$

Let us show that there is no version of this theorem for AR processes.

Example 2.4.2. Consider the case $m_1 = 1, m_2 = 2$ and define the $AR(1)$ process X_n as

$$X_n = \begin{pmatrix} x_{1n} \\ x_{2n} \end{pmatrix} = \begin{pmatrix} ax_{1n-1} + w_{1n} \\ bx_{2n-1} + w_{2n} \end{pmatrix}$$

Let $T = [1 \quad 1] : \mathbb{R}^2 \to \mathbb{R}$. Then $Y_n = T(X_n) = ax_{1n-1}+bx_{2n-1}+w_{1n}+w_{2n}$. Evidently, unless $a = b$, Y_n is not autoregressive.

The next proposition shows that the vector-valued ARMA class is closed with respect to linear transformations.

Proposition 2.4.3. *[45, Corollary 11.1.2] Let $X = (X_n, n \in \mathbb{Z})$ be an m_2-dimensional $ARMA(p,q)$ process, and $T = [t_{ij}]_{i,j} \neq 0$ be a real $m_1 \times m_2$ matrix of rank m_1. Then $Y_n = T(X_n)$ is an m_1-dimensional $MA(p^*, q^*)$ process with $p^* \leqslant m_2 p$ and $q^* \leqslant (m_2 - 1)p + q$.*

This theorem gives upper bounds for the ARMA orders of a linearly transformed ARMA process. For instance, if X_n is a $AR(p) = ARMA(p, 0)$ process, a linear transformation $Y_n = (T(X_n))$ has a $ARMA(p^*, q^*)$ representation. For some linear transformations, q^* will be zero. However, there are transformations of a finite order $AR(p)$ process that do not admit a finite order AR representation, as in Example 2.4.2, but just a mixed $ARMA$ representation.

In Chapter 5 we will present a characterization result regarding the conditions that guarantees that a linear transformation of a vector AR process is again an AR process both in finite and in infinite dimension.

2.5 $C([0,1])$-valued autoregressive processes

In [15] the authors introduced the model for the prediction of functional time series, where observations are assumed to be continuous random functions.

Consider a functional time series $X = (X_n, n \in \mathbb{Z})$, and let $x_n(\cdot) \in C([0,1])$ be a realization of the corresponding random process. In practice, the curves $x_n(\cdot)$ are usually recorded as high-dimensional vectors with highly correlated entrances, exactly as in our case with supply and demand curves. Then, the need of dimension reduction techniques that take into account the continuous nature of the data arises.

For the prediction of $x_{n+1}(\cdot)$ the whole curves $x_n(\cdot)$ can be replaced with the p most relevant evaluations $x_n(t_1), x_n(t_2), \dots, x_n(t_p)$. The problem of selection of the points $t_1, t_2, \dots, t_p \subset [0,1]$ under a suitable optimality criterion is commonly known as variable selection. Although this technique leads to a finite dimensional vector, the problem is fully functional, since the definition of this criterion is based on the whole curves.

The standard assumption for the process $X = (X_n, n \in \mathbb{Z})$ are:

1. The random variable $\sup\{|X_n(s)|,\ s \in [0,1]\}$ has finite variance. In this case each evaluation $X_n(s)$, for $s \in [0,1]$, also has finite variance.

2. X_n is a centered stationary stochastic process (i.e. $\mathbb{E}X_n = 0$)

Also we will use the following notations. $T_p = (t_1, t_2, \dots, t_p) \in [0,1]^p$ is the vector of the points; $f(T_p)$ is understood to be the column vector with coordinates $f(t_j)$. The covariance matrix of the random variables $X_n(t_1), X_n(t_2), \dots, X_n(t_p)$ indexed by T_p is Σ_{T_p}. The vector of lagged-covariance $c_1(\cdot, T_p)$ has coordinates $(\mathrm{cov}(X_1(\cdot), X_0(t_j)))_{j=1,\dots,p}$. The set Θ_p is the compact subset of $[0,1]^p$ defined as follows:

$$\Theta_p = \{T_p = (t_1, t_2, \dots, t_p) \in [0,1]^p : \, t_{i+1} - t_i \leqslant \delta,\ i = 1, \dots\ p\},$$

where $0 < \delta < 1$ is some fixed number. The following model is proposed in [15]

$$X_n(\cdot) = \sum_{j=1}^{p} \alpha_j(\cdot) X_{n-1}(t_j) + \varepsilon_n(\cdot), \tag{2.5.1}$$

where $\alpha_j(\cdot)$ are continuous functions in $[0,1]$ and ε_n is a strong $C([0,1])$-valued white noise pointwisely uncorrelated with X_n. That is, all the curves depend on the same set of points regardless of the index n. After finding the relevant points T_p and optimal functions $(\alpha_1(s), \alpha_2(s), \dots, \alpha_p(s))$ Equation (5.4.2) is a p-dimensional AR(1) model.

For the optimality criterion for variable selection we will define the following operators:

$$q(T_p;\alpha_1,\dots\alpha_p)=\mathbb{E}\left[\left(X_n(s)-\sum_{j=1}^{p}\alpha_j(s)X_{n-1}(t_j)\right)^2\right], \tag{2.5.2}$$

where the coefficients $\alpha_j(s)$ depend on the points $t_1,t_2,\dots,t_p$. Then, integrating q^2 over s leads to

$$Q(T_p)=\int_0^1 \min_{\alpha_j(s)\in\mathbb{R}} q(T_p;\alpha_1,\dots\alpha_p)^2(s)ds. \tag{2.5.3}$$

This function Q can now be minimized with respect to T_p.

Theorem 2.5.1. *[15, Proposition 1] Let $X=(X_n,n\in\mathbb{Z})$ be a stationary process such that $\mathbb{E}[(\sup|X_n(s)|)^2]<\infty$. Suppose that it can be expressed as in Equation* (5.4.2) *with $\sum_{i=1}^{p}\|\alpha_i\|<1$ and $\mathbb{E}\|\varepsilon_n^2\|<\infty$. Then*

$$\arg\min_{T_p\in\Theta_p} Q(T_p)=\arg\max_{T_p\in\Theta_p} Q^0(T_p), \tag{2.5.4}$$

where

$$Q^0(T_p)=\int_0^1 c_1(s,T_p)'\,\Sigma_{T_p}^{-1}\,c_1(s,T_p)ds, \tag{2.5.5}$$

and the optimal functions are given by

$$(\alpha_1(s),\alpha_2(s),\dots,\alpha_p(s))=\Sigma_{T_p}^{-1}\,c_1(s,T_p). \tag{2.5.6}$$

The optimality criterion defined by Q^0 is simple to implement in practice.

Let us go on to estimation from the sample. Suppose that we have a sample $x_1,\dots,x_m$ of size m drawn from a process satisfying the assumptions of Theorem 2.5.1. The usual estimator of the covariance function is

$$\widehat{c}_r(s,t)=\frac{1}{m-1}\sum_{i=1}^{m-1}x_{i+r}(s)x_i(t). \tag{2.5.7}$$

Then, the natural estimator for the functions $Q_0(T_p)$ is

$$\widehat{Q}_m^0(T_p)=\int_0^1 \widehat{c}_1(s,T_p)'\,\Sigma_{T_p}^{-1}\,\widehat{c}_1(s,T_p)\,ds, \tag{2.5.8}$$

where $\widehat{c}_1(\cdot,T_p)=(\widehat{c}_1(\cdot,t_1),\dots,\widehat{c}_1(\cdot,t_p))'$. According to Theorem 2.5.1 the most relevant points are

$$\widehat{T}_p=\arg\max_{T_p\in\Theta_p}\widehat{Q}_m^0(T_p). \tag{2.5.9}$$

Due to computational limitations, this optimization is not feasible even for relatively small values of p. Therefore, a *greedy approximation* is carried out. The function Q^0 can be decomposed in a way that directly suggests an iterative approximation to this optimization problem. If the vector T_{p+1} is such that it contains all the entries of T_p plus a new one $t_{p+1} \in [0, 1]$, the $\widehat{Q}_m^0(T_{p+1})$ can be expressed as

$$\widehat{Q}_m^0(T_{p+1}) = \widehat{Q}_m^0(T_p) + \frac{\int_0^1 (\widehat{c}_1(s, T_p)' \Sigma_{T_p}^{-1} \widehat{c}_0(t_{p+1}, T_p) - \widehat{c}_1(s, t_{p+1}))^2 \, ds}{\widehat{c}_0(t_{p+1}, t_{p+1}) - \widehat{c}_0(t_{p+1}, T_p)' \Sigma_{T_p}^{-1} \widehat{c}_0(t_{p+1}, T_p)}.$$

Notice that this quotient is easy to compute under the assumption that all the covariance matrices $\Sigma_{T_p}^{-1}$ are invertible. However, for some real data sets the condition of the invertibility of $\Sigma_{T_p}^{-1}$ may not be satisfied. If the data is not invertible, it can be always preprocessed to remove the conflicting points of the grid. This would not affect the efficiency of the method, since these points would be linearly dependent of the others, so their information would be redundant.

Chapter 3

Radial basis function interpolation

3.1 Historical remarks

For what concerns approximation theory, the historical and theoretical foundation of meshless methods lies in the concept of positive definite functions or, more in general, positive definite kernels. Their development can be traced, for example, back to the work of J. Mercer (1909) [48], a fellow of Trinity College at Cambridge University. Many positive definite functions are nowadays classified as *Radial Basis Functions*. Perhaps one of the most fundamental contributions, namely characterizations of positive definite functions in terms of Fourier transforms, were made a few years later by Salomon Bochner [8] and Iso Schoenberg [63].

The initial motivation for radial basis function (RBF) methods came from geodesy, mapping, and meteorology. RBF methods were first studied by Roland Hardy, an Iowa State geodesist, in 1968, when he developed one of the first effective methods for the interpolation of scattered data [34]. He suggested what he called the multiquadric method for applications in cartography because he was not satisfied with the results of polynomial interpolation. RBF methods were developed to overcome the structure requirements of existing numerical methods. Multiquadric radial basis function is only one of many existing radial basis function.

Then, in 1979, Richard Franke published a study of multiquadric radial basis function method for scattered data interpolation problem [30]. Later in

1986 Charles Micchelli, an IBM mathematician, developed the theory behind the multiquadric method [49]. Micchelli made the connection between scattered data interpolation and positive definite functions. He proved that the system matrix for the multiquadric method is invertible, which means that the RBF scattered data interpolation problem is well-posed. The contributions of Bochner and Schoenberg were used by Micchelli as the starting point of his proofs.

During the next years, research in RBF methods has rapidly grown. RBF methods are now considered an effective way to solve partial differential equations, to represent topographical surfaces as well as other intricate three-dimensional shapes, having been successfully applied in such diverse areas as climate modeling, facial recognition, topographical map production, car and aircraft design, ocean floor mapping, and medical imaging. RBF methods have been actively developed over the last 40 years. Now RBF methods are an active area of mathematical research, as many open questions still remain.

3.2 The scattered data interpolation problem

Interpolation and approximation techniques are used in solutions of many engineering problems. Given a set of N distinct *data points* (or nodes) $X_N = \{x_i : i = 1, 2, \ldots, N\}$ arbitrarily distributed on a domain $\Omega \subset \mathbb{R}^n$ and a set of *data values* (or function values) $Y_N = \{y_i : i = 1, 2, \ldots, N\} \subset \mathbb{R}$. The data interpolation problem consists in finding a function $s : \Omega \to \mathbb{R}$ such that

$$s(x_i) = y_i,\ i = 1, \ldots, N.$$

If the data points at which the values are taken do not lie on a uniform or regular grid and they are in a large amount, then the process is called *scattered data interpolation*. The interpolation and approximation of unorganized scattered data is still a difficult problem.

In this thesis we are going to use the data about supply bids from the Italian electricity market for the period starting on 01/01/2013 and ending on 05/02/2018 (data from the GME website **www.mercatoelettrico.org**). Notice that the size of these data is very large, due the number of offers for each load period, and hence not easy to handle. For each hour of the day, the original data published by GME consist of information corresponding to a single supplier and reported in a XML table format, where every row represents a single offer with its own date, trader name, awarded price, awarded quantity. For example, for the single hour of the first day we have 351 units of information about price and quantity of offered electricity (see Table 1).

Table 3.1: Data from the Italian electricity market

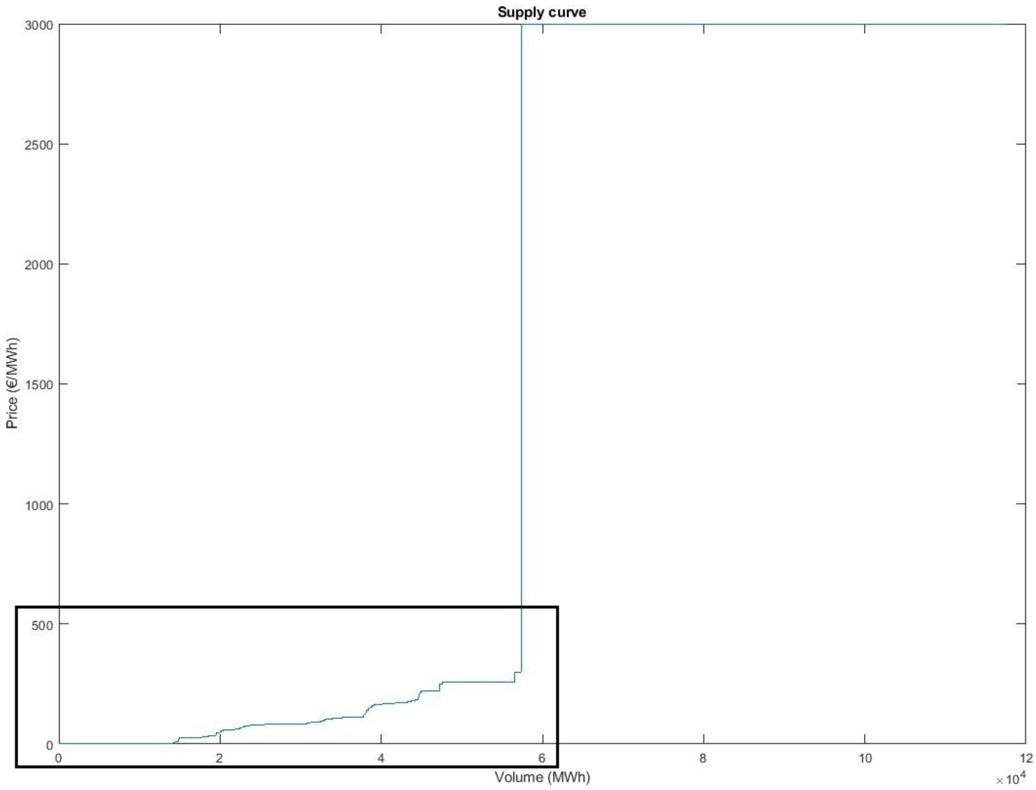
Date	Hour	Volume	Price
01-01-13	1	14117.32	0
01-01-13	1	52	0.01
01-01-13	1	66	1
01-01-13	1	15	2
01-01-13	1	15	5
01-01-13	1	150	8
01-01-13	1	18	9
01-01-13	1	8	9.01
01-01-13	1	8	9.02
01-01-13	1	6.006	9.03
01-01-13	1	2.004	9.04
01-01-13	1	2.994	9.05
01-01-13	1	2.14	9.06
...	...	...	...

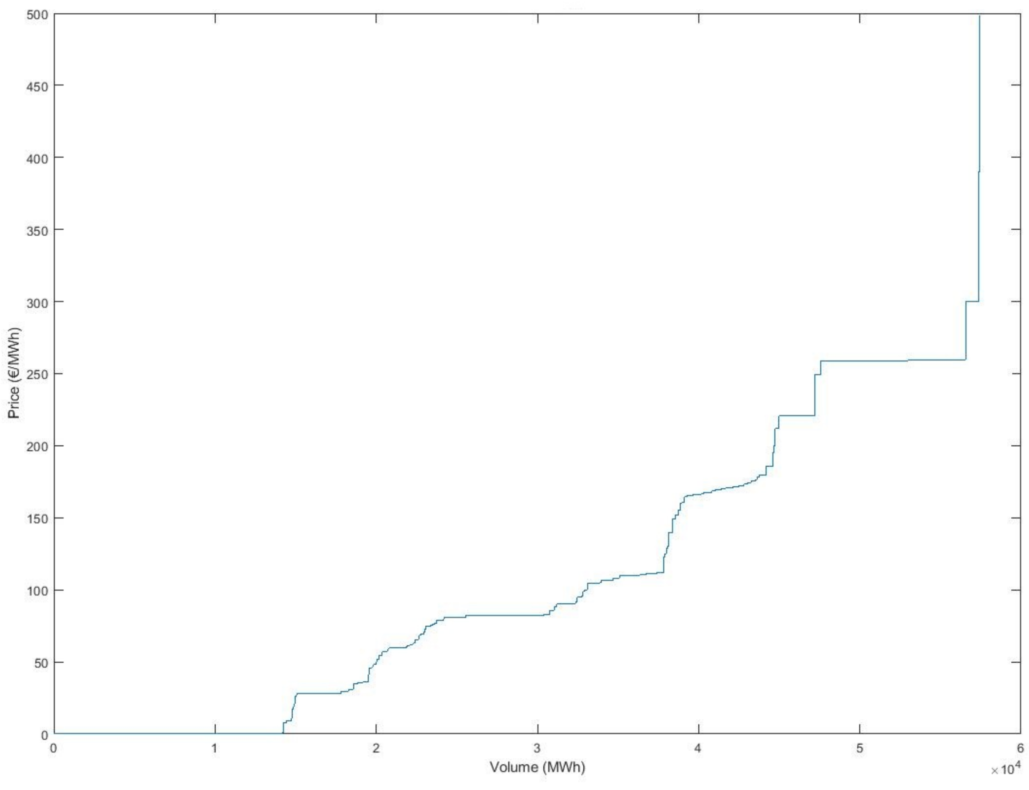
So, to analyze the period from 01/01/2013 to 05/02/2018, we need to deal with more than 16 million of data. In Figure 3.1 we present the supply curve corresponding to the first hour of the first day of the analyzed period. The first problem of our work is to present the information about electricity prices in a efficient and parsimonious way.

Let us review briefly the most popular methods for the interpolation problem.

- **Polynomial interpolation** is the interpolation of a given data set by the polynomial of lowest possible degree that passes through the points

Figure 3.1: Supply curve





of the dataset. For given data sites $x_1 < x_2 < \ldots < x_N$ and function values $y_1, \ldots, y_N$ there exists exactly one polynomial $p \in \pi_{N-1}(\mathbb{R})$ that interpolates the data at the data sites. Therefore the space $\pi_{N-1}(\mathbb{R})$ depends neither on the data sites nor on the function values but only on the number of points.

Runge's phenomenon (1901) shows that for high values of N, the interpolation polynomial may oscillate wildly between the data points. Evidently, the polynomial interpolation does not suit for our problem, because of the large amount of data.

- **Spline interpolation**. It is a well-established fact that a large data set is better dealt with splines than with polynomials. An aspect to notice in contrast to polynomials is that the accuracy of the interpolation process using splines is not based on the polynomial degree but on the spacing of the data sites. In particular, cubic splines are widely used to fit a smooth continuous function through discrete data.

  A cubic spline is a spline constructed of piecewise third-order polynomials which pass through a set of N control points. The second derivative of each polynomial is commonly set to zero at the endpoints, since this provides a boundary condition that completes the system of $N-2$ equations.

Notice that for all methods, the interpolant s is expressed as a linear combination of some basis functions B_i , i.e. $s(t) = \sum_{k=1}^{d} c_k B_k(t)$. The basis functions in polynomial interpolation does not depend on the data points. Another approach is to use a basis **which depends on the data points**.

3.3 Positive definite functions

The scattered data interpolation problem leads to the solution of a linear system of the form $Ax = y$. The solution of the system requires that the matrix A is non-singular. It is enough to know in advance that the matrix is positive definite. We need to introduce the concept of **positive definite functions** and **conditionally positive definite functions**.

3.3.1 Unconditionally positive definite functions

Definition 3.3.1. A real-valued function $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called *positive semi-definite* if , for all $m \in \mathbb{N}$ and for any set of pairwise distinct points $x_1, x_2, \ldots, x_m$, the $m \times m$ matrix

$$A = (\Phi(x_i - x_j))_{i,j=1}^m$$

is positive semi-definite, i.e. for every column vector z of m real numbers the scalar $z^T A z \geqslant 0$. The function $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called (strictly) *positive definite* if the matrix A is positive definite, i.e. for every non-zero column vector z of m real numbers the scalar $z^T A z > 0$.

Notice that, if Φ is positive semi-definite, then $\Phi(x) = \Phi(-x)$, $\Phi(0) \geqslant 0$, $|\Phi(x)| \leqslant \Phi(0)$ for all $x \in \mathbb{R}^n$.

Remark 3.3.2. Unfortunately, for historical reasons there is an alternative terminology around in the literature: other authors call a function positive definite if the associated matrices are positive semi-definite and strictly positive definite if the matrices are positive definite. We do not follow this historical approach here, keeping the terminology from [71].

The most important property of positive semi-definite matrices is that their eigenvalues are positive and so is its determinant.

One of the most celebrated results on positive semi-definite functions is their characterization in terms of Fourier transforms, which was established by Bochner [8].

Theorem 3.3.3 (Bochner's characterization)**.** *A continuous function $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is positive semi-definite if and only if it is the Fourier transform of a finite nonnegative Borel measure μ on $\mathbb{R}^n$, i.e.*

$$\Phi(x) = \int_{\mathbb{R}^n} e^{-ix^T w} d\mu(w), \quad x \in \mathbb{R}^n.$$

The proof of this theorem can be found in [8] or in the book [71, p. 70]. The Bochner representation is the most simple way to prove that a function is positive definite, as is the case of the following examples: e^{-x^2}, $e^{-|x|}$, $\frac{1}{1+x^2}$. Indeed,

$$e^{-x^2} = \int_{\mathbb{R}} e^{-ixt} d\mu(t) \ \text{ for } \ d\mu(t) = \frac{1}{2\sqrt{\pi}} e^{-t^2/4} dt;$$

$$e^{-|x|} = \int_{\mathbb{R}} e^{-ixt} d\mu(t) \ \text{ for } \ d\mu(t) = \frac{1}{\pi} \frac{1}{1+t^2} dt;$$

$$\frac{1}{1+x^2} = \int_{\mathbb{R}} e^{-ixt} d\mu(t) \ \text{ for } \ d\mu(t) = \frac{1}{2} e^{-|t|} dt;$$

Another useful characterization for positive semi-definite univariate function was given by Schoenberg in 1938 in terms of completely monotone functions.

Definition 3.3.4. A continuous function $\phi : [0, \infty) \to \mathbb{R}$ is called *completely monotone* on $[0, \infty)$ if

1. $\phi \in C^\infty(0, \infty)$;
2. $(-1)^k \phi^{(k)}(r) \geqslant 0$ for all $r \geqslant 0$, for $k = 0, 1, \ldots$.

For example, e^{-r}, $e^{-\sqrt{r}}$, $\frac{1}{1+r}$, are completely monotone functions.

Theorem 3.3.5 (Schoenberg's characterization). *Let $\phi : [0, \infty) \to \mathbb{R}$ be a continuous function which is additionally in $C^\infty((0, +\infty))$ and $\Phi : \mathbb{R}^n \to \mathbb{R}$ be a function such that $\Phi(x) = \phi(\|x\|_2^2)$. Then Φ is positive semi-definite if and only if ϕ is completely monotone on $[0, \infty)$.*

The proof is again in [71, p. 93].

3.4 Radial basis functions

Consider a set of N distinct data points $\{x_i\}_{i=1}^N \subset \mathbb{R}^n$ and a set of data values $\{y_i\}_{i=1}^N \subset \mathbb{R}$. We want to find a function $s : \mathbb{R}^n \to \mathbb{R}$ such that $s(x_i) = y_i$, $i = 1, \ldots, N$. Moreover, we want to find a basis for the solution, which depends on the data points. One simple way to do this is to choose a fixed function $\phi : \mathbb{R} \to \mathbb{R}$ and to form the interpolant as

$$s(x) = \sum_{i=1}^{N} \alpha_i \phi(\|x - x_i\|),$$

where the coefficients α_i are determined by the interpolation conditions $s(x_i) = y_i$. Therefore, the scattered data interpolation problem leads to the solution of a linear system

$$A\alpha = y, \text{ where } A_{i,j} = \phi(|x_i - x_j|). \tag{3.4.1}$$

Definition 3.4.1. A function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is called *radial* if there exists a function $\phi : [0, \infty) \to \mathbb{R}$, so that $\Phi(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{c}\|)$ for some point c, called a center.

So, a radial function is a real-valued function whose value depends only on the distance from the center $\mathbf{c}$. The norm is usually given by the Euclidean one; although other distance functions are also possible. A radial function has the advantage of a very simple structure. Sums of radial basis functions are typically used to approximate given functions. This approximation process can also be interpreted as a simple kind of neural network; this was the context in which they originally surfaced, in work by David Broomhead and David Lowe in 1988.

Solvability of the system (3.4.1) is guaranteed if Φ is positive semi-definite. Hence, if we choose the basis consisting of positive semi-definite radial functions, we would always have a well-posed interpolation problem.

Here are some standard radial basis function in dimension 1.

Let $\varepsilon > 0$ denote a shape parameter, $r = \|x\|_2$.

Positive definite radial function.

- Gaussian: $\phi(r) = e^{-(\varepsilon r)^2}$.
- Inverse multiquadric: $\phi(r) = \dfrac{1}{\sqrt{1 + (\varepsilon r)^2}}$.
- Matérn C^2: $\phi(r) = e^{-\varepsilon r}(\varepsilon r + 1)$.
- Matérn C^4: $\phi(r) = e^{-\varepsilon r}(\varepsilon^2 r^2 + 3\varepsilon r + 3)$.
- Wendland C^2: $\phi(r) = (1 - \varepsilon r)_+^4(4\varepsilon r + 1)$.
- Wendland C^4: $\phi(r) = (1 - \varepsilon r)_+^6(35\varepsilon^2 r^2 + 18\varepsilon r + 3)$.

In kernel-based methods, how to handle the scaling or the choice of the shape parameter is a well-documented but still an open problem. Variably scaled kernels (VSKs) were introduced in [16] with the aim to give a new technique to handle the problem of the choice of the scale or shape parameter in kernel-based interpolation problems. There, the authors consider native spaces whose kernels allow for a change the kernel scale of a d-variate interpolation problem locally, depending on the requirements of the application.

It is well-known that kernels on $\mathbb{R}^n$ can be scaled be a positive factor δ:

$$K(x, y; \delta) := K(x/\delta, y/\delta).$$

Variably scaled kernels were further developed in [58], [59]. VSKs were already used also in neural networks problems [52] and for approximating the solution of elliptic partial derivative problems [21]. In [59] the author showed that VSKs are a useful tool also for recovering unknown non-regular functions from set of scattered data.

3.5 Reproducing kernel Hilbert space

A reproducing kernel Hilbert space (RKHS) provides a practical and elegant structure to solve optimization problems in function spaces. We need to introduce the concept of RKHS which which plays an important role in approximation theory.

Let $\Omega \in \mathbb{R}^n$ be an arbitrary nonempty set.

Definition 3.5.1. A function $K : \Omega \times \Omega \to \mathbb{R}$ is *symmetric and positive definite* (SPD) if for all $m \in \mathbb{N}$ and for any set of pairwise distinct points $x_1, x_2, \dots, x_m \subset \Omega$, the $m \times m$ matrix

$$A = (K(x_i, x_j))_{i,j=1}^m$$

is symmetric and positive definite (i.e. for every non-zero column vector z of m real numbers the scalar $z^T Az > 0$). A function $K : \Omega \times \Omega \to \mathbb{R}$ is *symmetric and positive semi-definite* (nonnegative) if for all $m \in \mathbb{N}$ and for any set of pairwise distinct points $x_1, x_2, \ldots, x_m \subset \Omega$, the $m \times m$ matrix

$$A = (K(x_i, x_j))_{i,j=1}^{m}$$

is symmetric and positive semidefinite (i.e. for every non-zero column vector z of m real numbers the scalar $z^T Az \geqslant 0$).

We say that $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is *translation invariant* if $K(x, y) = K(x - t, y - t)$ for all $t, x, y \in \mathbb{R}^n$. In this case $K(x, y) = K(x - y, 0)$, so K can be viewed as a function on $\mathbb{R}^n$. Conversely, every positive definite function $\Phi : \mathbb{R}^n \to \mathbb{R}$ (see Definition 3.3.1) gives rise to a kernel that is translation invariant:

$$K(x, y) = \Phi(x - y).$$

Some examples of SPD translation invariant kernels are:

- Gaussian kernel: $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}, \quad x, y \in \mathbb{R}^n, \sigma > 0.$
- Inverse multiquadric kernel: $K(x, y) = \frac{1}{\sqrt{1+(\varepsilon\|x-y\|)^2}}, \quad x, y \in \mathbb{R}^n.$
- Matérn C^2 kernel: $K(x, y) = e^{-\varepsilon\|x-y\|}(\varepsilon\|x - y\| + 1), \quad x, y \in \mathbb{R}^n.$
- Wendland C^2 kernel: $K(x, y) = (1-\varepsilon\|x-y\|)_+^4(4\varepsilon\|x-y\|+1), \quad x, y \in \mathbb{R}^n.$

Let H be a Hilbert space of real-valued functions on Ω.

Definition 3.5.2. We say that H is *a reproducing kernel Hilbert space* if, for all $x \in \Omega$, the evaluation functional $L_x : f \to f(x)$ for all $f \in H$ is continuous at any f in H or, equivalently, if L_x is a bounded operator on H.

Definition 3.5.3. A function $K : \Omega \times \Omega \to \mathbb{R}$ is called *a reproducing kernel* for a Hilbert space H if

1. For every $x \in \Omega$ the functional $K_x := K(x, \cdot) \in H$;
2. For every $x \in \Omega$ and for every $f \in H \quad f(x) = \langle f, K_x \rangle_H$.

In fact, Definitions 3.5.2 and 3.5.3 are equivalent.

Proposition 3.5.4. *Suppose that H is a Hilbert space of functions $f : \Omega \to \mathbb{R}$. Then H is a reproducing kernel Hilbert space if and only if H has a reproducing kernel.*

Proof. Suppose that H has a reproducing kernel K. Then the reproducing property gives

$$|L_x(f)| = |f(x)| = |\langle f, K_x\rangle_H|.$$

Using the Cauchy-Schwarz inequality we can estimate

$$|L_x(f)| = |\langle f, K_x\rangle_H| \leqslant \|f\|\|K_x\|.$$

So, for all $x \in \Omega$ the functional $L_x : f \to f(x)$ for all $f \in H$ is continuous. Consequently, H is a reproducing kernel Hilbert space.

Now let us show that, conversely, every reproducing kernel Hilbert space has a unique reproducing kernel. The Riesz representation theorem implies that for all x in Ω there exists a unique element K_x of H with the reproducing property,

$$f(x) = L_x(f) = \langle f,\ K_x\rangle \quad \forall f \in H.$$

Since K_x is itself a function in H, it holds that for every y in Ω there exist a $K_y \in H$ such that

$$K_x(y) = \langle K_x,\ K_y\rangle.$$

This allows us to define the reproducing kernel of H as a function $K : \Omega \times \Omega \to \mathbb{R}$ by

$$K(x, y) = \langle K_x,\ K_y\rangle. \tag{3.5.1}$$

Clearly $K_x := K(x, \cdot) \in H$ and $f(x) = \langle f, K_x\rangle_H$. Thus, H has a reproducing kernel K. □

Let us give a few key examples.

Example 3.5.5 (Non-example)**.** $L^2[0, 1]$ is not RKHS.

The easiest way to demonstrate this fact is to construct a sequence $\{f_n\} \in L^2[0,1]$ such that $\lim_{n\to\infty} \|f_n\| = 0$ and $f_n(x_0) \neq 0$ for some fixed point $x_0 \in [0,1]$. Define $f_n(x) = (-nx+1)_+$.

Then, evidently, $\lim\limits_{n\to\infty} \|f_n\| = 0$ and $f_n(0) = 1$ for all n, So, the evaluation functional at 0 is not continuous.

Example 3.5.6 (L^2 on a discrete set)**.** Let X be a discrete set of points $\{x_i\} \subset \mathbb{R}$. Recall that the Dirac measures δ_{s_i} is defined by

$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{if } a \notin A \end{cases}$$

for any Lebesgue measurable set A. Choose the sequence of positive real numbers $a_1, a_2, \ldots$ and consider the measure

$$\mu = \sum_i a_i \delta_{x_i}.$$

Then $L^2(X,\mu)$ is RKHS. In this case the reproducing kernel $K : X \times X \to \mathbb{R}$ for $L^2(X,\mu)$ is

$$K(x_i, x_j) = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Example 3.5.7 (Sobolev space)**.** Consider the Sobolev space $\mathcal{H}^1[0,M]$ consisting of absolutely continuous functions $f : [0,M] \to \mathbb{R}$ whose distributional derivative lies in $L^2[0,M]$. $\mathcal{H}^1[0,M]$ has the inner product

$$\langle f, g\rangle_{\mathcal{H}^1} = \langle f, g\rangle_{L^2} + \langle f', g'\rangle_{L^2}.$$

We can demonstrate that an evaluation functional at any point is continuous. Indeed, for every $a \in [0,M]$ via integration by parts we have

$$\int_a^M f(t)dt = f(t)(t-M)\Big|_a^M - \int_a^M f'(t)(t-M)dt.$$

Therefore, for all $a \in [0,M)$

$$f(a) = \frac{1}{M-a}\left(\int_a^M f(t)dt + \int_a^M f'(t)(t-M)dt\right).$$

Then we can estimate the point evaluation functional at $a \in [0, M)$ using the Hölder inequality:

$$|f(a)| \leqslant \frac{1}{M-a}\left(\sqrt{\int_a^M 1dt}\sqrt{\int_a^M f^2(t)dt} + \sqrt{\int_a^M (t-M)^2 dt} \cdot \sqrt{\int_a^M (f'(t))^2 dt}\right)$$
$$\leqslant \frac{\sqrt{M-a}}{M-a}\left(\sqrt{\int_a^M f^2(t)dt} + \sqrt{\int_a^M (f'(t))^2 dt}\right) \leqslant \frac{1}{\sqrt{M-a}}\|f\|_{\mathcal{H}^1}.$$

Now consider the case of the evaluation functional at the point M.

$$|f(M)| \leqslant f(0) + \int_0^M |f'(t)|dt \leqslant f(0) + \sqrt{\int_0^M 1dt}\sqrt{\int_0^M (f'(t))^2 dt}$$
$$\leqslant \frac{1}{\sqrt{M}}\|f\|_{\mathcal{H}^1} + \sqrt{M}\|f\|_{\mathcal{H}^1} \leqslant \frac{M+1}{\sqrt{M}}\|f\|_{\mathcal{H}^1}.$$

So, we have shown that $|L_a(f)| = |f(a)| \leqslant C_a\|f\|$ for all $a \in [0, M]$.

Let us find the kernel function. The kernel $K : [0, M] \times [0, M] \to \mathbb{R}$ of the space $\mathcal{H}^1$ must exist and for all $x \in [0, M]$, $f \in \mathcal{H}^1[0, M]$ should satisfy

$$f(x) = \langle f(\cdot), K(x, \cdot)\rangle_{\mathcal{H}^1}. \tag{3.5.2}$$

From now on we keep x fixed and use only derivatives with respect to y. We can rewrite (3.5.2) as follow:

$$f(x) = \int_0^M f(y)K(x, y)dy + \int_0^M f'(y)K'(x, y)dy. \tag{3.5.3}$$

As this equation must hold for all $f \in L^2[0, M]$ we have to assume that $K(x, y)$ has a derivative discontinuity at $y = x$, and we split the integral there. Denote $K'_+(x, x)$ and $K'_-(x, x)$ right and left derivatives with respect to y at $y = x$ (i.e. $K'_+(x, x) = \lim\limits_{y\to x^+} \frac{K(x,y)-K(x,x)}{y-x}$, $K'_-(x, x) = \lim\limits_{y\to x^-} \frac{K(x,y)-K(x,x)}{y-x}$).

Using integration by parts on the second integral we can write

$$\int_0^M f'(y)K'(x,y)dy = \int_0^x f'(y)K'(x,y)dy + \int_x^M f'(y)K'(x,y)dy =$$

$$f(y)K'(x,y)\Big|_0^x - \int_0^x f(y)K''(x,y)dy + f(y)K'(x,y)\Big|_x^M - \int_x^M f(y)K''(x,y)dy =$$

$$-\int_0^M f(y)K''(x,y)dy + f(x)(K'_-(x,x) - K'_+(x,x))$$

$$- f(0)K'(x,0) + f(M)K'(x,M).$$

Substituting this expression in (3.5.3) we get

$$\begin{aligned} f(x) = &\int_0^M f(y)\left(K(x,y) - K''(x,y)\right)dy + \\ &f(x)(K'_-(x,x) - K'_+(x,x)) - f(0)K'(x,0) + f(M)K'(x,M). \end{aligned} \tag{3.5.4}$$

Thus, to find the kernel function we need to solve the boundary-value problem

$$K(x,y) - K''(x,y) = 0 \text{ for all } x,y \in [0,M], x \neq y \tag{3.5.5}$$

$$K'(x,0) = 0 \text{ for all } x \in [0,M] \tag{3.5.6}$$

$$K'(x,M) = 0 \text{ for all } x \in [0,M] \tag{3.5.7}$$

$$K'_-(x,x) - K'_+(x,x) = 1 \text{ for all } x \in [0,M] \tag{3.5.8}$$

The differential equation (3.5.5) has the general solution

$$K(x,y) = A(x)e^y + B(x)e^{-y}. \tag{3.5.9}$$

It remains to find coefficient functions $A(x), B(x)$ for which $K(x,y)$ satisfies (3.5.6)–(3.5.8). Denote $K'_-(x,x) := \alpha(x), K'_+(x,x) := \beta(x)$. We consider two cases separately.

Case 1: $y \leqslant x$ with $x > 0$. $K(x,y)$ have to satisfy

$$\begin{cases} K'(x,0) = A(x) - B(x) = 0, \\ K'(x,x) = A(x)e^x - B(x)e^{-x} = \alpha(x). \end{cases}$$

Therefore,

$$A(x) = \tfrac{\alpha(x)}{e^x - e^{-x}}, \quad B(x) = \tfrac{\alpha(x)}{e^x - e^{-x}}.$$

So, substituting in (3.5.9) we obtain

$$K(x,y) = \alpha(x)\tfrac{e^{y}+e^{-y}}{e^{x}-e^{-x}} = \alpha(x)\tfrac{\cosh(y)}{\sinh(x)}. \tag{3.5.10}$$

Case 2: $x \leqslant y$ with $x < 1$. Similarly, $K(x,y)$ have to satisfy

$$\begin{cases} K'(x,M) = A(x)e^{M} - B(x)e^{-M} = 0, \\ K'(x,x) = A(x)e^{x} - B(x)e^{-x} = \beta(x). \end{cases}$$

Therefore,

$$A(x) = \beta(x)\tfrac{e^{-M}}{e^{x-M}-e^{-(x-M)}}, \quad B(x) = \beta(x)\tfrac{e^{M}}{e^{x-M}-e^{-(x-M)}}.$$

So, substituting in (3.5.9) we obtain

$$K(x,y) = \beta(x)\tfrac{e^{y-M}+e^{-(y-M)}}{e^{x-M}-e^{-(x-M)}} = \beta(x)\tfrac{\cosh(y-M)}{\sinh(x-M)}. \tag{3.5.11}$$

Now, from (3.5.8) and the fact that $\lim\limits_{y\to x^{+}} K(x,y) = \lim\limits_{y\to x^{-}} K(x,y)$, we obtain the system for functions $\alpha(x)$ and $\beta(x)$:

$$\begin{cases} \alpha(x) - \beta(x) = 1, \\ \alpha(x)\tfrac{\cosh(x)}{\sinh(x)} - \beta(x)\tfrac{\cosh(x-M)}{\sinh(x-M)} = 0, \end{cases}$$

which results in

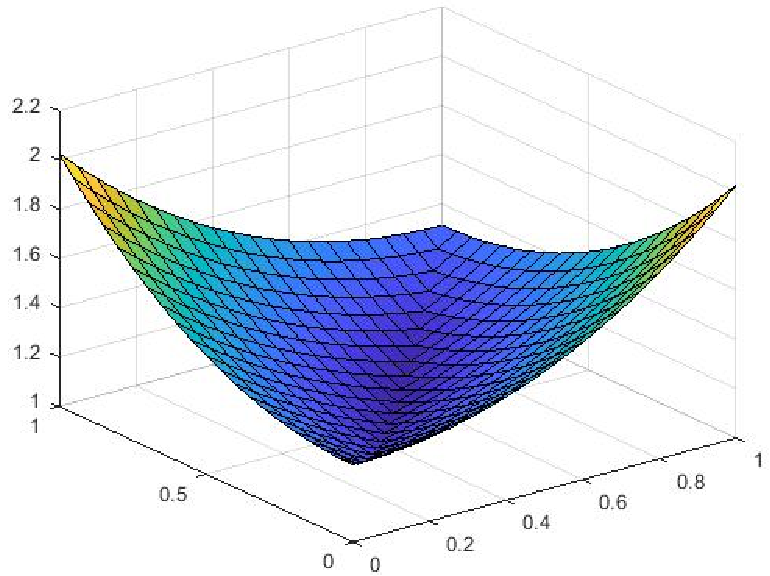
$$\begin{aligned} \alpha(x) &= \tfrac{\sinh(x)\cosh(x-M)}{\sinh(x)\cosh(x-M)-\sinh(x-M)\cosh(x)} = \tfrac{\sinh(x)\cosh(x-M)}{\sinh(M)} \\ \beta(x) &= \tfrac{\sinh(x-M)\cosh(x)}{\sinh(x)\cosh(x-M)-\sinh(x-M)\cosh(x)} = \tfrac{\sinh(x-M)\cosh(x)}{\sinh(M)}. \end{aligned}$$

Finally, together with (3.5.10) and (3.5.11) we obtain the result

$$K(x,y) = \begin{cases} \frac{\cosh(x-M)\cosh(y)}{\sinh(M)} & \text{if } x \leqslant y \\ \frac{\cosh(x)\cosh(y-M)}{\sinh(M)} & \text{if } x \geqslant y. \end{cases}$$

3.5.1 Native space

From the definition of the the reproducing kernel it is easy to see that $K : \Omega \times \Omega \to \mathbb{R}$ is symmetric and positive semi-definite. Namely, we know that



h!

Figure 3.2: The local kernel for $H^1[0,1]$.

$K(x,y) = \langle K_x,\ K_y\rangle_H$, so K is symmetric. Moreover, for $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, and $x_1, \ldots, x_n \in \Omega$ we have

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) = \left\langle \sum_{i=1}^{n} c_i K_{x_i}, \sum_{j=1}^{n} c_j K_{x_j} \right\rangle_H \geqslant 0,$$

so K is positive semi-definite.

The Moore-Aronszajn theorem goes in the other direction: it states that every symmetric, positive semi-definite kernel defines a unique reproducing kernel Hilbert space. Notice that the region $\Omega \in \mathbb{R}^n$ can be quite arbitrary except that it should contain at least one point.

Theorem 3.5.8 (Moore-Aronszajn)**.** *Suppose $K : \Omega \times \Omega \to \mathbb{R}$ is a symmetric positive definite kernel. Then there is a unique (up to isometry) Hilbert space $\mathscr{N}_K$ of functions on Ω for which K is a reproducing kernel. More precisely*

1. *For every $x \in \Omega$ the function $K_x(\cdot) = K(x, \cdot) \in \mathscr{N}_K$;*
2. *For every $x \in \Omega$ and for every $f \in \mathscr{N}_K$*

$$f(x) = \langle f, K_x \rangle_{\mathscr{N}_K}$$

This theorem first appeared in Aronszajn's Theory of Reproducing Kernels [3], although he attributes it to Eliakim Hastings Moore.

The associated function space $\mathcal{N}_K$ that has a given kernel K as its reproducing kernel is called the **native Hilbert space of a positive definite kernel** K. [62, Theorem 2.2] gives a description of a native space for (strictly) positive definite function.

Theorem 3.5.9. *Every symmetric positive definite function $K : \Omega \times \Omega \to \mathbb{R}$ has a unique native Hilbert space $\mathcal{N}_K(\Omega)$. It is the closure of the pre-Hilbert space*

$$H_K(\Omega) := \operatorname{span}\{K(\cdot, y) : y \in \Omega\}$$

under the inner product

$$\langle K(\cdot, x), K(\cdot, y)\rangle_{\mathcal{N}_K} = K(x, y) \text{ for all } x, y \in \Omega.$$

The elements of the native space can be interpreted as functions from Ω to $\mathbb{R}$ via the reproducing formula

$$f(x) = \langle f, K(\cdot, x)\rangle_{\mathcal{N}_K}.$$

So, there is a one-to-one correspondence between symmetric, positive definite kernel on Ω and Hilbert space of real-valued functions on Ω with the continuous evaluation functional.

One of the most difficult problems in the theory of RKHSs is starting with a positive definite function, K to give a concrete description of the space $H(K)$. We can refer to this as **the reconstruction problem**. However, there are some useful characterizations of the native spaces.

Theorem 3.5.10 (Characterization in terms of Fourier transform)**.** *Suppose $\Phi \in C(\mathbb{R}) \cap L_1(\mathbb{R})$ is a real-valued strictly positive definite function. Consider a translation invariant kernel $K(x, y) = \Phi(x - y)$. Then the native space $\mathcal{N}_K$ is given by*

$$\mathcal{N}_K(\mathbb{R}) = \{f \in C(\mathbb{R}) \cap L_2(\mathbb{R}) : \hat{f}/\sqrt{\hat{\Phi}} \in L_2(\mathbb{R})\},$$

and the native space inner product can be written as

$$\langle f, g\rangle_{\mathcal{N}_K} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\hat{f}(t)\overline{\hat{g}(t)}}{\hat{\Phi}(t)} dt.$$

In particular, every $f \in \mathcal{N}_K(\mathbb{R})$ *can be recovered from its Fourier transform* $\hat{f} \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$*.*

This result shows that in the case, when $\Omega = \mathbb{R}$ and the kernel is translation invariant, the native space actually consists of smooth functions. The proof of this fact can be found in [71, p. 139-141].

Another interesting characterization of the native space can be given in terms of the eigenfunctions of some linear operator associated with the reproducing kernel. Namely, the Mercer theorem provides a series representation for continuous kernels on compact domain.

Theorem 3.5.11 (Mercer)**.** *Let* $K : [a, b] \times [a, b] \to \mathbb{R}$ *be a continuous, symmetric, positive semi-definite kernel. Consider a linear operator* $T_K : L_2[a, b] \to L_2[a, b]$ *associated to* K*:*

$$[T_K(\varphi)](x) = \int_a^b K(x, t)\varphi(t)\, dt. \tag{3.5.12}$$

Then there is an orthonormal basis $\{\varphi_i\}_{i=1}^{\infty}$ *of* $L_2[a, b]$ *consisting of eigenfunctions of* T_K *such that the corresponding sequence of eigenvalues* $\{\lambda_i\}_{i=1}^{\infty}$ *is nonnegative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on* $[a, b]$ *and* K *has the representation*

$$K(x, y) = \sum_{j=1}^{\infty} \lambda_j\, \varphi_j(x)\, \varphi_j(y), \tag{3.5.13}$$

where the convergence is absolute and uniform.

Corollary 3.5.12 (Characterization in terms of eigenfunctions)**.** *Let* $K : [a, b] \times [a, b] \to \mathbb{R}$ *be a continuous, symmetric, positive semi-definite kernel,* $\{\varphi_i\}_{i=1}^{\infty}, \{\lambda_i\}_{i=1}^{\infty}$ *be the eigenfunctions and the eigenvalues of* T_K*. Then the native space* $\mathcal{N}_K$ *is given by*

$$\mathcal{N}_K[a, b] = \left\{ f \in L_2[a, b] : \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \left| \langle f, \varphi_i \rangle_{L_2[a,b]} \right| < \infty \right\},$$

and the native space inner product can be written as

$$\langle f, g \rangle_{\mathcal{N}_K} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle f, \varphi_i \rangle_{L_2[a,b]} \langle g, \varphi_i \rangle_{L_2[a,b]}\,.$$

It is one of the interesting topics in the theory of RKHS to deduce properties of the native space from properties of its reproducing kernel K, like continuity, measurability, differentiability ([64, Section 4.4]). For instance, continuity of K on $\Omega \times \Omega$ implies that all functions in the native space $\mathcal{N}_K$ are continuous on Ω.

At the end of this section, let us introduce a class of particularly important RKHS – universal RKHS.

Definition 3.5.13. Let Ω be a compact metric space, $K : \Omega \times \Omega \to \mathbb{R}$ be a continuous, symmetric, positive semi-definite kernel. $K : \Omega \times \Omega \to \mathbb{R}$ is called *universal* if its native space $\mathcal{N}_K$ is dense in $C(\Omega)$ with respect to uniform norm.

It is possible to prove that the following kernels are universal [64, Corollary 4.58]:

- Gaussian kernel: $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}, \quad x, y \in \mathbb{R}^n, \sigma > 0.$
- Exponential kernel: $K(x, y) = e^{\|x-y\|}, \quad x, y \in \mathbb{R}^n.$

Chapter 4

Prices prediction with supply and demand curves

In deregulated electricity markets, the study of price prediction is equally important for producers, buyers, investors and other load serving bodies for various reasons. These includes, among others, the cash flow analysis, least cost planning, integrated resource planning, financial procurement, optimal bidding strategies, regulatory rule-making and demand side management.

Instead of directly modeling the electricity price as it is usually done in time series or data mining approaches, we are going to model and utilize its true source: the sale and purchase curves of the electricity exchange.

4.1 Meshless approximation of supply and demand curves

Let us briefly notice some features of supply and demand curves that are relevant for our modeling:

- By construction, the curves are monotone.
- The values attained by the supply curve are roughly clustered around **layers**, corresponding to different production technologies. In Italy they are non-dispatchable renewables, gas, coal, hydro, oil.
- The fact that renewables are the first ones make the supply curve intrinsically "meshless".
- Demand is much more inelastic than supply.

So, we are dealing with a scattered data interpolation problem. We have a large amount of points (each point represents price and amount of electricity) that we want to approximate. We can formalize this problem as follows.

Given a set of N distinct *data points* $X_N = \{x_i : i = 1, 2, \ldots, N\}$ arbitrarily distributed on a domain $\Omega \subset \mathbb{R}$ and a set of *data values* (or function values) $Y_N = \{y_i : i = 1, 2, \ldots, N\} \subset \mathbb{R}$, the data interpolation problem consists in finding a function $s_f : \Omega \to \mathbb{R}$ such that

$$s_f(x_i) = y_i,\ i = 1, \ldots, N. \tag{4.1.1}$$

The idea of meshless approximation with radial basis functions is to find an approximant of f in the following form:

$$s_f(x) := \sum_{i=1}^{N} \alpha_i \phi(\|x - x_i\|)$$

where:

- the coefficients α_i and the **centers** x_i are to be chosen so that the interpolant is as near as possible as the original function f;
- $\phi : \mathbb{R} \to \mathbb{R}$ is a **radial basis function** (RBF).

Notice that the radial basis function $\phi \geqslant 0$, with $\alpha_i \geqslant 0$, so

$$\sum_{i=1}^{M} \alpha_i \phi(\|x - x_i\|) \geqslant 0.$$

As we need to approximate piecewise constant monotone function from $[0, M]$ to $\mathbb{R}^+$, we decided to use the integrals of RBF. Namely, we want to find an approximant of the form

$$s_f(t) = \int_0^t \sum_{i=1}^{M} \alpha_i \phi(\lambda_i \|x - x_i\|)\, dx = \sum_{i=1}^{M} \alpha_i \int_0^t \phi(\lambda_i \|x - x_i\|)\, dx$$

where λ_i is a shape parameter for every center x_i.

4.2 Approximation by Gauss error function

Let $F(x)$ be a function which corresponds to the supply curve (i.e. piecewise non-decreasing constant function from $[0, M]$ to $\mathbb{R}^+$). We need to find a function $G(x)$, such that the difference between F and G is reasonably small. The derivative of $F(x)$ in the sense of distribution is the sum of Dirac delta functions centered in the "jumps" of the supply curve. Also we know that the Dirac delta distribution can be written as limit of Gaussians:

$$\delta(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{t^2/2\sigma^2}.$$

Therefore, it seems natural to search for $G(x)$ as a linear combination of functions

$$g(x) = A \cdot \frac{2}{\sqrt{\pi}} \int\limits_0^{C(x-B)} e^{-t^2}\, dt + D, \tag{4.2.1}$$

which are called *the Gauss error function.*

The error function is a special non-elementary function of sigmoid shape that occurs in probability, statistics, and partial differential equations describing diffusions. The standard error function is defined as:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

Let us denote

$$h(x) = \begin{cases} 1 & \text{if } x \geqslant 0, \\ -1 & \text{if } x < 0. \end{cases}.$$

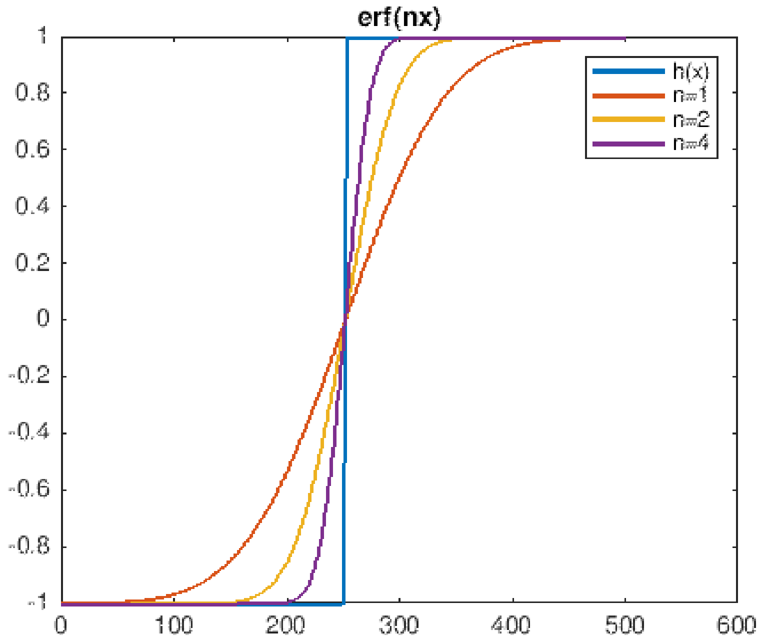
Notice that any supply curve can be expressed as a linear combination of functions $h(x - a)$ up to a constant. So, the problem of approximation the supply curve leads to the problem of approximation of $h(x - a)$ by error functions. In this subsection we are going to give an estimation of the difference between the step function $h(x)$ and Gauss error function.

From the picture we can see that $\operatorname{erf}_n(x) = \operatorname{erf}(n \cdot x)$ gets closer to $h(x)$ as n becomes bigger. So, our first task is to examine, in which sense erf_n converges to h. We are going to check four types of convergence:

1. Uniform convergence;

2. Pointwise almost everywhere convergence;

Figure 4.1: Gauss error function



3. Convergence in measure;

4. L_2 convergence on a real line;

For the first three items the answer is immediate. There is no uniform convergence, because $|\operatorname{erf}_n(0) - h(0)| = 1$ for all n. But for every $x \neq 0$ $\operatorname{erf}_n(x) \to h(x)$, so erf_n converges to h almost everywhere. And therefore, $\operatorname{erf}_n \to h$ in measure. To obtain the answer about $L_2(\mathbb{R})$ convergence we need to use some additional theory.

One of the related functions is *the complementary error function*, which is defined as

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\,dt.$$

To obtain an estimation for $\|h - \mathrm{erf_n}\|_{L^2(\mathbb{R})}$, we will use the following known fact about the complementary error function from [1]:

Lemma 4.2.1.

$$\int_0^{+\infty} \operatorname{erfc}^2(x)dx = \frac{2-\sqrt{2}}{\sqrt{\pi}}.$$

Proposition 4.2.2. *Consider the functions*

$$h(x) = \begin{cases} 1 & \textit{if } x \geqslant 0, \\ -1 & \textit{if } x < 0. \end{cases} \quad \textit{and} \ \operatorname{erf}_n(x) = \frac{1}{\sqrt{\pi}} \int_{-nx}^{nx} e^{-t^2}\, dt.$$

Then for every $n \in \mathbb{N}$

$$\| \operatorname{erf}_n - h \|_{L_2(\mathbb{R})} = \sqrt{\frac{2(2-\sqrt{2})}{n\sqrt{\pi}}}. \tag{4.2.2}$$

Proof. We can write

$$h(x) - \operatorname{erf}_n(x) = \begin{cases} 1 - \operatorname{erf}(nx) & \text{if } x \geqslant 0, \\ -(1 - \operatorname{erf}(-nx)) & \text{if } x < 0. \end{cases}.$$

It means that $h(x) - \operatorname{erf}_n(x) = \operatorname{sign}(x) \cdot \operatorname{erfc}(|nx|)$, and so,

$$(h(x) - \operatorname{erf}_n(x))^2 = \operatorname{erfc}^2(|nx|).$$

Therefore

$$\begin{aligned} \| \operatorname{erf}_n - h \|^2_{L_2(\mathbb{R})} &= \int_{-\infty}^{\infty} \operatorname{erfc}^2(|nx|)\, dx = 2 \int_0^{\infty} \operatorname{erfc}^2(nx)\, dx \\ &= \frac{2}{n} \int_0^{\infty} \operatorname{erfc}^2(y)\, dy = \frac{2}{n} \cdot \frac{2-\sqrt{2}}{\sqrt{\pi}} \end{aligned}$$

The last equality is obtained from Lemma 4.2.1, and this ends the proof. □

The goal of the next theorem is to show that any supply curve (piecewise constant function with a finite number of segments) can be approximated by a combination of error functions in the sense of L_2 convergence.

Theorem 4.2.3. *Any piecewise constant function can be approximated by the linear combination of error functions in the sense of $\| \cdot \|_{L_2(\mathbb{R})}$.*

More precisely, if we have a function of the form

$$F(x) = \sum_{i=1}^{k} a_i h(x - b_i) + c_i + D,$$

then for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that, for

$$G(x) = \sum_{i=1}^{k} a_i \operatorname{erf}(N \cdot (x - b_i)) + c_i + D,$$

it holds $\|F - G\|_{L_2} < \varepsilon$.

Proof. Without loss of generality we can assume that $D = 0$. Fix $\varepsilon > 0$. Let us denote

$$f_i(x) = a_i h(x - b_i) + c_i.$$

So, $F(x) = \sum_{i=1}^{k} f_i(x)$. For each i consider the sequence of functions

$$g_{i,n}(x) = a_i \operatorname{erf}(n(x - b_i)) + c_i.$$

Then, from (4.2.2) we have the estimation

$$\|f_i - g_{i,n}\|_{L_2} = a_i \|\operatorname{erf}_n - h\|_{L_2} = a_i \sqrt{\frac{2(2 - \sqrt{2})}{n\sqrt{\pi}}}.$$

We can choose N_i such that $\|f_i - g_{i,N_i}\|_{L_2} < \frac{\varepsilon}{k}$. Therefore, taking $N = \max_i N_i$, we obtain

$$\|f_i - g_{i,N}\|_{L_2} < \frac{\varepsilon}{k} \text{ for all } i. \tag{4.2.3}$$

Now take $G(x) = \sum_{i=1}^{k} a_i \operatorname{erf}(N \cdot (x - b_i)) + c_i$. Then we can estimate

$$\|F - G\|_{L_2} = \left\| \sum_{i=1}^{k} f_i - \sum_{i=1}^{k} g_{i,N} \right\|_{L_2} \leqslant \sum_{i=1}^{k} \|f_i - g_{i,N}\|_{L_2} \overset{(4.2.3)}{<} k\frac{\varepsilon}{k} = \varepsilon.$$

□

Evidently, any supply curve and any demand curve can be approximated by a combination of error functions, which is the integral of a normalized Gaussian function. The standard error function is defined as:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2}\, dt = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2}\, dt.$$

Since we want to approximate monotone curves we came up with the idea to use the integral of radial basis function. In order to find unknown coefficients α_i, λ_i, x_i we need to solve global minimization problem:

$$\min_p \|s_f(x_i, p) - y_i\|_2^2,$$

where $p = (\alpha_i, \lambda_i, x_i)_{i=1,\dots,N}$ and

$$s_f(t,p) := \sum_{i=1}^{M} \alpha_i \int_0^t \phi(\lambda_i \|x - x_i\|)\, dx$$

and $\phi(t) = (\mathrm{erf}(t) + 1)/2$ is the primitive of a Gaussian kernel. However, this optimization problem is very heavy, as it is a nonlinear and nonconvex minimization over $p \in \mathbb{R}^{3M}$ ($M \simeq 150$).

We divide our global problem in simpler subproblems, with lower dimensionality, so that the final result is faster. Let us to describe our method in some more details for the supply curve.

First, we divided the y-axis into M equal intervals, and approximate the supply function on that interval exactly with one basis function $\rightsquigarrow$ M 3-dimensional optimization problems. However, this has the huge drawback that a huge jump concentrates on itself, keeping uselessly many components. Then we divide the y-axis into M intervals $[p_i, p_{i+1}]$, where the p_i correspond to the greatest quantity Q_i offered, i.e. to the largest "plateaus" on the bidding curve; again, we approximate the supply function on that interval exactly with one basis function. On each part we need to fine only 3 coefficients. For the realization of our algorithm we are using standard function `lsqcurvefit` from MatLab Optimization Toolbox.

For optimizing the numerical procedure we solved some parts of the optimization problem by ourselves: when the interval $[p_i, p_{i+1}]$ contains just one jump, then

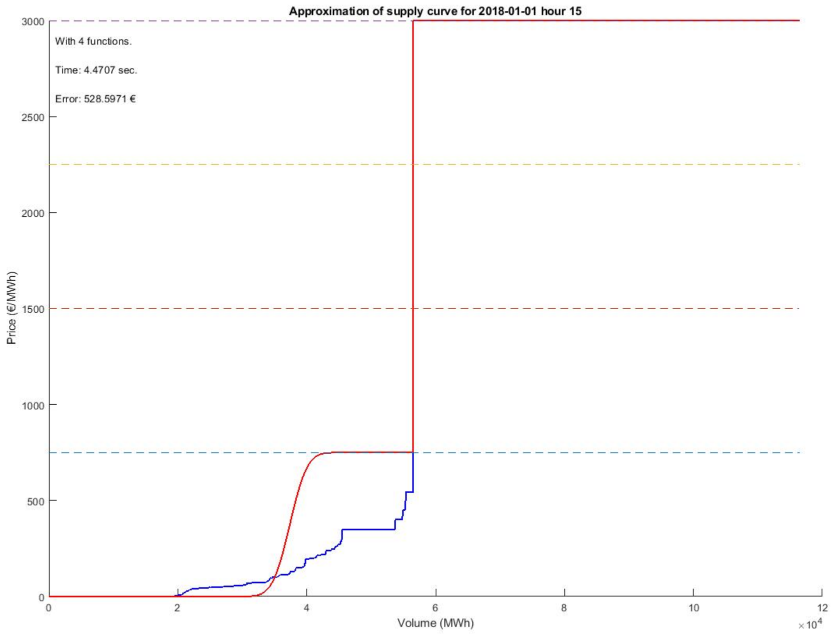
$$\alpha_i := f(p_{i+1}) - f(p_i)$$

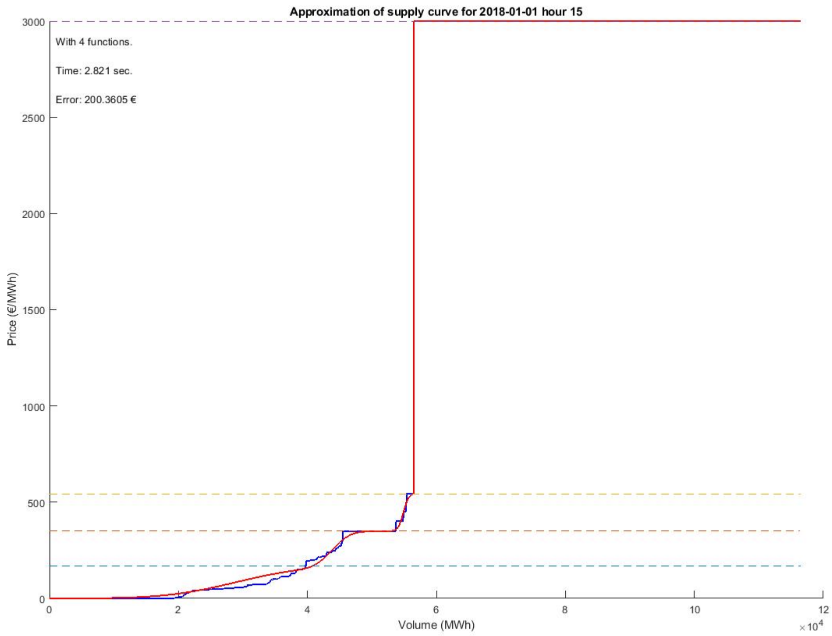
for any kernel function ϕ with unit integral.

4.3 Data set

In our work we are using the data about supply bids from the Italian electricity market from the GME website www.mercatoelettrico.org. We consider time period from 01.01.2017 to 31.12.2017. These data are in aggregated form, i.e. bids coming from different agents but with the same price are aggregated

Figure 4.2: Method 1 and Method 2





in the price layer. Even in this form, we are dealing with the massive amount of data. For instance, there were observed **2 800 687** offer and **558 926** bid layers during this period.

So, it means, that on average there are 324 offer and 65 bid layers for each hour of the year, which corresponds to one supply curve and one demand curve respectively.

Table 4.1: Data

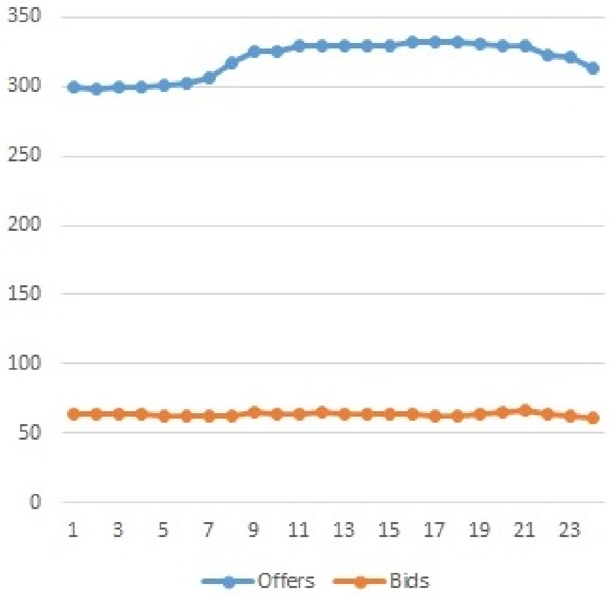
Date	Hour	Volume (MW)	Price (Euro)
01-01-2017	1	13392.7	0
01-01-2017	1	25	0.1
01-01-2017	1	113.8	1
01-01-2017	1	11	3.5
01-01-2017	1	270.3	5
01-01-2017	1	0.5	6
....	......	....	....
31-12-2017	24	370	554.2
31-12-2017	24	352	554.3
31-12-2017	24	365	554.5
31-12-2017	24	97	700
31-12-2017	24	60000	3000

It is a known fact that the dynamics of electricity trade displays a set of characteristics: external weather conditions, dependence of the consumption on the hour of the day, the day of the week, and time of the year. Variation in prices are all dependent on the principles of demand and supply. First of all, on the day-ahead market the energy is traded on an hourly basis and this means that the prices can and will vary per hour. For example, at 9:00 a.m. there could be a price peak, while at 4:00 a.m. prices could be only half of the peak price. Second, the weekly seasonal behaviour matters. Usually, it is necessary to differentiate between the two weekend days (Saturday and Sunday), the first business day of the week (Monday), the last business day of the week (Friday) and the remaining business days. Thirdly, electricity spot prices display a strong seasonal pattern. For instance, demand increases in summer, as consumers turn their air conditioners on, and also in winter because of electric heating in housing.

As far as the number of offers (or bids) affects directly the complexity of approximation, we decided to explore the relationship between the number of bids and offers and such a characteristics as the hour of the day, the day of the week, and the month of the year. Based on the dependence between this three factors and electricity prices we could expect that some hours, days have much less offers and bids than another one. This analysis is presented on

Figures 4.3 – 4.5.

The main conclusion that we have made is that there is no direct relationship between the number of offer and bid layers and the hour of the day, the day of the week, and the time of the year. In particular, during 24 hour of the day the number of offer layers varies between 299 and 332, and the number of bid layers varies between 61 and 66. With regard to dependence of the day of the week the number of offer layers varies between 310 and 320, and the number of bid layers varies between 55 and 68. Based on this observation we decided to chose the same number of basis functions independently of the hour of the day, the day of the week, and the time of the year.

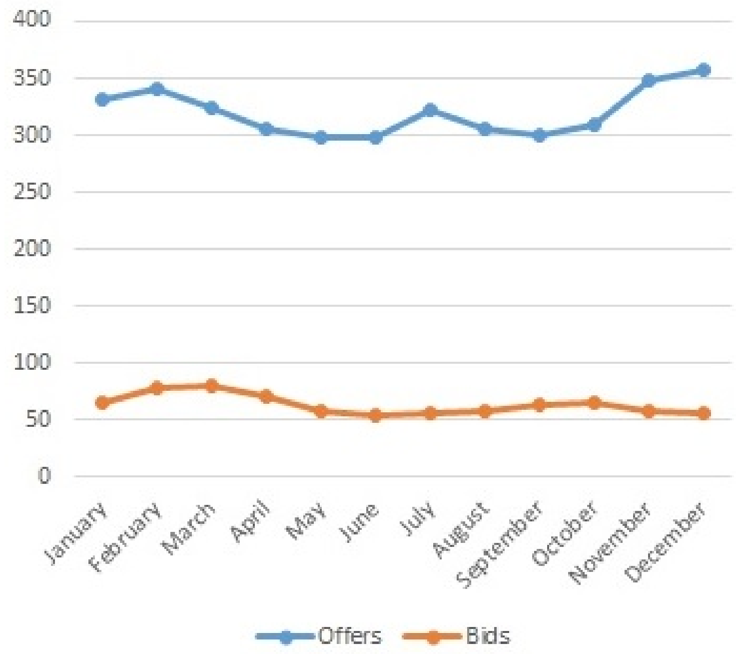


Hour	Number		Hour	Number	
	of offers	of bids		of offers	of bids
1	300	64	**13**	329	64
2	299	64	**14**	329	64
3	300	64	**15**	330	64
4	300	64	**16**	332	64
5	301	63	**17**	332	63
6	303	63	**18**	332	63
7	307	62	**19**	331	64
8	318	63	**20**	329	65
9	325	65	**21**	329	66
10	326	64	**22**	323	64
11	329	64	**23**	321	63
12	329	65	**24**	314	61

Figure 4.3: Hour dependence of the number of offer and bid layers

Month	Number	
	of offers	of bids
Sunday	310	55
Monday	310	56
Tuesday	322	68
Wednesday	324	67
Thursday	326	68
Friday	327	68
Saturday	329	68

Figure 4.4: Weekly dependence of the number of offer and bid layers



Month	Number	
	of offers	of bids
January	331	65
February	341	79
March	324	81
April	305	72
May	298	57
June	298	54
July	322	55
August	305	58
September	300	64
October	309	66
November	348	58
December	357	57

Figure 4.5: Monthly dependence of the number of offer and bid layers

4.4 Numerical experiments

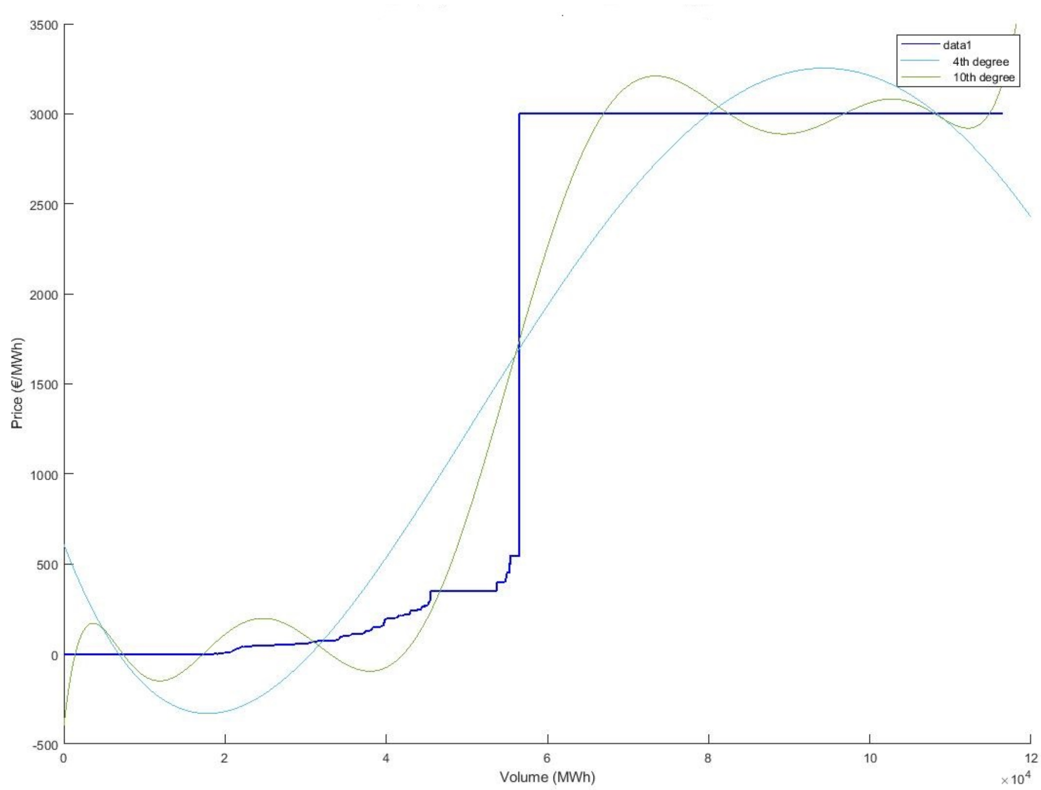
Since the maximum market clearing price for the period under review (i.e. from 01.01.2017 to 31.12.2017) is 350 €, in all the experiments we restricted ourselves to a maximum price 400 €. In Figure 4.6 we demonstrate that the approximation by polynomials does not suit to our problem. In Theorem 4.2.3 we have showed that we can approximate supply curve with a linear combination of error functions. Now we want to implement this into practice using MatLab. First of all, we care about

- accuracy of the approximation;
- running time.

Notice that Runge's phenomenon (1901) shows that for high values of N, the interpolation polynomial may oscillate wildly between the data points. Besides, the polynomial interpolation does not guarantee of monotonicity of the curves (see Figure 4.6).

For the realization of our algorithm we are using standard function `lsqcurvefit` from MatLab Optimization Toolbox and functions `main`,

Figure 4.6: Approximation of supply curve with polynomials



`datainterpolation`, `onestepdata`.

In the function `main` we download the data from a text file and choose the number of basis function M. The result of function `datainterpolation` is the coefficients a_i, b_i, c_i of the function

$$G(x) = \sum_{i=1}^{k} a_i(\mathrm{erf}(c_i \cdot (x - b_i)) + 1). \tag{4.4.1}$$

Here for the calculation convenience we are using $\{\mathrm{erf}(c_i \cdot (x - b_i)) + 1\}$ instead of $\{\mathrm{erf}(c_i \cdot (x - b_i))\}$, as our data values are never negative.

The `lsqcurvefit` function solves nonlinear data-fitting problems in least-squares sense. Suppose that we have data points $X_N = \{x_i : i = 1, 2, \ldots, N\}$ and data values $Y_N = \{y_i : i = 1, 2, \ldots, N\} \subset \mathbb{R}$ and we want to find a function f such that $f(x_i) \approx y_i$, $i = 1, \ldots, N$. We can consider the family of functions $\{f(x, p) : p \in \mathbb{R}^k\}$, depending of some parameter $p \in \mathbb{R}^k$. Let $p_0 \in \mathbb{R}^k$ be an "initial guess" such that $f(x_i, p)$ is reasonably close to y_i.

The function `lsqcurvefit` starts at p_0 and finds coefficients p from some neighborhood of p_0 to best fit the data set Y_N:

$$\min_p \|f(x_i, p) - y_i\|_2^2.$$

Notice that this function works well only if the number of parameters $(p_1, \ldots, p_k)$ is not very big. That is why we are forced to divide our problem into many local problems.

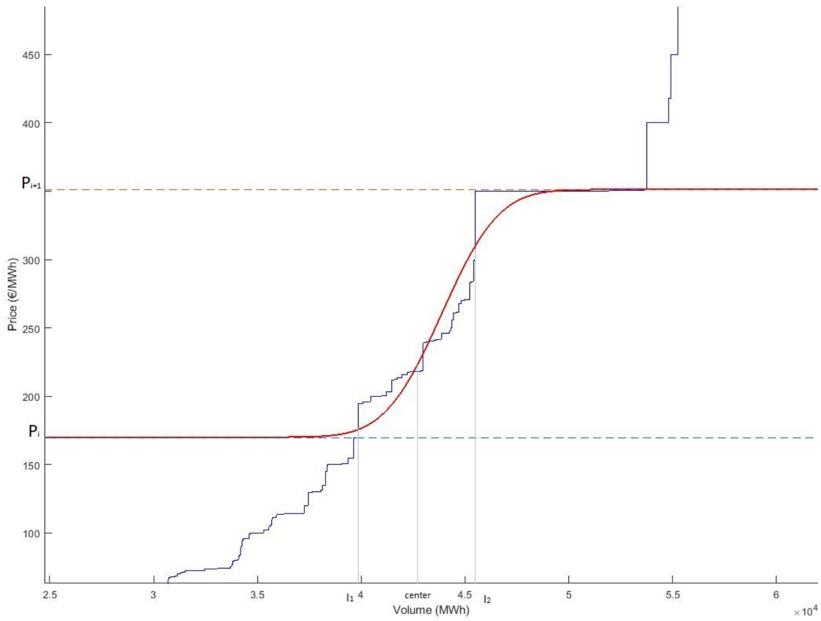
After we choose the number of basis function M, we want to divide our problem into M sub-problems. Then each part of the supply curve must be approximated by one error function. Our first attempt (Method 1) was just to divide y-axis uniformly into M equal intervals (see Figure 4.2, A). However this approach is ineffective, as "jumps" of supply curve can be bigger that the length of these intervals. To resolve this problem we created a simple algorithm that finds the points $P_1, \ldots, P_M$ on the y-axis such that our supply curve takes the value exactly P_i on some non-trivial interval (see Figure 4.2, B). Then M times we resolve the same optimization problem for the values of the supply curve between P_i and P_{i+1} using function `lsqcurvefit`. The function `onestepdata` gives for each step the initial point p_0.

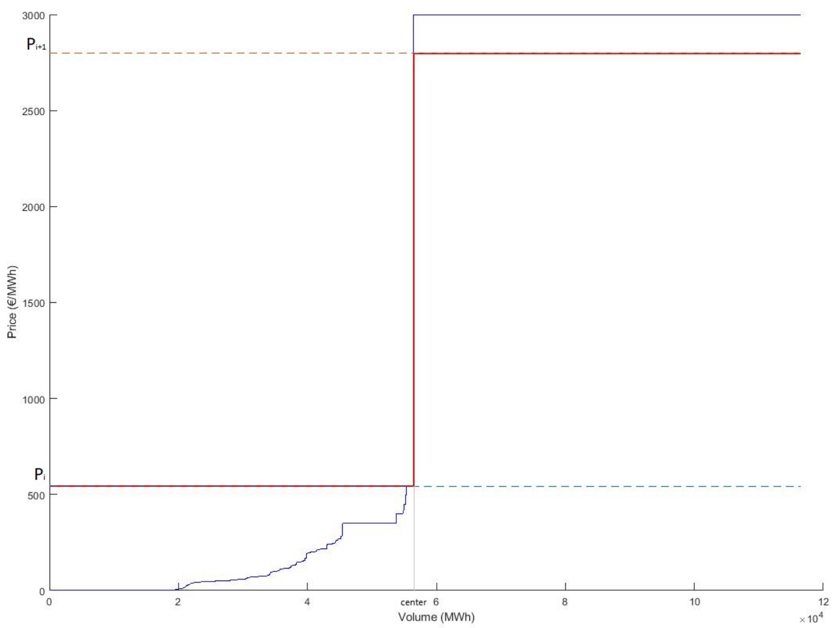
A summary of the results is shown in Table 4.2. For all experiments we proceed with the data for period from 01.01.2017 to 31.12.2017. We used different number of basis function to approximate supply and demand curves, and then compared the equilibrium price, which was received as intersection of approximants (P_{appr}), with the correct equilibrium price (P). We did this for each hour of each day, and then computed the average value of $|P - P_{appr}|$ (Error) for all 8 664 hours of the year and the maximum value of $|P - P_{appr}|$ (Max error).

This empirical results show that the accuracy of our approximation is good enough, if we use 5 basis function for the demand curve and 15 basis function for the supply curve. Then the increase in the number of functions leads to more time-consumption, but the increase of the accuracy is less significant.

As a last step we analyzed the stability of the coefficients for the case, when

Figure 4.7: Local interpolation by one error function with `lsqcurvefit` function





we approximate the supply curve with 10 basis functions and the demand curve with 5 basis functions for the same period of time.

$$S(x) = \sum_{i=1}^{10} A_i(\mathrm{erf}(C_i \cdot (x - B_i)) + 1) \text{ and } D(x) = \sum_{i=1}^{5} E_i(\mathrm{erf}(K_i \cdot (x - L_i)) + 1).$$

From Table 4.3 we can see that these coefficients do not have a stable behavior (namely, maximum values, minimum values and mean values are presented).

Table 4.2: Results of numerical experiment

Number of functions		Results		
For demand	For supply	Error	Max error	Running time
5	5	3.9 €	28.6 €	69 min.
5	10	2.2 €	14.9 €	82 min.
5	15	1.5 €	11.1 €	103 min.
5	20	1.3 €	9.1 €	110 min.
5	25	1.2 €	9.3 €	135 min.
5	30	1.2 €	9.4 €	159 min.
5	35	1.2 €	9.8 €	177 min.
5	40	1.2 €	9.6 €	190 min.
5	45	1.2 €	9.6 €	199 min.
5	50	1.2 €	9.6 €	207 min.
10	5	3.9 €	39.5 €	100 min.
10	10	2.1 €	14.9 €	128 min.
10	15	1.4 €	8.9 €	146 min.
10	20	1.2 €	9.1 €	162 min.
10	25	1.1 €	9.5 €	183 min.
10	30	1.1 €	9.3 €	199 min.
10	35	1.0 €	9.4 €	223 min.
10	40	0.98 €	9.8 €	241 min.
10	45	0.98 €	9.6 €	255 min.
10	50	0.98 €	9.6 €	273 min.

Although the values attained by the supply curve are clustered around layers, which correspond to different production technologies, we came to the conclusion that we have no chance to choose these coefficients uniformly for all curves, but we need to calculate them for all supply and demand curves.

Figure 4.8: Supply curve approximated with 10 basis functions

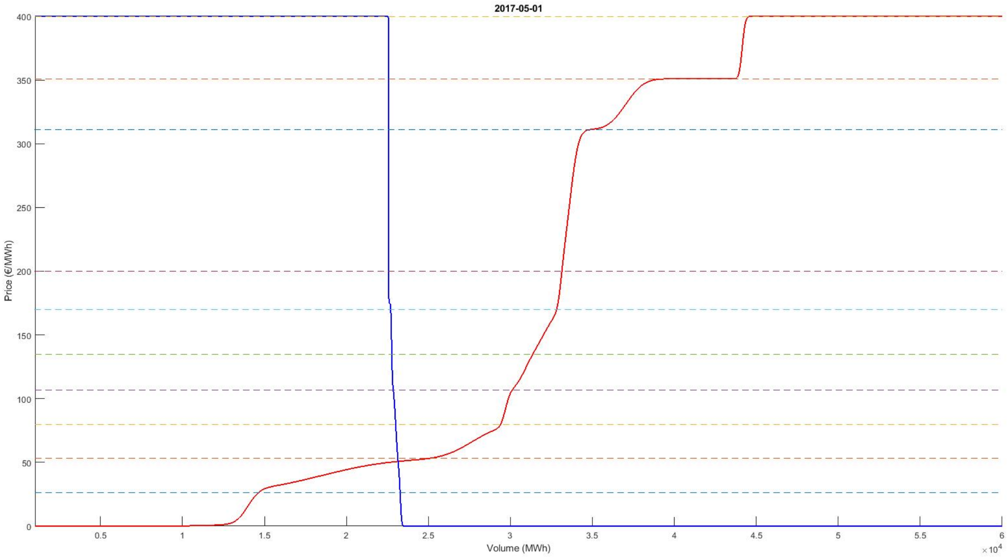


Table 4.3: Stability of the coefficients

	Min	Mean	Max
Coeffitients for supply curve			
A_1	10	14.76981	18
A_2	10.5	15.15519	21
A_3	10.5	15.21438	19.5
A_4	11	15.53944	22
A_5	11	16.8968	27.5
A_6	12.5	20.44287	27
A_7	14.5	22.15457	33
A_8	19	29.69132	57.5
A_9	17	24.48784	48
A_{10}	21	25.64777	50
Coeffitients for demand curve			
E_1	12	30.95154	37.5
E_2	25	34.31039	58.5
E_3	25	36.24469	50
E_4	33	40.19715	50
E_5	50	58.29623	75

So, in this section we presented a parsimonious way to represent supply and demand curves, using a mesh-free method based on Radial Basis Functions. Using the tools of functional data analysis, we are able to approximate the original curves with far less parameters than the original ones. Namely, in order to approximate piece-wise constant monotone functions, we are using the combination of the integral of a normalized Gaussian function.

4.5 Price and demand forecasting based on supply and demand curves

Our main goal in this section is to forecast next-day electricity demand and prices using approximated supply and demand curves and to compare different modeling techniques. The classical models do not explain the relationships between market clearing price and different influential factors that can be essential in the problem of price prediction. To this purpose, we want to compare commonly used autoregressive models, based just on the clearing price, with ours, based on supply and demand curves. For this test, we are using again the data about supply bids from the Italian electricity market considering the time period from 01.01.2017 to 31.12.2017. In particular, our training set includes data from 01.01.2017 to 31.10.2017, while the test set which is used for forecasting to test the performance of the model on out-of-sample data is from 01.11.2017 to 31.12.2017. We will consider a linear parametric autoregressive (AR) model for univariate price prediction and functional autoregressive (FAR) models for the prediction of supply and demand curves.

We performed electricity price forecasting using six different methods: autoregressive model of order 1 with (SAR(1)) and without seasonality (AR(1)) for the closing price; functional autoregressive model of order 1 applied to the modeled supply and demand curves, where for the representation of demand curve we used one basis function and for the representation of supply curve we used 5 or 10 functions (FAR(1) (5 functions) and FAR(1) (10 functions), respectively) together with the corresponding seasonal models (SFAR(1) (5 functions) and SFAR(1) (10 functions), respectively). In all the seasonal versions, dummy variables corresponding to weekdays were introduced. These models were applied to each market hour separately.

While formulations of AR(1) and SAR(1) models for the closing prices are quite standard (thus we do not give details on them here), we feel that a description of our implementation of FAR(1) and SFAR(1) models for supply and demand curves are needed. We considered the simplified representation of the supply curve $S_{d,h}(x)$ with M basis functions, and the demand curve $D_{d,h}(x)$ with one basis function, at day d and hour h, keeping the shape parameter constantly equal to 1

$$\begin{aligned} S_{d,h}(x) &= \sum_{i=1}^{M} A_{d,h,i} \cdot (\mathrm{erf}((x - B_{d,h,i})) + 1), \quad M = 5 \text{ or } M = 10, \\ D_{d,h}(x) &= 200 \cdot \mathrm{erf}((x - L_{d,h})) + 1. \end{aligned}$$

Then we provide a model for the process $X_{d,h} = (X^1_{d,h}, X^2_{d,h}, \ldots, X^{2M}_{d,h})$, where

$$\begin{aligned} X^i_{d,h} &= A_{d,h,i}, \qquad i = 1, \ldots, M-1, \\ X^{i+M-1}_{d,h} &= B_{d,h,i}, \; i = 1, \ldots, M, \\ X^{2M}_{d,h} &= L_{d,h}. \end{aligned}$$

Notice that, as we restricted ourselves to a maximum price (and so the maximum of supply and demand curves) of 400 €, we need to exclude the parameter $A_{d,h,M}$ from the model, as it is linearly dependent on others. The considered time series model FAR(1) for $X_{d,h}$ for each hour h is given by

$$X_{d,h} = \nu_d + \Phi_d X_{d,h-1} + \varepsilon_{d,h}$$

with the $2M \times 2M$ matrix Φ_d, and the $2M$-dimentional vector ν_d as parameters, and $\varepsilon_{d,h}$ as error term. We assume that the error process $\varepsilon_{d,h}$ is a $2M$-dimensional white noise process.

For modeling the day of the week impact in SFAR(1) models we define additionally function $W(d)$ that gives a number that corresponds to the weekday of day d ($W(d) = 1$ for a Sunday, for a Monday $W(d) = 2$ up to $W(d) = 7$ for a Saturday), and the weekday indicators

$$W_k(d) = \begin{cases} 1, \text{if } W(d) = k \\ 0, \text{if } W(d) \neq k \end{cases} .$$

We introduced parameters $D_{d,h,k}$ for the weekday effect. Thus, the corresponding SFAR(1) model for $X_{d,h}$ for each hour h and is written, in terms of coefficients, as

$$X_{d,h} = \nu_d + \Phi_d X_{d,h-1} + \sum_{k=1}^{7} W_k(d) D_{d,h,k} + \varepsilon_{d,h}.$$

We compared the results obtained with our functional approach with corresponding univariate price prediction. Three different summary measures, namely, mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) were used to evaluate the out-of-sample forecasting performance. Let us denote E_{dh} and $\hat{E}_{dh}$ the observed and the predicted values for day d, $\quad d = 1, \ldots, T = 61$ and hour h, $\quad h = 1, \ldots, 24$. We computed

$$MAE = \frac{\sum_{i=1}^{T} |E_{ih} - \hat{E}_{dh}|}{T}, \quad h = 1, \ldots, 24;$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T} (E_{ih} - \hat{E}_{dh})^2}{T}}, \quad h = 1, \ldots, 24;$$

$$MAPE = \frac{\sum_{i=1}^{T} |E_{ih} - \hat{E}_{dh}|/E_{ih}}{T}, \quad h = 1, \ldots, 24;$$

Table 4 provide summary statistics of errors for the forecasting of next-day electricity price. In order to facilitate the comparison between different methods we plot the errors for each of the six methods on Figures 4.9, 4.10 and 4.11.

As expected, SAR(1) performs better than AR(1). Surprisingly, instead, functional autoregressive models without seasonality gives better results than corresponding seasonal models. By comparing functional autoregressive models with 5 and 10 functions we can see similar results, so increasing the number of parameters does not lead to the improvement of the prediction accuracy. These two outcomes could be possibly due to overfitting effects. These results shows that we should use FAR(1) (5 functions) as this method is less time-consuming than the one with 10 functions. Finally, our method FAR(1) (5 functions) gives considerably more accurate results compared to the SAR(1) model for all hours. In particular, not only SAR(1) gives an average of the MAPE equal to

16.51% while FAR(1) (5 functions) gives 14.98%, but we can see that FAR(1) (5 functions) performs significantly better than SAR(1) on every single hour. Also comparing MAE and RMSE we obtain similar results.

Due to the superior performance of FAR(1) (5 functions) method, we also conducted prediction of electricity demand with just three methods: AR(1), SAR(1), and FAR(1) (5 functions). Table 5 provide summary statistics of errors for the forecasting of next-day electricity demand also represented in Figures 4.12, 4.13, 4.14. In this case AR(1) gives an average of the mean absolute percentage error 12.82%, SAR(1) gives 11.33% and FAR(1) (5 functions) gives 10.04%. Moreover, FAR(1) (5 functions) for the demand forecasting again gives more accurate results compared to the AR(1) model for all hours and also compared to the SAR(1) model. The same is true for MAE and RMSE.

Figure 4.9: Mean absolute error for price forecasting.

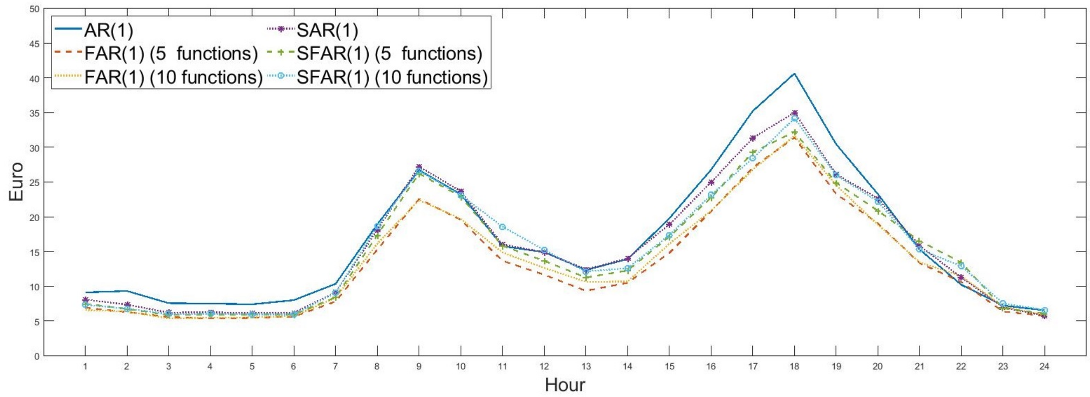


Figure 4.10: Root mean square error for price forecasting.

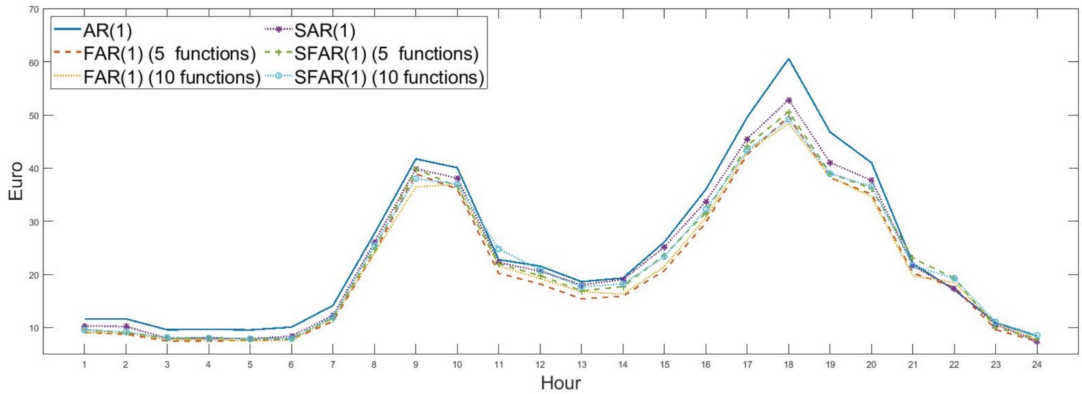


Figure 4.11: Mean absolute percentage error for price forecasting.

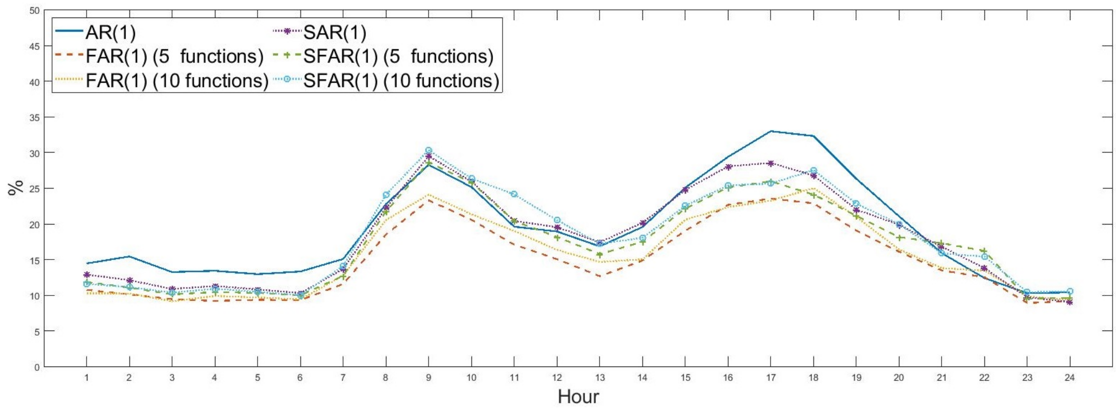


Table 4.4: Price prediction accuracy statistics.

Model	Hour	MAE Euro	RMSE Euro	MAPE %	Hour	MAE Euro	RMSE Euro	MAPE %
AR(1)	**1**	9.12	11.59	14.47	**13**	12.33	18.62	16.86
FAR(1) (5 functions)		6.92	9.06	10.77		9.35	15.37	12.68
FAR(1) (10 functions)		6.57	9.07	10.28		10.66	16.76	14.68
SAR(1)		8.1	10.31	12.92		12.43	18.04	17.43
SFAR(1) (5 functions)		7.5	9.60	11.86		11.27	16.91	15.77
SFAR(1) (10 functions)		7.32	9.55	11.62		12.14	17.67	17.22
AR(1)	**2**	9.32	11.64	15.45	**14**	13.92	19.31	19.58
FAR(1) (5 functions)		6.29	8.75	10.14		10.51	15.91	14.88
FAR(1) (10 functions)		6.39	9.03	10.27		10.68	16.28	15.05
SAR(1)		7.37	10.20	12.13		14.02	19.01	20.17
SFAR(1) (5 functions)		6.74	9.06	11.09		12.28	17.70	17.51
SFAR(1) (10 functions)		6.80	9.13	11.16		12.60	18.23	18.07
AR(1)	**3**	7.58	9.58	13.26	**15**	19.78	26.10	25.10
FAR(1) (5 functions)		5.59	7.47	9.47		14.80	20.73	19.05
FAR(1) (10 functions)		5.39	7.76	9.18		16.12	21.49	20.60
SAR(1)		6.22	7.97	10.89		18.91	25.16	24.76
SFAR(1) (5 functions)		5.88	7.98	10.13		17.14	23.63	22.16
SFAR(1) (10 functions)		6.00	8.13	10.39		17.37	23.29	22.55
AR(1)	**4**	7.51	9.67	13.44	**16**	26.77	35.98	29.41
FAR(1) (5 functions)		5.36	7.48	9.22		20.78	29.76	22.68
FAR(1) (10 functions)		5.48	7.74	9.91		20.86	30.67	22.36
SAR(1)		6.27	8.02	11.31		24.95	33.72	28.07
SFAR(1) (5 functions)		5.96	8.07	10.46		22.76	31.65	25.11
SFAR(1) (10 functions)		6.05	8.04	10.93		23.15	32.27	25.41
AR(1)	**5**	7.41	9.55	12.97	**17**	35.21	49.61	33.00
FAR(1) (5 functions)		5.47	7.55	9.38		27.07	42.61	23.56
FAR(1) (10 functions)		5.54	7.50	9.71		26.78	43.08	23.27
SAR(1)		6.17	7.92	10.83		31.34	45.55	28.56
SFAR(1) (5 functions)		5.94	7.86	10.34		29.29	44.22	25.99
SFAR(1) (10 functions)		5.95	7.85	10.49		28.40	43.37	25.66
AR(1)	**6**	8.01	10.06	13.34	**18**	40.62	60.62	32.32
FAR(1) (5 functions)		5.65	7.76	9.32		31.41	49.74	22.90
FAR(1) (10 functions)		5.70	7.66	9.44		31.65	48.41	25.03
SAR(1)		6.19	8.36	10.31		35.01	52.87	26.79
SFAR(1) (5 functions)		5.95	7.95	9.99		32.21	50.63	24.12
SFAR(1) (10 functions)		5.95	7.89	10.02		34.17	49.10	27.56

Table 4: Price prediction accuracy statistics.

Model	Hour	MAE Euro	RMSE Euro	MAPE %	Hour	MAE Euro	RMSE Euro	MAPE %
AR(1)	**7**	10.35	14.15	15.09	**19**	30.43	46.78	26.33
FAR(1) (5 functions)		7.80	11.15	11.60		23.27	38.21	19.06
FAR(1) (10 functions)		8.36	11.77	12.77		24.53	38.40	20.99
SAR(1)		9.13	12.29	13.70		26.13	41.06	22.01
SFAR(1) (5 functions)		8.42	11.75	12.74		24.91	39.04	21.16
SFAR(1) (10 functions)		9.07	11.99	14.11		25.98	38.99	22.84
AR(1)	**8**	18.91	27.67	22.79	**20**	23.26	41.01	21.08
FAR(1) (5 functions)		15.27	24.08	18.51		19.08	35.09	16.11
FAR(1) (10 functions)		16.13	23.97	20.55		18.85	34.54	16.43
SAR(1)		18.14	26.08	22.25		22.62	37.70	19.90
SFAR(1) (5 functions)		17.37	24.80	21.74		20.87	36.16	18.16
SFAR(1) (10 functions)		18.60	25.46	24.09		22.22	36.56	20.00
AR(1)	**9**	26.71	41.73	28.29	**21**	15.29	22.04	15.91
FAR(1) (5 functions)		22.56	38.91	23.33		13.34	20.24	13.49
FAR(1) (10 functions)		22.40	36.46	24.12		13.47	19.69	13.83
SAR(1)		27.24	39.85	29.50		15.85	21.60	16.80
SFAR(1) (5 functions)		26.21	39.85	28.64		16.51	23.04	17.30
SFAR(1) (10 functions)		26.62	38.01	30.35		15.28	21.66	15.88
AR(1)	**10**	23.25	40.09	25.17	**22**	10.21	17.07	12.41
FAR(1) (5 functions)		19.58	36.07	20.59		10.61	17.56	12.54
FAR(1) (10 functions)		19.70	36.96	21.37		11.45	18.48	13.46
SAR(1)		23.68	38.14	25.96		11.25	17.26	13.79
SFAR(1) (5 functions)		22.92	36.55	25.82		13.43	19.27	16.23
SFAR(1) (10 functions)		23.12	37.02	26.37		12.91	19.23	15.43
AR(1)	**11**	15.77	22.79	19.62	**23**	7.23	10.92	10.31
FAR(1) (5 functions)		13.66	20.20	17.10		6.37	9.62	8.92
FAR(1) (10 functions)		14.88	21.57	19.02		7.09	10.76	9.87
SAR(1)		16.04	22.24	20.42		6.97	10.54	9.72
SFAR(1) (5 functions)		15.82	22.02	20.42		6.84	10.00	9.54
SFAR(1) (10 functions)		18.59	24.78	24.15		7.55	11.02	10.45
AR(1)	**12**	14.92	21.56	18.95	**24**	6.55	8.37	10.43
FAR(1) (5 functions)		11.67	18.19	15.07		5.74	7.36	9.23
FAR(1) (10 functions)		12.63	19.20	16.37		5.83	7.79	9.35
SAR(1)		14.90	20.61	19.56		5.74	7.41	9.08
SFAR(1) (5 functions)		13.66	19.76	18.19		6.07	7.88	9.67
SFAR(1) (10 functions)		15.23	20.83	20.54		6.58	8.50	10.55

Figure 4.12: Mean absolute error for demand forecasting.

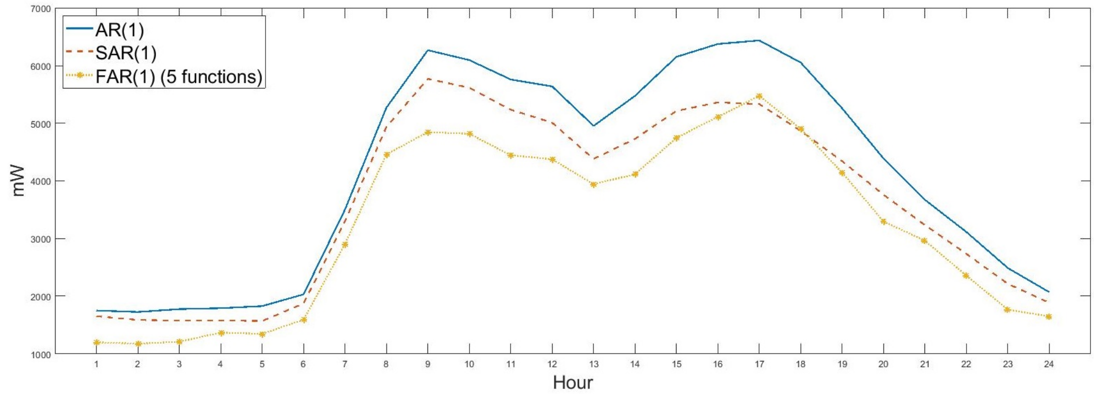


Figure 4.13: Root mean square error for demand forecasting.

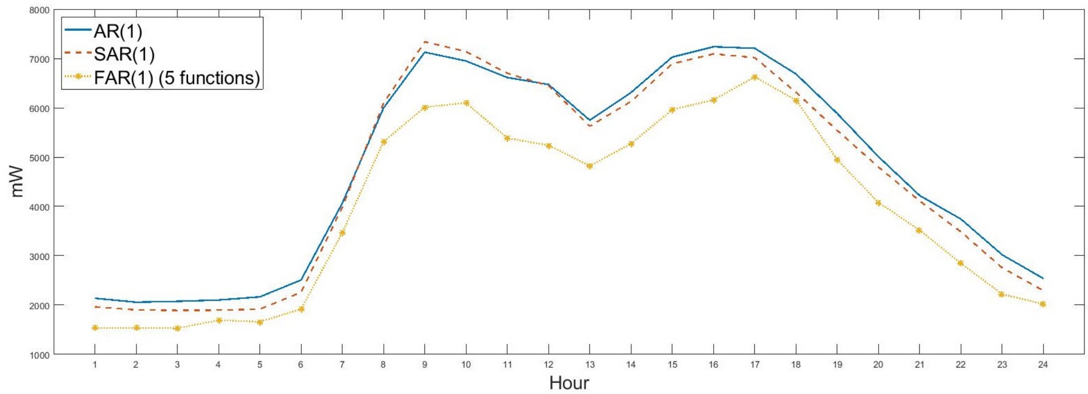


Figure 4.14: Mean absolute percentage error for demand forecasting.

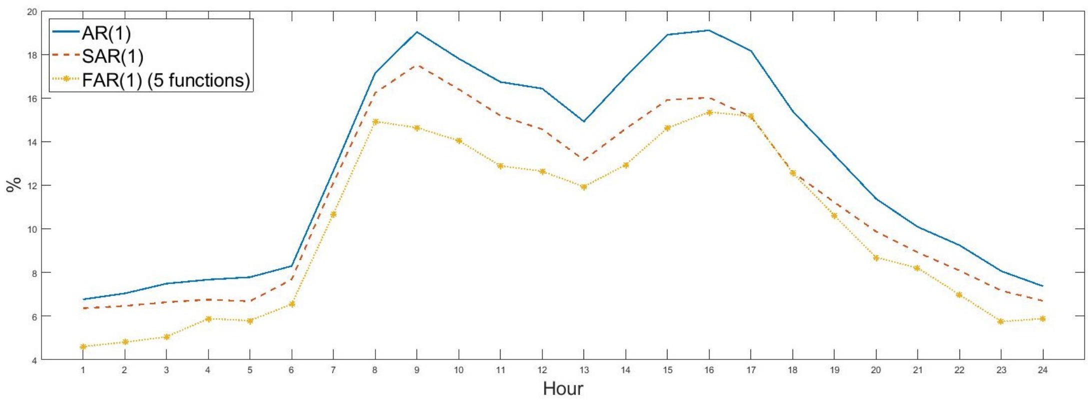


Table 5: Demand prediction accuracy statistics.

Model	Hour	MAE mW	RMSE mW	MAPE %	Hour	MAE mW	RMSE mW	MAPE %
AR(1)	**1**	1749	2134	6.7575	**13**	4955	5750	14.9272
SAR(1)		1650	1960	6.349		4381	5629	13.1578
FAR(1) (5 functions)		1197	1534	4.5906		3941	4824	11.9352
AR(p)	**2**	1723	2054	7.0308	**14**	5477	6313	16.9839
SAR(1)		1584	1897	6.458		4731	6132	14.596
FAR(1) (5 functions)		1173	1531	4.7919		4117	5272	12.9417
AR(p)	**3**	1773	2071	7.4811	**15**	6154	7033	18.9165
SAR(1)		1573	1887	6.6261		5214	6897	15.9249
FAR(1) (5 functions)		1206	1523	5.0349		4747	5964	14.6323
AR(p)	**4**	1789	2098	7.6647	**16**	6378	7243	19.1183
SAR(1)		1576	1892	6.7444		5364	7098	16.0223
FAR(1) (5 functions)		1365	1690	5.8724		5113	6164	15.3625
AR(p)	**5**	1825	2162	7.7768	**17**	6439	7211	18.1779
SAR(1)		1566	1912	6.6696		5334	7022	15.1515
FAR(1) (5 functions)		1343	1656	5.7846		5476	6630	15.1782
AR(p)	**6**	2029	2505	8.2893	**18**	6055	6690	15.406
SAR(1)		1870	2254	7.6712		4869	6316	12.6027
FAR(1) (5 functions)		1589	1913	6.5373		4905	6158	12.5699
AR(p)	**7**	3502	4065	12.6681	**19**	5259	5887	13.3915
SAR(1)		3299	3987	12.125		4341	5543	11.2166
FAR(1) (5 functions)		2903	3464	10.6912		4142	4948	10.6123
AR(p)	**8**	5272	6000	17.1605	**20**	4387	5009	11.3723
SAR(1)		4938	6097	16.2434		3758	4798	9.8719
FAR(1) (5 functions)		4461	5316	14.9315		3294	4079	8.6847
AR(p)	**9**	6270	7132	19.0464	**21**	3670	4220	10.0831
SAR(1)		5772	7345	17.5386		3231	4109	8.909
FAR(1) (5 functions)		4847	6016	14.6453		2960	3516	8.1968
AR(p)	**10**	6098	6954	17.8199	**22**	3111	3741	9.2407
SAR(1)		5618	7145	16.4139		2732	3483	8.0774
FAR(1) (5 functions)		4818	6103	14.0522		2353	2842	6.9697
AR(p)	**11**	5761	6618	16.7452	**23**	2485	3016	8.0533
SAR(1)		5236	6704	15.1949		2211	2752	7.1536
FAR(1) (5 functions)		4445	5389	12.8906		1764	2216	5.7381
AR(p)	**12**	5641	6476	16.4409	**24**	2067	2534	7.3591
SAR(1)		5011	6449	14.573		1883	2289	6.6889
FAR(1) (5 functions)		4377	5240	12.6463		1646	2015	5.8758

Chapter 5

Supply and demand curves as stochastic processes

Our next goal is to consider supply and demand curves as stochastic processes. As a functional space in this case we can consider the space, which contains all monotone bounded functions from $[0, M]$ to $[0, P]$, where $M = 60000$ MWh ans $P = 3000$ Euro/MWh. As far as the real data about supply and demand are discrete (there exist a minimum size of quantities of electricity for the supply offers and the demand bids) we are able to consider supply and demand curves either as piece-wise constant curves or as continuous piece-wise linear curves. In principle, it is an infinite dimensional subset of $L^2([0, M])$ or $H^1([0, M])$. However, market operators allow discrete minimum increases, or ticks, both for quantities as for prices. Then, in our model the dimension is finite[1]. Though finite, this is a huge number to implement in the numerical model, so we will consider the stochastic processes in an abstract Hilbert space.

In order to deal with the huge amount of bid data, we studied linear transformations of multivariate stochastic processes. It is known fact that a linear transformation of a vector ARMA process is again an ARMA process. Instead, a linear transformation of a finite order AR(p) process does not admit in general a finite order AR representation, but just a mixed ARMA representation. In this chapter we obtain a characterization result regarding the conditions that guarantee that a linear transformation of a vector AR process is again an AR

[1] For Italian Electricity Market the ticks are 1 kWh for quantities (i.e. 0.001 MWh) and 0.01 Euro/MWh for prices. Thus, the dimension of the model is 60000000.

process both in finite and in infinite dimension. We will then apply them to the model of Ziel and Steinert from [75].

5.1 Motivation

Let us reformulate the model of Ziel and Steinert mentioned in Section 1.3 in terms of linear transformation of multivariate stochastic process. They use a time series model for the bid volume processes $X_{S,t}^{(c)}$ and $X_{D,t}^{(c)}$ for each price class c. The original bid volume processes are $V_{S,t}(p)$ and $V_{D,t}(p)$ for each possible price $p \in P = \{p_1, p_2, \ldots, p_{n-1}, p_n\}$, where $p_1 = -500, p_2 = -499.9, \ldots, p_n = 3000$, thus $n = 35001$. So, we can say that the stochastic processes

$$V_{S,t} = (V_{S,t}(p_1), V_{S,t}(p_2), \ldots, V_{S,t}(p_n)),$$
$$V_{D,t} = (V_{D,t}(p_1), V_{D,t}(p_2), \ldots, V_{D,t}(p_n)),$$

are processes with values in $\mathbb{R}^n$, which represents the information about the whole supply and demand curves. More precisely, the sale and purchase curves are characterized by

$$S_t(p) = \sum_{i:p_i \leqslant p} V_{S,t}(p_i), \text{ and } D_t(p) = \sum_{i:p_i \geqslant p} V_{D,t}(p_i).$$

In order to reduce the dimensionality of the problem, Ziel and Steinert define price classes for supply and demand curves as $C_S = (c_1, c_2, \ldots, c_m)$ and $C_D = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_m)$, where $-500 = c_1 < c_2, \ldots < c_m = 3000$ and $3000 = \tilde{c}_1 > \tilde{c}_2 > \ldots > \tilde{c}_m = -500$. In such a way, m is a new dimension for the studied processes and it is much less than n, for instance in their paper they put $m = 16$. The price classes are given by

$$P_S(c_1) = \{-500\}, P_S(c_2) = (c_1, c_2] \cap P, \ldots, P_S(c_m) = (c_{m-1}, c_m] \cap P$$
$$P_D(c_1) = \{3000\}, P_D(\tilde{c}_2) = [\tilde{c}_2, \tilde{c}_1) \cap P, \ldots, P_S(c_m) = [\tilde{c}_m, \tilde{c}_{m-1}) \cap P.$$

So, instead of considering the processes $V_{S,t}$ and $V_{D,t}$, they study

$$X_{S,t} = (X_{S,t}(c_1), X_{S,t}(c_2), \ldots, X_{S,t}(c_m)),$$
$$X_{D,tm} = (X_{D,t}(\tilde{c}_1), X_{D,t}(\tilde{c}_2), \ldots, X_{D,t}(\tilde{c}_m)),$$

where

$$X_{S,t}(c_i) = \sum_{i:p_i \in P_S(c_i)} V_{S,t}(p_i) \text{ and } X_{D,t}(\tilde{c}_i) = \sum_{i:p_i \in P_D(\tilde{c}_i)} V_{D,t}(p_i)$$

We can assume that $C_S \subset P$ and $C_S \subset P$, rounding, if necessary, c_i and $\tilde{c}_i$ to one decimal place. Therefore, we can state that

$$X_{S,t} = T_S(V_{S,t}) \text{ and } X_{D,t} = T_D(V_{D,t})$$

where $T_S, T_D : \mathbb{R}^n \to \mathbb{R}^m$ are linear continuous operators such that

$$T_S(x_1, x_2, \ldots, x_n) = \left(\sum_{i=1}^{k_1} x_i, \sum_{i=k_1+1}^{k_2} x_i, \ldots, \sum_{i=k_{m-1}+1}^{n} x_i \right) \text{ and}$$

$$T_D(x_1, x_2, \ldots, x_n) = \left(\sum_{i=1}^{m_1} x_i, \sum_{i=k_1+1}^{m_2} x_i, \ldots, \sum_{i=m_{m-1}+1}^{n} x_i \right).$$

So, the processes $(X_{S,t})$ and $(X_{D,t})$ are linear transformations of the processes $(V_{S,t})$ and $(V_{D,t})$. Notice, that in practice, the original variables of interest are often transformed before their generation process is modeled.

As we already said, for modeling the electricity price Ziel and Steinert follow a simple regression approach described in [46], [73], [76]. So, in this case, the initial processes $(V_{S,t})$ and $(V_{D,t})$, and the transformed processes $(X_{S,t})$ and $(X_{D,t})$ are vector-valued processes and $(X_{S,t})$ and $(X_{D,t})$ are assumed to be autoregressive. We asked ourselves the following question: Suppose that (V_t) is $\mathbb{R}^n$-valued process and $(V_t) \in AR(p)$, i.e.

$$V_t = A_1 V_{t-1} + A_2 V_{t-2} + \ldots + A_p V_{t-p} + W_t,$$

where A_i are $(n \times n$ coefficient matrices and W_t is an $(n \times 1)$ zero-mean white noise vector process. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear continuous operator. Consider the $\mathbb{R}^m$-valued process $X_t = T(V_t)$. Can we state that $(X_t) \in AR(p)$? If $X_t = B_1 V_{t-1} + B_2 V_{t-2} + \ldots + B_p V(t-p) + Z_t$, what is the connection between A_i, B_i, W_t and Z_t? Could this result be generalized to infinite dimensional cases?

5.2 Linear transformation of ARH(1) processes

In this section we are going to obtain some partial results regarding the conditions that guarantee that a linear transformation of a AR process is again an AR process in the most general form – in infinite dimension. Recall that the space of all bounded linear operators between two Hilbert spaces H_1 and H_2 is denoted by $L(H_1, H_2)$, and $L(H)$ denote the space of continuous linear operators from H to H. First, let us prove the following lemma.

Lemma 5.2.1. *Let H_1 and H_2 be two Hilbert spaces and $T \in L(H_1, H_2)$. Suppose that $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is a H_1-valued white noise (see Definition 2.3.4) with covariance operator $C \in L(H_1)$, and*

$$\vartheta_n = T(\varepsilon_n),\ n \in \mathbb{Z}$$

Then $\vartheta = (\vartheta_n, n \in \mathbb{Z})$ is a H_2-valued white noise with covariance operator $C' = T \circ C \circ T^* \in L(H_2)$.

Proof. From Property 2.1.1 we have that $\vartheta \in L^2_{H_2}$ and

$$\mathbb{E}\vartheta_n = \mathbb{E}T(\varepsilon_n) = T(\mathbb{E}\varepsilon_n) = 0.$$

Recall that for $T : H_1 \to H_2$ there exists adjoint operator $T^* : H_2 \to H_1$ fulfilling $\langle Th_1, h_2\rangle_{H_2} = \langle h_1, T^*h_2\rangle_{H_1}$ (existence and uniqueness of this operator follows from the Riesz representation theorem). Since

$$\begin{aligned} C_{\vartheta_n}(h) &= \mathbb{E}\left[\langle \vartheta_n, h\rangle \vartheta_n\right] \\ &= \mathbb{E}\left[\langle T(\varepsilon_n), h\rangle T(\varepsilon_n)\right] \\ &= T\mathbb{E}\left[\langle \varepsilon_n, T^*(h)\rangle \varepsilon_n\right], \end{aligned}$$

so, C_{ϑ_n} does not depend on n. Also, ϑ_n are pairwise orthogonal. In fact, for any $x, y \in H$ and $n \neq m$

$$\begin{aligned} \mathbb{E}\left[\langle \vartheta_n, x\rangle \langle \vartheta_m, y\rangle\right] &= \mathbb{E}\left[\langle T(\varepsilon_n), x\rangle \langle T(\varepsilon_m), y\rangle\right] \\ &= \mathbb{E}\left[\langle \varepsilon_n, T^*x\rangle \langle \varepsilon_m, T^*y\rangle\right] = 0. \end{aligned}$$

It remains to show that $0 < \mathbb{E}[\|\vartheta_n\|^2] < \infty$ does not depend on n. This follows from the fact that $C_{\vartheta_n} = C_\vartheta$ does not depend on n. Indeed, let $\{h_i\}_{i=1}^{\infty}$

be orthonormal basis of H_2, then $\vartheta_n = \sum_{i=1}^{\infty}\langle \vartheta_n, h_i\rangle h_i$. Therefore,

$$\begin{aligned}
\mathbb{E}[\|\vartheta_n\|^2] = \mathbb{E}[\langle \vartheta_n, \vartheta_n\rangle] &= \mathbb{E}\left[\left\langle \vartheta_n, \sum_{i=1}^{\infty}\langle \vartheta_n, h_i\rangle h_i \right\rangle\right] \\
&= \sum_{i=1}^{\infty}\mathbb{E}\left[\langle\langle \vartheta_n, h_i\rangle\vartheta_n, h_i\rangle\right] = \sum_{i=1}^{\infty}\langle \mathbb{E}[\langle \vartheta_n, h_i\rangle\vartheta_n], h_i\rangle \\
&= \sum_{i=1}^{\infty}\langle C_\vartheta(h_i), h_i\rangle .
\end{aligned}$$

So, $\vartheta = (\vartheta_n, n \in \mathbb{Z})$ is a H_2-valued white noise. □

Remark 5.2.2. In the vector-valued case (n-dimensional or infinite-dimensional) we can always consider autoregressive processes of order 1 without loss of generality. Recall that for a Hilbert space H and a constant $p \in \mathbb{N}$ H^p is the product of Hilbert spaces

$$\underbrace{H \otimes H \ldots \otimes H}_{p}$$

with scalar product

$$\langle (x_1, x_2, \ldots, x_p), (y_1, y_2, \ldots, y_p)\rangle_{H^p} = \langle x_1, y_1\rangle_H + \langle x_2, y_2\rangle_H + \ldots + \langle x_p, y_p\rangle_H .$$

Suppose that $X_t \in ARH(p)$:

$$X_t = A_1X_{t-1} + A_2X_{t-2} + \ldots + A_pX_{t-p} + W_t.$$

Then we can define a new process $\widehat{X_t}$ putting

$$\widehat{X_t} = (X_t, X_{t-1}, \ldots, X_{t-p+1})';$$

$$A = \begin{pmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & I & 0 \end{pmatrix}$$

So, $\widehat{W_t} = (W_t, 0, \ldots, 0)'$ is $\underbrace{H \otimes H \ldots \otimes H}_{p} = H^p$-valued white noise by Lemma 5.2.1, and $\widehat{X_t} = A\,\widehat{X_{t-1}} + \widehat{W_t}$ is $ARH^p(1)$.

By the previous remark, all the results concerning $ARH(p)$ processes can be obtained from result, concerning $AR\tilde{H}(1)$, with $\tilde{H}$ suitable Hilbert space (namely, $\tilde{H} = H^p$). For this reason we are going to formulate our results for autoregressive processes of order 1.

Theorem 5.2.3. *Let H_1 and H_2 be two Hilbert spaces and $T \in L(H_1, H_2)$. Consider a zero-mean $ARH_1(1)$ process $X = (X_n, n \in \mathbb{Z})$ with values in H_1, satisfying, for all $n \in \mathbb{Z}$, the equation*

$$X_n = \rho(X_{n-1}) + \varepsilon_n,$$

where $\rho \in L(H_1)$ denotes the autocorrelation operator of the process X. Let $Y_t = T(X_t)$. Then the following are equivalent:

I. There exists $\vartheta \in L(H_2)$ such that

$$T\rho = \vartheta T \text{ on } \operatorname{span}\{X_n\}. \tag{5.2.1}$$

II. $Y = (Y_n, n \in \mathbb{Z})$ is an $ARH_2(1)$ and $Y_t = \vartheta Y_{t-1} + \xi_t$, $\vartheta \in L(H_2)$.

Proof. The following sequence of equality shows that condition (5.2.1) is sufficient for (II):

$$\begin{aligned} Y_t = T(X_t) &= T\rho X_{t-1} + T\varepsilon_t \\ &= \vartheta T X_{t-1} + T\varepsilon_t \\ &= \vartheta Y_{t-1} + \xi_t, \end{aligned}$$

where $\xi_t = T\varepsilon_t$ is a zero-mean white noise according to Lemma 5.2.1 and $\vartheta \in L(H_2)$ is the autocorrelation operator of process Y.

Conversely, if Y_t has the representation $Y_t = \vartheta Y_{t-1} + \xi_t$, therefore $\vartheta T X_{t-1} + \xi_t = T\rho X_{t-1} + T\varepsilon_n$, so (5.2.1) holds. □

Remark 5.2.4. Notice that necessary condition for equation (5.2.1) to be true is that for any $x \in \ker(T) \cap \operatorname{span}\{X_n\} \Rightarrow \rho(x) \in \ker(T)$. It means that $\ker(T) \cap \operatorname{span}\{X_n\}$ is an invariant subspace of ρ restricted on subspace $\operatorname{span}\{X_n\}$.

In the case that the operator T is invertible, obviously, condition (5.2.1) holds with $\vartheta := T\rho T^{-1}$. For operators that are not invertible, various types of generalized inverses exist in the literature. Before introducing the concept of pseudo-inverse operator, we briefly recall some basic facts concerning the orthogonal projection of a Hilbert space onto a closed subset.

Recall that the operator $P \in L(H, H)$ is called a projector onto $\tilde{H} \subset H$ if $P(H) \subset \tilde{H}$ and $Px = x$ for all $x \in \tilde{H}$. An element $h \in H$ is said to be orthogonal to the subset $X \subset H$ if h is orthogonal to all the elements of X. The set of all elements orthogonal to the subset X is called the orthogonal complement to X and is denoted by $X^{\perp}$. It is known fact that, if X is a closed linear subspace of the Hilbert space H, then H decomposes into the direct sum of the subspaces X and $X^{\perp}$. Since $H = X \bigoplus X^{\perp}$, there exists a bounded projector P onto the subspace X with $\ker P = X^{\perp}$. The orthogonal projection onto a closed subspace $M \subset H$ is the bounded linear operator $P : H \to H$ such that for each $x = m + m' \in H$ $(m \in M, m' \in M^{\perp})$, $P(x) = m$. A projector $P \in L(H)$ is an orthogonal projector if and only if P is a self-adjoint operator [36, Section 12.2].

Definition 5.2.5. Let H_1, H_2 be Hilbert spaces, and suppose that $T \in L(H_1, H_2)$. The pseudo-inverse of T (if it exist) is an element $T^+ \in L(H_2, H_1)$ such that

$$\begin{aligned} TT^+x = x \text{ for } x \in \operatorname{range}(T); \\ \ker(T^+) = \operatorname{range}(T)^{\perp}; \\ \operatorname{range}(T^+) = \ker(T)^{\perp}. \end{aligned}$$

It turns out that, in contrast to the finite dimensional setting, not every continuous linear operator has a continuous linear pseudo-inverse in this sense. Those that do are precisely the ones whose range is closed in H_2 [33, Theorem 2.4].

Example 5.2.6. Consider the operator

$$A = \operatorname{diag}(1, 1/2, 1/3, \ldots) \in L(\ell_2).$$

We can see that $\operatorname{range}(A) = \{y \in \ell_2 : \sum_{i=1}^{\infty} i^2 y_i^2 < \infty\}$ is not closed in ℓ_2, as the limit point $(1, 1/2, 1/3, \ldots) \notin \operatorname{range}(A)$. So, there is no pseudo-inverse of the operator A. Indeed, the only possible candidate would be the operator $B = \operatorname{diag}(1, 2, 3, \ldots)$, which is unbounded.

We collect some properties of T^+ and its relationship to T [20].

Proposition 5.2.7. *Let H_1, H_2 be Hilbert spaces, and $T \in L(H_1, H_2)$ have closed range. Then the following holds:*

1. *TT^+ is the orthogonal projection of H_2 onto $\operatorname{range}(T)$.*
2. *T^+T is the orthogonal projection of H_1 onto $\operatorname{range}(T^+)$.*
3. *T^* has closed range, and $(T^*)^+ = (T^+)^*$.*
4. *On $\operatorname{range}(T)$ the operator T^+ is given explicitly by*
$$T^+ = T^*(TT^*)^{-1}.$$
5. *T^+ satisfies to Moore-Penrose equations*
$$TT^+T = T; \quad T^+TT^+ = T^+; \quad (TT^+)^* = TT^+; \quad (T^+T)^* = T^+T. \tag{5.2.2}$$

Remark 5.2.8. For the finite dimensional case, it has been shown [53] that if the four equations (5.2.2) are considered as equations for the unknown matrix T^+, then these equations have a unique solution which is called the Moore-Penrose inverse. The pseudo-inverse defined in Definition 5.2.5 is therefore an extension of the Moore-Penrose inverse in Hilbert space.

From Proposition 5.2.7 and Theorem 5.2.3 we obtain the following result.

Corollary 5.2.9. *Let $T : H_1 \to H_2$ be a linear continuous operator between two Hilbert spaces with closed range in H_2. Consider a zero-mean $ARH(1)$ process $X = (X_n, n \in \mathbb{Z})$ with values in H_1, satisfying, for all $n \in \mathbb{Z}$, the equation*
$$X_n = \rho(X_{n-1}) + \varepsilon_n,$$
where $\rho \in L(H_1)$. Let $\ker(T)$ be an invariant subspace of ρ (i.e. $\rho(\ker(T)) \subset \ker(T)$).

Then $Y_n = T(X_n)$ is an $ARH_2(1)$ with values in R_T, and the process Y_n has the representation:

$$Y_n = \vartheta Y_{n-1} + T\varepsilon_n \text{ with } \vartheta := T\rho T^*(TT^*)^{-1}.$$

Proof. Let us denote $\tilde{H} :=$ range(T). $Y_n = T(X_n)$ has values in $\tilde{H} \subset H_2$. Let us define $\vartheta = T\rho T^*(TT^*)^{-1} \in L(\tilde{H})$, which is exactly $T\rho T^+$ on $\tilde{H}$.

T^+T is the orthogonal projection of H_1 onto range(T^+). Since

$$H_1 = \text{range}(T^+) \bigoplus (\text{range}(T^+))^\perp = \text{range}(T^+) \bigoplus \ker(T),$$

we can write any $x \in H_1$ as $x = y + z$ with $y \in$ range(T^+) and $z \in \ker(T)$. Then

$$\vartheta T(y) = T\rho T^+ T(y) = T\rho I(y) = T\rho(y)$$
$$\vartheta T(z) = 0 \text{ and } T\rho(z) = 0, \text{ as } \rho(\ker(T)) \subset \ker(T).$$

So, as $T\rho = \vartheta T$, according to Theorem 5.2.3, $Y = (Y_n, n \in \mathbb{Z})$ is an $ARH_2(1)$. □

5.3 Linear transformation of VAR(1) processes

Now we are going to reformulate the necessary and sufficient condition in Theorem 5.2.3 for the case $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$.

Theorem 5.3.1. *Let X_t be an n-dimensional $AR(1)$ process with the representation*

$$X_t = AX_{t-1} + W_t,$$

$T : \mathbb{R}^n \to \mathbb{R}^m$, be linear transformation and $Y_t = T(X_t)$. Then the following are equivalent:

I. There exists a $(m \times m)$ matrix B such that

$$BT = TA.$$

II. $Y = (Y_n, n \in \mathbb{Z})$ is an m-dimensional $AR(1)$ and $Y_t = BY_{t-1} + Z_t$, where $Z_t = TW_t$ is a zero-mean white noise vector process.

Remark 5.3.2. If T has linearly independent rows, then $B = TAT^+$, where $T^+ = T^T(TT^T)^{-1}$ is the Moore-Penrose inverse of T (see Definition **??**), and

$$Y_t = BY_{t-1} + Z_t.$$

Corollary 5.3.3. *Let X_t be an n-dimensional $AR(p)$ process with the representation*

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \ldots + A_p X_{t-p} + W_t,$$

and $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

Then $Y_t = T(X_t)$ is an m-dimensional $AR(p)$ if and only if there exist $(m \times m)$ matrices B_i such that

$$B_i T = T A_i \quad \textit{for all } i = 1, \ldots, p. \tag{5.3.1}$$

Moreover, if T has linearly independent rows, the process Y_t has the representation:

$$Y_t = TA_1T^+Y_{t-1} + TA_2T^+Y_{t-2} + \ldots + TA_pT^+Y_{t-p} + TW_t, \tag{5.3.2}$$

where T^+ is the Moore-Penrose inverse of T.

Proof. The statement can be deduced from Remark 5.2.2 and Theorem 5.3.1, but the easiest way to obtain this result is the straightforward proof. Indeed, the following sequence of equality shows that condition (5.3.1) is necessary and sufficient:

$$\begin{aligned} Y_t = T(X_t) &= TA_1X_{t-1} + TA_2X_{t-2} + \ldots + TA_pX_{t-p} + TW_t \\ &= B_1TX_{t-1} + B_2TX_{t-2} + \ldots + B_pTX_{t-p} + TW_t \\ &= B_1Y_{t-1} + B_2Y_{t-2} + \ldots + B_pY_{t-p} + Z_t, \end{aligned}$$

where $Z_t = TW_t$ is a zero-mean white noise vector process. So, Y_t is an m-dimensional $AR(p)$ process. If the rows of T are linearly independent, the Moore-Penrose inverse of T can be expressed as

$$T^+ = T^T(TT^T)^{-1}$$

so that $TT^+ = I$, and, therefore, $B_i = TA_iT^+$, so, the process Y_t has the representation (5.3.2). □

Now we are going to apply this result for the model of Ziel and Steinert in order to guarantee that the transformation of the initial process belongs to the class of autoregressive processes. For any matrix A let R_i^A denote its i-th row and let C_j^A denote its j-th column.

In the model of Ziel and Steinert we start from the stochastic processes with values in $\mathbb{R}^n$, which represents the information about the whole supply curve:

$$V_{S,t} = (V_{S,t}(p_1), V_{S,t}(p_2), \ldots, V_{S,t}(p_N)),$$

and then we define the modified process with values in $\mathbb{R}^m$

$$X_{S,t} = (X_{S,t}(c_1), X_{S,t}(c_2), \ldots, X_{S,t}(c_M)),$$

such that

$$X_{S,t} = T_S(V_{S,t})$$

where T_S is an $(m \times n)$ matrix $(m < n)$ with columns

$$C_{k_{s-1}}^{T_S} = C_{k_{s-1}+1}^{T_S} = C_{k_{s-1}+2}^{T_S} = \ldots = C_{k_s}^{T_S} = e_s, \quad 1 \leqslant s \leqslant m$$

where $0 = k_0 < k_1 < k_2 < \ldots < k_m = n$ and $\{e_1, e_2, \ldots, e_m\}$ is the standard basis of $\mathbb{R}^m$, i.e.

$$T_S = \begin{pmatrix} \underbrace{1 \ \ldots \ 1}_{k_1} & 0 \ \ldots \ 0 & 0 \ \ldots \ 0 & 0 \ \ldots \ 0 \\ 0 \ \ldots \ 0 & \underbrace{1 \ \ldots \ 1}_{k_2-k_1} & 0 \ \ldots \ 0 & 0 \ \ldots \ 0 \\ \vdots & & & \vdots \\ 0 \ \ldots \ 0 & 0 \ \ldots \ 0 & \underbrace{1 \ \ldots \ 1}_{k_{m-1}-k_{m-2}} & 0 \ \ldots \ 0 \\ 0 \ \ldots \ 0 & 0 \ \ldots \ 0 & 0 \ \ldots \ 0 & \underbrace{1 \ \ldots \ 1}_{n-k_{m-1}} \end{pmatrix} \tag{5.3.3}$$

The authors in [75] introduced $X_{S,d,h}^{(c)}$ as the bid supply volume process of price class $c \in C_S$, and $X_{D,d,h}^{(c)}$ as the bid demand volume process of price class

$c \in C_D$ at day d and hour h. Also they introduced the additional processes denoted by $X_{price,d,t}, X_{volume,d,t}, X_{generation,d,t}, X_{wind,d,t}, X_{solar,d,t}$ that represent the additional information that is available at the time where the auction will take place. For modeling the weekday impact they defined the weekday indicators

$$W_k(d) = \begin{cases} 1, & W(d) < k \\ 0, & W(d) \geqslant k, \end{cases}$$

where $W(d)$ is a function that gives a number that corresponds to the weekday of day d (without loss of generality, let $k = 1$ for a Monday, for a Tuesday $k = 2$ up to $k = 7$ for a Sunday).

To fully present the considered time series model, the object

$$X_{d,h} = (X_{1,d,h}, X_{2,d,h}, \dots, X_{M,d,h}) = ((X^{(c)}_{S,d,h})_{c \in C_S}, (X^{(c)}_{D,d,h})_{c \in C_D}, \\ X_{price,d,h}, X_{volume,d,h}, X_{generation,d+1,h}, X_{wind,d+1,h}, X_{solar,d+1,h})$$

was introduced. As the number of price classes for the supply side is $M_S = 16$ and the number of price classes for the demand side is $M_D = 16$, therefore, the dimension of $X_{d,h}$ is $M = M_S + M_D + 5 = 37$.

Then for each hour h the considered time series model of zero-mean process

$$Y_{d,h} = X_{d,h} - \mathbb{E}(X_{d,h}) = (Y_{1,d,h}, Y_{2,d,h}, \dots, Y_{M,d,h}).$$

The considered time series model for $Y_{m,d,h}$ for each hour h and $m \in \{1, \dots, M_S + M_D\}$ is given by

$$Y_{m,d,h} = \sum_{l=1}^{M} \sum_{j=1}^{24} \sum_{k \in I_{m,h}(l,j)} \phi_{m,h,l,j,k} Y_{l,d-k,j} + \sum_{k=2}^{7} \psi_{m,h,k} W_k(d) + \varepsilon_{m,d,h} \qquad (5.3.4)$$

with the side constraint $0 = \sum \psi_{m,h,k} W_k(d)$, with parameters $\phi_{m,h,l,j,k}$ and $\psi_{m,h,k}$ as lag sets of lags and $\varepsilon_{m,d,h}$ as error term. We assume that the error process $\varepsilon_{m,d,h}$ is i.i.d. with constant variance $\sigma^2_{m,h}$. The introduced parameters $\phi_{m,h,l,j,k} Y^l_{d-k,j}$ model the linear autoregressive impact and $\psi_{m,h,k}$ the day of the week effect. Lag sets c are defined as follow:

$$I_{m,h} = \begin{cases} \{1,2,\ldots,36\}, & m = l \text{ and } h = j \\ \{1,2,\ldots,8\}, & (m = l \text{ and } h \neq j) \text{ or } (m \neq l \text{ and } h = j) \\ \{1\}, & m \neq l \text{ and } h \neq j \end{cases}$$

Thus, the process $Y_{m,d,h}$ of price class m at day d and hour h can depend on the values of the past 36 days of price class m at hour h, it is only allowed to depend on the value of another process at another hour one with a maximum lag of 1, and in all other cases a maximum lag of eight is possible. The considered model is basically a simple regression approach model. In order to rewrite (5.3.4) as autoregressive model of order 1 we can define vector of larger dimension, namely $M \times 24 \times 35 := M_1$

$$Y_d = \left((Y_{i,d,h})_{\substack{i=1,\ldots,M \\ j=1,\ldots,24}}, (Y_{i,d-1,h})_{\substack{i=1,\ldots,M \\ j=1,\ldots,24}}, \ldots, (Y_{i,d-35,h})_{\substack{i=1,\ldots,M \\ j=1,\ldots,24}} \right).$$

Then

$$Y_d = \Phi_d Y_{d-1} + \sum_{k=2}^{7} \Psi_k W_k(d) + \varepsilon_d, \tag{5.3.5}$$

with parameters $\Phi_d, \Psi_k \in \mathbb{R}^{M_1}$ as lag sets of lags and $\varepsilon_d \in \mathbb{R}^{M_1}$ as error term.

Example 5.3.4. Let us calculate the Moore-Penrose inverse of the $(m \times n)$ matrix T given by (5.3.3).

$$T^+ = \begin{pmatrix} \frac{1}{k_1-k_0} & 0 & 0 & \ldots\ldots\ldots & 0 & 0 \\ \vdots & & & & & \vdots \\ \frac{1}{k_1-k_0} & 0 & 0 & \ldots\ldots\ldots & 0 & 0 \\ 0 & \frac{1}{k_2-k_1} & 0 & \ldots\ldots\ldots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & \frac{1}{k_2-k_1} & 0 & \ldots\ldots\ldots & 0 & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \ldots\ldots\ldots & 0 & \frac{1}{k_m-k_{m-1}} \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \ldots\ldots & 0 & \frac{1}{k_m-k_{m-1}} \end{pmatrix},$$

i.e. the j-th column of T^+ has exactly $k_j - k_{j-1}$ non-zero elements, all equals to $\frac{1}{k_j-k_{j-1}}$. It is straightforward to check that TT^+ is the $(m \times m)$ identity matrix and T^+T is self-adjoint, so T^+ satisfies Definition **??**.

Now we want to formulate a result, which gives sufficient and necessary condition for the specific operator which appears in the model of Ziel and Steinert.

Proposition 5.3.5. *Let T be an $(m \times n)$, matrix ($m < n$) with columns*

$$C^T_{k_{s-1}} = C^T_{k_{s-1}+1} = C^T_{k_{s-1}+2} = \ldots = C^T_{k_s} = e_s, \quad 1 \leqslant s \leqslant m$$

where $0 = k_0 < k_1 < k_2 < \ldots < k_m = n$ and $\{e_1, e_2, \ldots, e_m\}$ is the standard basis of $\mathbb{R}^m$, i.e. T is as in (5.3.3). Let $A = \{a_{ij}\}_{i,j=1}^n$ be an $(n \times n)$ matrix. Then there exists a $(m \times m)$ matrix $B = \{b_{ij}\}_{i,j=1}^m$ such that

$$BT = TA,$$

if and only if the following condition holds: for every index $1 \leqslant d \leqslant m$, whenever j and l are such that $k_{p-1}+1 \leqslant j \leqslant k_p$ and $k_{p-1}+1 \leqslant l \leqslant k_p$ for some $1 \leqslant p \leqslant m$, then

$$\sum_{i=k_{d-1}+1}^{k_d} a_{ij} = \sum_{i=k_{d-1}+1}^{k_d} a_{il}. \tag{5.3.6}$$

Moreover, in this case

$$b_{ij} = \sum_{l=k_{i-1}+1}^{k_i} a_{lk_j}.$$

Proof. The proof is straightforward. From the expressions

$$BT = \begin{pmatrix} \underbrace{b_{11} \; \ldots \; b_{11}}_{k_1} & b_{12} \; \ldots \; b_{12} & \ldots & b_{1m} \; \ldots \; b_{1m} \\ b_{21} \; \ldots \; b_{21} & \underbrace{b_{22} \; \ldots \; b_{22}}_{k_2-k_1} & \ldots & b_{2m} \; \ldots \; b_{2m} \\ \vdots & & & \vdots \\ b_{m1} \; \ldots \; b_{m1} & b_{m2} \; \ldots \; b_{m1} & \ldots & \underbrace{b_{mm} \; \ldots \; b_{mm}}_{n-k_{m-1}} \end{pmatrix} =$$

$$TA = \begin{pmatrix} \sum_{i=1}^{k_1} a_{i1} & \sum_{i=1}^{k_1} a_{i2} & \ldots & \sum_{i=1}^{k_1} a_{in} \\ \sum_{i=k_1+1}^{k_2} a_{i1} & \sum_{i=k_1+1}^{k_2} a_{i2} & \ldots & \sum_{i=k_1+1}^{k_2} a_{in} \\ \vdots & & & \vdots \\ \sum_{i=k_{m-1}+1}^{k_m} a_{i1} & \sum_{i=k_{m-1}+1}^{k_m} a_{i2} & \ldots & \sum_{i=k_{m-1}+1}^{k_m} a_{in} \end{pmatrix}$$

we obtain that (5.3.6) holds.

Conversely, suppose now that A satisfies (5.3.6). Then we can define the matrix $B = \{b_{ij}\}_{i,j=1}^m$ as follows:

$$b_{ij} = \sum_{l=k_{i-1}+1}^{k_i} a_{lk_j}$$

in order to have $BT = TA$. □

Remark 5.3.6. In particular, this condition holds if the first k_1 columns of A are the same, the following $k_2 - k_1$ are the same, and so on, until the last $k_m - k_{m-1}$ columns. However, Proposition 5.3.5 gives slightly weaker condition, as the following example shows:

$$T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Evidently, A satisfies (5.3.6), although the first two columns of A are not the same, and B is obtained as $b_{11} = 1, b_{12} = 2, b_{21} = 1, b_{22} = 1$.

5.4 An infinite-dimensional formulation of Ziel-Steinert's X-model

The supply and demand curves are characterized by

$$S_t(p) = \sum_{\substack{x \in P \\ x \leqslant p}} V_{S,t}(x) \text{ and } D_t(p) = \sum_{\substack{x \in P \\ x \geqslant p}} V_{D,t}(x) \text{ for } p \in P.$$

Here $t = (d, h)$, where d denotes day and h denotes hour. Assume that

$$S_{d,h}(p) = \int_{-500}^{p} s_{d,h}(z) dz \text{ for } p \in P \text{ and } D_{d,h}(p) = \int_{p}^{3000} d_{d,h}(z) dz \text{ for } p \in P, \tag{5.4.1}$$

where $s_{d,h}, d_{d,h} \in L^2([-500, 3000])$. Then we can state that the processes $S_{d,h}, D_{d,h}$ take values in the Hilbert space $H^1([-500, 3000])$. Recall that $H^1([-500, 3000])$ consists of $f \in L^2([-500, 3000])$ whose distributional derivative f' lies in $L^2([-500, 3000])$ and has the inner product

$$\langle f, g \rangle_{H^1} = \langle f, g \rangle_{L^2} + \langle f', g' \rangle_{L^2}.$$

From the general theory of Sobolev spaces we know that $H^1([-500, 3000]) \subset C([-500, 3000])$ and it is a reproducing kernel Hilbert space (see Example 3.5.7). Recall that the following model was proposed in [15]

$$X_n(\cdot) = \sum_{j=1}^{p} \alpha_j(\cdot) X_{n-1}(t_j) + \varepsilon_n(\cdot), \tag{5.4.2}$$

where $\alpha_j(\cdot)$ are continuous functions in $[0, 1]$ and ε_n is a strong $C([0, 1])$-valued white noise pointwisely uncorrelated with X_n. That is, all the curves depend on the same set of points regardless of the index n.

Now let us make a connection between the model described in (5.4.2) and the model of Ziel and Steinert.

Theorem 5.4.1. *Suppose that H is RKHS consisting of functions $f : [a, b] \to \mathbb{R}$, the points $t_1, t_2, \ldots, t_p \in [a, b]$ are fixed, and the functions $\alpha_1, \alpha_2, \ldots, \alpha_p \in C[a, b]$ are fixed. Let $X = (X_n, n \in \mathbb{Z})$ be a zero-mean $ARH(1)$ process, satisfying, for all $n \in \mathbb{Z}$, the equation*

$$X_n = \rho(X_{n-1}) + \varepsilon_n,$$

where

$$\rho(f)(\cdot) = \sum_{j=1}^{p} \alpha_j(\cdot) f(t_j),$$

and $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is a H-valued white noise. Let $T : H \to \mathbb{R}^p$ be the linear continuous operator defined as

$$T(f) = (f(t_1), f(t_2), \ldots, f(t_p)).$$

Then the process $Y_n = T(X_n), n \in \mathbb{Z}$ is $VAR(1)$ with values in $\mathbb{R}^p$ and dynamics

$$Y_n = \vartheta Y_{n-1} + T(\varepsilon_n), \text{ where}$$

$$\vartheta = \begin{pmatrix} \alpha_1(t_1) & \alpha_2(t_1) & \ldots & \alpha_p(t_1) \\ \vdots & & & \vdots \\ \alpha_p(t_1) & \alpha_p(t_2) & \ldots & \alpha_p(t_p) \end{pmatrix}$$

Proof. According to Theorem 5.2.3 the process $Y = (Y_n, n \in \mathbb{Z})$ is autoregressive if and only if there exists a linear continuous operator $\vartheta : \mathbb{R}^p \to \mathbb{R}^p$ such that

$$T\rho = \vartheta T. \tag{5.4.3}$$

This condition holds for the operator ϑ such that for every $(b_1, \ldots, b_p) \in \mathbb{R}^p$

$$\vartheta(b_1, b_2, \ldots, b_p) = \left(\sum_{j=1}^{p} \alpha_j(t_1) b_j, \sum_{j=1}^{p} \alpha_j(t_2) b_j, \ldots, \sum_{j=1}^{p} \alpha_j(t_p) b_j \right).$$

Indeed, for any $f \in H$

$$\begin{aligned}(T\rho)(f) &= T(\rho f) = T\left(\sum_{j=1}^{p} \alpha_j(\cdot) f(t_j)\right) \\ &= \left(\sum_{j=1}^{p} \alpha_j(t_1) f(t_j), \ldots, \sum_{j=1}^{p} \alpha_j(t_p) f(t_j)\right) \\ &= \vartheta(f(t_1), \ldots, f(t_p)) = \vartheta(Tf) = (\vartheta T)(f).\end{aligned}$$

Evidently, ϑ is a linear continuous operator with $\|\vartheta\| = \max\limits_{i=1,\ldots,p} |\sum_{j=1}^{p} \alpha_j(t_i)|$. □

Example 5.4.2. Consider $H = L^2(X, \mu)$, where X be a discrete set of points $\{x_i\} \subset [0,1]$ and the measure $\mu = \sum_i \delta_{x_i}$. Suppose that a finite $\tilde{X} = \{t_1, t_2, \ldots, t_p\} \subset X$ is fixed. Let $T : H \to \mathbb{R}^p$ be a linear continuous operator such that

$$T(f) = (f(t_1), f(t_2), \ldots, f(t_p)).$$

Then the pseudo-inverse of T is given by

$$T^+(y_1, y_2 \ldots, y_p) = y_1 \mathbf{1}_{\{t_1\}} + y_2 \mathbf{1}_{\{t_2\}} + \ldots + y_p \mathbf{1}_{\{t_p\}}.$$

Indeed, $TT^+ = I$ on $\mathbb{R}^p$; $\operatorname{range}(T) = \mathbb{R}^p$, $\ker(T^+) = \{0\}$, so $\ker(T^+) = \operatorname{range}(T)^\perp$ and $\operatorname{range}(T^+) = \ker(T)^\perp = \{f : f|_{X/\tilde{X}} = 0\}$.

Example 5.4.3. Consider the Sobolev space $H^1[0, M]$ consisting of absolutely continuous functions $f : [0, M] \to \mathbb{R}$ whose derivative lies in $L^2[0, M]$ with the inner product

$$\langle f, g \rangle_{H^1} = \langle f, g \rangle_{L^2} + \langle f', g' \rangle_{L^2}.$$

Recall that $f(x) = \langle f(\cdot), K(x, \cdot) \rangle$ for any $x \in [0, M]$, and the kernel function $K : [0, M] \times [0, M] \to \mathbb{R}$ (see Example 3.5.7) is given by

$$K(x, y) = \begin{cases} \frac{\cosh(x-M)\cosh(y)}{\sinh(M)} & \text{if } x \leqslant y \\ \frac{\cosh(x)\cosh(y-M)}{\sinh(M)} & \text{if } x \geqslant y. \end{cases}.$$

Suppose that a finite set $\{t_1, t_2, \ldots, t_p\} \subset [0, M]$ is fixed. Consider the same operator $T : H \to \mathbb{R}^p$ such that

$$T(f) = (f(t_1), f(t_2), \ldots, f(t_p)).$$

We have range$(T) = \mathbb{R}^p$, so the pseudo-inverse of T exists. Let us find T^+. First, notice that the adjoint operator is given by

$$T^*(y_1, y_2 \dots, y_p) = y_1 K(t_1, \cdot) + y_2 K(t_2, \cdot) + \dots + y_p K(t_p, \cdot).$$

Then

$$\begin{aligned} T^+(y_1, y_2 \dots, y_p) &= T^*(TT^*)^{-1}(y_1, y_2 \dots, y_p) \\ &= z_1 K(t_1, \cdot) + z_2 K(t_2, \cdot) + \dots + z_p K(t_p, \cdot), \end{aligned}$$

where $z = Cy$ and C is the inverse matrix of

$$TT^* = \begin{pmatrix} K(t_1, t_1) & K(t_1, t_2) & \dots & K(t_1, t_p) \\ \vdots & & & \vdots \\ K(t_p, t_1) & K(t_p, t_2) & \dots & K(t_p, t_p) \end{pmatrix}.$$

Now consider the processes $S_{d,h}, D_{d,h}$ defined in (5.4.1) with values in the Hilbert space $H^1([-500, 3000])$. In order to define an auto-regressive model of order 1, similarly to (5.3.5), we need to consider elements of "biggest" dimension, namely,

$$\begin{aligned} X_d =& ((S_{d,h})_{j=1,\dots,24}, (S_{d-1,h})_{j=1,\dots,24}, \dots, (S_{d-35,h})_{j=1,\dots,24}, \\ & (D_{d,h})_{j=1,\dots,24}, (D_{d-1,h})_{j=1,\dots,24}, \dots, (D_{d-35,hj=1,\dots,24})), \end{aligned}$$

which takes values in the product space $\left(H^1([-500, 3000])\right)^{24\times 35\times 2}$.

Let us now extend Theorem 5.4.1 for the case when stochastic processes takes values in a product space.

Theorem 5.4.4. *Suppose that H is a Hilbert space consisting of functions $f : [a, b] \to \mathbb{R}$, H^M is a product space, $t_1, t_2, \dots, t_p \in [a, b]$ are fixed, and the set of functions $\{\alpha_i^j\}_{\substack{i=1,\dots,p \\ j=1,\dots,M}} \in C[a, b]$ is fixed. Let $X = (X_d, d \in \mathbb{Z})$ be a zero-mean $ARH^M(1)$ process, satisfying, for all $d \in \mathbb{Z}$, the equation*

$$X_d = \rho(X_{d-1}) + \varepsilon_d,$$

where $\varepsilon = (\varepsilon_d, d \in \mathbb{Z})$ *is a* H^M*-valued white noise,* $\rho \in L(H^M)$*, and for any* $(f_1, \ldots, f_M) \in H^M$ *jth coordinate of* $\rho(f_1, \ldots, f_M)$ *is*

$$g_j(\cdot) = \sum_{i=1}^{p} \alpha_i^j(\cdot) f_j(t_i).$$

Let $T : H^M \to (\mathbb{R}^p)^M$ *be a linear continuous operator such that*

$$T(f_1 \ldots, f_M) = ((f_1(t_i))_{i=1,\ldots,p}, (f_2(t_i))_{i=1,\ldots,p}, \ldots, (f_M(t_i))_{i=1,\ldots,p}).$$

Then the process $Y_d = T(X_d), d \in \mathbb{Z}$ *is an* $AR(1)$ *with values in* $(\mathbb{R}^p)^M$*.*

Proof. In order to prove the statement, according to Theorem 5.2.3, we need to show that there exist a linear continuous operator $\vartheta : (\mathbb{R}^p)^M \to (\mathbb{R}^p)^M$ such that

$$T\rho = \vartheta T. \tag{5.4.4}$$

This condition holds for the operator ϑ such that for every $b = ((b_i^1)_{i=1,\ldots,p}, \ldots, (b_i^M)_{i=1,\ldots,p}) \in (\mathbb{R}^p)^M$

$$\vartheta(b) = \left(\left(\sum_{j=1}^{p} \alpha_j^1(t_i) b_j^1 \right)_{i=1,\ldots,p}, \ldots, \left(\sum_{j=1}^{p} \alpha_j^M(t_i) b_j^M \right)_{i=1,\ldots,p} \right).$$

Indeed, for any $(f_1, \ldots, f_M) \in H^M$

$$(T\rho)(f_1, \ldots, f_M) = T\left(\sum_{i=1}^{p} \alpha_i^1(\cdot) f_1(t_i), \ldots, \sum_{i=1}^{p} \alpha_i^M(\cdot) f_M(t_i) \right)$$
$$= \left(\left(\sum_{i=1}^{p} \alpha_i^1(t_k) f_1(t_i) \right)_{k=1,\ldots p}, \ldots, \left(\sum_{i=1}^{p} \alpha_i^M(t_k) f_M(t_i) \right)_{k=1,\ldots p} \right)$$
$$= \vartheta((f_1(t_i))_{i=1,\ldots p}, \ldots, (f_M(t_i))_{i=1,\ldots p}) = (\vartheta T)(f_1, \ldots, f_M)$$

Evidently, ϑ is linear continuous operator with $\|\vartheta\| = \max_{\substack{k=1,\ldots,p \\ j=1,\ldots,M}} |\sum_{i=1}^{p} \alpha_i^j(t_k)|$, so $Y = (Y_d, d \in \mathbb{Z})$ is AR(1). $\square$

Conclusion and further research

The liberalization of electricity sector introduced a new field of research. Accurate modeling and forecasting of different variables related to the market e.g. prices, demand, production etc. became more crucial due to market structure. Thus, accurate forecasting is very important issue for an efficient management of power grid. In the past, various techniques have been developed both for price and demand prediction with different levels of complexity and final performance. This thesis addressed the issue of forecasting electricity demand and prices following to a relatively new modeling technique based on functional data analysis. The main results are presented in Chapters 4 and 5.

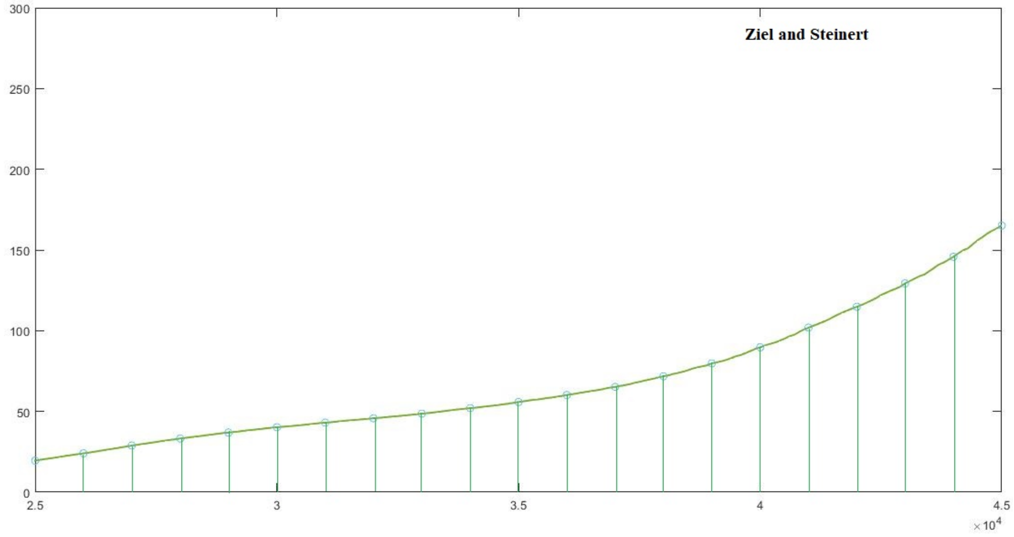
Chapter 4 focused on the parsimonious way for representing supply and demand curves, using a mesh-free method based on radial basis functions. The real data about supply and demand bids from the Italian day-ahead electricity market showed that there is no direct relationship between the number of offer and bid layers and the hour of the day, the day of the week, and the time of the year. Based on this observation, we decided to choose the same number of basis functions independently of these three seasonality modes. The numerical results showed that the accuracy of our approximation is good enough, if we use 5 basis function for the demand curve and 10 basis function for the supply curve, and then the increase in the number of functions leads to more time-consumption, but the increase of the accuracy is less significant.

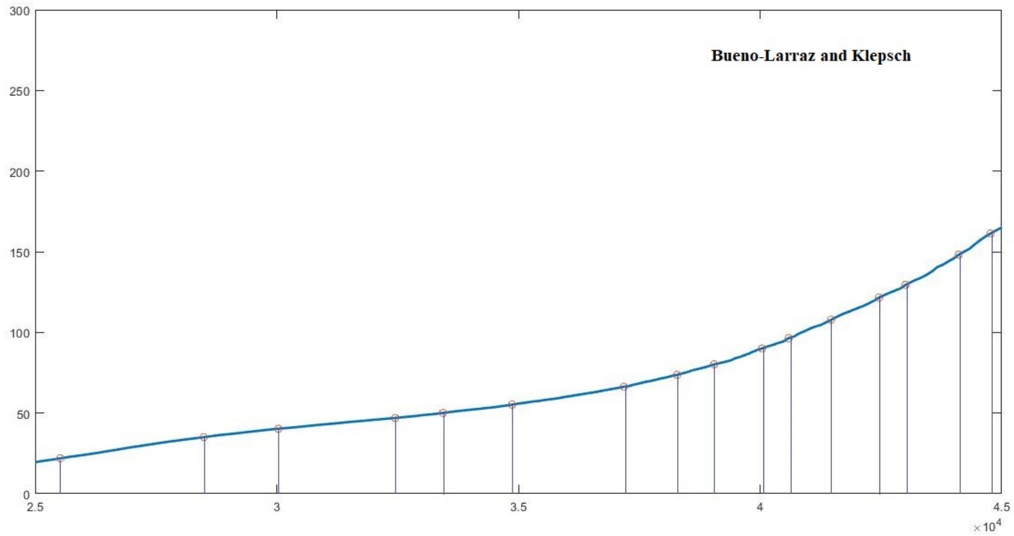
We also tested this new approach with the aim of forecasting supply and demand curves and finding the intersection of the predicted curves in order to obtain the market clearing price. In assess the goodness of our method, we compared it with models with similar complexity in terms of dependence of the past, but only based on the clearing price. Our forecasting errors are smaller compared with these univariate models. In particular, our analyses show that our multivariate approach leads to better results than the univariate one in terms of error measures like MAE, MAPE and RMSE.

In **Chapter 5** we considered supply and demand curves as stochastic processes with values in a functional space. We obtained a characterization result regarding the conditions that guarantees that a linear transformation of a vector AR process is again an AR process both in finite and in infinite dimension, and we applied these results to the model of Ziel and Steinert from [75].

We also found out that the model of Ziel and Steinert is a particular case of the model proposed in paper [15]. In particular, in [75] the authors applied a simple dimension reduction procedure to the price formation process that is computational manageable. It means that for the prediction of $x_{n+1}(\cdot)$ the whole curves $x_n(\cdot)$ is replaced with the p most relevant evaluations $x_n(t_1), x_n(t_2), \ldots, x_n(t_p)$. The problem of the selection of the most relevant points $t_1, t_2, \ldots, t_p \subset [0,1]$ is commonly known as variable selection problem. In [15] the authors showed how to find relevant points of the curves in terms of prediction accuracy. Applying the algorithm proposed in [15], we made an observation that the point used by Ziel and Steinert are not optimal in this sense (see Figure 5.4). So, one of the possibility for further research could be to add the optimal choice of the points into approach proposed by Ziel and Steinert.

Figure 5.1: Supply curve with chosen relevant points





Appendix A: Matlab Code

```
% Task I: CURVES APPROXIMATION

%create array of day between some dates
tic;
d1=datenum('2017-01-01','yyyy-mm-dd');
d2=datenum('2017-01-01','yyyy-mm-dd');
d=d1:d2;
quant_of_days=length(d);
s = [num2str(M_supply),' and ',num2str(M_demand)];
%approximate up to this price
% create columns for results
Eq_price_array=zeros(quant_of_days*24,1);
Eq_quant_array=zeros(quant_of_days*24,1);
Eq_price_approx=zeros(quant_of_days*24,1);
Eq_quant_approx=zeros(quant_of_days*24,1);
Error_price_array=zeros(quant_of_days*24,1);
Error_quantity_array=zeros(quant_of_days*24,1);
Hour_name = zeros(quant_of_days*24,1,'int8');
Date_name = strings(quant_of_days*24,1);
R=strings(quant_of_days*24,1);
% create tables for statistics of coeffitient
Supply_coeff = zeros(24*length(d), M_supply);
Demand_coeff = zeros(24*length(d), M_demand);
counter=0;
for k=1:length(d)% cycle for each day of the year
  filename1 = ['C:\Users\maria\Work\Energy market\Matlab\2017 Offers and ...
      Bids\',datestr(d(k),'yyyy-mm-dd'),'-OFF.txt'];
  filename2 = ['C:\Users\maria\Work\Energy market\Matlab\2017 Offers and ...
      Bids\',datestr(d(k),'yyyy-mm-dd'),'-BID.txt'];
  datestr(d(k),'yyyy-mm-dd')
  T_OFFERS = readtable(filename1);
  T_BIDS = readtable(filename2);
for hour = 1 : 1 : 24 % cycle for each hour of the day
      numer_row=24*counter+hour;
      %1.read the information about this hour to pOff,pBid, qOff, qBid
      indexes=find(T_BIDS.Hour==hour);
      pBid=zeros(length(indexes),1);
      qBid=zeros(length(indexes),1);
      for j=1:length(indexes)
          pBid(j)=T_BIDS.Price(indexes(j));
          qBid(j)=T_BIDS.Quantity(indexes(j));
      end
      indexes=find(T_OFFERS.Hour==hour);
      qOff=zeros(length(indexes),1);
      pOff=zeros(length(indexes),1);
      for j=1:length(indexes)
         pOff(j)=T_OFFERS.Price(indexes(j));
         qOff(j)=T_OFFERS.Quantity(indexes(j));
      end
      Max_price=400;%pOff(length(pOff));
      %2.calculate equilibrium price and quantity
      [P_eq, Q_eq] = Equilibrium(pOff,pBid,qOff,qBid);
```

```
%3.approximation of supply curve
 Matrix_coeff_supply = Approx_coeff(pOff,qOff,M_supply,Max_price);
%4.approximation of demand curve
pBid2=zeros(length(pBid)+1,1);
qBid2=zeros(length(qBid)+1,1);
pBid2(1)=0;
for j=2:length(pBid2)
 pBid2(j)=pBid(j-1);
 qBid2(j)=qBid(j-1);
end
qBid2(1)=abs(sum(qBid)-sum(qOff));
Matrix_coeff_demand = Approx_coeff(pBid2,qBid2,M_demand,Max_price);
%create functions as vectors
x=transpose(1:1:sum(qBid2));
Supply_approx = zeros(length(x),1);
for n = 1:M_supply
    Supply_coeff(24*counter+hour,n)=Matrix_coeff_supply(n,1);
    Supply_approx=Supply_approx+Matrix_coeff_supply(n,1)*(erf((x-Matrix_coeff_supply(n,2))/M
end
Demand_approx = zeros(length(x),1);
for n = 1:M_demand
    Demand_coeff(24*counter+hour,n)=Matrix_coeff_demand(n,1);
    Demand_approx=Demand_approx+Matrix_coeff_demand(n,1)*(erf((x-Matrix_coeff_demand(n,2))/M
end
Demand_approx = flipud(Demand_approx);
for n = 2:(length(qOff))
   qOff(n) = qOff(n)+qOff(n-1);
end
fOff=zeros(length(x),1);
j=1;
i=1;
while (i≤length(x)) && (j≤length(qOff))
   while (i≤length(x)) && (x(i)≤qOff(j))
     fOff(i)=pOff(j);
     i=i+1;
   end
j=j+1;
end
pBid = flipud(pBid);
qBid = flipud(qBid);
for n = 2:(length(qBid))
   qBid(n) = qBid(n)+qBid(n-1);
end
fBid=zeros(length(x),1);
j=1;
i=1;
while (i≤length(x)) && (j≤length(qBid))
    while (i≤length(x)) && (x(i)≤qBid(j))
       fBid(i)=pBid(j);
       i=i+1;
    end
    j=j+1;
end
[M,I]=min(abs(fOff-fBid));
P_eq=1/2*(fOff(I)+fBid(I));
Q_eq=I;
%5.intersection of approximated curves
```

```
        [A,Ind]=min(abs(Supply_approx-Demand_approx));
        P_approx=1/2*(Supply_approx(Ind)+Demand_approx(Ind));
        Q_approx=x(Ind);
        %graphic of results
        figure1 = figure;
        axes1 = axes('Parent',figure1);
        hold(axes1,'all');
        xlabel('Volume (MWh)')
        ylabel('Price (P,/MWh)')
        plot(x, fOff,'g',x,fBid,'r','linewidth',1.5)
        plot(x,Supply_approx,'r','linewidth',1.5); hold on;
        plot(x,Demand_approx,'b','linewidth',1.5); hold on;
        axis([0 60000 0 3500]);
        %write results in row of the table
        Eq_price_array(numer_row)=P_eq;
        Eq_quant_array(numer_row)=Q_eq;
        Eq_price_approx(numer_row)=P_approx;
        Eq_quant_approx(numer_row)=Q_approx;
        Error_price_array(numer_row)=abs(P_approx-P_eq);
        Error_quantity_array(numer_row)=abs(Q_approx-Q_eq);
        Hour_name(numer_row)=hour;
        Date_name(numer_row)=datestr(d(k),'yyyy-mm-dd');
        end %end cycle for each hour of the day
counter=counter+1;
end
time=toc;
Max_error=max(Error_price_array);
Mean_error = mean(Error_price_array);
R(1)=['time = ',num2str(time),' sec.'];
R(2)=['max_er = ',num2str(Max_error)];
R(3)=['mean_er = ',num2str(Mean_error)];
Table_results = ...
    table(Date_name,Hour_name,Eq_price_array,Eq_quant_array,Eq_price_approx,Eq_quant_approx,Error_p
namefile=['Expiriment with ',s,'.xlsx'];
writetable(Table_results,namefile);
%for statistics
        Table1CoeffDemand = table(Demand_coeff);
        namefile=['Coeffitients for demand with ',s,' functions.xlsx'];
        writetable(Table1CoeffDemand,namefile);
        Table1CoeffSupply = table(Supply_coeff);
        namefile=['Coeffitients for supply with ',s,' functions.xlsx'];
        writetable(Table1CoeffSupply,namefile);

close all
%FUNCTIONS
function [output1,output2] = Equilibrium(pOff,pBid,qOff,qBid)
        pBid = flipud(pBid);%bids need to be sort from
        qBid = flipud(qBid);%the biggest price to the least
        qBid2=zeros(length(qBid),1);
        qBid2(1)=qBid(1);
        for j=2:length(qBid)
            qBid2(j)=qBid(j)+qBid2(j-1);
        end
        qOff2=zeros(length(qOff),1);
        qOff2(1)=qOff(1);
        for j=2:length(qOff)
            qOff2(j)=qOff(j)+qOff2(j-1);
```

```
        end
        p_between_supl_dem=pBid(1)-pOff(1);
        j=1;
        while p_between_supl_dem > 0
            j=j+1;
            q=qBid2(j);
            i=1;
            while q>qOff2(i)
                i=i+1;
            end
            p_between_supl_dem=pBid(j)-pOff(i);
        end
        if p_between_supl_dem < 0
           Q_eq=qBid2(j-1);
           P_eq=pOff(i);
        end
        if p_between_supl_dem == 0
           Q_eq=qBid2(j);
           P_eq=pOff(i);
        end
      output1=P_eq;
      output2=Q_eq;
end

function [output1] = Approx_coeff(price,quant,M_supply,Max_price)
[p,q] = Simplier_data(price,quant,Max_price);%to make less data
qq=q;
for n = 2:length(qq)
    qq(n) = qq(n)+qq(n-1);
end
M=M_supply;
Q=qq(length(qq));% amount of electicity
%[Matrix_coeff, time, price_array]=datainterpolation(M,p,q);
an_matr=zeros(4,M+1);
%1 - price, 2 - amount of this price
%3 - where this price finishes 4 - number of jumps before this price
an_matr(1, M+1)=p(length(p));
an_matr(3, M+1)=qq(length(qq));
an_matr(2, M+1)=q(length(q));
an_matr(4,M+1)=1;% to count first and last price
an_matr(4,1)=1;
i=1;
for num = 0:M-1
    Price_jump=an_matr(1,M-num+1)/(M-num);
    j=0;
    while(i<=length(p)) && (p(length(p)-i)>an_matr(1,M-num+1)-Price_jump)
        i=i+1;
        j=j+1;
    end
    an_matr(1,M-num)=p(length(p)-i);
    an_matr(2, M-num)=q(length(q)-i);
    an_matr(3, M-num)=qq(length(q)-i);
    an_matr(4,M-num+1)=an_matr(4,M-num+1)+j;
end
Matrix_coeff = zeros(M,3);
for i = 1:M
 P1=an_matr(1,i);
```

```
 P2=an_matr(1,i+1);
 J=an_matr(4,i+1);
 center=an_matr(3, i);
 if J == 1
    Matrix_coeff(i,:)=[(P2-P1)/2,center,0.5];
 end

 if J > 1
     h=4;
    [datax,dataf,center]=onestepdata(P1,P2,h,p,q);
    a1=(P2-P1)/2;
    a2=center;
    a3=1/3000;
    a4=P1;
    z0 = [a2 a3];
    F = @(z,zdata)a1*(erf(z(2)*(datax-h*z(1)))+1)+a4;
    [z,resnorm,¬,exitflag,output] = lsqcurvefit(F,z0,datax,dataf);%optimization
    Matrix_coeff(i,:)=[a1,h*z(1),1/z(2)];
 end
end
output1=Matrix_coeff;
end
function [output1, output2] = Simplier_data(pOff,qOff,Max_price)
      q1=qOff;
      p1=min(round(pOff),Max_price);
      m=1;% m is a number of different prices
      for i=2:length(p1)
          if (p1(i)≠p1(i-1))
              m=m+1;
          end
      end
      p = zeros(m,1);
      q = zeros(m,1);
      k=1;
      for i=1:m
           p(i)=p1(k);
           q(i)=q1(k);
           k=k+1;
           while (k⩽length(p1)) && (p1(k)==p1(k-1))
                q(i)=q(i)+q1(k);
                k=k+1;
           end
      end
      q(length(q))=min(q(length(q)),10000);
output1=p;
output2=q;
end
```

```
% Task II: PRICE PREDICTION
%One variavle forecast
filename = ['Eq_quant2017.xlsx'];%with weekday dummy
MAPE=zeros(24,1);
MAE=zeros(24,1);
RMSE=zeros(24,1);
for hour=1:1:24
Table = readtable(filename);
data=table2array(Table(:,hour));
data_0=data(1:304);
E=zeros(7,1);% 1 - sunday, 2 - monday, ... 7 - saturday
for day=1:1:7
i=day;          %first day - Sunday
count=0;
while i < (304+1)
     E(day)=E(day)+data_0(i,:);
      i=i+7 ;
      count=count+1;
end
E(day)=E(day)/count;
end
% make E=0
for day=1:1:7
i=day;          %first day - Sunday
    while i < (304+1)
              data_0(i)=data_0(i)-E(day);
              i=i+7;
    end
end
Mdl = arima(1,0,0);
EstMdl = estimate(Mdl,data_0(1:304));
Coef_ar=cell2mat(EstMdl.AR);
constant = EstMdl.Constant;
Result=zeros(61,1);
Error=zeros(61,1);
Error_percent=zeros(61,1);
Date_name = strings(61,1);
D = datetime(2017,11,30);
for i=1:1:61
      Past=data(303+i)-E(weekday(D+i-1));
      Forecast=EstMdl.Constant+Coef_ar*Past+E(weekday(D+i));
      error=abs(data(304+i)-Forecast);
      Result(i)=Forecast;
      Error(i)=error;
      Error_percent(i)=Error(i)*100/data(304+i);
      Date_name(i)=datestr(D+i,'yyyy-mm-dd');
end
Table_results = table(Date_name,data(305:365),Result,Error,Error_percent);
namefile=['ForecastEq_price_2month_hour',int2str(hour),'.xlsx'];
writetable(Table_results,namefile);
MAPE(hour)=mean(Error_percent);
RMSE(hour)=sqrt(mean(Error.^2));
MAE(hour)=mean(Error);
end
```

```
% Multivariable forecast
M_supply=10; %number of basis function 5,10,15,20
for hour=1:1:24
s = num2str(M_supply);
namefile=['Supply_coeff with ',s,'functions_Hour',int2str(hour),'.xlsx'];
%Table = readtable('Supply_coeff with 5functions_Hour12.xlsx');
Table = readtable(namefile);
data_centers=table2array(Table(:,M_supply+2:2*M_supply+1));
data_prices=table2array(Table(:,2:M_supply));
%s
Mdl_centers = varm(M_supply,7);
EstMdl_centers = estimate(Mdl_centers,data_centers(1:334,:));
%2 Forecast
Result=zeros(31,2*M_supply);
Error=zeros(31,2*M_supply);
Date_name = strings(31,1);
D = datetime(2017,12,01);
for i=1:1:31
    %centers
    Forecast_centers=EstMdl_centers.Constant;
    for j=1:1:7
        Past=transpose(data_centers(334+i-j,:));
        Forecast_centers=Forecast_centers+EstMdl_centers.AR{j}*Past;
    end
    error=abs(data_centers(334+i,:)-transpose(Forecast_centers));
    Result(i,M_supply+1:2*M_supply)=transpose(Forecast_centers);
    Error(i,M_supply+1:2*M_supply)=error;
    Date_name(i)=datestr(D,'yyyy-mm-dd');
    %prices
    Result(i,M_supply)=350;
    D=D+1;
end
%prices
for j=1:1:(M_supply-1)
data=data_prices(:,j);
Mdl = arima(7,0,0);
EstMdl = estimate(Mdl,data(1:334));
Coef_ar=cell2mat(EstMdl.AR);
constant = EstMdl.Constant;
    for i=1:1:31
    Past=[data(333+i);data(332+i);data(331+i);data(330+i);data(329+i);data(328+i);data(327+i)];
    Forecast=EstMdl.Constant+Coef_ar*Past;
    error=abs(data(334+i)-Forecast);
    Result(i,j)=Forecast;
    Error(i,j)=error;
    end
end
Table_results = table(Date_name,Result,Error);
namefile=['Forecast_Supply',s,'fun_Hour',int2str(hour),'.xlsx'];
writetable(Table_results,namefile);
%for futher step
Table_results = table(Date_name,Result);
namefile=['Forecast_CoefSupply',s,'fun_Hour',int2str(hour),'.xlsx'];
writetable(Table_results,namefile);
end

PRICES=zeros(24*61,2);
```

```
MAE=zeros(24,1);
RMSE=zeros(24,1);
MAPE=zeros(24,1);
M_supply=5;
for hour=1:1:24
s = num2str(M_supply);
namefile=['Supply_coeff2 with ',s,'functions_Hour',int2str(hour),'.xlsx'];
%Table = readtable('Supply_coeff with 5functions_Hour12.xlsx');
Table = readtable(namefile);
K=2*M_supply;
data=zeros(365,K);
data(:,1:M_supply)=table2array(Table(:,2:M_supply+1));
data(:,M_supply+1:K)=table2array(Table(:,M_supply+2:K+1))/1000;
namefile=['Eq_quant2017.xlsx'];
Table = readtable(namefile);
data(:,K)=table2array(Table(:,hour))/1000;%the last center is the equilibrium quantity ...
    of electricity
Mdl = varm(K,1);
EstMdl = estimate(Mdl,data(1:304,:));
Coef_ar=cell2mat(EstMdl.AR);
constant = EstMdl.Constant;
Result=zeros(61,1);
Error=zeros(61,1);
Error_percent=zeros(61,1);
Date_name = strings(61,1);
D = datetime(2017,11,01);
for i=1:1:61
    Past=data(303+i,:);
    Forecast=EstMdl.Constant+Coef_ar*transpose(Past);
    Price = data(304+i, M_supply);
    Result(i)=Forecast(M_supply);
    Error(i)=abs(Result(i)-Price);
    Error_percent(i)=Error(i)*100/Price;
    PRICES((i-1)*24+hour,1)=Price;
    PRICES((i-1)*24+hour,2)=Result(i);
    Date_name(i)=datestr(D,'yyyy-mm-dd');
    D=D+1;
end
%prediction of demand
for i=1:1:61
    Past=data(303+i,:);
    Forecast=EstMdl.Constant+Coef_ar*transpose(Past);
    Demand = data(304+i,K)*1000;
    Result(i)=Forecast(K)*1000;
    Error(i)=abs(Result(i)-Demand);
    Error_percent(i)=Error(i)*100/Demand;
    Date_name(i)=datestr(D,'yyyy-mm-dd');
    D=D+1;
end
MAPE(hour)=mean(Error_percent);
RMSE(hour)=sqrt(mean(Error.^2));
MAE(hour)=mean(Error);
Table_results = table(Date_name,data(305:365, M_supply),Result,Error,Error_percent);
namefile=['Forecast_two_month_Second_method_hour',int2str(hour),' with ...
    ',int2str(K),'parameters.xlsx'];
writetable(Table_results,namefile);
end
```

Appendix B: SQL Code

```
SET GLOBAL innodb_buffer_pool_size=402653184;
USE electicity_offer_bid;

CREATE  TABLE all_offer_bid (day VARCHAR(50), hour INT, quantity FLOAT, price FLOAT, ...
    type VARCHAR(50));

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/all_offers.txt' INTO ...
    TABLE all_offers FIELDS TERMINATED BY ';' LINES TERMINATED BY "\n";

SELECT * FROM all_offer_bid;

SELECT hour, quantity, price FROM all_offer_bid WHERE day="2014-01-01" AND type LIKE ...
    '%OFF%';
SELECT hour, quantity, price FROM all_offer_bid WHERE day="2014-01-01" AND type LIKE ...
    '%BID%';

CREATE TABLE 2014-01-01-OFF AS SELECT hour, quantity, price FROM all_offer_bid WHERE ...
    day="2014-01-01" AND type LIKE '%OFF%';

CREATE TABLE new_table AS SELECT hour, quantity, price FROM all_offer_bid WHERE ...
    day="2014-01-01" AND price < 1 AND type LIKE '%OFF%' AND hour=1;

SELECT * INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/name.xls' FIELDS ...
    TERMINATED BY '\t' LINES TERMINATED BY '\n' FROM new_table;

STR_TO_DATE("August 10 2017", "%M %d %Y");

 count number of rows
SELECT COUNT(*) FROM all_bids2017;
SELECT COUNT(*) FROM all_offers2017 WHERE daydate LIKE '2017-12%';
SELECT COUNT(*) FROM all_bids2017 WHERE day LIKE '2017-12%';
 string to date
SELECT STR_TO_DATE(2017-01-01, '%Y-%m-%d') FROM  all_offers2017;

 declare @my_date datetime
 set @my_date = '20170101'
 while @my_date < '20171231'
 begin
    CREATE  TABLE bid2018-02-05-BID AS SELECT day, hour, quantity, price FROM all_bids ...
        WHERE day LIKE '2013%';
    set @my_date = dateadd(dd, 1, @my_date)
 end

CREATE  TABLE test(day DATE, hour INT, quantity FLOAT, price FLOAT);
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/test.txt' INTO TABLE ...
    test FIELDS TERMINATED BY ';' LINES TERMINATED BY "\n";
INSERT INTO test (day, hour, quantity, price) VALUES ('2017-01-01', 1,0,0);
INSERT INTO test (day, hour, quantity, price) VALUES (STR_TO_DATE(2017-01-01, ...
    '%Y-%m-%d'), 1,0,0);

 declare @my_date datetime
 set @my_date = '20170101'
```

```
while @my_date < '20171231'
begin
    INSERT INTO test (day, hour, quantity, price) VALUES ('2017-01-01', 1,0,0);;
    set @my_date = dateadd(dd, 1, @my_date)
end

delimiter #
create procedure cikl1()
begin

declare v_max int unsigned default 1000;
declare v_counter int unsigned default 0;

  truncate table foo;
  start transaction;
  while v_counter < v_max do
    insert into foo (val) values ( floor(0 + (rand() * 65535)) );
    set v_counter=v_counter+1;
  end while;
  commit;
end #

delimiter ;

UPDATE all_offers2017 SET day = str_to_date( day, '%Y-%m-%d' );

CONVERT VARCHR TO DATE IN THE TABLE+

ALTER TABLE all_offers2017 ADD COLUMN daydate DATE AFTER day;
UPDATE all_offers2017 SET daydate = STR_TO_DATE( day, '%Y-%m-%d');
SELECT * INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server ...
    8.0/Uploads/all_offers20172.txt' FIELDS TERMINATED BY '\t' LINES TERMINATED BY ...
    '\n' FROM all_offers2017;
DESCRIBE  all_offers2017;
ALTER TABLE all_offers2017 DROP COLUMN day;

SELECT 1 day

SELECT * FROM all_offers2017 WHERE daydate='2017-01-01';

CREATE PROCEDURE

DELIMITER //
DROP PROCEDURE IF EXISTS save_day_test//
CREATE PROCEDURE save_day_test()
BEGIN
DECLARE my_date DATE DEFAULT '20170101';
WHILE my_date < '20170103' DO
DECLARE my_name VARCHAR(10) DEFAULT
DROP TEMPORARY TABLE IF EXISTS tmp_deals;
CREATE TEMPORARY TABLE tmp_deals
SELECT hour, quantity, price FROM all_offers2017 WHERE daydate=my_date and hour=1;
SELECT * INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/"my_name"-OFF.txt'
FIELDS TERMINATED BY ';' LINES TERMINATED BY '\n' FROM tmp_deals;
set my_date = ADDDATE(my_date, INTERVAL 1 DAY);
END WHILE;
END//
```

```
DELIMITER;

CALL save_day_test();

DELIMITER //
DROP PROCEDURE IF EXISTS count_week_days//
    CREATE PROCEDURE count_week_days(OUT param1 INT)
    BEGIN
    DECLARE num INT DEFAULT 0;
    param1=num;
    WHILE my_date <= '2017-01-02' DO
        SELECT COUNT(*) INTO param1 FROM all_offers2017 WHERE daydate=my_date;
        set my_date = ADDDATE(my_date, INTERVAL 7 DAY);
    END WHILE;
END;
//
DELIMITER ;

CALL count_week_days(@a);
SELECT @a;

DELIMITER //
DROP PROCEDURE IF EXISTS save_day_bids//
    CREATE PROCEDURE save_day_bids()
    BEGIN
DECLARE my_date DATE DEFAULT '2017-01-01';
WHILE my_date <= '2017-12-31' DO
      SET @file_date=CAST(my_date AS CHAR);
      DROP TEMPORARY TABLE IF EXISTS tmp_deals;
      CREATE TEMPORARY TABLE tmp_deals
      SELECT hour, quantity, price FROM all_bids2017 WHERE day=@file_date;
      SET @tmp_sql= CONCAT("SELECT 'Hour', 'Quantity', 'Price' UNION ALL
      SELECT * INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server ...
          8.0/Uploads/",@file_date,"-BID.txt'
      FIELDS TERMINATED BY ';' LINES TERMINATED BY '\n' FROM tmp_deals");
      PREPARE s1 FROM @tmp_sql;
      EXECUTE s1;
      DEALLOCATE PREPARE s1;
      set my_date = ADDDATE(my_date, INTERVAL 1 DAY);
END WHILE;
END;
//
DELIMITER ;

call save_day_bids();
```

Bibliography

[1] Abramowitz, M. , and Stegun, I. A., emphRepeated Integrals of the Error Functio. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, 1972, 299-300,.

[2] Aneiros, G., Vilar, J.M., Cao, R., San Roque, A.M., *Functional prediction for the residual demand in electricity spot markets*. IEEE Trans. Power Syst. 28 (4), 2013, 4201-4208.

[3] Aronszajn, N., *Theory of Reproducing Kernels*, Transactions of the American Mathematical Society. 68 (3), 1950, 337-404.

[4] Barlow, M.T. *A diffusion model for electricity prices*. Mathematical Finance 12, 2002, 287-298.

[5] Berrendero, J. R., Bueno-Larraz, B., Cuevas, A. *An RKHS model for variable selection in functional linear regression*. Journal of Multivariate Analysis, 170 , 2019, 25-45.

[6] Bessembinder, H., Lemmon, M.L. emphEquilibrium pricing and optimal hedging in electricity forward markets. Journal of Finance LVII, 2002, 1347-1382.

[7] Beutier, F. J., *The operator theory of the pseudo-inverse. I: Bounded operators*, J. Math. Anal. Appl. 10, 31, 1965, 451-470.

[8] Bochner, S., Lectures on Fourier Integrals, Princeton University Press, 1959. (Translated by Morris Tenenbaum and Harry Pollard.)

[9] Bosq, D. Nonparametric statistics for stochastic processes, second ed. of Lecture Notes in Statistics. Springer-Verlag, New York, 1998.

[10] Bosq, D. Linear Processes in Function Spaces. Lecture Notes in Statistics, Springer, New York, 2000.

[11] Bosq, D., Blanke, P., Inference and Prediction in Large Dimensions. John-Wiley & Sons Ltd. 2007.

[12] Bosq, D., *Sufficiency and efficiency in statistical prediction.* Statist. Probab. Lett. 77, 3, 2007, 280-287.

[13] Bosq, D., *General linear processes in Hilbert spaces and prediction.* J. Statist. Plann. Inference 137(3), 2007, 879-894.

[14] Brockwell, P.J. and Davis, R.A. Time series: theory and methods 2nd edn. Springer-Verlag, New York, 1991, 187-190.

[15] Bueno-Larraz, B., Klepsch, J., *Variable Selection for the Prediction of* $C[0,1]$*-Valued Autoregressive Processes using Reproducing Kernel Hilbert Spaces.* Journal of Technometrics, Taylor & Francis, 61(2), 2018 , 139-153.

[16] Bozzini, M., Lenarduzzi, L., Rossini, M., Schaback, R., *Interpolation with variably scaled kernels.* IMA Journal of Numerical Analysis, 35(1), 2015, 199-219.

[17] Buescu, J. , Paixao, A.C., *Complex Variable Positive Definite Functions.* Complex Anal. Oper. Theory, 8, 2014, 937-954.

[18] Cartea, A., Villaplana, P., *Spot price modeling and the valuation of electricity forward con-tracts: The role of demand and capacity.* Journal of Banking and Finance, 32 (12), 2008, 2502-2519.

[19] Cavoretto, R. , De Rossi, A. , Perracchione, E., *Optimal Selection of Local Approximants in RBF-PU Interpolation.* J. Sci. Comput. 74(1), 2018, 1-22.

[20] Christensen, O. An Introduction to Frames and Riesz Bases. Birkhauser, 2016, 56-58.

[21] De Marchi, S., MartГnez, A. ,Perracchione, E. ,Rossini, M., *RBF-based partition of unity method for elliptic PDEs: Adaptivity and stability issues via VSKs.* Journal of Scientific Computing., 2018, 1вЂ“24

[22] Diestel, J. ,Uhl, J. J.. Vector Measures. American Mathematical Society, Providence, R.I., 1977.

[23] Dunford, N., Schwartz, J. T.. Linear Operators. I. General Theory. Interscience Publishers, New York, 1958.

[24] Düker, M.-C., *Limit theorems for Hilbert space-valued linear processes under long range dependence*. Stochastic Processes and their Applications, 128, 5, 2018, 1439-1465.

[25] Eichler, M., Sollie, J., Tuerk, D., *A new approach for modelling electricity spot prices based on supply and demand spreads.* Conference on Energy Finance 2012 Trondheim, Norway., 1-4.

[26] Engl, H. W. and Nashed, M. Z., *New extremal characterizations of generalized inverses of linear operators.* J. Math. Anal. Appl. 82 , 2, 1981 , 566-586.

[27] Fasshauer, G.E., *Positive definite kernels: past, present and future.* Dolomite Research Notes on Approximation 4, 2011, 21-63.

[28] Fasshauer, G. E., Qi Ye, *Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators*, Numerische Mathematik 119, 2011, 585-611.

[29] Fasshauer, G.E., McCourt, M.J.. Kernel-based Approximation Methods Using MATLAB, World Scientific, Singapore, 2015.

[30] Franke, R., A critical comparison of some methods for interpolation of scattered data, Technical Report NPS, 1979, 53-79.

[31] Friedman, J., Hastie, T., Hofling, H., Tibshirani, R., *Pathwise coordinate optimization.* Ann. Appl. Stat. 1 (2), 2007, 302-332.

[32] Grenander, U.. Probabilities on Algebraic Structures. Wiley, NewYork, 1963.

[33] Hagen, R., Roch, S., Silbermann, B.. C*-algebras and Numerical Analysis. (Section 2.1.2.)CRC Press, 2001.

[34] Hardy, R. L., *Multiquadric equations of topography and other irregular surfaces.* Journal of Geophysical Research, 76(8), 1971, 1905-1915.

[35] Hastie, T., Tibshirani, R., Wainwright M., Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press, 2015.

[36] Kadets V.. A Course in Functional Analysis and Measure Theory. Springer, 2018.

[37] Kitaqawa, G. , *An Algorithm for Solving the Matrix Equation* $X = FXF' + S$. International Journal of Control., 25 (5), 1977, 745-753.

[38] Knorn, F.. M-code LaTeX Package (https://www.mathworks.com/matlabcentral/fil m-code-latex-package), MATLAB Central File Exchange. Retrieved March 2, 2020.

[39] Kubrusly, C.S. , *A Note On The Lyapunov Equation For Discrete Linear Systems In Hilbert Space* , Appl. Math. Lett., 2, 4, 1989, 349-352.

[40] Kulakov, S. V., Ziel F., *Determining the Demand Elasticity in a Wholesale Electricity Market*, arXiv:1903.11383.

[41] Ledoux, M. , Talagrand, M. Probability in Banach Spaces: Isoperimetry and Processes. Springer Science & Business Media, 1991.

[42] Liebl, D. *Modeling and Forecasting Electricity Spot Prices: A Functional Data Perspective.* The Annals of Applied Statistics , 7, 3, 2013, 1562-1592.

[43] Longstaff, F.A., Wang, A.W., *Electricity forward prices: a high-frequency empirical analysis.* Journal of Finance LIX, 2004, 1877-1900.

[44] Lutkepohl, H., Linear transformations of vector ARMA processes. Journal of Econometrics, Elsevier, 26(3), 1984, 283-293.

[45] Lutkepohl, H. New introduction to multiple time series analysis. New York: Springer Science & Business Media, 2005.

[46] Maciejowska, K., Nowotarski, J., Weron, R., *Probabilistic forecasting of electricity spot prices using factor quantile regression averaging*. Int. J. Forecast. 32 (3), 2016, 957-965.

[47] Mas, A., Pumo, B. Linear processes for functional data. F. Ferraty, Y. Romain (Eds.), The Oxford Handbook of Functional Data, Oxford, 2010.

[48] Mercer, J., *Functions of positive and negative type and their connection with the theory of integral equations*, Phil. Trans. Royal Society 209, 1909, 415-446.

[49] Micchelli, C.A., *Interpolation of scattered data: Distance matrices and conditionally positive definite functions*. Constr. Approx. 2, 1986, 11-22.

[50] Nowotarski, J., Weron, R., *Recent advances in electricity price forecasting: A review of probabilisticforecasting*. Renew. Sustain. Energy Rev., 81, 2018, 1548-1568.

[51] Paulsen, V., and Raghupathi, M.. An Introduction to the Theory of Reproducing Kernel Hilbert Spaces. Cambridge: Cambridge University Press, 2016.

[52] Pazouki, Allaei, S.S., Hossein Pazouki, and Moller, D.P.F., *Adaptive learning algorithm for RBF neural networks in kernel spaces*. In Proceedings of the International Joint Conference on Neural Networks, 2016, 4811-4818.

[53] Penrose, R.A., *A Generalized Inverse for Matrices*. Proceedings of Cambridge Philosophical Society, 51, 1955, 406-413.

[54] Pirrong, C., Jermakyan, M. Valuing power and weather derivatives on a mesh using finite difference methods. Energy Modelling and the Management of Uncertainty, Risk Publications, 1999.

[55] Prajapati, A. K. , Srivastava S. K. and Narain A., *Electricity Price Forecasting: A Bibliographical Review*. International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, 345-349.

[56] Prakasa, R., *Characterization of Gaussian distribution on a Hilbert space from samples of random size.* Journal of Multivariate Analysis, 132, 2014, 209-214.

[57] Prakasa, R. *Chebyshev´s inequality for Hilbert-space-valued random elements.* Statistics & Probability Letters, 80, 11-12, 2010, 1039-1042.

[58] Romani, L., Rossini, M., Schenone, D., *Edge detection methods based on RBF interpolation*, J.Comput.Appl.Math., 349, 2019, 532-547.

[59] Rossini, M., *Interpolating functions with gradient discontinuities via variably scaled kernels.* Dolom. Res. Notes Approx.,11, 2018, 3-14.

[60] Runge, C., *Uber empirische Funktionen und die Interpolation zwischen aquidistanten Ordinaten*, Zeitschrift fur Mathematik und Physik . 46, 1901, 224-243.

[61] Schaback, R., *Native Hilbert Spaces for Radial Basis Functions I.* New Developments in Approximation Theory, 132,1999, 255-282.

[62] Schaback, R., *A unified theory of radial basis functions: Native Hilbert spaces for radial basis functions II.* Journal of computational and applied mathematics 121 (1-2), 2000, 165-177.

[63] Schoenberg, I. J., *Metric spaces and completely monotone functions*, Ann. of Math. 39, 1938, 811-841.

[64] Steinwart, I., Christmann, A., Support Vector Machines, Springer-Verlag New York, 2008.

[65] Shah, I. Modeling and Forecasting Electricity Market Variables (2016) [PhD Thesis].

[66] Shin, H., Hsing T., *Linear prediction in functional data analysis.* Stochastic Processes and their Applications, 122, 11, 2012, 3680-3700.

[67] Tibshirani, R., *Regression shrinkage and selection via the lasso.* J. R. Stat. Soc. Ser.B Stat Methodol. 58(1), 1996, 267вЂ“288.

[68] Tutaj, E., *An example of a reproducing kernel Hilbert space*. Complex Anal. Oper. Theory 13, 2019, 193вЂ“221.

[69] Tutaj, E., *Some particular norm in the Sobolev space $H_1[A, B]$*. Complex Anal. Oper. Theory 13, 2019, 1931вЂ“1947.

[70] Ugurlu, U., Oksuz, I., Tas O., *Electricity Price Forecasting Using Recurrent Neural Networks*. Energies 11(5), 2018, 1255.

[71] Wendland, H.. Scattered Data Approximation, Cambridge Monogr. Appl. Comput. Math., vol. 17, Cambridge Univ. Press, Cambridge, 2005.

[72] Weron, R.. Modeling and forecasting electricity loads and prices: a statistical approach. Chichester: Wiley, 2006.

[73] Weron, R., Misiorek, A., *Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models*. Int. J. Forecast. 24 (4), 2008, 744-763.

[74] Weron, R., *Electricity price forecasting: A review of the state-of-the-art with a look into the future*. Int. J. Forecast, 30(4), 2014, 1030 - 1081.

[75] Ziel, F., Steinert, R., *Electricity price forecasting using sale and purchase curves: the X-Model*. Energy Econ., 59, 2016, 435-454.

[76] Ziel, F., *Forecasting electricity spot prices using lasso: on capturing the autoregressive intraday structure*. IEEE Trans. Power Syst., 31(6), 2016, 4977-4987.

[77] Ziel, F., Steinert, R., *Probabilistic mid- and long-term electricity price forecasting*. Renewable and Sustainable Energy Reviews 94, 2018, 251-266.