Technical University of Denmark

DTU

Inverse Optimization and Forecasting Techniques Applied to Decision-making in Electricity Markets

Saez Gallego, Javier; Madsen, Henrik

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Saez Gallego, J., & Madsen, H. (2017). Inverse Optimization and Forecasting Techniques Applied to Decision-making in Electricity Markets. Kgs. Lyngby: Technical University of Denmark (DTU). (DTU Compute PHD-2016; No. 425).

DTU Library
Technical Information Center of Denmark

Inverse Optimization and Forecasting Techniques Applied to Decision-making in Electricity Markets

Javier Sáez Gallego

Summary (English)

This thesis deals with the development of new mathematical models that support the decision-making processes of market players. It addresses the problems of demand-side bidding, price-responsive load forecasting and reserve determination. From a methodological point of view, we investigate a novel approach to model the response of aggregate price-responsive load as a constrained optimization model, whose parameters are estimated from data by using inverse optimization techniques.

The problems tackled in this dissertation are motivated, on one hand, by the increasing penetration of renewable energy production and smart grid technologies in power systems, that is expected to continue growing in the coming years. Non-dispatchable electricity generation cannot ensure a certain production at all times, since it depends on meteorological factors. Also, smart grid technologies are affecting the consumption patterns that the load traditionally exhibited. On the other hand, this thesis is motivated by the decision-making processes of market players. In response to these challenges, this thesis provides mathematical models for decision-making under uncertainty in electricity markets.

Demand-side bidding refers to the participation of consumers, often through a retailer, in energy trading. Under the smart-grid paradigm, the demand bids must reflect the elasticity of the consumers to changes in electricity price. Traditional forecasting models are typically not able to reflect this elasticity, hence we propose two novel approaches to estimate market bids. Both approaches are data-driven and take into account the uncertainty of future factors, as, for example, price. In both cases, demand-side bids that comprise a price-energy term decrease the expected imbalances and also increase the profit of retailers

participating in electricity markets.

In the field of load forecasting, this thesis provides a novel approach to model time series and forecast loads under the real-time pricing setup. The relationship between price and aggregate response of the load is characterized by an optimization problem, which is shaped by a set of unknown parameters. Such parameters are estimated from data by using an inverse optimization framework. The usability of the proposed method is studied and we conclude that inverse-optimization-based modeling is a computationally attractive method that outperforms the forecasting capabilities of traditional time series models.

Regarding the reserve determination, the special characteristics of the Danish power system do not allow for co-optimizing the unit commitment and reserve requirements. Hence, we propose a probabilistic framework, where the reserve requirements are computed based on scenarios of wind power and load forecast errors and power plant outages. The solution of the stochastic optimization models increases the safety of the overall system while decreases the associated reserve costs, with respect to the method currently used by the Danish TSO.

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in partial fulfilment of the requirements for acquiring a Ph.D. degree.

The thesis deals with three decision-making problems in the context of electricity markets and smart grids: demand-side bidding, price-responsive load forecasting and reserve determination. The developed solutions comprise techniques from time series analysis, stochastic optimization and inverse optimization. Various evaluation studies consider the positive impact of the proposed approaches.

The thesis consists of a summary report and four research papers, documenting the research conducted over the period of January 2013 to July 2016.

Kgs. Lyngby, 29-July-2016

Javier Sáez Gallego

Acknowledgments

List of Publications

Scientific Research Publications in This Thesis

A J. Saez-Gallego, J. M. Morales, H. Madsen and T. Jónsson, "Determining reserve requirements in DK1 area of Nord Pool using a probabilistic approach". In: *Energy*, Volume 74, Pages 682-693, September 2014.

B J. Saez-Gallego, J. M. Morales, M. Zugno, and H. Henrik, "A data-driven bidding model for a cluster of price-responsive consumers of electricity". In: *IEEE Transactions on Power Systems*, February, 2016 .

C J. Saez-Gallego, M. Kohansal, A. Sadeghi-Mobarakeh and J. M. Morales "Optimal Price-energy Demand Bids for Aggregate Price-responsive Loads". Submitted to *IEEE Transactions on Power Systems*, 2016.

D J. Saez-Gallego and J. M. Morales, "Short-term Forecasting of Price-responsive Loads Using Inverse Optimization". Submitted to *IEEE Transactions on Smart Grid*, 2016.

Dissemination of Research at Conferences

The developed work during the course of the PhD study have been presented in the following conferences.

E J. Saez-Gallego, J. M. Morales, H. Madsen and T. Jónsson, "Optimal

Spinning Reserve by taking advantage of probabilistic forecasting," In: *Danish Wind Industry Annual Event*, 2014.

F J. Saez-Gallego, J. M. Morales, H. Madsen and T. Jónsson. "Determining reserve requirements in DK1 area of Nord Pool using a probabilistic approach". In: *International Conference on Computational Management Science*, 2014.

G J. Saez-Gallego, J. M. Morales, M. Zugno, and H. Henrik, "A data-driven bidding model for a cluster of price-responsive consumers of electricity,". In: *EURO Glasgow. 27th European Conference on Operational Research*, July, 2015.

H J. Saez-Gallego, J. M. Morales, M. Zugno, and H. Henrik, "A data-driven bidding model for a cluster of price-responsive consumers of electricity,". In: *INFORMS Annual Meeting*, November, 2015.

Contents

Part I

Summary Report

CHAPTER 1

Introduction

The problems solved in this thesis are ultimately motivated by the increasing share of electricity produced by renewable means, such as wind and solar power production. Traditionally, the supply in power systems has been provided mostly by nuclear, coal, or gas plants. These types of plants are dispatchable, meaning that their output can be scheduled and modified as required. Renewable energy, on the other hand, is a non-dispatchable source of energy, since its production is subject to meteorological factors.

As the share of electricity produced by renewables increases, several challenges must be faced. Amongst the vast number of problems to be solved in relation to the integration of renewable energy in power systems [1], this thesis tackles three of them. The three problems are optimal bidding, load forecasting, and reserve determination. They relate to decision-making in electricity markets, and share the common objective of finding the most optimal decision under uncertain future factors. Examples of such uncertain factors are the load and the price. The need for making a decision now, whose future effect is only partially known in the future, calls for advanced methodologies as the ones elaborated in this thesis.

1.1   Thesis Objectives and Contributions

In this dissertation, we start by giving an introduction to the different electricity markets in the Nordic countries, focusing specifically on the Danish case. The functioning of the markets and the bidding process play an important role in the integration of renewable energy, since all players must interact with electricity markets to buy or sell electricity. Paper A deals with reserve capacity markets and the optimal reserve schedule. The integration of demand response in current power systems is also analyzed. Papers B, C and D of this thesis tackle two problems that arise in the case where demand response and dynamic pricing coexist.

From a methodological point of view, the objective of this thesis is to develop new models that combine concepts from time series analysis and operations research to support the decision making. The summary report of this thesis provides an overview of the mathematical tools used. One of the main contributions is the use of a constrained optimization model to characterize the price response of the aggregate load. This novel concept is proved to be relevant in Papers B and D of this dissertation. Stochastic optimization models are also provided in Papers A and C for reserve determination and optimal bidding, respectively. In general, the developed methods represent a mixture between operations research models and time series models.

Application-wise, the objectives and contributions of this thesis are threefold:

- **Optimal bidding in electricity markets**. We consider the case where consumers of electricity are equipped with a smart grid meter and communication devices that receive a price associated with a period of time. An aggregator or retailer buys energy from the wholesale electricity market on behalf of its pool of consumers. The aggregator must submit a bid to the market that reflects the response of the pool, specially to changes in price. In this thesis, we tackle the bidding problem, and propose two solutions to it. In Paper B, the optimal bidding problem is based on the minimization of expected imbalances, while in Paper C the bids are used to maximize the aggregator's profit.

- **New load forecasting techniques under the smart grid paradigm**. We develop a novel methodology to forecast the aggregate load of a pool of price-responsive consumers. By price-responsive consumers, we understand consumers that change their profile of consumption depending on the price of electricity during the considered period. An example of such a consumer is a household with an energy management control system with a dynamic pricing contract. The proposed methodology is based on inverse optimization,

and its usefulness is benchmarked against traditional time series forecasting models. The work is presented in Paper D.

- **Reserve determination**. One of the challenges that arise when managing power systems with high penetration of renewable energy production is the determination of electricity reserves. In brief, reserves are "extra" capacity that is ready to be activated if necessary. A well-known rule, called *n-1 rule*, sets the minimum amount of manual reserves to be the capacity of the largest online generator. With an increased penetration of renewable energy, deterministic rules might not be the most optimal. The solution proposed in Paper A leverages the cost of not supplying the load versus the cost of scheduling reserves. The methodology considers the uncertainty of wind production, power plant outages and load, to provide a schedule for the optimal level of reserves.

1.2 Thesis Structure

This thesis is structured as follows.

Part I is subdivided in chapters and consists on a summary report outlining the main contributions of this thesis. Chapter 2 provides an overview of the electricity markets, focused on the case of the Nordic Countries and Denmark. Chapter 3 presents the mathematical tools that have been developed and used in this thesis. In Chapter 4, the applications problems are presented with the corresponding research results. Chapter 5 provides conclusions and perspectives.

Part II consists of the publications that contribute to this thesis, summarized as follows:

Paper A is a journal article published in *Energy*. It consists of a stochastic optimization model to optimally schedule the electricity reserves in the western power system area of Denmark, based on the uncertainty related to the load, the wind power production, and to the outages of power plants.

Paper B is a journal article published in *IEEE Transactions on Power Systems*. An inverse optimization method is used for estimating the optimal complex market bid, relative to a pool of price-responsive consumers.

Paper C is a journal article submitted to *IEEE Transactions on Power Systems*. This publication presents both analytical and computational bidding models for a retailer that aggregates users that respond to dynamic prices.

Paper D has been submitted for consideration to *IEEE Transactions on Smart Grid*. It presents a practical solution method for generalized inverse optimization problems, which is then applied to short-term load forecasting. A case study based on the simulation of the price-response behavior of a pool of buildings equipped with a heat pump is built and analyzed.

Electricity Markets

This chapter starts by providing with an overview of the special characteristics of electricity that make it different to other types of goods, serving as a motivation for the need of electricity markets. Then, in Section 2.2, we introduce the structure of the electricity markets in the Nordic countries, focusing in the specific case of Denmark. We give a short introduction to the day-ahead market and the real-time market, also called *Elspot* and *balancing* market. Furthermore, in Section 2.2.1, we motivate the need for electricity reserves and present the current regulations in this regard. Finally, in Section 2.3, we introduce the concept of demand response and the role of the *aggregator*.

2.1 The Characteristics of Electricity

Electricity is one of the most necessary elements in a contemporary society. It is used by millions of people in their houses and offices to empower appliances, cool down or heat the air, by factories to power up their production machinery, transport goods, and power up computers, amongst other uses. As most of the commodities, as for example food and electronic objects, consumers and producers of electricity buy and sell electricity through a market. However, the characteristics of the electricity markets are different to the markets where other

goods are traded. These differences are caused by the nature of the electricity itself, which are summarized as follows:

1. The physical system works much faster than any other market. Electricity can be transported long distances in a much faster way than other commodities. Moreover, it requires a special and expensive infrastructure of cables and transformers.

2. The energy produced by power plants is pooled on the way to the consumer, making the consumer unable to determine from which power plant the electricity comes from.

3. The electricity must be supplied at the same time when it is consumed. This also means that the production and consumption must be the same at all times.

4. The demand has been traditionally very inelastic, implying that consumers generally do not change their consumption depending on the price. The first reason why this occurs is because electricity dot not have a direct substitute product that consumers can switch to if electricity becomes too expensive. The second reason is that small consumers are traditionally not affected by prices changes instantly. This fact could change in the future with the implementation of smart grid technologies.

The need for an organized electricity market structure arises mostly by the fact that, for technical reasons, consumers cannot determine who produced their electricity. Similarly, suppliers cannot determine who is consuming their production. A market managed by an independent entity allows to exchange payments and ensure secure and reliable supply of and demand for electricity. In the remaining of this chapter we provide an overview of the current structure in Denmark.

2.2   Market Structure in Denmark

Denmark has been integrated into the Nordic Power exchange area since 2000, trading energy through the Nord Pool Spot market [2] since then. The Nord Pool Spot AS market is an organization that offers both day-ahead and intraday markets to its members, 380 companies from 20 countries. Nord Pool is owned by the Nordic transmission system operators from Norway (Statnett SF), Sweden (Svenska kraftnät), Finand (Fingrid Oy), Demark (Energinet.dk), and

the Baltic transmission system operators (Elering, Litgrid and Augstsprieguma tikls).

Denmark is split into two regions. The West Denmark grid area, also called DK1, covers the Jutland peninsula, Fyn island, and the rest of the islands to the west of the Great Belt. The East Denmark area, also called DK2, covers the island to the East of Fyn. A map illustrating the grid areas is shown in Figure 2.1

Depending on the time when the electricity is traded, one could distinguish between several electricity markets. The market with the longest trading horizon is the financial market, arranged by *Nasdaq Commodities* [4]. Financial contracts are used for price hedging and risk management, and there is no physical delivery for financial power market contracts, only cash settlement. Moreover, technical conditions such as grid congestion and access to capacity are not taken into consideration when entering financial contracts.

After the financial markets are closed, the next market to be cleared is the reserve capacity market, where players receive payments for having various types of reserve capacity available in face of contingency. This market is introduced below in Section 2.2.1. The next market to be cleared is the day-ahead market, also called *Elspot*, where physical power is traded on a daily basis. The day-ahead market is explained in Section 2.2.2. On a shorter horizon, at the intraday market or *Elbas* market, buyers and sellers can trade volumes closer to real time, in order to bring the market back in balance, as explained in Section 2.2.3. The regulating or *real-time* market, introduced in Section 2.2.4, closes 45 minutes before operational time and is the last market before the electricity is actually exchanged.

2.2.1   Reserve Capacity Market

The reserve capacity market guarantee that enough back-up generation is available in case of equipment failure, drastic fluctuations of production from intermittent sources and sudden demand changes [5]. In Denmark, the Transmission System Operator, namely Energinet.dk, is responsible for organizing reserve capacity markets. Energinet.dk pays the providers of the reserve services and recover the cost from the users trough taxes. Producers are paid for the availability of the energy, even though they might not be used at the operational time. At the market closure, the TSO collects bids from producers willing to provide reserve capacity, and selects them by a cost merit-order procedure

The reserve capacity market is settled independently of and before the day-

The Danish electricity transmission system
ultimo 2012

Legend

- 150 kV- or 132 kV-substation
- 220 kV-substation
- 400 kV-substation
- Power station
- Offshore wind farm

······ 150 kV- or 132 kV-cable
—— 150 kV- or 132 kV-overheadline
······ 220 kV-cable
—— 220 kV-overheadline
······ 400 kV-cable
—— 400 kV-overheadline
······ HVDC cable
—— HVDC overheadline

DK1

DK2

Norway
1,000 MW

Sweden
740 MW

Nordjyllandsværket
Vendsysselværket
Vester Hassing
Ferslev

Tjele

Idomlund

Trige
Studstrupværket

Malling

Askær

Landerupgård
Skærbækværket
Endrup
Esbjerg-
værket

Fynsværket
Kingstrup
Fraugde

Kassø
Enstedværket

Germany
1,200 MW

Hovegård
Glentegård
Svanemølleværket
Amagerværket
Ishøj
H. C. Ørsted Værket
Avedøreværket
København

Sweden
1,300/1,700 MW

Gørløse

Asnæsværket
Hørslev
Bjæverskov

Germany
600 MW

Sweden

60 kV
Østkraft

Figure 2.1: Map of Denmark. The West Denmark DK1 gird region is high-
lighted in orange while East Denmark or DK2 is colored in green.
The main transmission lines and power plants are included as well
[3]

ahead energy market, implying that, at the moment of scheduling reserves, no information about which units will be online is known. Currently, the reserve levels in Denmark are computed by deterministic rules such as allocating an amount of reserve equal to the capacity of the largest unit online, the so-called *n-1 rule* [6, 7]. In a system with high penetration of renewable intermittent production, such deterministic rules might not be the most optimal. This issue has been addressed in Article A of this thesis.

Depending on certain technical conditions, Energinet.dk buys several kinds of reserves: primary reserves, secondary reserves, manual reserves, black-start capability, short-circuit power, reactive reserves and voltage control reserves. All those types are commonly named *ancillary services*. The scope of this dissertation includes primary, secondary, and manual reserves. Their technical differences, as explained in [7], are outlined in the subsections below.

2.2.1.1 Primary Reserves

The primary reserve regulation ensures that the balance between production and consumption is restored after a deviation from the 50Hz of frequency. The first half of the reserve must be activated within 15 seconds, while the second half must be fully supplying within 30 seconds. The reserve must be supplied for maximum 15 minutes. Energinet.dk buys two types of primary reserve, upwards regulation power and downward regulation power, in case of under frequency or over frequency respectively. An auction is held once a day for the coming day of operation. Bids are sent before 15:00, stating an hour-by-hour volume and price having the 24-hour period divided into six equally sized blocks. In 2011 the quantity of the primary reserves sums up to +/- 27MW, having the option of buying +/-90MW from other European transmission system operator as well as from East Denmark area, the Nordic countries and Germany.

2.2.1.2 Secondary Reserves

The secondary reserve serves two proposes. One is to release the primary reserve which has been activated and the other is to restore any imbalances on the interconnections to follow the agreed plan. The requested energy must be supplied within 15 minutes and it can be supplied by a combination of unit in operation and fast-start units. It consists of upward and downward regulation that can be provided by several of production or consumption units. Energinet.dk currently buys approximately +/- 90MW on a monthly basis, based on a recommendation from the ENTSO-E RG Continental Europe organization.

2.2.1.3   Manual Reserves

The manual reserves are sometimes referred as tertiary reserves, and their purpose is to relieve the secondary reserve in the event of minor imbalances, ensuring that the demand is fulfilled in the event of outages, restrictions affecting productions plants and international connections. Manual reserves are typically less costly than secondary reserves. The reserve must be supplied in full within 15 minutes of activation, so often it is players with fast start units such as gas turbines or local CHP plants who can provide this reserve.

Energinet.dk activates the reserve by manually ordering upward and downward regulation to the suppliers. The method used to determine the requirements for reserve is known as the *n-1 rule*: setting the minimum amount of manual reserves to be the capacity of the largest online generator.

2.2.2   Day-ahead market

The day-ahead market, also known as *Elspot market*, allows Nordic market participants trade power contracts for next-day physical delivery. At 12:00 CET each day, bids for either purchase or sale are collected and at 12:42 CET, or later, the hourly prices are announced to the market participants. From 00:00 CET of the next day, power contracts are physically delivered and power is provided to the buyer.

Each participant submits volume bids to the market in MWh. There are four different types of bids, summarized as follows:

1. **Single hour order**. The member specifies the purchase and/or sales volume for each hour. The member can submit two different types of single-hour bid. One is a price independent order, where the member receives the specified volume for all hours. The other type of hourly bid is price dependent, and may consist of up to 62 price steps in addition to the current ceiling and floor price limits set by Nord Pool Spot. This is the most flexible product and, in fact, the largest share of the bids belong to this group.

2. **Block order**. It consists of a specified volume and price for at least three consecutive hours within the same day. These bids can be fully accepted, fully rejected or partially accepted, depending on the specified constraints from the supplier.
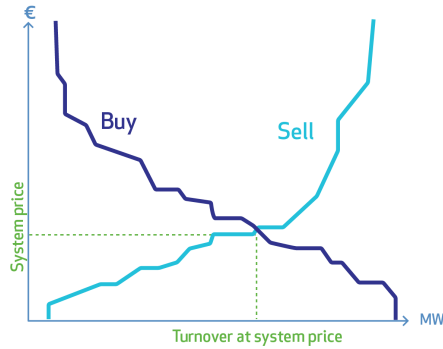
Figure 2.2: Aggregate supply and demand curves, taken from [2].

3. **Exclusive group**. It is a cluster of sell and/or buy blocks out of which only one block can be activated.

4. **Flexible order**. The user can define which energy volume they would be willing to sell or buy at a specified order price limit. In addition to the volume and the price limit, the duration and time interval within the delivery day must be specified.

At 12:00 CET, Nordpool collects all the bids and aggregates them into a supply and a demand curve. An example of these curves is depicted in Figure 2.2. The system price is calculated as the intersection between the supply and demand curve.

Whenever there is congestions in the transmission lines, the system price is not applied to all areas. In this case, some areas will sell more and purchase less, while others areas will do the opposite. The different prices are calculated by aggregating the curves in each areas and hence obtaining a new equilibrium in each one.

2.2.3 Elbas

Elbas is a continuous market where power trading takes place until one hour before the power is delivered. Members submit bids stating how much power they want to sell and buy and at what price. Trading is then set based on a first-come, first-served basis between a seller and a buyer. West Denmark joined this market in 2008. Since energy already traded on the Elspot market is higher

prioritized than energy traded on the Elbas market, transactions between areas where transmission capacities are already fully utilized are not allowed.

This market is specially interesting for stochastic producers, as wind power producers, and for retailers of aggregate demand response unit, because the trading horizon is shorter than for the day-ahead market. The shorter horizon allows players to use more accurate information about the actual volume compare to what they know one day in advance.

2.2.4   Balancing Market

The Balancing market, sometimes referred as *real-time market*, is cleared up to 45 minutes prior to the upcoming delivery hour [8]. It allows producers and consumer to alter their plans with a very short notice. The balancing market is divided in two submarkets: the regulating power market, and the balancing power market. Regulating power is bought or sold by Energinet.dk, according to the regulating bids submitted by the players. The balancing market happens after the delivery hour, and imbalances have been quantified. Then, Energinet.dk buys/sells balancing power to neutralize imbalance incurred by the players.

In the regulating market, we distinguish between two types of bids. Up-regulating bids refer to the increase of production or, similarly, decrease of consumption. Down-regulating bids, on the other hand, refer to the decrease of production or the increase of consumption. Player must be able to fully activate a given bid in maximum 15 minutes from receipt of the activation order.

The price in the regulating market is determined according to the day-ahead price, in a way that buying/selling regulating volume is always more expensive/cheaper than doing so in day-ahead price. This is commonly known as *two-price market*, and it is designed in order to encourage scheduling one's production or consumption as much in advance as possible. Other regulating markets around the world, as the one organized by the California Independent System Operator [9], are *one-price* market allowing for players to arbitrage. This case has been studied in Paper C of this thesis.

2.3   Demand Response

The way the power system is operated traditionally can shortly be characterized as consumption-based production. This means that the production is adjusted to the needs of the consumers, so that that power plants increase or decrease their production according to the load. However, in a future scenario where the share of electricity produced by non-dispatchable renewable resources increase drastically, the traditional consumption-based production will be much more costly to achieve, as more fast-start power plants will be needed to counteract the imbalances of the wind and solar producers. An alternative solution is to switch the strategy from a consumption-based production to a production-based consumption. This commonly called *Demand-Side Management* (DSM) or *Demand Response*. Under the demand-response setup, consumers, together with small producing units, are called *Distributed Energy Resources* (DER).

Demand response can be defined as the intentional alteration of the power consumption profile by an end-user in response to an external stimulus [10]. Often, it is the TSO, or another market entity, that manages the stimulus in order to achieve a desired goal. In general, the market entity that aggregates the DERs is called *aggregator* and its goal could be, for example, to shift, curtail, or defer load.

Depending on the communication flow between aggregator and the DERs, we differentiate between two ways of managing a portfolio of units:

- **Direct control**. The aggregator sends the DERs a consumption schedule and expects the DERs to follow it. In return, the DER sends their consumption plan. On-line consumption data is observed. The consumer is rewarded economically when participating in such demand response programs.

- **Indirect control**. The aggregator decides on a price to bill its portfolio of DERs and communicates it to each of them in advance or online. Afterwards, each DER decides on how much to consume and when depending on the billing price and also on their own preferences.

Papers B, C and D of this thesis deal with the latter case. These papers tackle some of the problems that arise by the fact that DER are *flexible*, meaning that they are capable of modifying their consumption depending on the price of electricity. The flexibility of a DER is understood as its willingness to shift their consumption from high priced hours to lower priced hours.

Generally, the flexibility of each DER is too small to provide a service to the electric power system. However, by aggregating several DERs, it is possible
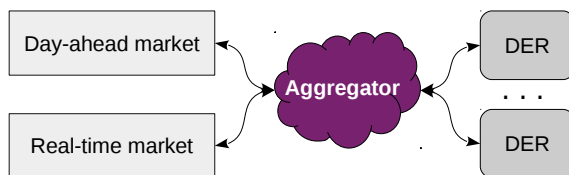
Figure 2.3: Representation of the role of the aggregator as a link between markets and DERs.

to reach volumes large enough for bidding into the markets. This concept is commonly known as Virtual Power Plant (VPP), as it enables small DERs to provide grid services similar to conventional power plants [11]. The aggregator must then interact with the electricity markets on behalf of the clients in order to provide for energy. As a summary, the role of the aggregator is depicted in Figure 2.3. On the one hand, the aggregator buys and sells electricity from the markets. On the other, it sells or buy from the DERs. Depending on the characteristics of the portfolio of DERs and the control strategy, the aggregator can interact with one or many markets.

The work presented in this thesis closely relates to the decision-making process of an aggregator of price-responsive units. Papers B and C deal with the bidding problem of an aggregator, while Paper D deal with forecasting the aggregate load of a price-response pool of DERs.

Mathematical Foundations

In this chapter we give an introduction to the mathematical models used to solve the challenging applications presented in Chapter 4.

The mathematical tools are divided into two main categories. In the first category, explained in Section 3.1, we present some of the time-series related models used for this dissertation. The second category of mathematical tools relate to operations research and they are introduced in Section 3.2. We elaborate on linear problems, optimality conditions, bilevel programming and finally on inverse optimization.

3.1 Time Series Related Models

Time series analysis comprises methods for analyzing data that are sampled along time, and the goal is to infer information about its nature. Time series forecasting aims to predict future values of a process, based on historical information from the process of interest and possibly other processes related to it.

Let us first denote some of the basic elements of a time series model. To begin with, let $\{Y_t\}$ denote the stochastic process that we aim to model, and Y_t the

random variable associated to such process at time t. Often, Y_t is called *dependent variable*. Here we consider the case where $\{Y_t\}$ is an univariate process for brevity. The extension for multivariate processes is analogue and the reader is referred to [12, Ch. 9] for a detailed explanation on that matter. Let us define also a set of r stochastic processes, called X_r, and their associated random variables $X_{r,t}$, i.e. a random vector indexed by the time t, to be the explanatory or *independent* variables. Independent variables are sometimes called *regressors* or *exogenous inputs*. A time series model aim to explain the characteristics of the dependent variable Y_t by the use of past values of Y_t, together with information of the independent variable X_t. The relationship between Y_t and X_t is modeled trough a set of unknown parameters that must be estimated using historical data.

In the following we present two time series model. In Section 3.1.1 we introduce a parametric model based on linear relationship between the dependent and independent variables. Afterwards, in Section 3.1.2, we present a non-parametric approach to estimate conditional densities. Finally, in Section 3.1.3, we combine both time series models and elaborate on a methodology for generating scenarios of future realizations of a stochastic process.

3.1.1   ARIMAX Models

In this section we introduce the Auto Regressive Integrated Moving Average model with Exogenous variables (ARIMAX). This model represents a stochastic process and ultimately forecast its future realizations. Generally speaking, ARIMAX models are composed of four elements. The first group alludes to lagged terms of the series, the co-called "auto regressive" terms. The "integrated" label comes from the fact that the series is differentiated to become stationary. The third element refer to the lags of the forecast errors, or "moving average" terms. The fourth and last indicates that external regressors or "exogenous variables" are used to explain some part of the variability of the series of interest.

In mathematical terms, the time varying process Y_t is an ARIMA process if described as [12]

$$Y_t + \sum_{p=1}^{P} \theta_p Y_{t-p} = \mu + \epsilon_t + \sum_{q=1}^{Q} \psi_q \epsilon_{t-q} + \sum_{r=1}^{R} \gamma_r X_{r,t}. \tag{3.1}$$

The symbols in Equation (3.1) have the following interpretation:

P Number of autoregressive terms.
Q Number of moving average terms.
R Number of regressors.
μ Intercept term.
ϵ_t Gaussian white noise error at time t.
$X_{r,t}$ Regressor r at time t.
θ_p Coefficient relative to the lagged observation y_{t-p}.
ψ_q Coefficient relative to the lagged error term at time $t - q$.
γ_r Coefficient relative to the independent variable $x_{r,t}$.

The parameters P and Q define the order of the $ARIMA$ model. The intercept μ and the coefficients $\boldsymbol{\theta}, \boldsymbol{\psi}$ and $\boldsymbol{\gamma}$ are commonly estimated from data by minimizing the Sum of Square Error (SSE), written as follows:

$$\text{SSE}(\mu, \boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\gamma}) = \sum_{t=1}^{T} (y_t - \widehat{y}_{t|t-1})^2 \tag{3.2}$$

where $\widehat{y}_{t|t-1}$ is the estimated value of the dependent variable, given as the outcome from (3.1) using $\mu, \boldsymbol{\theta}, \boldsymbol{\psi}$ and $\boldsymbol{\gamma}$ as coefficients. Also, when computing $\widehat{y}_{t|t-1}$ from (3.1), all the information up to time $t - 1$ is known.

An integrated ARMAX model, also called ARIMAX, is a generalization of (3.1), where the output variable is differenced in the following way:

$$z_t = (1 - B)^d y_t \tag{3.3}$$

where B is the lag or backshift operator. In the special case where $d = 1$, equation (3.3) is equivalent to transforming the output variable y_t by subtracting it with the previous value such that $y'_t = y_t - y_{t-1}$. Such transformation if often used in practice in order to turn a non-stationary variable into a stationary one.

The ARIMAX model presented above can be extended to account for seasonal patterns, as daily or weekly patterns. The explanation is left out of this dissertation for brevity but the reader is referred to [12] for a detailed description.

In Paper B and D we use an ARIMAX model to model the aggregate load of a pool of households, and predict their response with respect to weather variables and price of electricity. Furthermore, in Paper C we used an ARIMA model to model the price of electricity in the day-ahead market.

3.1.2 Kernel Density Estimation

Kernel density estimation is a non-parametric method for estimating the probability density function of a random variable. Let $(x_1, x_1, \cdots, x_n)$ be an independent and identically distributed sample drawn from some distribution with an unknown density f. The kernel density estimation of f, evaluated at x, is given by

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_i) \tag{3.4}$$

where $K_h(x)$ is a weight function, often called *kernel*, and $h > 0$ is a smoothing parameter, often called *bandwidth*. The kernel function $K_h(x)$ must be a symmetric non-negative function that integrates to one. An example of a kernel function is the uniform kernel, where $K_h(x) = 1/h$ if $a \leq x \leq a + h$, and $K_h(x) = 0$ otherwise. Another commonly used kernel is the Gaussian kernel, where $K_h(x)$ is equal to the standard normal density function with standard deviation of h. In general, the bandwidth of the kernel is chosen by minimizing a criteria, for example, the mean integrated square error. Often, a more suited approach to real-life problem involves performing cross-validation. The reader is referred to [13] and [14] for further information about optimal choice of the bandwidth.

Next, we use a slightly modified version of the kernel density estimator that has been used in Paper C of this thesis. Here, each of the observations $(x_1, x_1, \cdots, x_n)$ has an assigned weight, denoted by $\boldsymbol{w} = (w_1, w_1, \cdots, w_n)$, such that $\sum_{i=1}^{n} w_i = 1$. Then, the kernel density estimation writes as follows:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} w_i K_h(x - x_i). \tag{3.5}$$

This modification of the kernel estimator allows to give different weight or "importance" to observations. Indeed, the weights $\boldsymbol{w}$ can potentially be calculated by a kernel. In this case, this approach could be seen as a two-step kernel density estimation. This approach to estimate the density function has a very powerful application, shown in Paper C of this thesis, that is to model the load, conditioned to a price reference. In such a case, the weights allows us to give greater importance to observations with a similar price.

3.1.3   Generation of Scenarios

In this section we elaborate on a methodology to generate scenarios from a stochastic process, conditioned on another stochastic process. Scenarios are often used in stochastic optimization problems due to the fact that they represent the development of a stochastic variable along time. Moreover, a collection of scenarios must be consistent with the innovation errors of the series. In Paper C of this thesis we use cross-correlated scenarios of load and price as input for a stochastic optimization problem.

In brief, the proposed method starts by transforming the input data to a normal distribution using a non-parametric transformation. Then, we compute its covariance, and finally, generate random correlated Gaussian errors that are transformed back to the original distribution. The estimated multivariate non-parametric distribution is conditioned to a given value of another random variable.

The advantage of the proposed approach is threefold. First, we do not make any assumption on the distribution of the data we model, due to the non-parametric nature of the kernel density estimation. Second, we do not make assumption on the relationship between the variable of interest, and the one we condition on. This means, the response of the variable of interest to the given one can be non-linear. Finally, the proposed approach is computationally attractive and big datasets can be quickly processed.

For illustrative purposes, below we show an example of a scenario generation algorithm, which outcome is a scenario of load, for every hour of the day, conditioned to a price. The example below relates to the contribution given in Paper C of this dissertation.

Example 1 (Generation of price-responsive load scenarios)   Let $x_t^{(j)}$ be aggregate measured load of a price-responsive pool of loads, where $j = 1, \ldots, J$ is the index relative to the considered day, and $t = 1, \ldots, 24$ the index relative to the hour of the day. Let us assume also that, together with the measured load, we have a record of the price of electricity that the load paid for, denoted by $\pi_t^{(j)}$. The problem consists on generating scenarios of load, conditioned to a given price reference $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_{24})$.

The first step is to weight the historical observations, so that the days when the price resemble to the given one $\tilde{\boldsymbol{\pi}}$ weight more. Denote the weights by $\boldsymbol{w} = (w^{(1)}, \ldots, w^{(J)})$, with one weight for each day, and such that $\sum_{j=1}^{J} w^{(j)} = 1$. The reader is referred to Paper C for a practical approach to compute such

weights.

The scenario-generation procedure consists of the following seven steps:

1. For each hour of the day, we compute a non-parametric estimation of the density of the price-responsive load conditional on a price trajectory $\tilde{\pi}$. We achieve this by computing the kernel density estimator at hour t, with the weights $w^{(j)}$, in an analogue way as in (3.5):

$$\hat{f}_t(x|\tilde{\pi}) = \frac{1}{J} \sum_{j=1}^{J} w^{(j)} K_h(x - x_t^{(j)}), \qquad (3.6)$$

where $K_h(x)$ is a kernel (non-negative function that integrates to one and has mean zero), h is its bandwidth, and $x_t^{(j)}$ is the observed load at time t and day j.

2. Using $\hat{f}_t(x|\tilde{\pi})$ from Step 1, we compute the cumulative density function, called $\hat{F}_t(x|\tilde{\pi})$.

3. The transformed load values $y_t^{(j)} = \hat{F}_t(x_t^{(j)}|\tilde{\pi})$, for every hour t, follow a uniform distribution $U(0,1)$. Then, we normalize the load data through the transformation $z_t^{(j)} = \Phi^{-1}(y_t^{(j)})$, where $\Phi^{-1}(Y)$ is the probit function. Consequently, $(z_t^{(1)}, \ldots, z_t^{(J)}) \equiv Z_t \sim N(0,1)$.

4. We estimate the variance-covariance matrix Σ of the transformed load Z, relative to the 24 hours of the day. One could do it recursively as in [15].

5. Using a multivariate Gaussian random number generator, we generate a realization of the Gaussian distribution $\tilde{Z} \sim N(0, \Sigma)$.

6. We use the inverse probit function to transform $\tilde{Z}$ to a uniform distribution, that is, $\tilde{Y} = \Phi(\tilde{Z})$.

7. Finally, we obtain a scenario of load by transforming back $\tilde{Y}$ using the inverse cumulative density function from Step 2, that is, $\tilde{x}_t = \hat{F}_t^{-1}(\tilde{Y}_t|\tilde{\pi}), \forall t$. Numerically, we use a smoothing spline to interpolate $\hat{F}_t^{-1}(\tilde{Y}_t|\tilde{\pi})$. The scenario is finally given by $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_{24})$.

The procedure outlined above generates a scenario of load conditioned on the price $\tilde{\pi}$. Steps 5 to 7 are repeated as needed if more scenarios of load per price are desired. An example of the generated scenarios of day-ahead price is given in Figure 3.1. By graphical inspection we conclude that the scenarios of day-ahead price are a plausible representation of the actual day-ahead price and its uncertainty.
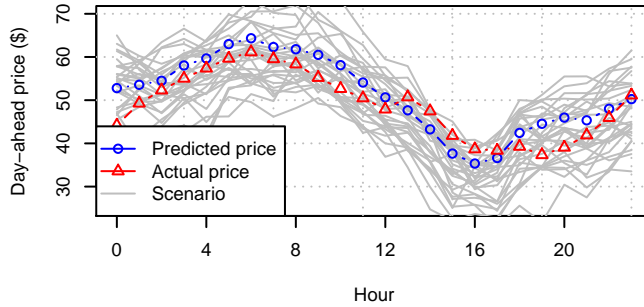
Figure 3.1: Actual price, point forecast and generated scenarios for the day-
ahead price.

3.2 Operations Research Fundamentals

In this section we present a brief introduction to some topics from the field of operations research. We start by formulating one of the most fundamental models: linear problems and their corresponding dual problem. Then, we elaborate on bilevel programming and Mathematical Problems with Equilibrium Constraints (MPECs), which is a widely used family of models in energy systems. Finally, we introduce the concept of inverse optimization.

3.2.1 Linear Programming and Duality

Linear programming consists on choosing the values of a decision variable that maximizes a linear function subject to some linear constraints. In mathematical notation, a linear problem is written as follows:

$$\underset{\boldsymbol{x}}{\text{Maximize }} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \tag{3.7a}$$

$$\text{subject to } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \tag{3.7b}$$

where $\boldsymbol{x}$ represents the decision variables, $\boldsymbol{c}$ the cost vector, $\boldsymbol{A}$ the matrix of coefficients that define the left-hand side of the constraints, and $\boldsymbol{b}$ the vector defining the right-hand side of the constraints.

The solution to Problem (3.7) is often denoted as $\boldsymbol{x}^*$, and can be obtained by, for example, using the *simplex* method [16] or by *interior-point* methods [17]. From the implementation perspective, one can solve this type of problems by using computer software, as for example, GAMS [18] and CPLEX [19].

The *dual* problem of (3.7) is defined as follows:

$$\underset{\boldsymbol{\lambda}}{\text{Minimize}} \; \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} \tag{3.8a}$$

$$\text{subject to } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{c} \tag{3.8b}$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0}. \tag{3.8c}$$

Each dual variable, denoted in each element of $\boldsymbol{\lambda}$, is relative to a constraint from the linear problem (3.7), also referred as *primal* problem. Dual variables can be interpreted as *shadow price* or *marginal cost*. In other words, dual variables represent the change in the objective function of the primal problem (3.7a) when the right-hand side of the constraint (3.7b) is marginally increased.

There are two main theorems that relate the solution of the primal and the dual problems. The **weak duality theorem** states that the objective function value of the dual at any feasible solution is always greater than or equal to the objective function value of the primal at any feasible solution. The **strong duality theorem** states that if the primal has an optimal solution, then the dual also has also an optimal solution and the value of the objective functions from the primal and the dual are the same. The reader is referred to [20] for rigorous proofs of the theorems and further properties.

As a consequence of the weak and strong duality theorems, the following is also satisfied: if the dual problem is unbounded, then the primal is infeasible. Similarly, if the primal is unbounded, the dual must be infeasible.

3.2.2   Mathematical Programs with Equilibrium Constraints

Mathematical Programs with Equilibrium Constraints (MPEC) are a type of optimization problems with nested structure, i.e., there is a an optimization problem inside another optimization problem. The inner problem is sometimes referred as *lower-level problem*, and the outer problem as *upper-level problem*.

In the following, we present an example of an MPEC where both the upper and lower level problems are linear:

$$\underset{\boldsymbol{x},\boldsymbol{y}}{\text{Minimize}} \; \boldsymbol{c}_x^{u\mathrm{T}}\boldsymbol{x} + \boldsymbol{c}_y^{u\mathrm{T}}\boldsymbol{y} \tag{3.9a}$$

$$\text{subject to } \boldsymbol{A}_x^u \boldsymbol{x} \leq \boldsymbol{b}_x^u \tag{3.9b}$$

$$\boldsymbol{A}_y^u \boldsymbol{y} \leq \boldsymbol{b}_y^u \tag{3.9c}$$

$$\boldsymbol{x} \in \left\{ \begin{array}{ll} \underset{\boldsymbol{z}}{\text{Maximize}} & \boldsymbol{c}_x^{\mathrm{T}}(\boldsymbol{y})\boldsymbol{z} \\ & \boldsymbol{A}_x \boldsymbol{z} \leq \boldsymbol{b}_x(\boldsymbol{y}) \end{array} \right\}. \tag{3.9d}$$

The formulation above includes two optimization problems. The lower-level problem relates to the lower-level decision variable $\boldsymbol{x}$. The coefficients of the objective function of the lower-level problem are $\boldsymbol{c}_x(\boldsymbol{y})$, and the right-hand side of the constraints are $\boldsymbol{b}_x(\boldsymbol{y})$. In this example, both $\boldsymbol{c}_x(\boldsymbol{y})$ and $\boldsymbol{b}_x(\boldsymbol{y})$ depend on the upper-level variable $\boldsymbol{y}$. However, from the perspective of the lower-level problem, they are parameters. The matrix $\boldsymbol{A}_x$ is a given parameter.

In the example above, the upper-level problem minimizes a weighted sum of $\boldsymbol{x}$ and $\boldsymbol{y}$, subject to the constraints defined by the coefficient matrices $\boldsymbol{A}_x^u$ and $\boldsymbol{A}_y^u$, and the right-hand-side parameters $\boldsymbol{b}_x^u$ and $\boldsymbol{b}_x^u$. Note that the upper and lower level problems are interrelated, due to the fact that the constraints and the objective function of the upper-level problem depends on the lower-level variables. In addition, the constraints and the objective function of the lower-level problem, depend on the upper-level variables.

Problem (3.9) relates to two of the challenges tackled in this thesis. The models formulated in Paper B and Paper D are similar to Problem (3.9). In the remaining of this section, we elaborate on properties of such models and solution techniques.

In order to solve (3.9), the lower-level problem needs to be reformulated. In the following, we present two different reformulations for the lower-level problem (3.9d). The first one, based on the The Karush-Kuhn-Tucker (KKT) conditions, is related to Paper B in this dissertation. The second reformulation, based on the primal-dual properties of the lower-level problem, has been used in Paper D.

3.2.3 Reformulations Based on the Karush-Kuhn-Tucker Conditions

Consider the lower-level problem from (3.9). Recall that $\boldsymbol{x}$ (or $\boldsymbol{z}$) is the lower-level variable and both the coefficient of the objective function $\boldsymbol{c}_x(\boldsymbol{y})$ and the right-hand-side vector $\boldsymbol{b}_x(\boldsymbol{y})$ depend on the upper-level variable $\boldsymbol{y}$. From the point of view of the lower-level problem, $\boldsymbol{y}$ is a parameter. In other words, the upper-level variable enters the lower-level problem as a parameter and should treated as such. For this reason, the lower-level problem (3.9d) is analogue to the basic linear problem (3.7).

For notational simplicity, let us use the notation from the generic linear program (3.7) instead of the lower-level problem (3.9d). Its Lagrangian is defined as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}). \tag{3.10}$$

The Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality for problem (3.7), meaning that the solution to the linear problem (3.7) must satisfy the following:

$$c - A^{\mathrm{T}}\lambda = 0 \tag{3.11a}$$
$$Ax - b \leq 0 \tag{3.11b}$$
$$\lambda \geq 0 \tag{3.11c}$$
$$\lambda^{\mathrm{T}}(Ax - b) = 0. \tag{3.11d}$$

Equations (3.11a) are the stationarity conditions, equations (3.11a) are the primal feasibility constraints, equations (3.11c) are the dual feasibility constraints, and finally equation (3.11d) are the complementarity constraints.

The notation for constraints (3.11b), (3.11c) and (3.11d) can be compacted by using the *perpendicular* sign "$\perp$", which indicates perpendicularity between the vectors on the right side and on the left side of $\perp$. The constraints can be denoted as

$$Ax - b \leq 0 \perp \lambda \geq 0. \tag{3.12}$$

As a final comment, note that in the KKT conditions stated above, equations (3.11) are non linear due to the complementarity constraint (3.11d).

3.2.4   Reformulation Based on the Primal-dual Properties

Here we describe the primal-dual reformulation of the linear problem (3.7), that is analogue to the lower-level problem (3.9d). The solution of the primal and dual problems can be found by solving the following set of equations:

$$c^{\mathrm{T}}x = b^{\mathrm{T}}\lambda \tag{3.13a}$$
$$Ax - b \leq 0 \tag{3.13b}$$
$$A^{\mathrm{T}}\lambda = c \tag{3.13c}$$
$$\lambda \geq 0. \tag{3.13d}$$

Equation (3.13a) is the strong duality condition, derived from the strong duality theorem introduced in Section 3.2.1. To put it more simply, it enforces the primal and the dual objective functions to be equal. Equation (3.13b) refers to the constraints of the primal problem, and Equations (3.13c) and (3.13d) are the constraints of the dual problem.

The primal-dual reformulation of the lower-level problem in (3.9) has a computationally advantage over the KKT conditions, due to the fact that for linear problems, the primal-dual formulation is linear.

Consider now the MPEC problem defined by (3.9), where the lower-level problem has been reformulated by using the primal-dual properties (3.13). The MPEC now recast as a single-level non-linear problem. The reason is because of the multiplication of $\boldsymbol{c}_x(\boldsymbol{y})$ and $\boldsymbol{b}_x(\boldsymbol{y})$ in the strong-duality constraint. Recall that the coefficient of the objective function $\boldsymbol{c}_x(\boldsymbol{y})$ and the right-hand-side vector $\boldsymbol{b}_x(\boldsymbol{y})$ depend on the upper-level variable $\boldsymbol{y}$.

3.2.5   Static Inverse Optimization

An inverse optimization problem is a special case of a MPEC. Generally, it consists on inferring the value of the model parameters, i.e., the cost coefficients, the right-hand side vector, and the constraint matrix, given the optimal decision variables. In mathematical terms, an inverse optimization problems is defined as follows. Consider the linear problem (3.7), re-defined below

$$\underset{\boldsymbol{x}}{\text{Maximize }} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \tag{3.14a}$$

$$\text{subject to } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \tag{3.14b}$$

and let $\boldsymbol{x}'$ be an optimal solution. The inverse optimization problem consists on finding $\boldsymbol{c}$, $\boldsymbol{b}$ and $\boldsymbol{A}$, such that $\boldsymbol{x}'$ is a solution to (3.14). The linear problem above is sometimes referred as *forward problem* or *reconstruction problem*. It should be noted that in this thesis we deal with inverse linear problems, meaning that the reconstruction problems are linear programming problems.

Inverse optimization was first formally described by [21]. Their application focused on finding the minimum perturbation of the cost coefficient, such that the optimal decision variables are optimal. Recent work [22, 23, 24, 25, 26, 27, 28, 29, 28], including Papers B and D in this dissertation, have addressed inverse optimization from a broad range of perspectives. The work performed in this thesis differs with existing literature in two main aspects. First, we consider the right-hand side of the constraints of the reconstruction problem to be unknown. And second, we apply inverse optimization to time-series data by using external regressors, hence making the inverse models *dynamic*. The *static* formulation is introduced in this section, while the dynamic formulation is presented below in Section 3.2.6.

We divided the inverse optimization models into two categories, depending on the type of objective function that quantifies the "optimality" of $\boldsymbol{x}'$. In the first

category, the objective function consists of minimizing the distance between the input optimal decision variables and the optimal value resulting from the reconstruction problem. In this case, because the reconstructed value of x is found by solving an optimization problem itself, the problem is a bilevel problem. This is explained in detail in Section 3.2.5.1. In the second category, optimality is quantified by minimizing a duality gap. The inverse optimization based on minimization of a duality gap is introduced in Section 3.2.5.2

3.2.5.1   Minimizing a Norm

We start by showing an example of an inverse optimization problem, where the aim is to minimize the distance, measured by the L^1-norm $|| \cdot ||$, between the input $\boldsymbol{x}'$ and the variable $\boldsymbol{x}$. It belongs to the group of MPEC problem, and writes as follows:

$$\underset{\boldsymbol{x},\boldsymbol{\theta}}{\text{Minimize}} \ ||\boldsymbol{x}' - \boldsymbol{x}|| \tag{3.15a}$$

$$\text{subject to } \boldsymbol{D}^{\theta}\boldsymbol{\theta} \leq \boldsymbol{d}^{\theta} \tag{3.15b}$$

$$\boldsymbol{x} \in \left\{ \begin{array}{ll} \underset{\boldsymbol{z}}{\text{Maximize}} & \boldsymbol{c}^{\text{T}}\boldsymbol{z} \\ & \boldsymbol{A}\boldsymbol{z} \leq \boldsymbol{b} \end{array} \right\}. \tag{3.15c}$$

The variables that define the lower level problem are denoted, for brevity, as $\boldsymbol{\theta} = [\boldsymbol{c}, \boldsymbol{b}]$. In this example, we assume $\boldsymbol{A}$ to be know, even though it could potentially be estimated as well, i.e., coming from a model. Equations (3.15b) define some constraints on $\boldsymbol{\theta}$, that, for example, might be given by their nature. A typical example is the positiveness of the parameter, defined by $\boldsymbol{\theta} \geq 0$. The variable $\boldsymbol{x}$ is defined by the lower-level problem (3.15c), and its value is dependent on $\boldsymbol{\theta}$. Due to the bilevel nature of the inverse optimization problem (3.15), it provides with an estimate of the reconstructed input $\boldsymbol{x}$.

In this example, the objective function (3.15a) consists on minimizing a distance. If the objective function at the optimum is equal to zero, it implies that the input $\boldsymbol{x}'$ can be reconstructed perfectly. On the other hand, if the objective function is greater than zero, it means there does not exist a $\boldsymbol{\theta}$ such that $\boldsymbol{x}' = \boldsymbol{x}$. In other words, the input $\boldsymbol{x}'$ does not belong to any of the possible optimality regions of the lower-level problem (3.15c).

It is noteworthy to say that other types of norm could be used instead of the L^1-norm, depending on the nature of the problem of interest. The advantage of minimizing the L^1-norm with respect to other norms, is that it can be reformulated as a linear problem by introducing two auxiliary extra positive linear

variables. In mathematical terms, the absolute value is reformulated as follows:

$$\underset{\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-}{\text{Minimize}} \; \mathbf{1}^{\text{T}}(\boldsymbol{\alpha}^+ + \boldsymbol{\alpha}^-) \tag{3.16a}$$

$$\text{subject to} \; \boldsymbol{x}' - \boldsymbol{x} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^- \tag{3.16b}$$

$$\boldsymbol{D}^\theta \boldsymbol{\theta} \le \boldsymbol{d}^\theta \tag{3.16c}$$

$$\boldsymbol{x} \in \left\{ \begin{array}{cc} \underset{\boldsymbol{z}}{\text{Maximize}} & \boldsymbol{c}^{\text{T}}\boldsymbol{z} \\ & \boldsymbol{Az} \le \boldsymbol{b} \end{array} \right\} \tag{3.16d}$$

where $\mathbf{1}$ is a vector of ones of appropriate size.

The bilevel problem (3.16) recasts as a single-level optimization problem by replacing the lower-level problem with its optimality conditions. This reformulation allows us to solve the problem using commercial software. In general, the solution of the reformulated problem is nonlinear and non-convex. Below we present two possible reformulations, based on KKT conditions and on primal-dual properties.

Reformulation based on KKT conditions   The KKT conditions, that were previously introduced in Section 3.2.3, are used to reformulate the lower-level problem in (3.16) as follows:

$$\underset{\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\lambda},\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-}{\text{Minimize}} \; \mathbf{1}^{\text{T}}(\boldsymbol{\alpha}^+ + \boldsymbol{\alpha}^-) \tag{3.17a}$$

$$\text{subject to} \; \boldsymbol{x}' - \boldsymbol{x} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^- \tag{3.17b}$$

$$\boldsymbol{D}^\theta \boldsymbol{\theta} \le \boldsymbol{d}^\theta \tag{3.17c}$$

$$\boldsymbol{c} - \boldsymbol{A}^{\text{T}}\boldsymbol{\lambda} = \mathbf{0} \tag{3.17d}$$

$$\boldsymbol{Ax} - \boldsymbol{b} \le 0 \perp \boldsymbol{\lambda} \ge 0 \tag{3.17e}$$

Note that the problem above is non-linear due to the complementarity constraint (3.17e). For practical applications, the complementarity constraints (3.17e) are reformulated by including an extra vector of binary variables, denoted as $\boldsymbol{w}$, and two large constants terms M_1 and M_2. The so-called "*big-M*" reformulation writes as follows:

$$\boldsymbol{Ax} - \boldsymbol{b} \le \mathbf{0} \tag{3.18a}$$

$$\boldsymbol{\lambda} \ge \mathbf{0} \tag{3.18b}$$

$$\boldsymbol{Ax} - \boldsymbol{b} \ge -M_1\boldsymbol{w} \tag{3.18c}$$

$$\boldsymbol{\lambda} \le M_2(\mathbf{1} - \boldsymbol{w}) \tag{3.18d}$$

$$w \in \{0,1\} \quad \forall w \tag{3.18e}$$

Using the reformulation above, the inverse optimization problem (3.17) recast as a single-level Mixed-Integer Linear Program (MILP). Commercial software can be used to solve this type of problems.

The reformulation of the inverse optimization problem based on the KKT conditions has been studied in Paper B of this dissertation. Due to the large size of the problem, commercial software was not capable of finding a solution in a reasonable amount of time. For this reason, we developed a solution method, based on a relaxation of the KKT conditions. The relaxed inverse model is the following:

$$\underset{\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{w},\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-}{\text{Minimize}} \quad \mathbf{1}^{\mathrm{T}}(\boldsymbol{\alpha}^+ + \boldsymbol{\alpha}^- + L_1(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) + L_2\boldsymbol{\lambda}) \tag{3.19a}$$

$$\text{subject to} \quad \boldsymbol{x}' - \boldsymbol{x} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^- \tag{3.19b}$$

$$\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0} \tag{3.19c}$$

$$\boldsymbol{D}^{\theta}\boldsymbol{\theta} \leq \boldsymbol{d}^{\theta} \tag{3.19d}$$

where L_1 and L_2 are penalty parameters. The solution to (3.19) is not necessarily the same as for the original problem (3.17). However, our proposed approach leverages data to calculate the penalty parameters L_1 and L_2 in an optimal way. Further analysis on this can be found in Paper B of this dissertation.

Reformulation based on primal-dual properties Here we present a reformulation of the inverse optimization problem using the primal-dual properties introduced in Section 3.2.4. The reformulation is a non-linear problem and writes as follows:

$$\underset{\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-}{\text{Minimize}} \quad \mathbf{1}^{\mathrm{T}}(\boldsymbol{\alpha}^+ + \boldsymbol{\alpha}^-) \tag{3.20a}$$

$$\text{subject to} \quad \boldsymbol{x}' - \boldsymbol{x} = \boldsymbol{\alpha}^+ - \boldsymbol{\alpha}^- \tag{3.20b}$$

$$\boldsymbol{D}^{\theta}\boldsymbol{\theta} \leq \boldsymbol{d}^{\theta} \tag{3.20c}$$

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} \tag{3.20d}$$

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \mathbf{0} \tag{3.20e}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{c} \tag{3.20f}$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \tag{3.20g}$$

Constraints (3.20d), (3.20e), (3.20f) and (3.20g) are the primal-dual optimality conditions of the lower-level problem (3.15c). Moreover, constraint (3.20c) provides some a priori known constraints on $\boldsymbol{\theta}$. The norm-minimizing inverse optimization problem, based on reformulating the lower-level problem with its primal-dual properties, is non-linear due to the multiplications of two variables in the strong duality condition (3.20d).

3.2.5.2   Minimizing Duality Gap

In this section, we present an inverse optimization model where the measure of optimality of the input x' is a duality gap.

Below, we present an example that relates to Papers B and D in this thesis. For simplicity, in the example we consider the reconstruction problem (3.7), with the matrix of coefficients A and the right-hand side vector b to be known. Hence, the coefficient c is the only "unknown parameter" that needs to be estimated by the inverse model.

The objective function of the example is to minimize the duality gap of the reconstruction problem:

$$\underset{c,\epsilon,\lambda}{\text{Minimize}} \ \epsilon \tag{3.21a}$$

$$\text{subject to } c^{\mathrm{T}} x' + \epsilon = b^{\mathrm{T}} \lambda \tag{3.21b}$$

$$A^{\mathrm{T}} \lambda = c \tag{3.21c}$$

$$\lambda \geq 0. \tag{3.21d}$$

Equation (3.21b) corresponds to the relaxed strong duality conditions from the reconstruction problem (3.7), and equations (3.21c) and (3.21d) are its dual feasibility constraints.

Note that, in the example, we assume A and b to be known. For this reason, the primal constraints are not included in Problem (3.21). It should be noted that the techniques developed in Paper D of this dissertation provide a computationally attractive method to obtain b for the cases when is unknown.

Let us denote the solution to Problem (3.21) by c^* ad ϵ^*. If $\epsilon^* = 0$, the input x' can be obtained by solving the reconstruction problem (3.7) with c^*. If, on the other hand, $\epsilon^* > 0$, the input x' cannot be obtained by solving the reconstruction problem (3.7) with c^*.

The outcome of the primal-dual-based inverse optimization does not provide for an estimate of the variable x. However, it can be easily obtain a posteriori by solving the reconstruction problem (3.7) with c^* as cost coefficient.

3.2.6   Dynamic Inverse Optimization

In this section, we show an example of an inverse optimization model that accounts for time-dependent changes. In other words, the model is *dynamic*. This type of inverse-optimization models has been used in Paper B for optimal bidding and in Paper D to predict price-responsive electricity load.

We start from the premise that the choices made by a certain decision-maker (e.g., an aggregator of price-responsive power loads) at a certain time t, denoted by $\boldsymbol{x}_t$, are driven by the solution to the following linear optimization problem:

$$\text{RP}_t(\boldsymbol{p}_t|\boldsymbol{c},\boldsymbol{b}): \qquad \underset{\boldsymbol{x}_t}{\text{Maximize}} \ (\boldsymbol{c} - \boldsymbol{p}_t)^T \boldsymbol{x}_t \qquad\qquad (3.22\text{a})$$

$$\text{subject to } \boldsymbol{A}\boldsymbol{x}_t \leq \boldsymbol{b} \qquad\qquad (3.22\text{b})$$

$$\boldsymbol{x}_t \leq \boldsymbol{u} \qquad\qquad (3.22\text{c})$$

where $\boldsymbol{p}_t$ is a given time-varying input (e.g., the electricity price). Problem (3.22) is typically referred to as the *reconstruction problem.*

Now assume that the matrix of coefficients $\boldsymbol{A}$ and the right-hand side vector $\boldsymbol{u}$ are known and that we are able to observe the multivariate time series $\boldsymbol{X}'_T = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_T]$, which is presumed to be the solution to the reconstruction problem (3.22) at every time t. That is, $\boldsymbol{x}'_t$ represents the choices actually made by the decision-maker at time t. The basic goal of our inverse optimization approach is to infer the unknown parameter vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ from $\boldsymbol{X}'_T$ given $\boldsymbol{A}$, $\boldsymbol{u}$, and the series of measured inputs $\boldsymbol{p}_t$. To this end, one tries to find values for the unknowns $\boldsymbol{c}$ and $\boldsymbol{b}$ such that the observed choices $\boldsymbol{X}'_T$ are as optimal as possible for every problem (3.22).

When using inverse optimization for *forecasting* a time series, it is relevant to consider the case where we let the unknown parameter vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ vary over time so as to capture structural changes in the decision-making problem (3.22). To this end, we assume that we also observe a number of time-varying *regressors* $\boldsymbol{Z}_t$ that, to a lesser or greater extent, may affect the decision maker's choices. We then describe the unknown vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ as functions of those regressors by letting $\boldsymbol{c}_t = f_c(\boldsymbol{Z}_t)$ and $\boldsymbol{b}_t = f_b(\boldsymbol{Z}_t)$. In this way, functions $f_c(\cdot)$ and $f_b(\cdot)$ become decision variables in our estimation problem.

In this thesis, we consider $f_c(\cdot)$ and $f_b(\cdot)$ to be affine functions, i.e., $\boldsymbol{c}_t = f_c(\boldsymbol{Z}_t) = \boldsymbol{\beta}_c^{\mathrm{T}} \boldsymbol{Z}_t$ and $\boldsymbol{b}_t = f_b(\boldsymbol{Z}_t) = \boldsymbol{\beta}_b^{\mathrm{T}} \boldsymbol{Z}_t$, where $\boldsymbol{\beta}_c$ and $\boldsymbol{\beta}_b$ are the affine coefficients that relate $\boldsymbol{Z}_t$ with $\boldsymbol{c}_t$ and $\boldsymbol{b}_t$, respectively. Note that the past choices of the decision-maker, namely, $\boldsymbol{X}'_{t-1}$, can also be treated as regressors, in an analogous way as ARMAX time series models.

With the considerations from above in mind, we solve the following generalized inverse optimization problem:

$$\text{GIOP:} \underset{\boldsymbol{b},\boldsymbol{c},\boldsymbol{\beta}_c,\boldsymbol{\beta}_b,\boldsymbol{\phi},\boldsymbol{\lambda}}{\text{Minimize}} \sum_{t=1}^{T} \epsilon_t \tag{3.23a}$$

$$\text{subject to } \boldsymbol{b}_t^T \boldsymbol{\lambda}_t + \boldsymbol{u}^T \boldsymbol{\phi}_t - \epsilon_t = (\boldsymbol{c}_t - \boldsymbol{p}_t)^T \boldsymbol{x}_t' \quad \forall t \tag{3.23b}$$

$$[\boldsymbol{A}^T \ \boldsymbol{I}][\boldsymbol{\lambda}_t^T \ \boldsymbol{\phi}_t^T]^T = (\boldsymbol{c}_t - \boldsymbol{p}_t) \qquad \forall t \tag{3.23c}$$

$$\boldsymbol{A}\boldsymbol{x}_t' \leq \boldsymbol{b}_t \qquad \forall t \tag{3.23d}$$

$$\boldsymbol{c}_t = \boldsymbol{\beta}_c^{\mathrm{T}} \boldsymbol{Z}_t \qquad \forall t \tag{3.23e}$$

$$\boldsymbol{b}_t = \boldsymbol{\beta}_b^{\mathrm{T}} \boldsymbol{Z}_t \qquad \forall t \tag{3.23f}$$

$$\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t, \epsilon_t \geq 0 \qquad \forall t \tag{3.23g}$$

where $\boldsymbol{I}$ is the identity matrix of an appropriate size. The objective of optimization problem (3.23) is to minimize the sum over time of the duality gaps associated with the primal-dual reformulation of problem (3.22). Thus, when the objective function of GIOP is equal to zero, namely, the accumulated duality gap is zero, $\boldsymbol{x}_t'$ is optimal in $\mathrm{RP}_t(\boldsymbol{p}_t|\boldsymbol{c}_t, \boldsymbol{b}_t)$, $\forall t$. The first and second constraints in (3.23) are the relaxed strong duality condition and the dual problem constraints of (3.22), respectively. The third inequality represents the primal feasibility constraint involving the unknown right-hand side vector $\boldsymbol{b}_t$. The second primal constraint in (3.22) is, in contrast, omitted in (3.23), because it does not involve any decision variable in GIOP. Constraints (3.23e) and (3.23f) define the linear relationship between the regressors $\boldsymbol{Z}_t$ and $\boldsymbol{c}_t$ and $\boldsymbol{b}_t$, respectively.

The GIOP problem (3.23) is non-linear due to the multiplication of two decisions variables in (3.23b). In Paper D we provide a two-step estimation procedure with a view to minimizing the out-of-sample prediction error. Furthermore, the proposed two-step estimation procedure overcomes the nonconvexity and computational issues mentioned above regarding the solution to problem (3.23).

CHAPTER 4

Application Results

This chapter introduces three real-life problems that relate to decision-making under uncertainty in electricity markets. The first application arises by the increasing penetration of renewable energy in power systems. In this respect, in Section 4.1 we propose a stochastic optimization framework to optimally determine the level of ancillary reserves needed to safely operate power systems. The second application relates to the increasing importance of smart grid technologies and the necessary interaction of the price-responsive users with traditional electricity markets. In Section 4.2 we propose two methodologies to optimally compute demand-side bids to be submitted to the day-ahead market. The third application, introduced in Section 4.3, focuses on load forecasting under the real-time pricing paradigm.

4.1 Reserve Determination in Power Systems with High Penetration of Renewable Energy

Paper A is motivated by the expected increase of wind power penetration in Europe, where the European Union has set an ambitious target such that the EU will reach 20% share of energy from renewable sources by 2020. This paper focuses on Denmark, where this target is up to 30% [30]. In Denmark, in the

**Wind power capacity and wind powe
share of domestic electricity supply**
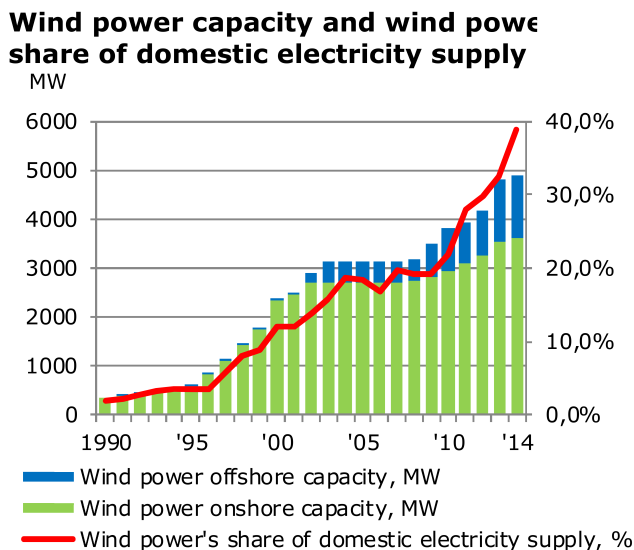
Figure 4.1: The share of wind power offshore and onshore capacity is shown in
blue and green bars, respectively. The share of domestic electric-
ity demand supplied by wind power production is shown in red.
Figure taken from [31].

year 2014, wind power production accounted for 38.8% of domestic electricity
supply, compared with 32.5% in 2013 and only 1.9% in 1990. The increasing
trend of installed wind power capacity and its importance relative to the total
domestic electricity consumption is shown in Figure 4.1.

As the share of electricity produced by renewables increases, several challenges
must be faced. Non-dispatchable electricity generation cannot ensure a certain
production at all times, but instead depends on meteorological factors. The
stochastic nature of such factors inevitably leads to forecast errors that will
likely result in producers deviating from their contracted power, thus causing
the system to be imbalanced. Such imbalances are mitigated by scheduling extra
capacity, also called *electricity reserves*, that are activated only when needed.
A detailed description of the main types of reserves has been given in this
dissertation in Section 2.2.1.

Currently, the TSO in Denmark schedules the amount of electricity reserve
by deterministic rules. A well-known rule, called *n-1 rule*, sets the minimum
amount of manual reserves to be the capacity of the largest on-line genera-
tor. With a relatively high penetration of renewable energy, deterministic rules

might not be the most optimal. Solutions call for methods capable of managing the uncertainty that wind power production and other stochastic variables induce into the system. In response to this challenges, in Paper A we propose two stochastic optimization models that account for future uncertainties and determine the optimal level of electricity reserves.

The methodology discussed in Paper A of this thesis is mainly targeted to the current structure of the Danish electricity market, where reserve markets are settled independently of and before the day-ahead energy market, implying that, at the moment of scheduling reserves, no information about which units will be online is known. At the market closure, the TSO collects bids from producers willing to provide reserve capacity, and selects them by a cost merit-order procedure. Most of the existing literature focuses on co-optimizing the unit commitment and reserve requirements at the same time [32, 33, 34, 35, 36]; such methods however cannot be applied under the current design of the Danish electricity market.

In paper A we propose two models for determining the optimal amount of total reserves to be scheduled, differing in the way risk is measured. One of the models is based on the Loss-Of-Load Probability (LOLP) and ensures that the probability of shedding load remains under a given threshold. The other proposed model calculates a trade-off between the cost of scheduling reserves with the cost of shedding load, by minimizing the Conditional Value at Risk (CVaR) of the total system cost. Both models are meant to be run before the clearing of the reserve market and can be used by the TSO to decide on how many MW of reserve should be scheduled. The reader is referred to Paper A for further elaboration on the mathematical description of the stochastic optimization models.

In Figure 4.2 we show an example of the results from the proposed model that minimizes the CVaR of the total system cost. The blue line is the data relative to the actual deployed reserves in Denmark West for every hour of the day. The red line represents the TSO's scheduled reserves, and finally the black lines represent the optimal reserves for different risk levels. In the upper plot, shedding load is set to be 500 €/MW, as opposed to the bottom plot, where shedding 1 MW of load cost 5000€. In both cases, the optimal reserves are higher than the TSO's decision. Nevertheless, the solution from the proposed models, when minimizing the CVaR, is from 3.38% cheaper with $\{\alpha = 0.99, V^{LOL} = 200$ €/MW$\}$, to 82.9% cheaper for $\{\alpha = 0.99, V^{\text{LOL}} = 5\,000$ €/MW$\}$, compared to the solution given by the TSO. Note that the CVaR method tends to schedule more reserves than the Danish TSO's solution, while at the same time the solution is cheaper, because shedding load is highly penalized by the coefficient V^{LOL}. Similar conclusions are drawn when using the LOLP model.
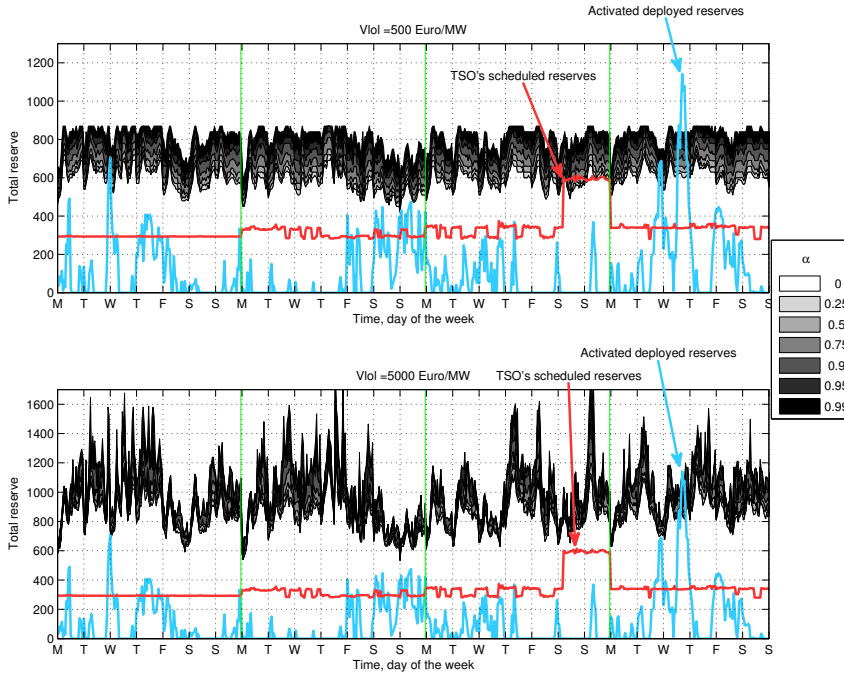
Figure 4.2: The reserve schedule using the CVaR methodology is displayed for $V^{\text{LOL}} = 500$ €/MW in the upper plot and for $V^{\text{LOL}} = 5\,000$ €/MW in the lower plot. The shaded areas in the background represent the solution of the CVaR model for different values of the risk aversion parameter α. The actual deployed reserve in DK1 and the reserve capacity scheduled by Energinet.dk (the Danish TSO) are depicted on top.

4.2 Demand Response Bidding in Electricity Markets

Demand response has been defined in Section 2.3 of this dissertation as the intentional alteration of the power consumption profile by an end-user in response to an external stimulus. In Chapter 2, we have seen that nowadays electricity is traded through markets, where players submit bids reflecting their needs. In this section, we tackle the challenge of linking both concepts: demand response *bidding* in electricity markets.

Amongst the many different forms of demand response [10, 37], this thesis focuses on the case when the consumers are *price-responsive*. Under the smart-grid paradigm, consumers are equipped with a smart grid meter and communication devices that allow them to receive dynamical prices. This pricing scheme is commonly named as Real-time Pricing (RTP) [38, 39], also introduced in Section 2.3 under the denomination of *indirect control* scheme. In the case of households, the dynamic pricing allows the residents to schedule their appliances and heating or cooling demand based on the price of the electricity and also on their own preferences [39, 40, 41, 42, 43].

We consider the case where a retailer, also called *aggregator*, bids to the day-ahead market on behalf of its price-responsive pool. Traditionally, retailers bid an inelastic curve that simply indicates the desired amount of power, independently of the market price. Nonetheless, because of the price-responsiveness of the consumers under the considered setup, inelastic bidding curves might not be the most adequate decision. Instead, a bidding curve with a price component, or *price-bid*, can reduce imbalances and increase the retailer's profit, as we show in Paper B and Paper C of this thesis.

The bidding curves we consider in this thesis are composed by, at least, a set of price and energy terms, forming an *price-energy* demand bid. Another way of interpreting the price term of the bid is as the *marginal utility* of the aggregate load. In both cases, the demand bid reflects the elasticity of the pool of consumers to changes in the electricity price. Demand-side bids are typically defined as monotonically decreasing block-wise functions. The bids are formed by a price-bid term, and bock width term, denoted by u_b and E_b, respectively, with a total of B terms. For a day-ahead price p^{DA}, the price-bid indicates that the total load will be equal to $\sum_b y_b E_b$ with

$$y_b = \begin{cases} 1 & \text{if } u_b \geq p^{DA} \\ 0 & \text{if } u_b < p^{DA} \end{cases}. \tag{4.1}$$

We developed two methodologies that provide for an optimal bid, differing in the way they measure the "goodness" of the estimated bid. The first methodology, introduced in Section 4.2.1, relates to Paper B from this thesis. It aims at finding the optimal bid that minimizes the expected imbalances between the scheduled load in the day-ahead and the actual realized load. The second methodology, explained in Section 4.2.2, relates to Paper C from this dissertation and focuses on finding the optimal bid that maximizes the retailer's profit.

4.2.1   Optimal Bidding By Minimizing Imbalances

The proposed approach presented in Paper B from this dissertation calculates the optimal demand-side bid by minimizing the expected imbalances between the scheduled load and the actual one. The distinctive features of our work with respect to the existing literature are the fourfold. First, we develop a novel approach to capture the price-response of the pool of flexible consumers in the form of a market bid using price-consumption data. Second, we propose a generalized inverse optimization framework to estimate the market bid that best captures the price-response of the pool. Third, we use machine-learning techniques to leverage auxiliary information on a set of features that may have predictive power on the consumption pattern of the cluster. Lastly, we test our methodology using data from a real-world experiment and compare its performance with state-of-the-art prediction models on the same dataset.

The considered complex market bid is analogue to traditional supply bid, and is defined by five elements:

$u_{b,t}$ Marginal utility corresponding to bid block b and time t.

$\underline{P}_t$ Minimum power consumption at time t.

$\overline{P}_t$ Maximum power consumption at time t

r_t^u Maximum load pick-up rate at time t, analogue to the ramp-up limits of a power generating unit.

r_t^d Maximum load drop-off rate at time t, analogue to the ramp-down limits of a power generating unit.

In the proposed method, the aggregate load is characterized by an optimization problem, called also *reconstruction problem*, which is defined by the bidding parameters. In Figure 4.3 we show a representation of the functioning of the reconstruction problem. On the x-axis, we show two time periods, denoted by
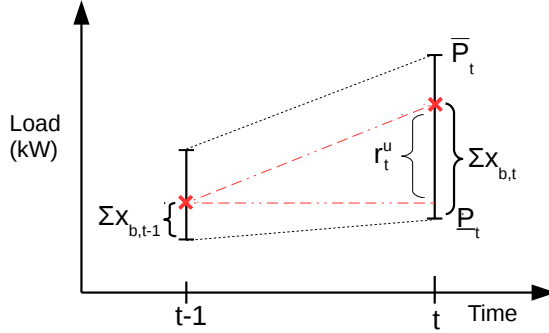
Figure 4.3: Representation of the reconstruction problem, representing the estimated complex bidding problem. On the x-axis,

$t-1$ and t. The interval defined by $\underline{P}_t$ and $\overline{P}_t$ indicates the range where the load, denoted by $\underline{P}_t + \sum x_{b,t}$, can possibly realize. The evolution of the load from time $t-1$ to time t is constrained by the pick-up and drop-off limits, denoted by r_t^u and r_t^d. In Figure 4.3, only the pick-up limit appears. Observe also that in the Figure 4.3, the load at time t is greater than at time $t-1$.

Given the technical constraints imposed by r_t^u, r_t^d, $\underline{P}_t$ and $\overline{P}_t$, and the marginal utility of the aggregate pool of consumers $u_{b,t}$, the reconstruction problem chooses the optimal load such that it maximizes the consumers'welfare:

$$\underset{x_{b,t}}{\text{Maximize}} \sum_{t=1}^{T} \left(\sum_{b=1}^{B} u_{b,t} x_{b,t} - p_t \sum_{b=1}^{B} x_{b,t} \right) \tag{4.2a}$$

subject to

$$x_t^{tot} = \underline{P}_t + \sum_{b=1}^{B} x_{b,t} \qquad \forall t \tag{4.2b}$$

$$-r_t^d \leq x_t^{tot} - x_{t-1}^{tot} \leq r_t^u \qquad \forall t > 1 \tag{4.2c}$$

$$0 \leq x_{b,t} \leq \frac{\overline{P}_t - \underline{P}_t}{B} \qquad \forall b, t \tag{4.2d}$$

where $x_{b,t}$ is the consumption assigned to the utility block b during the time t and p_t is the price of the electricity during time t. Equations (4.2b) define the total load as the minimum power consumption plus the load for each block. Equations (4.2c) impose a limit on the load pick-up and drop-off rates, respectively. The set of equations (4.2d) defines the size of each utility block to be equally distributed between the maximum and minimum power consumptions, with B being the total number of blocks. Constraint (4.2d) also enforces the

consumption pertaining to each utility block to be positive. Note that, by definition, the marginal utility is decreasing in x_t ($u_{b,t} \geq u_{b+1,t}$), so one can be sure that the first blocks will be filled first.

The parameters that shape the market bid, denoted by $\theta_t = \{u_{b,t}, r_t^u, r_t^d, \underline{P}_t, \overline{P}_t\}$, are estimated using an inverse optimization framework. The model resembles the inverse optimization model based on the minimization of an absolute value, as presented in Section 3.2.5.1. In mathematical terms, the inverse optimization problem is the following:

$$\underset{x_t^{tot}, \theta_t}{\text{Minimize}} \sum_{t=1}^{T} w_t \left| x_t^{tot} - x_t^{meas} \right| \tag{4.3a}$$

subject to

$$u_{b,t} \geq u_{b+1,t} \qquad \forall b, t \tag{4.3b}$$

$$\text{KKT conditions of (4.2)} \tag{4.3c}$$

where x_t^{meas} is the measured load at time t and w_t is a given weight for each time period.

Constraints (4.3b) are the upper-level constraints, ensuring that the estimated marginal utility must be monotonically decreasing. Constraints (4.3c) correspond to the KKT conditions of the bidding problem (4.2).

The objective function (4.3a) represents the mismatch or *estimation errors* between the measured load x_t^{meas} and the estimated load x_t^{tot} resulting from the reconstruction problem (4.2). The estimation errors could represent deviations of consumption from the contracted power in a forward (e.g., day-ahead) market, which must be settled by purchasing or selling energy in the balancing market. Therefore, when the objective function (4.3a) is equal to zero, the load is perfectly represented by the reconstruction problem, hence the contracted power will be precisely the realized load.

Indicator variables and external variables, often referred to as *features* of *regressors*, can be used to explain more accurately the parameters of the market bid that best represents the price-response of the pool of consumers. This approach is potentially useful in practical applications, as numerous sources of data can help better explain the consumers' price-response. Practical applications also need to include a set of robust constraints to ensure the consistency of the market bid for all possible realizations of the external variables and not only for the ones observed in the past. This is explained in detail in the appendix of Paper D.
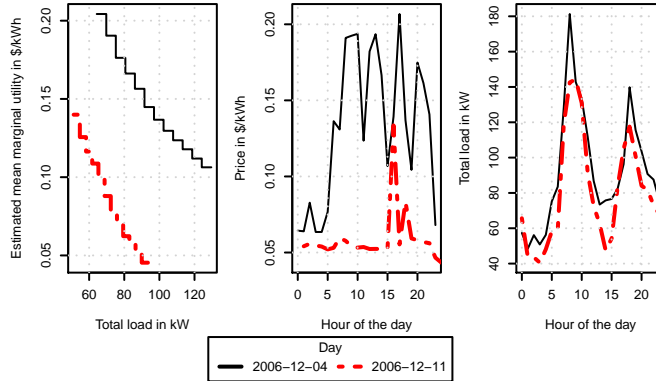
Figure 4.4: Averaged estimated block-wise marginal utility function for the inverse-optimization based model (left panel), price in \$/kWh (middle panel), and load in kW (right panel). The solid lines represent data relative to the 4th of December. Dashed lines represent data relative to the 11th of December.

The inverse optimization problem (4.3) is non-lineal due to the KKT conditions of (4.2). For realistic applications involving multiple time periods and/or numerous features, exact algorithms do not provide a solutions in a reasonable amount of time. In order to overcome the computationally difficulties, we propose a two-step solution strategy. The proposed solution method is a heuristic in the sense that it does not solve the inverse optimization problem to optimality. However, our relaxed problem calibrates a penalty term in order to minimize the *out-of-sample* prediction error. The reader is referred to Paper B for further elaboration on a practical approach to solve the inverse optimization problem above.

A case study with real data is performed using measurements relative to a pool of price-responsive users. The data relates to the Olympic Peninsula experiment, which took place in Washington and Oregon states between May 2006 and March 2007 [44]. The results show that the estimated complex bid successfully models the price-response of the pool of houses, in such a way that the mean absolute percentage error incurred when using the estimated market bid for predicting the consumption of the pool of houses is kept in between 14% and 22% for all the months of the test period. The performance results are in line with reported measures from state-of-the-art forecasting methods [45] and outperform traditional forecasting methods [46] for the same dataset.

An example of the estimated block-wise marginal utility function, averaged for the 24 hours of the day, is shown in the left plot of Figure 4.4. The solid line

corresponds to the 4th of December, when the price was relatively high (middle plot), as was the aggregate consumption of the pool of houses (right plot). The dashed line corresponds to the 11th of December and shows that the estimated marginal utility is lower, as is the price on that day.

4.2.2   Optimal Bidding By Maximizing the Retailer's Benefit

In Paper C we propose a bidding model that maximizes the profit of a retailer that aggregates price-responsive consumers. The retailer must decide, at the time of closure of the day-ahead market, the optimal price-energy bids to be submitted relative to the next operational day. When the day-ahead market is cleared, the retailer is assigned a certain scheduled load that depends on its submitted price-bid and also on the market price. Over the next day, as time goes on and the load is actually realized, the imbalances between the scheduled load at the day-ahead energy market and the actual load are purchased or sold in the real-time market.

The uncertain nature of the market prices and the response of the load to changes in price call for solutions that consider stochastic variables. Moreover, the solutions should take into account the aversion of the retailer towards risk. In response to these challenges, the contributions of the work shown in this section and in Paper C are the following.

1. We develop an analytic solution to the problem of finding optimal block-wise price-energy demand bids in the day-ahead market, when the retailer is risk-neutral. Moreover, we propose a mixed-integer linear programming solution approach to the case where the retailer is risk-averse.

2. The dynamic price-responsive behavior of consumers is modeled based on scenarios. The conditional probability of the load given a certain retail price trajectory is estimated using a non-parametric approach.

3. We assess the practicality of the proposed methodology by using data from a real-world experiment.

It is important to note that, under the considered setup in Paper C and in this section, the retail price is given exogenously. The leading reason for this consideration is the fact that the retail price must, to a certain extent, represent the true cost of electricity. This might not always be the case if the retail price is subject to the will of the retailer [47]. Under the considered setup, the retail
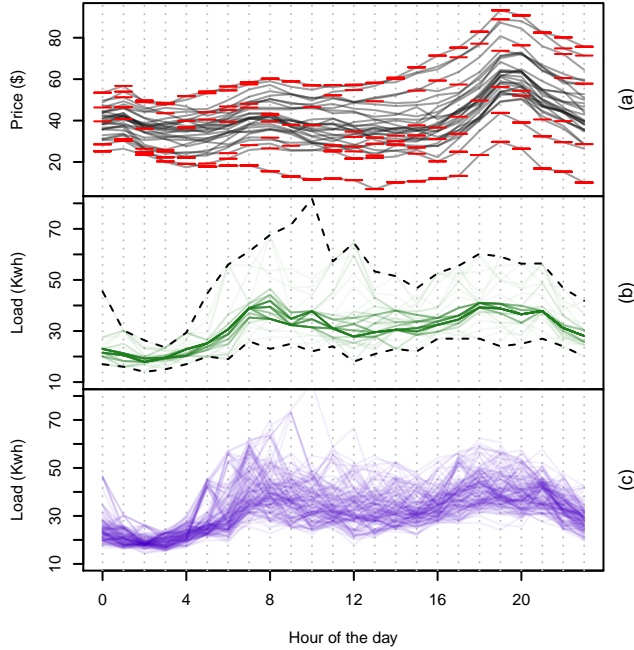
Figure 4.5: In (a), the day-ahead scenarios (lines) are shown together with the estimated price-bids (horizontal segments) by the risk-constrained model. The day-ahead purchase for each scenario, and the load in each scenario, are shown in (b) and (c), respectively.

price can be, for example, equal to the day-ahead price plus a proportional fee. This implies that the retailer's decisions affect only its revenue, and not the payments that its consumers face.

The responsiveness of the load to changes in price is represented by scenarios, which are generated from a dataset in a non-parametric way. The procedure used for generating scenarios using a non-parametric distribution is explained in detailed in Section 3.1.3 of this dissertation. The proposed approach to characterize uncertainty allows us to model the non-linear relationship between load and price, without making any assumption on the nature of the load. Moreover, utilizing scenarios of 24-hours allows us to describe realistic behaviors of the price-responsive load as, for example, load-shifting and load reduction. Another relevant advantage is that the scenario-generation technique is computationally effective, so large datasets can be quickly analyzed. An example of the scenarios of price-responsive load is shown in Figure 4.5(c).

We develop a mathematical model that maximizes the retailer's profit based on the revenue that it collects from its loads, the payments it makes to the day-ahead market, and the payments it makes or receives in the real-time market. We distinguish two cases, whether the retailer is risk-averse or not.

For the risk-unconstrained case, in Paper C we formulate two theorems that are used to calculate the optimal price-bid. These theorems allow us to compute the optimal bid without performing any optimization problem. The results show that, in the risk-unconstrained case, there is no extra benefit from bidding a curve. Moreover, the optimal bid in this case is used for arbitrage. Naturally, the profit of the risk-neutral retailer is maximized, with the downside that the variability of the profit is also high. This fact is confirmed in the case study.

For the risk-averse case we formulate a stochastic optimization model, that computes the optimal price-bid given a maximum level of allowed risk. In this case, the risk-aversion is modeled by the following chance constraints:

$$\mathrm{P}\left(X_t^D \in [(1-L)\,X_t, (1+L)X_t]\right) \geq \beta \quad \forall t \tag{4.4}$$

with X_t^D being the scheduled load in the day-ahead market and X_t the realized load. The risk-aversion is defined by two parameters. Parameter L represents the fraction of the load that the retailer purchases in the real-time market. Thus, a small allowed fraction indicates that the retailer is risk averse. Parameter β imposes a minimum to the probability of the risk constraint to be fulfilled.

The results from the risk-averse case show that a block-wise bidding curve successfully mitigate the risk, compared to the case where we trade based only on point forecast of average values. This is empirically shown in the case study and the results are summarized in Table 4.1. We show the mean (1st column) and the standard deviation (2nd) of the profit of the retailer for three benchmark models, using realistic data relative to the months of November and December. We observe that the simple model *ExpBid* where the expected value of the load is always purchased under-performs the rest of the models and, indeed, delivers a negative expected profit. The risk-constrained problem *LRisk* yields positive expected profit, with a variance greater than the *ExpBid* model but substantially lower than for the risk-unconstrained model *UncRisk*. The risk-unconstrained model *UncRisk*, as anticipated, provides the highest mean returns.

As a final remark, in Figure 4.6 we show the estimated bidding curves for two different hours of the day. We show the estimated optimal price-bid for a risk-neutral retailer and for a risk-constrained retailer in green and red, respectively. In accordance to the developed theorems, the risk-unconstrained retailer bids a flat curve, indicating that the retailer buys *all-or-nothing* in the day-ahead market. On the other hand, the risk-constrained retailer bids a decreasing curve that allows the retailer to adapt its purchase depending on the market price.

	Mean	Std. dev.
ExpBid	-1.78	34.52
LRisk	22.26	45.22
UncRisk	188.82	259.62

Table 4.1: Mean and standard deviation of the profit for the benchmarked models during November and December.
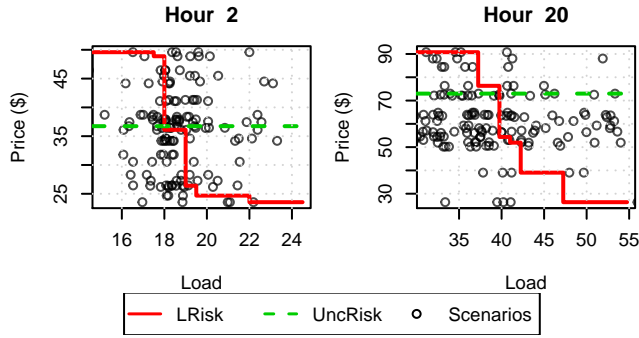
Figure 4.6: The left figure is relative to the 2nd of November, while the right figure is relative to the 1st of November.

4.3   Forecasting Price-responsive Load

Load forecasting has been subject to research for many years and, in fact, is still nowadays [1, 48, 49]. Prognoses of the load are used as a support to decision-making processes, allowing the decision-maker to intelligently act upon the uncertain nature of the load [5]. As an example, network operators use long-term load forecasts to plan for grid expansions, and short-term forecasting to mitigate possible congestions during high-demand hours. Also, load forecasting is used by electric utilities to minimize the costs of over- or under-contracting power in electricity markets.

As explained in Section 2.3, the role of demand response in power systems is rapidly increasing. This relatively new set of technologies are changing the patterns that the load traditionally exhibited. For this reason, forecasting load that is equipped with smart grid technology is a challenge that needs to be addressed.

In this Section, as well as in Section 4.2 above and in Paper D of this thesis,

we deal with the case where consumers of electricity are equipped with an Energy Management Control (EMC) system that allows them to receive a varying price of electricity along the day. The EMC has a controller that optimizes the consumption by scheduling the consumption at low-priced periods of time while complying with the consumer's preferences [39, 40, 41, 42, 43]. This type of demand response has been introduced in Section 2.3 under the denomination of *indirect control*, also referred as Real-Time Pricing (RTP) scheme [38, 39, 50, 51]. Under the indirect control setup, the load is considered to be *price-responsive*.

The research conducted in Paper D focuses on forecasting price-responsive load. The novelty of the proposed forecasting method with respect to the existing literature is that the aggregate consumption is described by an optimization problem, which is characterized by a set of unknown parameters. This optimization problem is called *forward* or *reconstruction* problem, and its decision variable is the estimated load at every time t, denoted by x_t. In this regard, Paper D of this thesis has been the first work to exploit inverse optimization for time series forecasting and, in particular, for load prediction.

The unknown parameters that define the reconstructed model are the marginal utility for every block b and time t, denoted by $u_{b,t}$, and the maximum and minimum power consumptions, denoted by $\overline{P}_t$ and $\underline{P}_t$, respectively. The parameter-estimation problem recast as an inverse optimization problem, and its goal is to estimate appropriate values for $u_{b,t}$, $\underline{P}_t$ and $\overline{P}_t$, such that the solution to the reconstruction problem serves as a good forecast of the future aggregate power consumption of the load.

For the reconstruction problem, we consider that the available information is the electricity price p_t, which is broadcast to every load in the pool. The aggregate response $\boldsymbol{x}_t$ of the loads to the price of electricity at time t is assumed to be the solution to the following optimization problem:

$$\underset{\boldsymbol{x}_t}{\text{Maximize}} \sum_{b=1}^{B} x_{b,t} \left(u_{b,t} - p_t \right) \tag{4.5a}$$

$$\text{subject to } 0 \leq x_{b,t} \leq E_{b,t} \qquad \forall t, b \tag{4.5b}$$

$$\underline{P}_t \leq \sum_{b=1}^{B} x_{b,t} \leq \overline{P}_t \quad \forall t. \tag{4.5c}$$

Problem (4.5) is a linear programming problem and fits into the generic formulation given in Problem (3.7) from Section 3.2.1. Its objective function (4.5a), to be maximized, represents the aggregate consumers' surplus or welfare, given as the product of the pool consumption and the difference between the marginal

utility and the electricity price. We consider a step-wise marginal utility curve made up of B blocks, each of a width $E_{b,t}$, as enforced by (4.5b), and a value $u_{b,t}$ for each block. The aggregate load of the pool at time t, given as $\sum_{b=1}^{B} x_{b,t}$, is bounded from below and above by $\underline{P}_t$ and $\overline{P}_t$, respectively, as expressed by (4.5c).

Based on the observed historical load x'_t and the price of electricity p_t, the inverse optimization method infers the appropriate values for $u_{b,t}$, $\underline{P}_t$ and $\overline{P}_t$, such that x'_t becomes as optimal as possible in the reconstruction problems (4.5) for every time t. To this end, the proposed inverse methodology is able to exploit external variables or *regressors* in an attempt to explain the variations of the unknown parameters over time. In this paper, we consider affine relationship between the regressors and $u_{b,t}$, $\underline{P}_t$ and $\overline{P}_t$. Hence, the inverse optimization problem boils down to finding the most optimal set of intercepts and affine coefficients that relate the regressors with the unknown parameters over time. The concept of dynamic inverse optimization has been introduced in Section 3.2.6.

The inverse optimization model we use to estimate the unknown parameters is similar to Problem (3.23). In brief, we seek a combination of the unknown parameters such that the duality gap of the reconstruction problem (4.5) is minimized. This approach arises several challenges summarized as follows:

- Non-linearity. The inverse optimization model is non-linear due to the product of two variables appearing in the strong duality condition of (4.5). From this point of view, a method capable of obtaining a good estimation of the unknown parameters in a reasonable amount of time is needed, even if such a solution may be suboptimal.

- Optimality: there might not exist a combination of the unknown parameters that make the historical load x'_t to be an optimal in (4.5).

- Feasibility: the historical load x'_t is not necessarily feasible in the reconstruction problem (4.5)

In order to overcome these difficulties, in Paper D we propose a solution approach that consists of two steps, where a linear programming problem is solved in each one:

1. *Feasibility problem* or bound estimation problem. It consists in finding a "good" feasible region for the reconstruction problem (4.5) in terms of prediction performance. The problem is linear and makes use of a penalty parameter that ensures the bounds to have good predictive power. The

penalty parameter is tuned by means of cross-validation in order to obtain the smallest out-of-sample prediction error [14, Ch. 7].

2. *Optimality problem* or marginal utility estimation problem. Given a set of bounds calculated by the bound estimation problem, we estimate the marginal utility for every block b and time t such that the observed load x_t' becomes as optimal as possible in the reconstruction problem (4.5).

A detailed elaboration on the proposed solution method is left out of this chapter for brevity. The reader is referred to Paper D of this thesis for further information on the solution method.

The proposed inverse optimization framework can be seen as a generalization of a linear time series model: the relationship between the load and the regressors is linear, but the relationship between the load and the price at time t is not. In other words, in the feasibility problem, we model the linear relationship between the load and the regressors, excluding the price at time t. The outcome of the feasibility problem is the power bounds $\widehat{\underline{P}}_t$ and $\widehat{\overline{P}}_t$. In the optimality problem we model the non-linear relationship between the load that falls inside the interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$, and the price at time t. Unlike in the proposed scheme, in a simple linear regression model, the relationship between the load and the price is given by an affine coefficient.

For the case study, we simulate two datasets using the work in [43]. In short, we simulate 100 different buildings equipped with heat pumps. The heat pumps in each building schedule their consumption by solving an Economic Model Predictive Control (EMPC) problem that minimizes the cost of energy plus a penalty term for not complying with a comfort temperature band. We simulate and aggregate the load of the buildings in two datasets. One dataset, called *no flex*, shows no response of the buildings to price. In the other dataset, called *flex*, the buildings are price-responsive and the relationship between the aggregate load and the price is non-linear, as shown in Figure 4.7.

The load is predicted 1-step ahead, and the performance of the proposed method is compared with the forecasting capability of an Auto Regressive Model with Exogenous Variables (ARMAX) [12]. An example of the predictions of the aggregate load for the *flex* dataset is displayed on the top of Figure 4.8. The estimated minimum and maximum load bounds are able to explain a certain part of the variability of the load. The remaining variability is explained by the relationship between the marginal utilities and the price. The predictions made by the ARMAX model are also able to anticipate the behavior of the load, but to a lesser extent. On the bottom plot of Fig. 4.8, the electricity price is displayed together with the estimated marginal utility blocks, for each hour of
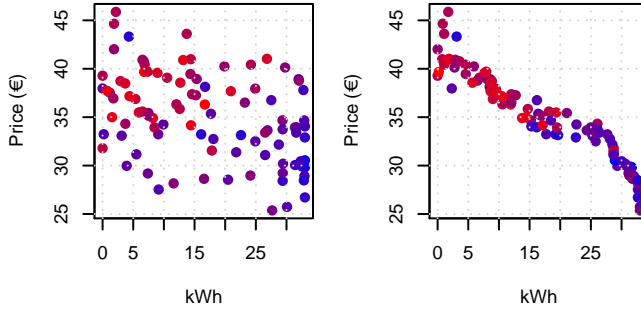
Figure 4.7: Price and aggregate load for the non-flexible (left) and flexible (right) cluster of buildings. Red and blue colors indicate high an low ambient temperature, respectively.

the test period. The magnitude and distribution of the marginal utilities change with time and capture the dynamic response of the load to the price.

Performance metrics computed over the test set are summarized in Table 4.2. Each row is relative to one of the three benchmark models. Columns 1 and 3 give information on the Normalized Root Mean Square Error (NRMSE), defined as

$$\text{NRMSE} = \frac{1}{x^{max} - x^{min}} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{b=1}^{B} \widehat{x}_{b,t} - x'_t \right)} \tag{4.6}$$

and columns 2 and 4 on the Symmetric Mean Absolute Percentage Error (SMAPE)

$$\text{SMAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|\sum_{b=1}^{B} \widehat{x}_{b,t} - x'_t|}{(|\sum_{b=1}^{B} \widehat{x}_{b,t}| - |x'_t|)/2}. \tag{4.7}$$

In Table 4.2 we also compare the performance of the proposed forecasting method using the two simulated data sets. On the left part, we show the performance measures relative to the *no flex* data set. The ARMAX and the *InvFor* models yield almost identical results in terms of NRMSE and SMAPE. The differences between the ARMAX and the *InvFor* stand out when used for predicting the *flex* data set. On the right side of Table 4.2, we see that our methodology outperforms the ARMAX with a NRMSE and a SMAPE 32% and 16.8% lower, respectively. The persistence model, as expected, exhibits the worst performance. We conclude that the non-linear relationship between the price and the load is well captured by the *InvFor* model.
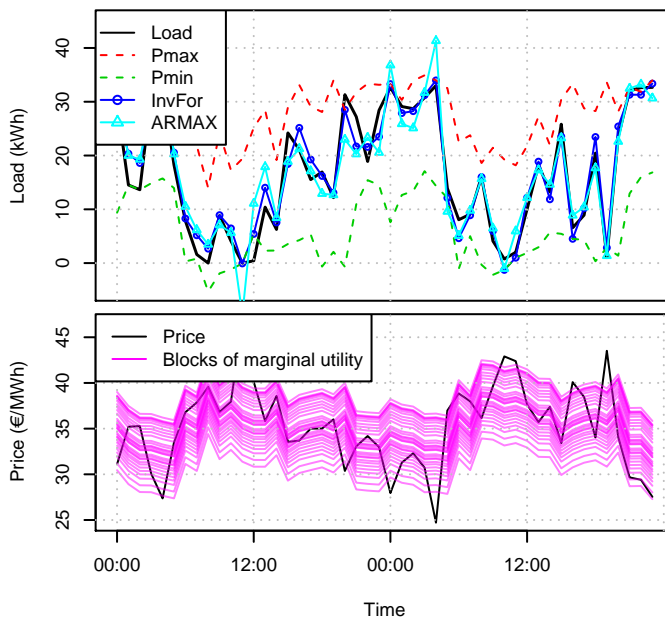
Figure 4.8: On the top, the actual load is displayed together with the predictions from the inverse-optimization methodology and the ARMAX model. On the bottom, the price is shown together with the estimated marginal utility blocks.

| | No Flex | | Flex | |
	NRMSE	SMAPE	NRMSE	SMAPE
Persistence	0.1727	0.1509	0.3107	-
ARMAX	0.10086	0.08752	0.13107	0.08426
InvFor	0.10093	0.0886	0.08903	0.07003

Table 4.2: Benchmark for the test set

CHAPTER 5

Conclusions and Perspectives

5.1 Contributions

In this dissertation we have provided a number of innovative contributions to
the state of the art. The contributions are novel from a methodological point of
view, and their applicability is rooted to decision making in power systems. This
thesis provides mathematical models that enhance the operation of electricity
markets and also of their players. In summary, the contributions of this thesis
are to the field of demand-side bidding, price-responsive load forecasting, reserve
determination and inverse optimization.

The first mayor contribution of the conducted research is on the development
of inverse optimization models applied to power system operations and time
series analysis. The use of constrained optimization models to characterize the
response of aggregate price-responsive load is a novel concept in the existing
literature, and has been proved to be relevant in Papers B and D of this dis-
sertation. From a mathematical perspective, the inverse optimization models
developed in Papers B and D pose several challenges to which we provide com-
prehensive answers. In both papers, we consider a generalized inverse optimiza-
tion scheme novel in the literature, where the unknown parameters appear both
in the objective function and in the constraints of the reconstruction problem.
To overcome the nonlinear and non-convex nature of the models, we provide

two different computationally attractive methods that approximate the solution
to the proposed generalized inverse optimization models by solving instead two
linear programming problems. In Paper B the solution method is based on
the relaxation of the KKT conditions, while in Paper D it relies on a duality-
gap minimization problem preceded by a feasibility problem. In both cases,
by solving linear problems instead of one single nonlinear one, the computa-
tional burden is significantly reduced, allowing us to include external variables
or *regressors* that, in practice, are shown to enhance the performance of the so-
lutions. A special case of such regressors is time and calendar variables, which
allow to model the time-varying characteristics of the variables of interest.

Optimal bidding, though a relatively old concept in the field of power systems
and load forecasting, gains a whole new dimension when smart grid and real-
time pricing is considered. Traditionally, a forecast of the consumption of a pool
of loads comprises the inelastic bid that is submitted to the wholesale electricity
market, indicating the desired purchase quantity. In Papers B and C of this
dissertation, we prove the usefulness of more complex bids that capture the
response of the pool of consumers to changes in price. These bids are analogue
to the traditional supply bids, hence they can easily be implemented in the
current market clearing algorithms. The main challenge that we overcome is how
to obtain *optimal* demand-side bids that best represent the complex behavior
of the load to varying price.

This thesis proposes two methodologies that provide for an optimal demand-side
bid in the case where a retailer, also called *aggregator*, bids to the day-ahead
market on behalf of its price-responsive pool of consumers. From the market-
bid perspective, the bids take the form of a price-bid curve, as in Paper C, and
possibly other technical constraints as in Paper B. The contributions differ in
the way they measure the "goodness" of the estimated bid and their usefulness
depend on whether the retailer's goal is to minimize its expected imbalances or
to maximize its profit. In both cases, the complex bids outperform simple bids
in terms of imbalances and retailer's profit.

In the field of load forecasting, Paper D gives a novel approach to model time
series and forecast price-responsive loads. Under the real-time pricing setup,
the impact of the price on the aggregate load for every time period is naturally
non-linear. This relationship is modeled by an optimization problem, which is
characterized by a set of unknown parameters. The parameter estimation pro-
cess boils down to solving a generalized inverse optimization problem. For the
case study, it is shown that the performance of the inverse-optimization-based
method outperforms traditional time series models, in particular, ARMAX mod-
els.

In paper A we consider the problem of determining the most appropriate levels

of electricity reserves to be scheduled in a market structure such as the ones in Denmark and the Nordic countries. A probabilistic framework is given, where the reserve requirements are computed based on scenarios of wind power and load forecast errors and power plant outages. The proposed solution differs with the existing work in that it is well-suited for power systems where the reserve market and the day-ahead energy market are cleared independently at different times and by different entities. We propose a stochastic optimization modeling framework that, without knowing which units will be online, is able to determine the optimal allocated reserves for the next operational day.

5.2   Perspectives and Future Research

The development of advanced metering and communication infrastructure in smart grid opens up an interesting and broad field of research. In addition, our energy needs are being increasingly covered by renewable energy production means. Having said that, current electricity markets are designed for a no-smart and no-renewable system. Research that is able to couple them is certainly needed. This thesis provides solutions to some of the existing problems, but certainly more needs to be done. Our focus has been to solve problems that arise under the current electricity market structure. Nevertheless, new market frameworks that facilitate the inclusion of smart grid technologies is a relevant topic to be studied in future work.

From a methodological point of view, the research conducted in this thesis opens up several lines of work. The most interesting one is the further application of inverse optimization modeling to power systems, whether it is for optimal bidding, for load forecasting, or any other possible application. The research conducted in this thesis has just touched the tip of the iceberg in regards to inverse optimization and time series. Exploiting large amounts of data, allowing for non-linear relationships by, for example, the use of splines, and obtaining robust solutions, are only three of the many improvements that need to be addressed. From a theoretical perspective, further mathematical foundations that characterize the solutions to generalized inverse optimization problems could help to understand better the underlying characteristics of this family of models.

Summarizing complex problems into simple optimization problems by using inverse optimization seems applicable to other fields than just power systems. Certainly, an exiting line of work would be to find such applications as for example transport, biology and finance.

In regards to optimal reserve determination, the proposed approach could be

enhanced by modeling the interconnections between grid areas and by modeling the possible reserve capabilities of demand under the smart grid paradigm. Furthermore, computing and allocating optimal reserves in a large interconnected system like the European one could be of interest to policy makers. From the perspective of the Danish power system, assessing the benefit of jointly determining the day-ahead energy and reserve dispatch versus their independent determination could shed some light on whether the market structure needs to be changed.

Bibliography

[1]   J. M. Morales et al. *Integrating renewables in electricity markets: Operational problems*. Vol. 205. Springer Science & Business Media, 2013.

[2]   *Nord Pool Spot AS*. URL: http://www.nordpoolspot.com/.

[3]   *Energinet.dk*. URL: http://energinet.dk.

[4]   *NASDAQ OMX Commodities Europe*. URL: http://www.nasdaqomx.com/commodities.

[5]   A. J. Conejo, M. Carrión, and J. M. Gonzalez Morales. *Decision Making Under Uncertainty in Electricity Markets*. International series in operations research and management science. Springer US, 2010.

[6]   Energinet.dk. *Ancillary Services to be Delivered in Denmark Tender Conditions*. Tech. rep. Oct. 2012.

[7]   Energinet.dk. *Ancillary services strategy*. 2011.

[8]   Energinet.dk. *Regulation C2 : The balancing market and balance settlement*. Tech. rep. Dec. 2008.

[9]   *California ISO Open Access Same-time Information System (OASIS)*. URL: http://oasis.caiso.com.

[10] N. O'Connell. "Approaches for Accommodating Demand Response in Operational Problems and Assessing its Value". PhD thesis. DTU, 2015.

[11] S. Harbo and B. Biegel. *Contracting Flexibility Services*. Tech. rep. iPower, 2012.

[12] H. Madsen. *Time Series Analysis*. Chapman & Hall, 2007.

[13] W. Hardle. *Applied nonparametric regression*. Cambridge Univ Press, 1990.

[14] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning.* Vol. 1. Springer series in statistics Springer, Berlin, 2001.

[15] P. Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind energy* 12.1 (2009), pp. 51–62.

[16] G. B. Dantzig. *Linear programming and extensions.* Princeton university press, 1998.

[17] N. Karmarkar. "A new polynomial-time algorithm for linear programming". In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing.* ACM. 1984, pp. 302–311.

[18] *General Algebraic Modeling System (GAMS).* URL: http://www.gams.com/.

[19] *IBM ILOG CPLEX.* URL: https://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/.

[20] J. Nocedal and S. Wright. *Numerical optimization.* Springer Science & Business Media, 2006.

[21] R. K. Ahuja and J. B. Orlin. "Inverse Optimization". In: *Operations Research* 49 (2001), pp. 771–783.

[22] C. Ruiz, A. J. Conejo, and S. A. Gabriel. "Pricing non-convexities in an electricity pool". In: *Power Systems, IEEE Transactions on* 27.3 (2012), pp. 1334–1342.

[23] T. C. Chan et al. "Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy". In: *Operations Research* (2014).

[24] A. Keshavarz, Y. Wang, and S. Boyd. "Imputing a convex objective function". In: *Intelligent Control (ISIC), 2011 IEEE International Symposium on.* Sept. 2011, pp. 613–619.

[25] A. Aswani, Z.-J. M. Shen, and A. Siddiq. "Inverse Optimization with Noisy Data". In: *arXiv preprint arXiv:1507.03266* (2015).

[26] P. M. Esfahani et al. "Data-driven Inverse Optimization with Incomplete Information". 2015.

[27] D. Bertsimas, V. Gupta, and I. C. Paschalidis. "Data-driven estimation in equilibrium using inverse optimization". In: *Mathematical Programming* 153.2 (2015), pp. 595–633.

[28] Z. Xu et al. "Data-Driven Pricing Strategy for Demand-Side Resource Aggregators". In: *IEEE Transactions on Smart Grid* 99 (2016).

[29] Q. Zhao et al. "Learning cellular objectives from fluxes by inverse optimization". In: 54th IEEE Conference on Decision and Control (CDC). 2015, pp. 1271–1276.

[30] *Energy Strategy 2050 - from coal, oil and gas to green energy*. Tech. rep. February. The Danish Ministry of Climate and Energy, 2011.

[31] Danish Energy Agency. *Energy Statistics 2014*. Tech. rep. 2014. URL: http://www.ens.dk/sites/ens.dk/files/info/tal-kort/statistik-noegletal/aarlig-energistatistik/energystatistics2014.pdf.

[32] F. Bouffard and F. Galiana. "An electricity market with a probabilistic spinning reserve criterion". In: *IEEE Transactions on Power Systems* 19.1 (2004), pp. 300–307.

[33] N. Amjady, J. Aghaei, and H. Shayanfar. "Market clearing of joint energy and reserves auctions using augmented payment minimization". In: *Energy* 34.10 (2009), pp. 1552–1559.

[34] J. Morales, A. Conejo, and J. Perez-Ruiz. "Economic Valuation of Reserves in Power Systems With High Penetration of Wind Power". In: *IEEE Transactions on Power Systems* 24.2 (2009), pp. 900–910.

[35] J. Warrington et al. "Robust reserve operation in power systems using affine policies". In: *IEEE 51st Annual Conference on Decision and Control (CDC)*. 2012, pp. 1111–1117.

[36] F. Partovi et al. "A stochastic security approach to energy and spinning reserve scheduling considering demand response program". In: *Energy* 36.5 (2011), pp. 3130–3137. ISSN: 0360-5442.

[37] H. Madsen et al. *Control of Electricity Loads in Future Electric Energy Systems, in Handbook of Clean Energy Systems*. Vol. 4. Wiley, 2014, pp. 2213–2238.

[38] A.-H. Mohsenian-Rad et al. "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid". In: *Smart Grid, IEEE Transactions on* 1.3 (2010), pp. 320–331.

[39] A. Conejo, J. Morales, and L. Baringo. "Real-Time Demand Response Model". In: *IEEE Transactions on Smart Grid* 1.3 (2010), pp. 236–242.

[40] J. M. Lujano-Rojas et al. "Optimum residential load management strategy for Real Time Pricing (RTP) demand response programs". In: *Energy Policy* 45 (2012), pp. 671–679. ISSN: 0301-4215.

[41] C. Chen, S. Kishore, and L. Snyder. "An innovative RTP-based residential power scheduling scheme for smart grids". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, pp. 5956–5959.

[42] H. Mohsenian-Rad and A. Leon-Garcia. "Optimal residential load control with price prediction in real-time electricity pricing environments". In: *IEEE Transactions on Smart Grid* 1.2 (2010), pp. 120–133.

[43]   R. Halvgaard et al. "Economic Model Predictive Control for building climate control in a Smart Grid". In: *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES* (2012), pp. 1–6.

[44]   D. Hammerstrom et al. *Pacific Northwest GridWise Testbed Demonstration Projects. Part I. Olympic Peninsula Project.* Tech. rep. Oct. 2007.

[45]   J. Hosking et al. "Short-term forecasting of the daily load curve for residential electricity usage in the Smart Grid". In: *Applied Stochastic Models in Business and Industry* 29.6 (2013), pp. 604–620.

[46]   O. Corradi et al. "Controlling Electricity Consumption by Forecasting its Response to Varying Prices". In: *IEEE Transactions on Power Systems* 28.1 (2013), pp. 421–429.

[47]   M. Zugno et al. "A bilevel model for electricity retailers' participation in a demand response market environment". In: *Energy Economics* 36 (2013), pp. 182–197.

[48]   G. Gross and F. D. Galiana. "Short-term load forecasting". In: *Proceedings of the IEEE* 75.12 (1987), pp. 1558–1573.

[49]   T. Hong and S. Fan. "Probabilistic electric load forecasting: A tutorial review". In: *International Journal of Forecasting* 32.3 (2016), pp. 914–938.

[50]   S. Borenstein. "The long-run efficiency of real-time electricity pricing". In: *The Energy Journal* (2005), pp. 93–116.

[51]   J. S. Vardakas, N. Zorba, and C. V. Verikoukis. "A survey on demand response programs in smart grids: pricing methods and optimization algorithms". In: *IEEE Communications Surveys & Tutorials* 17.1 (2015), pp. 152–178.

Part II

Publications

PAPER A

Determining Reserve Requirements in DK1 Area of Nord Pool Using a Probabilistic Approach

Authors:
Javier Saez-Gallego, Juan M. Morales, Henrik Madsen,Tryggvi Jónsson

Determining Reserve Requirements in DK1 Area of Nord Pool Using a Probabilistic Approach

Javier Saez-Gallego[1], Juan M. Morales[1], Henrik Madsen[1], and Tryggvi Jónsson[2]

Abstract

Allocation of electricity reserves is the main tool for transmission system operators to guarantee a reliable and safe real-time operation of the power system. Traditionally, a deterministic criterion is used to establish the level of reserve. Alternative criteria are given in this paper by using a probabilistic framework where the reserve requirements are computed based on scenarios of wind power forecast error, load forecast errors and power plant outages. Our approach is first motivated by the increasing wind power penetration in power systems worldwide as well as the current market design of the DK1 area of Nord Pool, where reserves are scheduled prior to the closure of the day-ahead market. The risk of the solution under the resulting reserve schedule is controlled by two measures: the Loss-of-Load Probability (LOLP) and the Conditional Value at Risk (CVaR). Results show that during the case study period, the LOLP methodology produces more costly and less reliable reserve schedules, whereas the solution from the CVaR method increases the safety of the overall system while decreasing the associated reserve costs, with respect to the method currently used by the Danish TSO.

A.1   Introduction

Electricity is a commodity that must be supplied continuously at all times at certain frequency. When this requirement is not fulfilled and there is shortage

---

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[2]Meniga ehf. Kringlan 5 103 Reykjavik, Iceland

of electricity, industrial consumers can face the very costly consequences of outages: their production being stopped or their systems collapsed. Households will experience high discomfort and losses too. From a different point of view, service interruptions also affect electricity producers as they are not able to sell the output of their power plants. Therefore, it is of high importance that the demand is always covered. The main tool for transmission system operators to avoid electricity interruptions is the allocation of operating reserves. In practice, scheduling reserves means that the system is operating at less than full capacity and the extra capacity will only be used in case of disturbances.

The term *operating reserves* is defined in this paper as "the real power capability that can be given or taken in the operating time frame to assist in generation and load balance and frequency control" [A1]. The types of reserves are differentiated by three factors: first, the time frame when they have to be activated ranging from few seconds to minutes; secondly, their activation mode, either automatically or manually; finally, by the direction of the response, upwards or downwards. Members of the European Network for System Operators for Electricity (ENTSO-E) and more specifically, the Danish Transmission System Operator (TSO), follow this classification criterion. Primary control is activated automatically within 15 seconds and its purpose is to restore the balance after a deviation of ± 0.2 mHz from the nominal frequency of 50Hz. Secondary control releases primary reserve and has to be automatically supplied within 15 minutes or 5 minutes if the unit is in operation. Manual reserve releases primary and secondary reserves and has to be supplied within 15 minutes. In Denmark, this type of reserve is often provided by Combined Heat-and-Power (CHP) plants and fast start units. The activated manual reserves are often referred as regulating power. This paper deals with the total upward reserve requirements, namely the sum of primary, secondary and manual reserves, neglecting the short-circuit power, reactive and voltage-control reserves. The result of the proposed optimization models, namely the schedule of reserves, refers to the total MW of upward reserve required. It is assumed that the reserve acts instantaneously to any generation deficit and no activation times are considered.

Currently the provision of reserve capacity in the DK1 area of Nordpool obeys the following rules, which can be found in the official documents issued by the Danish TSO [A2]. The requirements for primary and secondary reserve are $\pm 27 MW$ and $\pm 90 MW$, respectively. The provision of tertiary or manual reserves follows the recommendations in both the ENTSO-E Operation Handbook and the Nordic System Operation Agreement [A3], where it is stipulated that each TSO must procure the amount of tertiary reserves needed to cover the outage of a dimensioning unit in the system (the so-called $N-1$ *criterion*), be it a domestic transmission line, an international interconnection or a generating unit. The inspection of the historical data reveals that this criterion roughly results in an amount of tertiary reserve in between 300 and 600 MW.

The methodologies discussed in this article are mainly targeted to the current structure of the Danish electricity market, where reserve markets are settled independently of and before the day-ahead energy market, implying that, at the moment of scheduling reserves, no information about which units will be online is known. At the market closure, the TSO collects bids from producers willing to provide reserve capacity, and selects them by a cost merit-order procedure. Most of the existing literature focuses on co-optimizing the unit commitment and reserve requirements at the same time; such methods however cannot be applied under the current design of the Danish electricity market.

This paper is also motivated by the increasing penetration of wind power production in Europe and, in particular, in Denmark. As a matter of fact, the commission of the European Countries has set an ambitious target such that the EU will reach 20% share of energy from renewable sources by 2020, and in Denmark the target is 30% [A4]. As the share of electricity produced by renewables increases, several challenges must be faced. Non-dispatchable electricity generation cannot ensure a certain production at all times, but instead depends on meteorological factors. The stochastic nature of such factors inevitably leads to forecast errors that will likely result in producers deviating from their contracted power, thus causing the system to be imbalanced. Solutions call for methods capable of managing the uncertainty that wind power production and other stochastic variables induce into the system.

The main contributions of this paper are the following:

1. A probabilistic framework to determine the total reserve requirements independently to the generation power schedule in a power system with high penetration of wind production. The reserve levels in Denmark are currently computed by deterministic rules such as allocating an amount of reserve equal to the capacity of the largest unit online [A2, A3]. Another example is the rule used in Spain and Portugal, where the upward reserve is set equal to 2% of the forecast load plus the largest unit in the system. These rules are designed for systems with very low penetration of renewable energy and fairly predictable load, where the biggest largest need for reserve capacity arises from outages of large generation units. With the increasing share of renewables (and decentralized production in general) in the generation portfolio, renewables will naturally have a larger influence on the system imbalance. Hence, the non-dispatchable and uncertain nature of these plants needs to be accounted for when reserve power is scheduled [A5]. Previous studies perform a co-optimization of the energy and reserve markets, either in a deterministic manner [A6] or in a probabilistic way [A7, A8, A9, A10, A11, A12, A13]. However, these methods cannot be applied directly to the DK1 area of Nord Pool since the reserve

market and the day-ahead energy market are cleared independently at different times and by different entities. The methodology in [A14] is not suitable either as the Capacity Probability Table (COPT) refers to the units that are online; this information is not available to the Danish TSO at the time of clearing the reserve market.

2. A flexible scenario-based approach for modeling system uncertainty, which takes into account the limited predictability of wind and load, and plausible equipment failures. Moreover, the distributions from which the scenarios are generated are time-dependent, being the distributions of the scenarios of forecasts errors of load and wind power production non-parametric and correlated. The authors of [A9] characterize the uncertainty in the system only by scenarios of wind power forecast errors. Load and wind generation uncertainty is described in [A8, A15] by independent Gaussian distributions and not in a scenario framework. Other authors [A7, A15] use outage probabilities as a constant parameter for each unit and for each hour. The authors of [A14] represent the forecast error distributions of the load and wind generation by a set of quantiles, assuming both distributions are independent.

3. Equipment failures are modeled as the amount of MW that fail in the whole system due to the forced outages of generating unit. This way we can model and simulate simultaneous outages. Furthermore, the distribution of failures is dependent on time. Existing literature takes into account just one or two simultaneous failures [A7, A8, A15] or several [A14].

4. Two different methods for controlling the risk of the resulting capacity reserve schedule. The first one imposes a target on the probability of load shedding as in [A7], while the second one is based on the Conditional Value at Risk of the reserve cost distribution. The latter method minimizes the societal costs, while penalizing high cost scenarios given a certain level of risk aversion.

The remaining of the paper is organized as follows. Section II presents two different optimization models for reserve determination. Section III describes the methodology to generate scenarios of load forecast error, wind power forecast error and equipment failures, which altogether constitute the input information to the proposed reserve determination models. Section IV elaborates on the estimation of the cost of allocating and deploying reserves. Section V discusses the results and comments on the implications of applying the two reserve determination models to the Danish electricity market. Conclusions are summarized in Section VI.

A.2   Modeling Framework

This section presents two formulations for determining the reserve requirements in DK1, both of them solved using a scenario-based approach. The first limits the Loss-Of-Load Probability (LOLP), while the second one minimizes the Conditional Value at Risk (CVaR) of the cost distribution of reserve allocation, reserve deployment and load shedding. Both models are meant to be run to clear the reserve market and can be used by the TSO to decide on how many MW of reserve should be scheduled. In Denmark, where the study case in this paper is focused, the reserve market is cleared previous to and independently of the day-ahead energy market. This implies that the unit commitment problem is not addressed at the time when the reserve market closes and thus neither is it in this paper.

A.2.1   LOLP Formulation

The objective is to minimize the total cost of allocating reserves,

$$\underset{R_i}{\text{Minimize}} \sum_{i}^{M} \lambda_i R_i, \tag{A.1a}$$

where R_i is a variable representing the total amount of reserve assigned to producer i and λ_i is the price bid submitted to the reserve market by this producer. M is the total number of bids. The objective (A.1a) is subject to the following constraints

$$R_i \le R_i^{max} \qquad\qquad \forall i \tag{A.1b}$$

$$R^T = \sum_{i}^{M} R_i \tag{A.1c}$$

$$\text{LOLP} = \int_{R^T}^{\infty} f(z)\,dz \tag{A.1d}$$

$$\text{LOLP} \le \beta \tag{A.1e}$$

$$R^T \ge 0 \qquad\qquad \forall i. \tag{A.1f}$$

The set of inequalities (A.1b) indicates that the amount of reserve provided by producer i cannot be greater than its bid quantity. The total reserve to

be scheduled is defined in (A.1c) as the sum of the reserve contribution from each producer. The probability density function of balancing requirements is represented by $f(z)$, and hence the integral from $z = R^T$ to $z = \infty$ is the probability of not scheduling enough reserves to cover the demand, namely the loss-of-load probability, defined in (A.1d). It is constrained by a parameter target $\beta \in [0, 1]$ in Equation (A.1e), which is to be specified by the transmission system operator. The smaller β is, the more reserves are scheduled, as the LOLP is desired to be small. On the other hand, if β is equal to 1, no reserves are allocated at all.

The optimal solution to problem (A.1a) can be found analytically, under the assumption that the objective function (A.1a) is monotonically increasing with respect to the total scheduled reserve R^T (i.e., reserve capacity prices are non-negative) and because the LOLP is a decreasing function with respect to R^T (note that $f(z)$ is a density function and therefore, always non-negative). Indeed, under the above assumption, greater R^T implies greater costs, thus R^T is pushed as low as possible until the relation LOLP $= \beta$ is satisfied. Therefore, at the optimum, it holds that $\beta = \int_{R^T}^{\infty} f(z)\,dz$ or similarly $1 - \beta = F(R^T)$ being $F(Z) = P(Z \le z)$ the cumulative distribution function of Z (the required reserve). Finally, since β is a given parameter, then the solution is $R^{T^*} = F^{-1}(1 - \beta)$.

In practice, $f(z)$ can be difficult to estimate in a closed form; one way of dealing with this issue is to describe the uncertainty by scenarios. Let z_w be the reserve required to cover balancing needs in scenario w and π_w the associated probability of occurrence. Then the optimal solution to problem (A.1a) boils down to the quantile $1 - \beta$ of the scenarios. In other words, let $\hat{F}(Z) = P(Z \le z)$ be the empirical cumulative distribution function of the set of scenarios $\{z_w\}$ with $\hat{F} : (-\infty, \infty) \to (0, 1)$, then the analytical solution is $R^{T^*} = inf\{z \in (-\infty, \infty) : (1 - \beta) \le \hat{F}(z)\}$.

Finally, we define the Expected Power Not Served (EPNS) as the expected amount of MW of balancing power needed during one hour which cannot be covered by the scheduled reserves. It can be computed, once the total scheduled reserve R^{T^*} has been obtained, as

$$\text{EPNS} = \int_{R^{T^*}}^{\infty} z f(z)\,dz. \tag{A.2}$$

In the case where the uncertainty of reserve requirements is characterized by scenarios, the EPNS can be determined as EPNS $= \sum_{w \in S}(z_w - R^{T^*})\pi_w$, $S = \{w \in W : z_w > R^{T^*}\}$.

A.2.2 Conditional Value at Risk (CVaR) Formulation

The following reserve determination model corresponds to a two-stage stochastic linear program where each scenario is characterized by a realization of the stochastic variable Z "*reserve requirements*". Variable R^T represents the amount of MW that the TSO should buy at the reserve market. In the jargon of stochastic programming, this variable is referred to as a *first stage variable*, or equivalently, as a *here-and-now* decision, i.e., a decision that must be made before any plausible scenario z_w of energy shortage is realized. This models the fact that reserve capacity is to be scheduled before the scenarios of reserve requirement are realized. For their part, the *second stage variables*, or recourse variables, r_w^T and L_w, are relative to each scenario w, and represent the deployed regulating power and the MW of shed load, respectively. Consequently, during the real-time operation of the power system, once a certain scenario w of wind power production, load and equipment failures realizes, reserve is activated r_w^T or some load is shed (L_w). In such a way, the first stage of our stochastic programming model represents the reserve availability market and the second stage represents the reserve activation market. Finally, the probability of occurrence of each scenario is denoted by π_w.

The objective function to be minimized is the CVaR_α of the distribution of total cost. By definition, the Value-at-Risk at the confidence level α (VaR_α) of a probability distribution is its α-quantile, whereas the CVaR_α is the conditional expectation of the area below the VaR_α. The CVaR is known to have better properties than the VaR [A16] and hence, it is used in this paper. Parameter $\alpha \in [0, 1)$ represents the risk-aversion of the TSO, i.e. the greater α is, the more conservative the solution will be in terms of costs. The objective is to minimize the CVaR_α of the distribution of the total cost:

$$\underset{R^T, R_g, r_w^T, r_{gw}, L_w, \xi, \eta_w, Cost_w}{\text{Minimize}} \text{CVaR}_\alpha = \xi + \frac{1}{1-\alpha} \sum_{w=1}^{W} \pi_w \eta_w \tag{A.3a}$$

where ξ is, at the optimum, the $\alpha-$Value at Risk (VaR_α) and η_w is an auxiliary variable indicating the positive difference between the VaR and the cost associated with scenario w. The cost of each scenario, named Cost_w, is computed in (A.3b) as the sum of the cost of allocating and deploying reserve capacity plus the cost incurred by involuntary load shedding. The objective (A.3a) is subject to the following constraints:

$$\text{Cost}_{\text{w}} = \sum_{j=1}^{J} \lambda_j^{\text{cap}} R_j + \sum_{g=1}^{G} \lambda_g^{\text{bal}} r_{gw} + V^{\text{LOL}} L_w \qquad \forall w \qquad (\text{A.3b})$$

$$R^T = \sum_{j=1}^{J} R_j \qquad\qquad (\text{A.3c})$$

$$r_w^T = \sum_{g=1}^{G} r_{gw} \qquad \forall w \qquad (\text{A.3d})$$

$$0 \le R_j \le I_j^R \qquad \forall j \qquad (\text{A.3e})$$

$$0 \le r_{gw} \le I_g^r \qquad \forall g, w \qquad (\text{A.3f})$$

$$\text{Cost}_{\text{w}} - \xi \le \eta_w \qquad \forall w \qquad (\text{A.3g})$$

$$r_w^T \le R^T \qquad \forall w \qquad (\text{A.3h})$$

$$z_w - r_w^T \le L_w \qquad \forall w \qquad (\text{A.3i})$$

$$0 \le L_w, \eta_w \qquad \forall w. \qquad (\text{A.3j})$$

The first term in Equation (A.3b) represents the cost of allocating R^T MW of reserve capacity. The TSO has information about the marginal cost of allocating reserves at the closure time of the reserve market, as it is given by the bids submitted by producers to the reserve market. These bids, however, are treated confidentially and hence, were not available for the study case. Consequently, we estimate a cost function for the supply of reserve capacity from the historical series of clearing prices in the Danish reserve market. Naturally, this function must be monotonically increasing. The estimation of the parameters of this function is discussed in Section A.3. In order to keep formulation (A.3) linear, the marginal cost of reserve capacity is further approximated by a stepwise function consisting of J intervals of length I_j^R each, as indicated in (A.3e), and associated values λ_j^{cap}, which result from evaluating the estimated reserve cost function at the midpoint of each interval. The term $\sum_{j=1}^{J} \lambda_j^{\text{cap}} R_j$ represents thus the total cost of allocating R^T MW of reserves. Furthermore, the total allocated reserves are given by (A.3c). Note that the formulation would remain equal if the real bids were used instead of the estimated cost function. One could interpret λ_j^{cap} and I_j^R as the bid that producer j submit to the reserve market and R_j as the MW of reserve capacity provided this producer.

The second term of (A.3b) represents the reserve deployment cost. This cost is unknown at the time of clearing the reserve market and therefore, has to be estimated by as well. The estimation procedure is discussed in Section A.3. Similarly as before, λ_g^{bal} can be seen as the cost of deploying r_{gw} MW of reserve

in interval g and scenario w. The length of the intervals is I_g^r, as stated in (A.3f), having a total of G intervals. The total deployed reserve in scenario w is then given by (A.3d).

The third term of (A.3b) represents the cost of involuntary load curtailment. The parameter "Value of Lost Load" V^{LOL} expresses the societal cost of shedding 1 MWh of load. Often, the V^{LOL} is interpreted as the maximum price of upward regulation that is permitted to bid in the market, which in Denmark is 37 500 DKK/MWh or roughly 5 000 €/MWh. In Great Britain, the V^{LOL} is estimated to be from 1 400 £/MWh to 39 000 £/MWh depending on the type of consumer and the time of the year [A17]. A study performed on the Irish power system indicates that, on average, the V^{LOL} is 12.9 €/KWh [A18]. In this paper, a sensitivity analysis is performed to study how the parameter V^{LOL} affects the solution.

Constraint (A.3g) is used to linearly define the CVaR_α as in [A19]. Variable η_w is equal to zero if $Cost_w < \xi$, and equal to $Cost_w - \xi$ if $Cost_w \geq \xi$; in other words, η_w accounts for the difference between the cost in each scenario and the VaR_α when such a difference is positive. Equation (A.3h) indicates that the deployed reserve cannot be greater than the scheduled reserves. Equation (A.3i) is used to define the shed load L_w (or similarly, the lack of reserve). At the optimum, L_w is equal to zero if $z_w \leq R^T$, implying that $z_w = r_w^T$; when $z_w > R^T$, then L_w is equal to the difference between the reserve requirements and the deployed reserves, namely, $z_w - r_w^T$. In this case, the deployed reserve is equal to the scheduled capacity reserve $r_w^T = R^T$.

Once the CVaR problem has been solved, one can calculate the EPNS by multiplying the lacking reserve from each scenario L_w^* at the optimum by its probability of occurrence π_w, namely $\text{EPNS} = \sum_{w=1}^{W} \pi_w L_w^*$.

A.3 Cost Functions

This section elaborates on the estimation of the cost of allocating reserves and the cost of providing regulating power.

In practice, the bids that producers submit to the reserve market, that are used to define (A.3b), are available to the Danish TSO at the closure of the reserve market. Nevertheless, this information is not available to us for the case study presented in Section 5. Consequently, in order to adjust the optimization models to the available data and test the efficiency of such, the bids of producers are substituted by a cost function, being $g^R(z)$ the cost in €/MW of allocating of z

MW of upward reserve. This function is built from the series of clearing prices in the Danish reserve market, which is publicly available in [A20]. The price per MW of reserve capacity is assumed to be quadratic for simplicity, in particular, of the form $g^R(z) = az^2$. The coefficient $a = 1.25 \times 10^{-5}$ is estimated using least-square method.

A staircase linear approximation of $g^R(z)$ is then used in order to maintain formulation (A.3) linear. The reason for the choice of a staircase function is that, due to market rules, the aggregated bidding curve is also a staircase function. The feasible region of R^T is split into intervals of length $I_j^r = 30$ MW $\forall j$, ranging from 0 to an upper bound of R^T chosen to be 1890MW. For every interval, we compute the estimated marginal cost λ_j^{cap} at the mid point of the interval and set it to the height of each stair. Figure A.1 shows on the left the data points and the estimated curve of prices in Euro per MW of allocated reserve. The data appears very homoscedastic, for example, the variability around $R^T = 300$ MW is much lower than around $R^T = 450$ MW. Nevertheless, the curve is not intended to capture all the variability of the data but to represent a plausible aggregated bidding curve in the reserve market.

The cost of deploying reserves is a necessary input to Equation (A.3b) and must be estimated in practice, as it is unknown at the time of clearing the reserve market. We denote the marginal cost of deploying zMW of reserve by $g^r(z)$. In order to compute this cost, we approximate the clearing prices of the regulating market by a quadratic term plus an intercept, $g^r(z) = \mu + bz^2$. The parameters are estimated using the least-squares method and data relative to the clearing price of the regulating market in DK1 collected from [A20]. The regulating power traded versus the market price is displayed in dots in Figure A.1. The resulting estimates of the parameters are $\mu = 48.2$ and $b = 6 \times 10^{-4}$. In order to maintain the optimization problem (A.3) linear, $g^r(z)$ is linearized as a staircase function, which is shown in the right plot in Figure A.1. More complex functions could possibly be estimated, for example using time and other external factors as explanatory variables. This implementation is left for future work.

Lastly, it should be noted the difference in scale between the settlement prices of the two markets. On average, the price of allocating reserve is approximately 40 times lower than deploying them. Allocating reserve is cheaper as no energy is actually deployed but only the capacity is allocated.
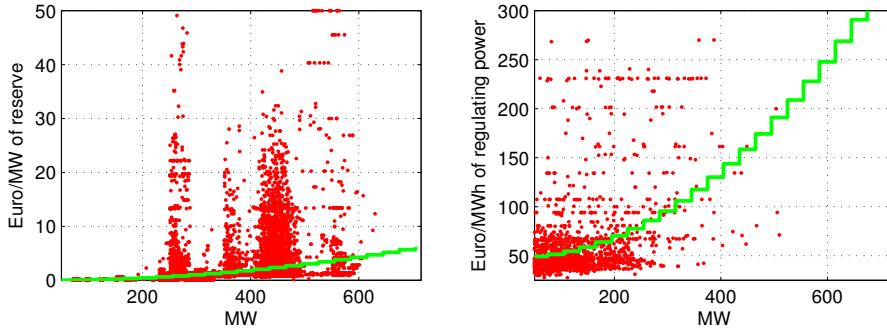
Figure A.1: On the left, the settlement price and the allocated reserves in the reserve market. Dots represent data and the staircase curve constitutes the estimated fit. On the right, the data relative to the settlement price of the regulating power market is shown in dots while the fitted stair-wise curve is displayed on top.

A.4   Scenarios of Reserve Requirements

The total reserve capacity that should be scheduled and allocated in advance is mainly affected by three factors or uncertainty sources: the forecast error of wind power production, the forecast error of electricity demand and the forced outages of power plants, namely failures of the plants that cause their production to stop. They are all taken into account in this paper.

Suppose that wind power production is the only source of uncertainty. We assume that wind power producers bid their expected production in the day-ahead market. If the actual wind power production is greater than what was expected, then there will be extra power to sell and hence a reduction in power supply (down-reserves) will be required to maintain the system balance; if, on the other hand, the realized wind is lower than the expected value, upward reserves will be required. In other words, if the forecasts were perfect and the errors equal to zero, no reserve would be needed. Likewise, as the forecast errors increase, more reserves are required to account for the possible mismatches between supply and demand. Similarly with the power load: it is assumed that the amount of power traded in the day-ahead energy market is equal to the expected power load demand, therefore positive errors imply upward reserve requirements while negative errors imply downward reserve requirements. The predicted outages of power plants lead directly to upward reserve requirements.

The probability distributions of the forecast errors and the power plant outages

can be combined into one by convolving them, resulting in a function which will represent the probability distribution of the combined balancing requirements $f(z)$. In this paper, we draw scenarios from each individual distribution and sum them up to produce scenarios characterizing the total reserve requirements in the DK1 area of Nord Pool. A scenario-based approach is chosen because the convolution of the probability distributions of the individual stochastic variables does not have a closed form and can be highly complex. The remaining of this section elaborates on how the individual scenarios are obtained.

A.4.1 Scenario Generation of Wind Power Production and Load Forecast Errors

In this subsection both the generation of scenarios of wind power production and load forecast errors are discussed. Scenarios from both stochastic variables are generated together to account for correlation between them.

Regarding the wind power production in DK1, point quantile forecast have been issued using a conditional parametric model, i.e., a linear model in which the parameters are replaced by a smooth unknown functions of one or more explanatory variables. The explanatory variables are on-line and off-line power measurements from wind turbines and numerical weather prediction of wind speed and wind direction. The functions are estimated adaptively. The errors are modeled as a sum of non-linear smooth functions of variables forecast by the meteorological model or variables derived from such forecasts. Further information about the employed modeling approach can be found in [A21, A22, A23].

The load in DK1 area has been modeled as a function of the temperature, the wind, and the solar radiation. The annual trend is modeled by a cubic B-spline basis with orthogonal columns. The daily variations are modeled as a combination of different sinusoids, one referring to each time of the day. The reader is referred to [A24] for a detailed description of the methodology used in this paper to model the electricity demand in DK1.

The scenarios of wind production and load are generated in pairs in order to account for their mutual correlation. Each scenario is composed by two variables and is built in three steps as in [A25]: first by a sample of a multivariate Gaussian distribution where the covariance matrix is estimated recursively as new observations are collected; then, by applying the inverse probit function of such sample, and finally by using the estimated inverse cumulative function of the desired variables.
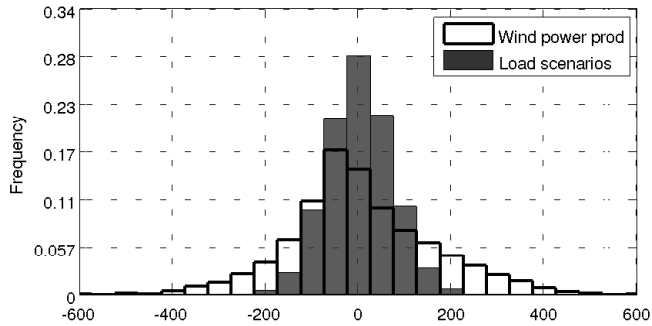
Figure A.2: The histogram of forecast error scenarios of wind power produc-
tion and load during one specified hour shown are shown in gray
and white, respectively.

Figure A.2 shows the distribution of the scenarios of forecast errors of wind
power production and load in gray and white respectively, during the 15th Dec
2011 from 13:00 to 13:59. Note that the distribution of the forecast error of
wind power is wider, indicating that, in general, wind power production has
a greater impact on reserve requirements than the load. On average, forecast
error scenarios of wind exhibit five times more variance than the load scenarios.
Finally, note that both distributions are centered around zero.

A.4.2   Scenario Generation of Power Plant Outages

The modeling of individual power plant outages requires historical data and
specific information on each power plant which might not always be available
to the TSO. Secondly, it requires computing an individual model for each unit,
thus increasing complexity significantly. Thirdly, it requires information about
which units will be on/off during the operation horizon, which is not available
at the clearing of the Danish reserve market. An alternative approach taken in
this paper is to model the total amount of MW that fail in the entire system
by aggregating all the units into one. The predicted MW failed in the entire
system depend on time and on the load. Historical data of power plant outages
can be found at the Urgent Market Messages service of Nord Pool [A26]. The
left plot in Figure A.3 shows the forced outages in MW during 2011. The area
inside the box corresponds to the MW failed in May, also zoomed in the right
plot. In the course of 2011, there was 92% of the hours where 0 MW failed;
during the rest of the hours, either an outage of a single unit, a partly outage
or simultaneous outages occurred.
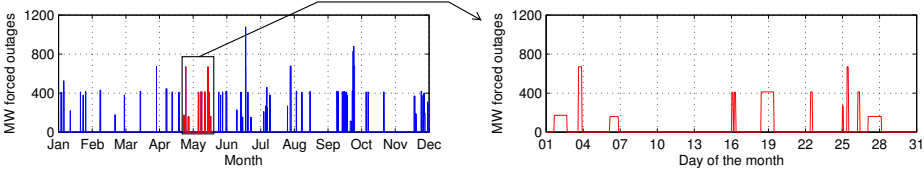
Figure A.3: Historical data of MW forced to fail in DK1. On the left, data relative to year 2011 is shown. The area inside the rectangle is zoomed on the right plot and refers to May 2011.

The procedure proposed in this paper to model power plant outages is comprised of two steps. In the first step, we model the presence or absence of an outage. In the second step, we model the amount of MW failed, conditioned on the fact that a failure occurred. In [A27] we explored alternative methodologies based on Hidden Markov Models that were proven to perform worse at predicting future outages.

The variable modeled in the first step Y_t is defined as

$$Y_t = \begin{cases} 1 & \text{if failure occurs at time t} \\ 0 & \text{otherwise.} \end{cases} \tag{A.4}$$

It is natural to assume that Y_t follows a Bernoulli distribution, $Y_t \sim bern(p_t)$, and therefore, it is appropriate to model Y_t as a Generalized Linear Model [A28]. The link function chosen is the *logit* function. The explanatory variables are the hour of the day, the day of the week and the month, all represented through sinusoidal curves. Sinusoidal terms of the form $k^{(1)}cos(2\pi\frac{hour_t}{24})$, $k^{(2)}cos(2\pi\frac{day_t}{7})$ and $k^{(3)}cos(2\pi\frac{month_t}{12})$ with $k^{(1)} = 1...24$, $k^{(2)} = 1...7$, and $k^{(3)} = 1...12$, are considered, also using the *sin* function. Only the most relevant were kept using a likelihood ratio test as in [A28]. The final model is

$$\eta_t = log\left(\frac{p_t}{1-p_t}\right) = \mu + \alpha_1 cos\left(\frac{2\pi day_t}{7}\right) + \alpha_2 sin\left(\frac{2\pi day_t}{7}\right) + \alpha_3 cos\left(2\frac{\pi month_t}{12}\right) +$$

$$\alpha_4 cos\left(5\frac{2\pi month_t}{12}\right) + \alpha_5 sin\left(\frac{2\pi month_t}{12}\right) + \alpha_6 sin\left(2\frac{2\pi month_t}{12}\right) +$$

$$\alpha_7 sin\left(3\frac{2\pi month_t}{12}\right) + \alpha_8 sin\left(4\frac{2\pi month_t}{12}\right) + \alpha_9 sin\left(5\frac{2\pi month_t}{12}\right). \tag{A.5}$$
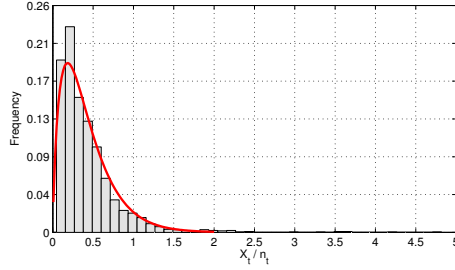
Figure A.4: Histogram of $\frac{x_t}{n_t}|y_t = 1$, namely the amount of MW failed divided by the load at time t, knowing that a failure has occurred. The curve represents the estimated Gamma distribution.

The reduced model indicates that the hour of the day is not significant when predicting the probability of an outage. The day of the week and the month are both significant variables. The parameters of the model are optimized using train data and updated everyday including data from the previous 24 hours during the test period.

The second stage of the model accounts for the amount of failed MW at time t, X_t, conditioned on the fact that a failure has occurred. Note that the more energy is demanded, the more power plants are online and more generators are subject to fail, meaning that the load n_t will affect our predictions of X_t. The histogram of $(\frac{X_t}{n_t}|Y_t = 1)$ depicted in Figure A.4 clearly resembles the density of a Gamma distribution. Thus, we assume that $(\frac{X_t}{n_t}|Y_t = 1) \sim Gamma(s_t, k)$, where k is the shape parameter, common for all observations, and s_t the scale parameter at time t.

The probability density function of a Gamma distribution is defined as

$$f(x) = \frac{1}{\Gamma(k)s_t^k}x^{k-1}e^{-\frac{x}{s_t}}, \qquad (A.6)$$

with mean $\mu_t = ks_t$ and variance $\sigma^2 = ks_t^2$. The canonical link for the gamma distribution is the inverse link $\eta = 1/\mu$ [A28]. As in the previous binary model, the explanatory variables are several sinusoidal curves. Several approximate χ^2-distribution tests were performed to disregard irrelevant terms. The final model only including the significant terms is

$$
\begin{aligned}
\eta_t = \frac{1}{\mu_t} = &\mu + \alpha_1 cos\left(\frac{2\pi hour_t}{7}\right) + \alpha_2 cos\left(2\frac{2\pi hour_t}{7}\right) + \alpha_3 sin\left(\frac{\pi hour_t}{12}\right) + \\
&\alpha_4 sin\left(2\frac{2\pi hour_t}{12}\right) + \alpha_5 sin\left(\frac{3\pi hour_t}{12}\right) + \alpha_6 cos\left(\frac{2\pi day_t}{12}\right) + \\
&\alpha_7 cos\left(2\frac{2\pi day_t}{12}\right) + \alpha_8 sin\left(\frac{2\pi day_t}{12}\right) + \alpha_9 sin\left(3\frac{2\pi day_t}{12}\right) + \\
&\alpha_{10} cos\left(\frac{2\pi month_t}{12}\right) + \alpha_{11} sin\left(2\frac{2\pi month_t}{12}\right) + \alpha_{12} sin\left(3\frac{2\pi month_t}{12}\right).
\end{aligned}
$$

$$\text{(A.7)}$$

When predicting the ratio X_t/n_t, the hour, the week day and the month are statistically significant.

Scenarios are generated in an iterative process. Every day at 9:00 am, the parameters of both models are updated including data from the previous day. At this time, 5000 scenarios for each hour of the next day are generated, i.e., with lead time ranging from 16 to 40 hours. Each scenario corresponds to an independent simulation of a Bernoulli multiplied by a Gamma simulated value and by the predicted load n_t.

A.5   Results and Discussion

The performance of the proposed reserve determination models is assessed by comparing the reserve capacity scheduled by the models and the reserve capacity actually deployed in the DK1 area of Nord Pool. The latter is calculated as the sum of the activated secondary reserve, regulating power produced in DK1 and the regulating power exchanged through the interconnections with neighboring areas. Data pertaining to the activated secondary reserve and the total volume of regulating power can be downloaded from [A26] and [A20], respectively. The regulating power exchanged through the interconnections with Germany, Norway, Sweden and East Denmark is estimated using data from [A20]. More specifically, it is computed by subtracting the total power scheduled for each interconnector in the day-ahead market from the actual power that eventually flows through it. Primary regulation was not considered due to unavailability of the data. However, conclusions would be barely affected by considering the primary regulation as it is comparatively very low. In the remainder of the paper, a shortage event is defined as an hour when the scheduled reserves are lower than the actual reserve deployed. In practice, a shortage event does not

necessarily imply that load is shed. In the Nordic market, producers who are scheduled to provide a certain capacity in the reserve market must place an offer of regulating power of the corresponding size in the regulating market, 45 minutes before operational time. However, other players who are not committed to provide reserves in the reserve market may still bid in the regulating market and thereby provide regulating power. This case is not considered in this paper.

The presented case study has been performed on data spanning three years. The training period of the scenario-generating models goes from the 1st January 2009 at 00:00 CET to the 30st June 2011 at 23:00 CET. The test period covers week 36 of year 2011 and weeks 3, 17 and 22 of 2012, all of them randomly selected.

Recall that scenarios are generated every day at 9:00 am with a lead time of 16 to 40 hours, namely the next operational day. The optimization models are run using the same lead time, as if they were to be solved at the clearing of the reserve market. The solutions to the LOLP and CVaR models are compared with the actual deployed reserves in DK1 during the four testing weeks.

It is worth stressing that the proposed models for reserve determination focus on the total reserve requirements, which are triggered by unexpected fluctuations in the load and in the wind power production, and by outages of power plants. We do not distinguish, therefore, between primary, secondary and tertiary reserve. In the case of the LOLP-formulation, it is up to the TSO to decide how to split the total reserve requirements into the different types of reserve that may be considered. Likewise, the CVaR-formulation can be easily tuned to represent the three types of reserves through the estimated cost functions. Indeed, if it is much easier for plants to participate in the tertiary reserve market, because providing tertiary reserve is cheaper than providing primary and secondary reserve, then the distinction between these should be made through the supply cost function for reserve, with the tertiary reserve being cheaper than the secondary and primary ones. If, on the contrary, it is much easier for the plants to participate in the tertiary reserve market for reasons that cannot be translated into costs, then the required primary and secondary reserve should be treated as input information in the CVaR-method and subtracted from the total reserve requirements.

A.5.1   LOLP-model Results

This subsection shows the results of applying the LOLP model introduced in Section A.2.1. The model was run during the four testing weeks and the simulation results are included in Figure A.5. The actual deployed reserves in DK1 and the total scheduled reserves by Energinet.dk are also shown in the figure.
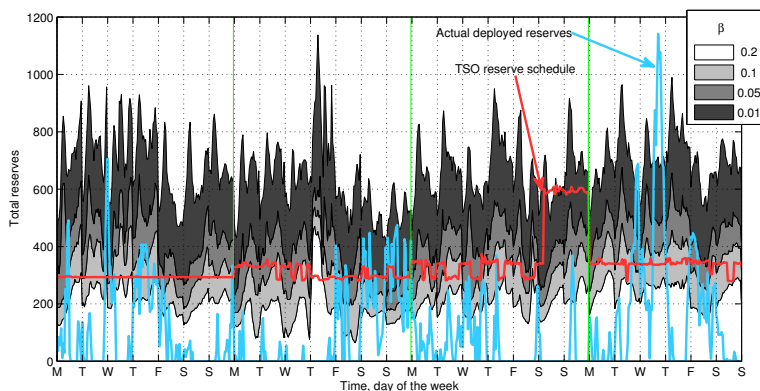
Figure A.5: The deployed reserves and the actual scheduled reserves by En-
erginet.dk are plotted as indicated in the text boxes. The shaded
areas in the background represent the solution of the LOLP model
for different values of β. Weeks are separated by vertical lines.

The shaded areas in the background represent the solution to the LOLP model
for different values of β. Weeks are separated by vertical lines. The amount
of scheduled reserve varies substantially depending on the level of uncertainty
of the wind, the load and the power plant outages. As an example, on the
Thursday of the fourth week, the prediction of load and wind power production
happened to be wrong, and up to 1 100 MW of regulating power were needed. In
this special case, both the LOLP solution and Energinet.dk's reserve scheduling
criterion led to a shortage event.

The reliability plot in Figure A.6 shows i.e. the desired or expected LOLP
(parameter β), against the observed LOLP, namely, the number of shortage
events divided by the time span. The whole data set was used to compute
this reliability plot. The ideal case is illustrated by the dashed line where both
quantities, expected and observed, are equal. The actual performance of the
model is represented by the continuous line, which is fairly close to the ideal
one, indicating that the expected probability of reserve shortage is well adjusted
to the observed one.

It is worth mentioning that the value of the parameter β is directly connected
to the reliability level of the underlying power system. With this in mind, a
simple rule for the TSO to decide on an appropriate value for this parameter
would read as follows: divide the number of hours in a year where shedding load
is tolerated to happen by the total number of hours in the year. This simple
rule would roughly indicate the probability that, in each hour, the need for up-
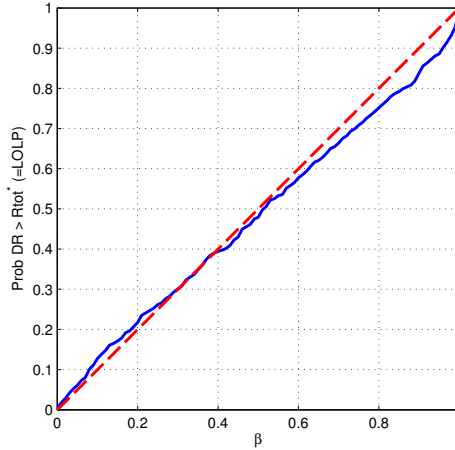
Figure A.6: Observed LOLP vs expected, namely parameter β

ward balancing power exceeds the scheduled reserves. The UCTE suggests that enough reserve should be scheduled to manage energy deviations in $99,9\%$ of all hours during the year [A29]. For example, if the TSO tolerates that there are around 96 hours during a year where some load might not be covered, then the resulting LOLP would be equal to 0.01. The case presented in [A14] is performed using LOLP=$\{1, 0.5, 0.1\}$. The authors in [A30] consider five scenarios of demand, where the LOLP is in between 0.005 and 0.016. Note, however, that the LOLP does not account for how many MW of load are shed or the cost of such load shedding events.

Next, we perform a sensitivity analysis to asses how changes in β affect the solution. We choose several plausible values of β which are displayed in the first column of Table A.1, and then compute the optimal reserve schedule for each of them. The four testing weeks are considered and the results shown are averaged by the number of days. The second column shows the numbers of shortage events. The cost of allocating the reserve given by the LOLP model is displayed in the third column. It is computed using the cost function $g^R(z)$, presented in Section A.3. The fourth column shows the cost of deploying the actual reserve computed using the function $g^r(z)$, which estimation is discussed in Section A.3. Lastly, the MW shed, or in other words, the number of MW of actual deployed reserve exceeding the LOLP solution, is presented in the fifth column.

Note that a decrease in the parameter β implies that the solution becomes more conservative and hence, more reserve will be scheduled. For this reason, as β decreases, the number of shortage events decreases too, at the expense

β	Shortage events	Alloc. cost in €×10^3	Deploy. cost in €×10^3	MW not covered
0.2	5.214	1.194	95.643	862.056
0.15	4.107	2.077	110.966	668.879
0.089	2.678	4.528	134.826	423.645
0.07	1.928	6.132	143.782	346.910
0.05	1.464	9.150	153.939	270.271
0.01	0.428	33.032	187.791	85.464

Table A.1: The first column shows several values of parameter β which is an input to the LOLP model. The second column presents the number of shortage events on average per day. The third column includes the cost of allocating the amount of reserves given by the LOLP model on average per day. The fourth column displays the cost of deploying the actual reserve requirements. The number of MW not covered by the scheduled reserve on average per day when using the LOLP solution are shown on the fifth column.

of increasing the allocation and deploy cost. On the other hand, the amount of MW not covered by the scheduled reserves, collated in the fourth column, decreases as β diminishes.

Next, we compare the solution to the LOLP model with the solution given by Energinet.dk. TSO's solution incurs 75 shortage events during the four testing weeks, or equivalently 2.67 shortage events per day. The estimated total cost of allocating reserve is 4 355 €; the estimated deployment cost is 132 350 €, and the MW not covered 422.39.

For the same number of shortage events, the LOLP gives a worse solution that Energinet.dk's solution: the allocation costs are 3.97% higher, the deployment cost 1.18% higher and the MW shed increase in 1.15 MW per day. This means that during the four testing weeks, the LOLP methodology underperforms the solution given by Energinet.dk in terms of reliability and economic efficiency. The main advantage that the LOLP method brings is the analogy of the parameter β with the probability of a shortage event to occur, which is a very easy risk measure to interpret. On the other hand, the method has two drawbacks. As discusses before, its solution does not depend on the cost of allocating reserves, namely on λ_i or on the estimated cost function $g(z)$ (as long as it is increasingly monotonic). Neither it depends on the cost of deploying reserves. The solution only depends on the parameter β, as the relation $LOLP = \beta$ in the optimization problem (A.1) will always be satisfied at the optimum, no matter what the cost is. The second disadvantage is that load shedding costs are not taken into

account. These drawbacks are overcome by the CVaR model, for which results are presented in the next subsection.

A.5.2 CVaR-Method Results

In this section we discuss the results of the CVaR-based reserve determination method, which has been presented in Section A.2.2, and compare them with the deployed reserves that were actually needed in DK1 during the simulation horizon.

The CVaR model needs as input two parameters which, in practice, are to be determined by the TSO: α, which controls the CVaR risk measure and represents the risk aversion of the TSO, and V^{LOL}, which accounts for the cost in € of shedding 1 MW of load. We performed a sensitivity analysis to determine how changes in these parameters affect the level of procured reserve. The model was run for values of $\alpha = \{0, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99\}$ and $V^{lol} = \{200, 500, 1\,000, 2\,000, 5\,000\}$ €/MW.

The cost of allocating and deploying reserve capacity, the cost of shedding load, and the total cost, are displayed on the upper-left, upper-right, lower-left and lower-right plot of Figure A.7. The cost is shown on the y-axis in €$\times 10^3$, while the risk parameter α is shown on the x-axis. Each line represents a cost of the reserve schedule solution for a certain V^{LOL}. All costs are averaged by the number of days in the test period. As the TSO becomes more risk averse, i.e., as α increases, the allocation and deployment costs increase, because a larger amount of reserve is procured. The same occurs as V^{LOL} increases, since shortage events become more penalized and more reserves are scheduled to avoid them. On the other hand, the cost of curtailing load, depicted in the left-lower plot, decreases as α increases, but does not necessarily increase as V^{LOL} increases. The reason for this is that, even though the amount of curtailed load decreases as V^{LOL} increase, the product $V^{\text{LOL}} \times L_w$ representing the cost of curtailing load in scenario w may still increase.

The total cost shown in the down-right subplot of Figure A.7 is computed by summing up the reserve allocation, the reserve deployment and the load shedding costs. In general, the total cost increases as the risk-aversion parameter α increases. However, this is not always the case when $V^{LOL} = \{1\,000, 2\,000, 5\,000\}$ €/MW. The reason for this discrepancy is that the generated scenarios do not represent the potential need for reserve capacity accurately enough. Adding more variables to the scenario representation of the reserve requirements, increasing the amount of scenarios or adding more weeks to the test period could solve this issue, in particular, they underestimate the amount of

upward balancing power that may potentially be required, as confirmed by the plot in Figure A.6. Finally, one should notice that changes in the total cost are mainly driven by changes in the V^{LOL}, while changes in α have smaller impact on the solution.

The scheduled reserves when $V^{\text{LOL}} = 500 \text{ €}/\text{MW}$ and $V^{\text{LOL}} = 5\,000 \text{ €}/\text{MW}$, over time, are displayed in shadowed areas in the upper and lower plot of Figure A.8. The actual deployed reserve and the reserve scheduled by Energinet.dk are drawn on top. Weeks are separated by vertical lines. It is interesting to note how the reserve schedule given by the CVaR_{α} based reserve determination model changes as V^{LOL} changes. When minimizing the CVaR of the cost distribution of reserve allocation and deployment, and load shedding costs, an increase in V^{LOL} makes the load shedding costs have more weight in the total costs, and hence the events of reserve shortage will be penalized to a larger extent. When V^{LOL} is low, those events are less relevant and the curves look more flat. Another reason for the flatness of the curves is the linear approximation of the cost functions $g^R(z)$ and $g^r(z)$, both introduced in Section A.3. The increase in cost when increasing the reserve in one unit is much higher when jumping from one step of the stepwise function to another, than when the function remains in the same step. The step lengths are defined in (A.3e) and in (A.3f). A finer linearisation of such cost functions by reducing the step length would solve this issue.

The number of interruption events and the amount of load that is involuntarily shed are further analyzed in Figure A.9, in the left and right plots, respectively. As α and/or V^{LOL} increase, both the number of interruptions and the MW shed on average per day decrease. Under the assumption that only the producers committed to the reserve market are allowed to participate in the regulating market, the Danish TSO's solution incurs 75 shortage events during the four testing weeks. On average per day, the Danish TSO's solution incurs 2.67 shortage events, with an estimated total cost of allocation equal to $4\,355$ €, an estimated deployment cost of $13\,2350$ €, and amount of load shed of 422.39 MW. The CVaR-method produces cheaper results in terms of total cost. The CVaR solution is from 3.38% cheaper with $\{\alpha = 0.99, V^{LOL} = 200 \text{ €}/\text{MW}\}$, to 82.9% for $\{\alpha = 0.99, V^{\text{LOL}} = 5\,000 \text{ €}/\text{MW}\}$, compared to the solution given by the TSO. Note that the CVaR method tends to schedule more reserves than the Danish TSO's solution, while at the same time the solution is cheaper, because shedding load is highly penalized by the coefficient V^{LOL}.

Figure 10 illustrates the so-called *efficient frontier* [A5]. This plot can be used by the TSO to choose an appropriate value for the risk-aversion parameter α according to its attitude towards risk. The efficiency frontier shows the expected total cost per day, namely, the expected cost of allocating and deploying reserve plus the load shedding cost per day, against the expected LOLP, that is, the expected probability of a load shedding event. The numbers along the curve
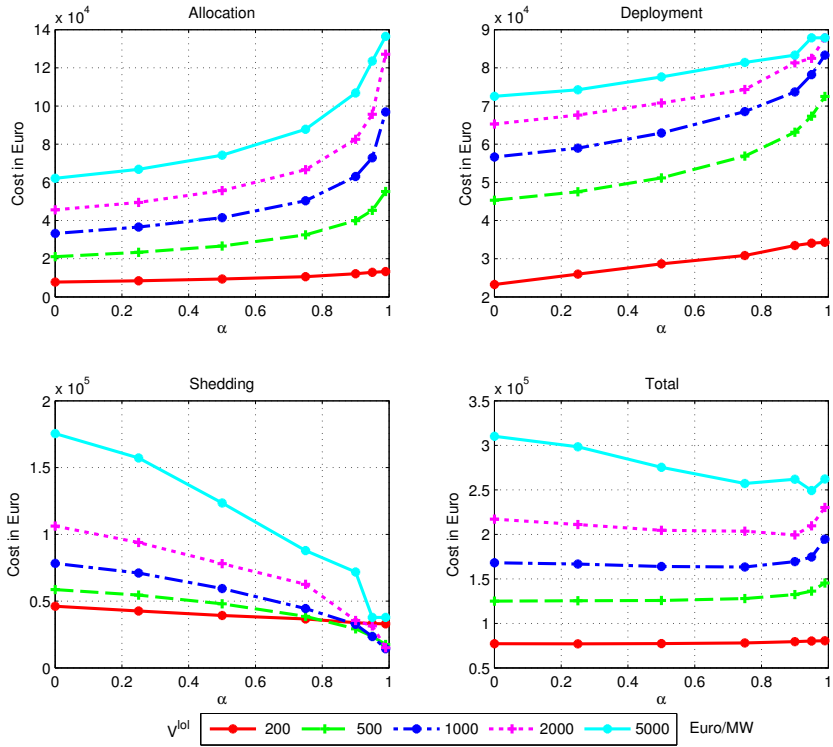
Figure A.7: Sensitivity analysis of the CVaR-based reserve determination model. The parameter α is displayed on the x-axis and the cost in €$\times 10^3$ on the y-axis. Each line represents the cost of the solution for a certain V^{LOL}. The cost of allocating and deploying reserve capacity, the cost of shedding load and the total cost are displayed on the upper-left, upper-right, lower-left and lower-right plot, respectively. All costs are averaged by the number of days in the test period.
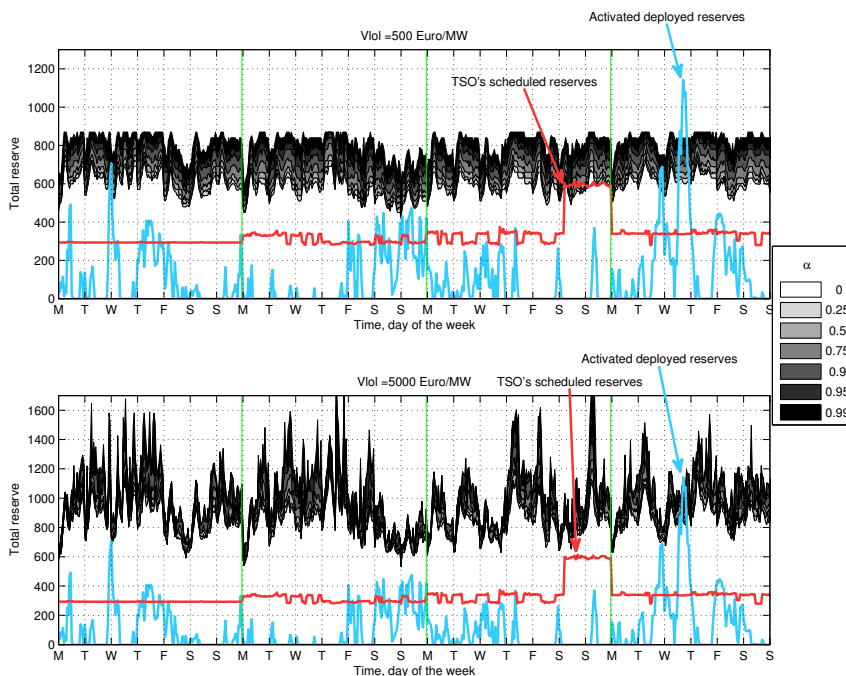
Figure A.8: The reserve schedule using the CVaR methodology is displayed
for $V^{\mathrm{LOL}} = 500$ €/MW in the upper plot and for $V^{\mathrm{LOL}} = 5\,000$
€/MW in the lower plot. The shaded areas in the background
represent the solution of the CVaR model for different values of
the risk aversion parameter α. The actual deployed reserve in
DK1 and the reserve capacity scheduled by Energinet.dk (the
Danish TSO) are depicted on top.
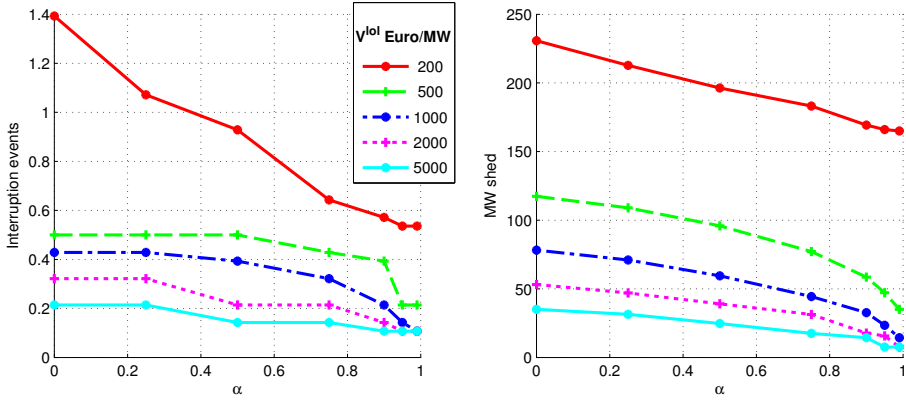
Figure A.9: In both plots, the reserve schedule, computed by the CVaR-based method, is compared to the actual deployed reserve in the DK1 area, for different values of V^{LOL} and α. On the left, we depict the average number of shortage events per day, and on the right, the average MW of load shedding.

indicate the value of the risk parameter α used to obtain such a point in the curve. The efficient frontier shown in Figure 10 has been determined for a V^{LOL} equal to $5\,000 \mbox{\euro}/$MW. Needless to say, the efficient frontier would change for different values of V^{LOL}, but the interpretation of the resulting curves would remain similar.

The TSO can thus use this efficient frontier to resolve the trade-off between desired or expected LOLP versus the expected total cost that such a level of reliability would entail. For example, the TSO can achieve a LOLP of 0.001 with an expected total cost of approximately $3 \times 10^5 \mbox{\euro}$ per day. To this end, the TSO should set the risk-aversion parameter α to 0.5. If, for instance, the LOLP is to be decreased down to 0.0002, the expected total cost would raise up to $3.2 \times 10^5 \mbox{\euro}$ per day. In that case, the parameter α should be set to 0.9.

The main advantage that the CVaR method offers over the LOLP method is that the TSO is able to input the cost of shedding load, V^{LOL}, in the model and, therefore, the reserve dispatch solution is dependent on it. On the contrary, the LOLP method depends only on the number of interruption events and their associated cost is not accounted for. Another advantage of the CVaR method is that the reserve schedule depends on the reserve costs, both the allocation and deployment cost. In a real set-up, the solution would depend on the bids from the producers, while the solution of the LOLP method is independent
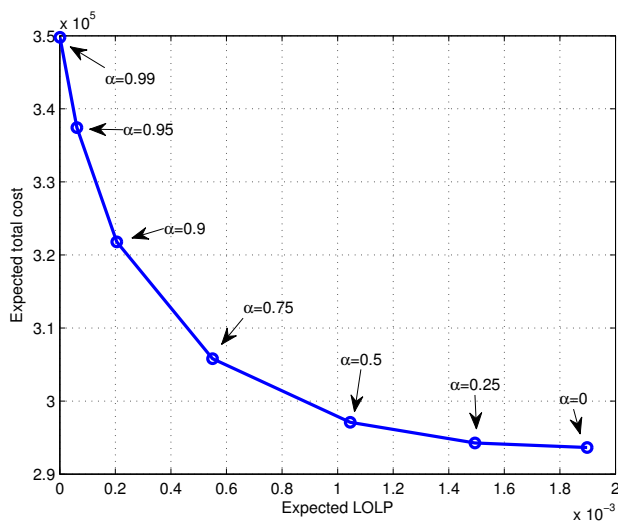
Figure A.10: Efficiency frontier plot of the CVaR model. The expected total cost, computed as the sum of the expected cost of allocating and deploying reserve, and the expected cost of shedding load for $V^{\mathrm{LOL}} = 5\,000$ €, is shown in the the x-axis. In the y-axis, the expected LOLP is displayed. The numbers along the curve indicate the value of the risk parameter α used to obtain each of the solutions.

of the reserve costs. Lastly, both optimization models are able to reflect the
risk aversion of the TSO through the risk parameters β or α. However, the
risk parameter α of the CVaR methodology does not have a straightforward
interpretation in real power systems, as compared to the parameter β from the
LOLP formulation, which has a direct physical interpretation.

A.6   Conclusion

In this paper we present two methods to determine the reserve requirements
using a probabilistic approach, suited for a market structure where the reserves
are scheduled independently of and before to the day-ahead energy market. This
is the case in the Nordic countries and, more specifically, in the DK1 area of
Nord Pool, under which the study case of this paper is framed. The first method
ensures that the LOLP is kept under a certain target. The second method
considers the costs of allocating and deploying reserve and of shedding load, and
minimizes the CVaR of the total cost distribution at a given confidence level
α. Both approaches are based on scenarios of potential balancing requirements,
induced by the forecast error of the wind power production, the forecast error
of the load, and the forced failures of the power plants in the power system.

The performance of the proposed reserve determination models is assessed by
comparing the resulting optimal scheduled reserves with the Danish TSO's solu-
tion approach and with the actual deployed reserves during four testing weeks,
in terms of costs and shortage events. The results from the case study show that
the LOLP method underperforms the Danish TSO's solution in terms of costs,
for the same shortage events. By using a CVaR risk approach, the cost of allo-
cating reserves is reduced from 3.38% to 82.9%, depending on the value of the
parameters of confidence level and value of lost load. The CVaR methodology
provides adequate levels of reserves.

Further studies should focus on the applicability of these methods to the Nordic
reserve market, by differentiating between types of reserves. This could be
achieved by modeling the amount of MWh of each type of reserve required at
every hour and the cost of allocating and activating each of them. Also, further
improvements should be done on the modeling of the failed MW in the whole
system. More specifically, time-dependencies could be modeled, since a power
plant is more likely to be off-line if the previous hour was off-line too. This could
be achieved by, for example, a non-homogeneous Hidden Markov Model, where
the transition probabilities between states depend on time and other external
variables.

Acknowledgement

References A

[A1]   M. Milligan et al. "Operating reserves and wind power integration: an international comparison". In: *proc. 9th International Workshop on large-scale integration of wind power into power systems.* 2010, pp. 18–29.

[A2]   Energinet.dk. *Ancillary Services to be Delivered in Denmark Tender Conditions.* Tech. rep. Oct. 2012.

[A3]   Energinet.dk. *Ancillary services strategy.* 2011.

[A4]   *Energy Strategy 2050 - from coal, oil and gas to green energy.* Tech. rep. February. The Danish Ministry of Climate and Energy, 2011.

[A5]   A. J. Conejo, M. Carrión, and J. M. Gonzalez Morales. *Decision Making Under Uncertainty in Electricity Markets.* International series in operations research and management science. Springer US, 2010.

[A6]   N. Amjady, J. Aghaei, and H. Shayanfar. "Market clearing of joint energy and reserves auctions using augmented payment minimization". In: *Energy* 34.10 (2009), pp. 1552–1559.

[A7]   F. Bouffard and F. Galiana. "An electricity market with a probabilistic spinning reserve criterion". In: *IEEE Transactions on Power Systems* 19.1 (2004), pp. 300–307.

[A8]   M. A. Ortega-Vazquez and D. S. Kirschen. "Estimating the spinning reserve requirements in systems with significant wind power generation penetration". In: *IEEE Transactions on Power Systems* 24.1 (2009), pp. 114–124.

[A9]   J. Morales, A. Conejo, and J. Perez-Ruiz. "Economic Valuation of Reserves in Power Systems With High Penetration of Wind Power". In: *IEEE Transactions on Power Systems* 24.2 (2009), pp. 900–910.

[A10] A. Kalantari, J. F. Restrepo, and F. D. Galiana. "Security-Constrained Unit Commitment With Uncertain Wind Generation : The Loadability Set Approach". In: *IEEE Transactions on Power Systems* (2012), pp. 1–10.

[A11] J. Warrington et al. "Robust reserve operation in power systems using affine policies". In: *IEEE 51st Annual Conference on Decision and Control (CDC)*. 2012, pp. 1111–1117.

[A12] E. Karangelos and F. Bouffard. "Towards Full Integration of Demand-Side Resources in Joint Forward Energy/Reserve Electricity Markets". In: *IEEE Transactions on Power Systems* 27.1 (2012), pp. 280–289. ISSN: 0885-8950.

[A13] F. Partovi et al. "A stochastic security approach to energy and spinning reserve scheduling considering demand response program". In: *Energy* 36.5 (2011), pp. 3130–3137. ISSN: 0360-5442.

[A14] M. Matos and R. Bessa. "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts". In: *IEEE Transactions on Power Systems* 26.2 (2011), pp. 594–603. ISSN: 0885-8950.

[A15] R. Doherty and M. O'Malley. "A new approach to quantify reserve demand in systems with significant installed wind capacity". In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 587–595. ISSN: 0885-8950.

[A16] G. C. Pflug. "Some remarks on the value-at-risk and the conditional value-at-risk". In: *Probabilistic constrained optimization*. Springer, 2000, pp. 272–281.

[A17] *The Value of Lost Load (VoLL) for electricity in Great Britain*. Tech. rep. London Economics, 2013.

[A18] E. Leahy and R. S. Tol. "An estimate of the value of lost load for Ireland". In: *Energy Policy* 39.3 (2011), pp. 1514–1520. ISSN: 0301-4215.

[A19] R. T. Rockafellar and S. Uryasev. "Optimization of conditional value-at-risk". In: *Journal of risk* 2 (2000), pp. 21–42.

[A20] *Energinet.dk*. URL: http://energinet.dk.

[A21] J. Kloppenborg Møller, H. Nielsen Aalborg, and H. Madsen. "Time-adaptive quantile regression". In: *Computational Statistics & Data Analysis* 52.3 (Jan. 2008), pp. 1292–1303. ISSN: 01679473.

[A22] H. A. Nielsen, H. Madsen, and T. S. Nielsen. "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts". In: *Wind Energy* 9.1-2 (2006), pp. 95–108. ISSN: 1099-1824.

[A23] T. S. Nielsen et al. *Prediction of regional wind power*. Tech. rep. 2002.

[A24] H. Nielsen, K. Andersen, and H. Madsen. *Empirisk bestemt model for elforbruget i Østdanmark*. Institut for Matematisk Modellering, Danmarks Tekniske Universitet, Lyngby, 1998.

[A25] P. Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind Energy* 12.1 (2009), pp. 51–62.

[A26]   *Nord Pool Spot AS*. URL: http://www.nordpoolspot.com/.

[A27]   J. Saez. "Determination of Optimal Electricity Reserve Requirements".
         MA thesis. Technical University of Denmark, 2012.

[A28]   H. Madsen and P. Thyregod. *Introduction to General and Generalized
         Linear Models*. CRC Press, 2011.

[A29]   Energinet.dk. *Appendix 1 : Load-Frequency Control and Performance.
         UCTE Operation Handbook*.

[A30]   E. Shayesteh, A. Yousefi, and M. P. Moghaddam. "A probabilistic risk-
         based approach for spinning reserve provision using day-ahead demand
         response program". In: *Energy* 35.5 (2010), pp. 1908–1915. ISSN: 0360-
         5442.

A Data-driven Bidding Model for a Cluster of Price-responsive Consumers of Electricity

Authors:
Javier Saez-Gallego, Juan M. Morales, Marco Zugno, Henrik Madsen

A Data-driven Bidding Model for a Cluster of Price-responsive Consumers of Electricity

Javier Saez-Gallego[1], Juan M. Morales[1], Marzo Zugno[2], Henrik Madsen[1]

Abstract

This paper deals with the market-bidding problem of a cluster of price-responsive consumers of electricity. We develop an inverse optimization scheme that, recast as a bilevel programming problem, uses price-consumption data to estimate the complex market bid that best captures the price-response of the cluster. The complex market bid is defined as a series of marginal utility functions plus some constraints on demand, such as maximum pick-up and drop-off rates. The proposed modeling approach also leverages information on exogenous factors that may influence the consumption behavior of the cluster, e.g., weather conditions and calendar effects. We test the proposed methodology for a particular application: forecasting the power consumption of a small aggregation of households that took part in the Olympic Peninsula project. Results show that the price-sensitive consumption of the cluster of flexible loads can be largely captured in the form of a complex market bid, so that this could be ultimately used for the cluster to participate in the wholesale electricity market.

Notation

The main notation used throughout the paper is stated below for quick reference. Other symbols are defined as required.

---

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.
[2]Nordea, DK-1401 Copenhagen, Denmark.

Indexes

t Index for time periods $t \in \{1 \ldots 24\}$.
b Index for bidding blocks $b \in \mathcal{B}$.
i Index for features $i \in \mathcal{I}$.

Lower-level variables

$x_{b,t}$ Estimated load at block b and time t.
λ_t^u Dual variable of pick-up limit constraint at time t.
λ_t^d Dual variable of drop-off limit constraint at time t.
$\overline{\psi}_{b,t}$ Dual variable of maximum power constraint for block b at time t.
$\underline{\psi}_{b,t}$ Dual variable of maximum power constraint for block b at time t.

Upper-level variables

e_t^+, e_t^- Auxiliary variables to model the absolute value of the estimation error.
$a_{b,t}$ Marginal utility corresponding to bid block b and time t.
$\underline{P}_t$ Minimum power consumption at time t.
$\overline{P}_t$ Maximum power consumption at time t.
r_t^u Maximum load pick-up rate at time t.
r_t^d Maximum load drop-off rate at time t.
a_b^0 Intercept relative to the estimation of the marginal utility corresponding to bid block b.
$\underline{P}^0$ Intercept relative to the estimation of the minimum power consumption.
$\overline{P}^0$ Intercept relative to the estimation of the maximum power consumption.
r^{u0} Intercept relative to the estimation of the maximum load pick-up rate.
r^{d0} Intercept relative to the estimation of the maximum load drop-off rate.
α_i^a Affine coefficient relative to the marginal utility for feature i.
α_i^u Affine coefficient relative to the maximum load pick-up rate for feature i.
α_i^d Affine coefficient relative to the maximum load drop-off rate for feature i.
$\alpha_i^{\overline{P}}$ Affine coefficient relative to the minimum power consumption for feature i.
$\alpha_i^{\underline{P}}$ Affine coefficient relative to the maximum power consumption for feature i.

Parameters

p_t Price of electricity at time t.

w_t Weight of observation at time t.

x^{meas} Measured load at time t.

$x_{b,t}^{meas'}$ Split of the measured load at block b and time t.

$Z_{i,t}$ Feature i at time t.

L   Penalization factor of the complementarity constraints.

E   Forgetting factor.

B.1   Introduction

We consider the case of a cluster of flexible power consumers, where *flexibility* is understood as the possibility for each consumer in the cluster to change her consumption depending on the electricity price and on her personal preferences. There are many examples of methods to schedule the consumption of individual price-responsive loads (see, e.g., [B1, B2, B3]). The portfolio of flexible consumers is managed by a retailer or *aggregator*, which bids in a wholesale electricity market on behalf of her customers. We consider the case where such a market accepts complex bids, consisting of a series of price-energy bidding curves, consumption limits, and maximum pick-up and drop-off rates. In this paper, we present a data-driven methodology for determining the complex bid that best represents the reaction of the pool of flexible consumers to the market price.

The contributions of this paper are fourfold. The first contribution corresponds to the methodology itself: *we propose a novel approach to capture the price-response of a pool of flexible consumers in the form of a market bid using price-consumption data.* In this work, the price is given as the result of a competitive market-clearing process, and we have access to it only from historical records. This is in contrast to some previous works, where the price is treated as a control variable to be decided by the aggregator or retailer. In [B4], for example, the relationship between price and consumption is first modeled by a Finite Impulse Response (FIR) function as in [B5] and peak load reduction is achieved by modifying the price. Similar considerations apply to the works of [B6, B7, B8, B9], where a bilevel representation of the problem is used: the lower-level problem optimizes the household consumption based on the broadcast electricity price, which is determined by the upper-level problem to maximize the aggregator's/retailer's profit. Another series of studies concentrate on estimating price-energy bids for the participation of different types of flexible loads in the wholesale electricity markets, for example, time-shiftable loads [B10], electric vehicles [B11] and thermostatically-controlled loads [B12]. Contrary to these studies, our approach is data-driven and does not require any assumption about the nature of the price-responsive loads in the aggregation. In that sense, our

work is more similar to [B13], where the satisfaction or the utility of users is estimated through historical data. The main differences are that we aim to minimize prediction errors instead of just estimating the utility, and that our estimated utility and further technical parameters defining a complex market bid may depend on time and external factors. Furthermore, we test our methodology on actual data obtained from a real-life experiment.

The second contribution lays in the estimation procedure: *we develop an inverse optimization framework that results in a bilevel optimization problem.* Methodologically, our approach builds on the inverse optimization scheme introduced in [B14], but with several key differences. First, we let the measured solution be potentially non-optimal, or even non-feasible, for the targeted optimization problem as in [B15, B16, B17, B18]. Second, we seek to minimize the out-of-sample prediction error through the use of a penalty factor L. Moreover, we extend the concept of inverse optimization to a problem where the estimated parameters may depend on a set of features and are also allowed to be in the constraints, and not only in the objective function. Lastly, the estimation procedure is done using a two-step algorithm to deal with non-convexities.

Third, *we study heuristic solution methods to reduce the computing times resulting from the consideration of large datasets of features for the model estimation.* We do not solve the resulting bilevel programming problem to optimality but instead we obtain an approximate solution by penalizing the violation of complementarity constraints following a procedure inspired by the work of [B19].

Finally, *we test the proposed methodology using data from a real-world experiment* that was conducted as a part of the *Olympic Peninsula Project* [B20].

It should be stressed that the proposed methodology aims to capture the price-response behavior of the pool of flexible consumers, and not to modify it. For this reason, we treat the electricity price as an exogenous variable to our model and not as a control signal to be determined. By means of our methodology, the consumers are directly exposed to the wholesale market price, without the need for the aggregator to artificially alter this price or to develop any trading strategy. Notwithstanding this, the price-response model for the pool of flexible consumers that we estimate in the form of a complex market bid could also be used by the aggregator to determine a series of prices such that the consumption of the pool pursues a certain objective.

B.2   Methodology

In this section, we describe the methodology to determine the optimal market bid for a pool of price-responsive consumers. The estimation procedure is cast as a bilevel programming problem. The upper level is the parameter-estimation problem and represents the aggregator, who aims at determining the parameters of the complex market bid such that the estimated absolute value of the prediction error is minimized. This bid can be directly processed by the market-clearing algorithms currently in place in most electricity markets worldwide. A detailed explanation is given in Section B.2.2. The estimated bid, given by the upper-level problem, is relative to the aggregated pool of consumers. The lower-level problem, explained in Section B.2.1, represents the price-response of the *whole* pool of consumers under the estimated bid parameters.

B.2.1   Lower-Level Problem: Price-response of the Pool of Consumers

The lower-level problem models the price-response of the pool of consumers in the form of a market bid, whose parameters are determined by the upper-level problem. The bid is given by $\theta_t = \{a_{b,t}, r_t^u, r_t^d, \underline{P}_t, \overline{P}_t\}$, which consists of the declared marginal utility corresponding to each bid block b, the maximum load pick-up and drop-off rates (analogues to the ramp-up and -down limits of a power generating unit), the minimum power consumption, and the maximum power consumption, at time $t \in \mathcal{T} \equiv \{t : t = 1 \ldots T\}$, in that order. If the whole aggregation of consumers behaves indeed as a utility-maximizer individual, the declared utility function represents the benefit that the pool of flexible consumers obtains from consuming a certain amount of electricity. In the more general case, the declared marginal utility, or simply the bidding curve, reflects the elasticity of the pool of consumers to changes in the electricity price. The declared marginal utility function together with the rest of parameters in (B.1) define a complex market bid that can be processed by the market-clearing algorithms used by most wholesale electricity markets worldwide, while representing, as much as possible, the price-response behavior of the aggregation of consumers.

The declared marginal utility $a_{b,t}$ at time t is formed by $b \in \mathcal{B} \equiv \{b : b = 1 \ldots B\}$ blocks, where all blocks have equal size, spanning from the minimum to the maximum allowed consumption. In other words, the size of each block is $\frac{\overline{P}_t - \underline{P}_t}{B}$. Furthermore, we assume that the marginal utility is monotonically decreasing as consumption increases, i.e., $a_{b,t} \geq a_{b+1,t}$ for all times t. Finally, the total consumption at time t is given by the sum of the minimum power demand plus

the consumption linked to each bid block, namely, $x_t^{tot} = \underline{P}_t + \sum_{b \in \mathcal{B}} x_{b,t}$.

Typically, the parameters of the bid may change across the hours of the day, the days of the week, the month, the season, or any other indicator variables related to the time. Moreover, the bid can potentially depend on some external variables such as temperature, solar radiation, wind speed, etc. Indicator variables and external variables, often referred to as *features*, can be used to explain more accurately the parameters of the market bid that best represents the price-response of the pool of consumers. This approach is potentially useful in practical applications, as numerous sources of data can help better explain the consumers' price-response. We consider the I external variables or features, named $Z_{i,t}$ for $i \in \mathcal{I} \equiv \{i : i = 1, \dots, I\}$, to be affinely related to the parameters defining the market bid by a coefficient α_i. This affine dependence can be enforced in the model by letting $a_{b,t} = a_b^0 + \sum_{i \in \mathcal{I}} \alpha_i^a Z_{i,t}$, $r_t^u = r^{u0} + \sum_{i \in \mathcal{I}} \alpha_i^u Z_{i,t}$, $r_t^d = r^{d0} + \sum_{i \in \mathcal{I}} \alpha_i^d Z_{i,t}$, $\overline{P}_t = \overline{P}^0 + \sum_{i \in \mathcal{I}} \alpha_i^{\overline{P}} Z_{i,t}$, and $\underline{P}_t = \underline{P}^0 + \sum_{i \in \mathcal{I}} \alpha_i^{\underline{P}} Z_{i,t}$. The affine coefficients $\alpha_i^a, \alpha_i^u, \alpha_i^d, \alpha_i^{\overline{P}}$ and $\alpha_i^{\underline{P}}$, and the intercepts $a_b^0, r^{u0}, r^{d0}, \underline{P}^0, \overline{P}^0$ enter the model of the pool of consumers (the lower-level problem) as parameters, together with the electricity price.

The objective is to maximize consumers' welfare, namely, the difference between the total utility and the total payment:

$$\underset{x_{b,t}}{\text{Maximize}} \sum_{t \in \mathcal{T}} \left(\sum_{b \in \mathcal{B}} a_{b,t} x_{b,t} - p_t \sum_{b \in \mathcal{B}} x_{b,t} \right) \tag{B.1a}$$

where $x_{b,t}$ is the consumption assigned to the utility block b during the time t, $a_{b,t}$ is the marginal utility obtained by the consumer in block b and time t, and p_t is the price of the electricity during time t. For notational purposes, let $\mathcal{T}_{-1} = \{t : t = 2, \dots, T\}$. The problem is constrained by

$$\underline{P}_t + \sum_{b \in \mathcal{B}} x_{b,t} - \underline{P}_{t-1} - \sum_{b \in \mathcal{B}} x_{b,t-1} \leq r_t^u \qquad t \in \mathcal{T}_{-1} \tag{B.1b}$$

$$\underline{P}_{t-1} + \sum_{b \in \mathcal{B}} x_{b,t-1} - \underline{P}_t - \sum_{b \in \mathcal{B}} x_{b,t} \leq r_t^d \qquad t \in \mathcal{T}_{-1} \tag{B.1c}$$

$$x_{b,t} \leq \frac{\overline{P}_t - \underline{P}_t}{B} \qquad\qquad b \in \mathcal{B}, t \in \mathcal{T} \tag{B.1d}$$

$$x_{b,t} \geq 0 \qquad\qquad b \in \mathcal{B}, t \in \mathcal{T}. \tag{B.1e}$$

Equations (B.1b) and (B.1c) impose a limit on the load pick-up and drop-off rates, respectively. The set of equations (B.1d) defines the size of each utility block to be equally distributed between the maximum and minimum power consumptions. Constraint (B.1e) enforces the consumption pertaining to each utility block to be positive. Note that, by definition, the marginal utility is

decreasing in x_t ($a_{b,t} \geq a_{b+1,t}$), so one can be sure that the first blocks will be filled first. We denote the dual variables associated with each set of primal constraints as $\lambda_t^u, \lambda_t^d, \overline{\psi}_{b,t}$ and $\underline{\psi}_{b,t}$.

Problem (B.1) is linear, hence it can be equivalently recast as the following set of KKT conditions [B21], where (B.2a)–(B.2c) are the stationary conditions and (B.2d)–(B.2g) enforce complementarity slackness:

$$-\lambda_2^u + \lambda_2^d - \underline{\psi}_{b,1} + \overline{\psi}_{b,1} = a_{b,1} - p_1 \qquad\qquad b \in \mathcal{B} \text{ (B.2a)}$$

$$\lambda_t^u - \lambda_{t+1}^u - \lambda_t^d + \lambda_{t+1}^d - \underline{\psi}_{b,t} + \overline{\psi}_{b,t} = a_{b,t} - p_t \qquad \forall b \in \mathcal{B}, t \in \mathcal{T}_{-1} \text{ (B.2b)}$$

$$\lambda_T^u - \lambda_T^d - \underline{\psi}_{b,T} + \overline{\psi}_{b,T} = a_{b,T} - p_T \qquad\qquad b \in \mathcal{B} \text{ (B.2c)}$$

$$\underline{P}_t + \sum_{b \in \mathcal{B}} x_{b,t} - \underline{P}_{t-1} - \sum_{b \in \mathcal{B}} x_{b,t-1} \leq r_t^u \perp \lambda_t^u \geq 0 \qquad\qquad t \in \mathcal{T}_{-1} \text{ (B.2d)}$$

$$\underline{P}_{t-1} + \sum_{b \in \mathcal{B}} x_{b,t-1} - \underline{P}_t - \sum_{b \in \mathcal{B}} x_{b,t} \leq r_t^d \perp \lambda_t^d \geq 0 \qquad\qquad t \in \mathcal{T}_{-1} \text{ (B.2e)}$$

$$x_{b,t} \leq \frac{\overline{P}_t - \underline{P}_t}{B} \perp \overline{\psi}_{b,t} \geq 0 \qquad\qquad b \in \mathcal{B}, t \in \mathcal{T} \text{ (B.2f)}$$

$$0 \leq x_{b,t} \perp \underline{\psi}_{b,t} \geq 0 \qquad\qquad b \in \mathcal{B}, t \in \mathcal{T}. \text{ (B.2g)}$$

B.2.2 Upper-Level Problem: Market-Bid Estimation Via Inverse Optimization

Given a time series of price-consumption pairs (p_t, x_t^{meas}), the inverse problem consists in estimating the value of the parameters θ_t defining the objective function and the constraints of the lower-level problem (B.1) such that the optimal consumption x_t resulting from this problem is as close as possible to the measured consumption x_t^{meas} in terms of a certain norm. The parameters of the lower-level problem θ_t form, in turn, the market bid that best represents the price-response of the pool.

In mathematical terms, the inverse problem can be described as a minimization problem:

$$\underset{x,\theta}{\text{Minimize}} \sum_{t \in \mathcal{T}} w_t \left| \underline{P}_t + \sum_{b \in \mathcal{B}} x_{b,t} - x_t^{meas} \right| \qquad\qquad \text{(B.3a)}$$

subject to

$$a_{b,t} \geq a_{b+1,t} \qquad\qquad b \in \mathcal{B}, t \in \mathcal{T} \qquad\qquad \text{(B.3b)}$$

$$\text{(B.2).} \qquad\qquad\qquad\qquad\qquad\qquad \text{(B.3c)}$$

Constraints (B.3b) are the upper-level constraints, ensuring that the estimated marginal utility must be monotonically decreasing. Constraints (B.3c) correspond to the KKT conditions of the lower-level problem (B.1).

Notice that the upper-level variables θ_t, which are parameters in the lower-level problem, are also implicitly constrained by the optimality conditions (B.2) of this problem, i.e., by the fact that $x_{b,t}$ must be optimal for (B.1). This guarantees, for example, that the minimum power consumption be positive and equal to or smaller than the maximum power consumption ($0 \leq \underline{P}_t \leq \overline{P}_t$). Furthermore, the maximum pick-up rate is naturally constrained to be equal to or greater than the negative maximum drop-off rate ($-r_t^d \leq r_t^u$). Having said that, in practice, we need to ensure that these constraints are fulfilled for all possible realizations of the external variables and not only for the ones observed in the past. We achieved this by enforcing the robust counterparts of these constraints [B22]. An example is provided in the appendix.

Parameter w_t represents the weight of the estimation error at time t in the objective function. These weights have a threefold purpose. Firstly, if the inverse optimization problem is applied to estimate the bid for the day-ahead market, the weights could represent the cost of balancing power at time t. In such a case, consumption at hours with a higher balancing cost would weigh more and consequently, would be fit better than that occurring at hours with a lower balancing cost. Secondly, the weights can include a forgetting factor to give exponentially decaying weights to past observations. Finally, a zero weight can be given to missing or wrongly measured observations.

The absolute value of the residuals can be linearized by adding two extra nonnegative variables, and by replacing the objective equation (B.3a) with the following linear objective function plus two more constraints, namely, (B.4b) and (B.4c):

$$\underset{x_t, \theta_t, e_t^+, e_t^-}{\text{Minimize}} \sum_{t=1}^{T} w_t(e_t^+ + e_t^-) \qquad\qquad \text{(B.4a)}$$

subject to

$$\underline{P}_t + \sum_{b \in \mathcal{B}} x_{b,t} - x_t^{meas} = e_t^+ - e_t^- \qquad\qquad t \in \mathcal{T} \qquad \text{(B.4b)}$$

$$e_t^+, e_t^- \geq 0 \qquad\qquad t \in \mathcal{T} \qquad \text{(B.4c)}$$

$$a_{b,t} \geq a_{b+1,t} \qquad\qquad t \in \mathcal{T} \qquad \text{(B.4d)}$$

$$\text{(B.2).} \qquad\qquad \text{(B.4e)}$$

In the optimum, and when $w_t > 0$, (B.4b) and (B.4c) imply that $e_t^+ = x_t - x_t^{meas}$ if $x_t \geq x_t^{meas}$, else $e_t^- = x_t^{meas} - x_t$. By using this reformulation of the absolute value, the weights could also reflect whether the balancing costs are symmetric or skewed. In the latter case, there would be different weights for e_t^+ and e_t^-.

To sum up, we have, on the one hand, problem (1), which represents the postulated price-response model for the pool of flexible consumers. This optimization problem, in turn, takes the form of a complex market bid that can be directly submitted to the electricity market. On the other hand, we have problem (4), which is an estimation problem in a statistical sense: it seeks to estimate the parameters of problem (1), that is, the parameters defining the complex market bid, by using the sum of the weighted absolute values of residuals as the loss function to be minimized. This problem takes the form of a bilevel programming problem.

Finally, it is worth pointing out that there is no obstacle to reformulating (B.3) as a least-squares estimation problem by the use of the L_2-norm to quantify the prediction error. However, the minimization of the absolute value of residuals (i.e., the minimization of the L_1-norm of the estimation error) allows interpreting the objective function of (B.3) as an energy mismatch (to be traded on a market closer to real time), while keeping the estimation problem linear.

B.3   Solution Method

The estimation problem (B.4) is non-linear due to the complementarity constraints of the KKT conditions of the lower-level problem (B.2). There are several ways of dealing with these constraints, for example, by using a non-linear solver [B23], by recasting them in the form of disjunctive constraints [B24], or by using SOS1 variables [B25]. In any case, problem (B.4) is NP-hard to solve and the computational time grows exponentially with the number of complementarity constraints. Our numerical experiments showed that, for realistic applications involving multiple time periods and/or numerous features, none of

these solution methods were able to provide a good solution to problem (B.4) in a reasonable amount of time. To tackle this problem in an effective manner, we propose the following two-step solution strategy, inspired by [B19]:

Step 1: Solve a linear relaxation of the mathematical program with equilibrium constraints (B.4) by penalizing violations of the complementarity constraints.

Step 2: Fix the parameters defining the constraints of the lower-level problem (B.1), i.e., $r^u, r^d, \underline{P}, \overline{P}, \alpha_i^a, \alpha_i^d, \alpha_i^{\overline{P}}$ and $\alpha_i^{\underline{P}}$, at the values estimated in Step 1. Then, recompute the parameters defining the utility function, $a_{b,t}$ and α_d^a. To this end, we make use of the primal-dual reformulation of the price-response model (B.1) [B15].

Both steps are further described in the subsections below. Note that the proposed solution method is a heuristic in the sense that it does not solve the bilevel programming problem (4) to optimality. However, data can be used to calibrate it (through the penalty parameter L) to minimize the *out-of-sample* prediction error. For the problem at hand, this is clearly more important than finding the optimal solution to (4), see Section B.3.3 for further details.

B.3.1 Penalty Method

The so-called penalty method is a convex (linear) relaxation of a mathematical programming problem with equilibrium constraints, whereby the complementarity conditions of the lower-level problem, that is, problem (B.1), are moved to the objective function (B.4a) of the upper-level problem. Thus, we penalize the sum of the dual variables of the inequality constraints of problem (B.1) and their slacks, where the slack of a "$\leq$"-constraint is defined as the difference between its right-hand and left-hand sides, in such a way that the slack is always nonnegative. For example, the slack of the constraint relative to the maximum pick-up rate (B.1b) is defined as $s_t = r_t^u - \underline{P}_t - \sum_{b \in \mathcal{B}} x_{b,t} + \underline{P}_{t-1} + \sum_{b \in \mathcal{B}} x_{b,t-1}$, and analogously for the rest of the constraints of the lower-level problem.

The penalization can neither ensure that the complementarity constraints are satisfied, nor that the optimal solution of the inverse problem is achieved. Instead, with the penalty method, we obtain an approximate solution. In the case study of Section B.4, nonetheless, we show that this solution performs notably well.

After relaxing the complementarity constraints (B.2d)–(B.2g), the objective function of the estimation problem writes as:

$$
\begin{aligned}
\underset{\boldsymbol{\Omega}}{\text{Minimize}} \quad & \sum_{t \in \mathcal{T}} w_t (e_t^+ + e_t^-) + \\
& L \Bigg(\sum_{\substack{b \in \mathcal{B} \\ t \in \mathcal{T}}} w_t \Big(\psi_{b,t}^{\overline{P}} + \psi_{b,t}^{\underline{P}} + \frac{\overline{P}_t - \underline{P}_t}{B} \Big) + \\
& \qquad \sum_{t \in \mathcal{T}_{-1}} w_t \Big(\lambda_t^u + \lambda_t^d + r_t^u + r_t^d \Big) \Bigg)
\end{aligned}
\tag{B.5a}
$$

with the variables being $\boldsymbol{\Omega} = \{x_t, \theta_t, e_t^+, e_t^-, \psi_t^{\overline{P}}, \psi_t^{\underline{P}}, \lambda_t^u, \lambda_t^d, \}$, subject to the following constraints:

$$
(\text{B.4b}) - (\text{B.4d}), (\text{B.1b}) - (\text{B.1e}), (\text{B.2a}) - (\text{B.2c}) \tag{B.5b}
$$

$$
\lambda_t^u, \lambda_t^d \geq 0 \qquad t \in \mathcal{T}_{-1} \tag{B.5c}
$$

$$
\overline{\psi}_{b,t}, \underline{\psi}_{b,t} \geq 0 \qquad b \in \mathcal{B}, t \in \mathcal{T}. \tag{B.5d}
$$

The objective function (B.5a) of the relaxed estimation problem is composed of two terms. The first term represents the weighted sum of the absolute values of the deviations of the estimated consumption from the measured one. The second term, which is multiplied by the penalty term L, is the sum of the dual variables of the constraints of the consumers' price-response problem plus their slacks. Note that summing up the slacks of the constraints of the consumers' price-response problem eventually boils down to summing up the right-hand sides of such constraints. The weights of the estimation errors (w_t) also multiply the penalization terms. Thus, the model weights violations of the complementarity constraints in the same way as the estimations errors are weighted.

Objective function (B.5a) is subject to the auxiliary constraints modeling the absolute value of estimation errors (B.4b)–(B.4c); the upper-level-problem constraints imposing monotonically decreasing utility blocks (B.4d); and the primal and dual feasibility constraints of the lower-level problem, (B.1b)–(B.1e), (B.2a)–(B.2c), and (B.5c)–(B.5d).

The penalty parameter L should be tuned carefully. We use cross-validation to this aim, as described in the case study; we refer to Section B.4 for further details.

Finding the optimal solution to problem (B.5) is computationally cheap, because it is a linear programming problem. On the other hand, the optimal solution to this problem might be significantly different from the one that we are actually looking for, which is the optimal solution to the original estimation problem (B.4). Furthermore, the solution to (B.5) depends on the user-tuned penalization parameter L, which is given as an input and needs to be decided beforehand.

B.3.2   Refining the Utility Function

In this subsection, we elaborate on the second step of the strategy we employ to estimate the parameters of the market bid that best captures the price-response of the cluster of loads. Recall that this strategy has been briefly outlined in the introduction of Section B.3. The ultimate purpose of this additional step is to re-estimate or refine the parameters characterizing the utility function of the consumers' price-response model (B.1), namely, a_b^0 and the coefficients α_i^a. In plain words, we want to improve the estimation of these parameters with respect to the values that are directly obtained from the relaxed estimation problem (B.5). With this aim in mind, we fix the parameters defining the constraints of the cluster's price-response problem (B.1) to the values estimated in Step 1, that is, to the values obtained by solving the relaxed estimation problem (B.5). Therefore, the bounds $\underline{P}_t, \overline{P}_t$ and the maximum pick-up and drop-off rates r_t^u, r_t^d are now treated as given parameters in this step. Consequently, the only upper-level variables that enter the lower-level problem (B.1), namely, the intersects a_b^0 of the various blocks defining the utility function and the linear coefficients α_i^a, appear in the objective function of problem (B.1). This will allow us to formulate the utility-refining problem as a linear programming problem.

Indeed, consider the primal-dual optimality conditions of the consumers' price-response model (B.1), that is, the primal and dual feasibility constraints and the strong duality condition. These conditions are also necessary and sufficient for optimality due to the linear nature of this model.

We determine the (possibly approximate) block-wise representation of the measured consumption at time t, x_t^{meas}, which we denote by $\sum_{b \in \mathcal{B}} x_{b,t}^{m'}$ and is given as a sum of B blocks of size $\frac{\overline{P}_t - \underline{P}_t}{B}$ each. In particular, we define $\sum_{b \in \mathcal{B}} x_{b,t}^{meas'}$ as follows:

$$\sum_{b \in \mathcal{B}} x_{b,t}^{m'} = \begin{cases} \overline{P}_t & \text{if } x_t^{meas} > \overline{P}_t , \ t \in \mathcal{T} \\ x_t^{meas} & \text{if } \underline{P}_t \leq x_t^{meas} \leq \overline{P}_t , \ t \in \mathcal{T} \\ \underline{P}_t & \text{if } x_t^{meas} < \underline{P}_t , \ t \in \mathcal{T}. \end{cases}$$

where each $x_{b,t}^{m'}$ is determined such that the blocks with higher utility are filled first. Now we replace x_t in the primal-dual reformulation of (B.1) with $\sum_{b\in\mathcal{B}} x_{b,t}^{m'}$. Consequently, the primal feasibility constraints are ineffective and can be dropped.

Once x_t has been replaced with $\sum_{b\in\mathcal{B}} x_{b,t}^{m'}$ in the primal-dual reformulation of (B.1) and the primal feasibility constraints have been dropped, we solve an optimization problem (with the utility parameters a_b and α_i^a as decision variables) that aims to minimize the weighted duality gap, as in [B15]. For every time period t in the training data set, we obtain a contribution (ϵ_t) to the total duality gap ($\sum_{t\in\mathcal{T}} \epsilon_t$), defined as the difference between the dual objective function value at time t minus the primal objective function value at time t. This allows us to find close-to-optimal solutions for the consumers' price-response model (B.1). Thus, in the case when the duality gap is equal to zero, the measured consumption, if feasible, would be optimal in (B.1). In the case when the duality gap is greater than zero, the measured consumption would not be optimal. Intuitively, we attempt to find values for the parameters defining the block-wise utility function such that the measured consumption is as optimal as possible for problem (B.1).

Hence, the utility-refining problem consists in minimizing the sum of weighted duality gaps

$$\underset{\substack{a_{b,t},\lambda_t^u,\lambda_t^d, \\ \psi_t^{\overline{P}},\psi_t^{\underline{P}},\underline{\psi}_{b,t},\overline{\psi}_{b,t},\epsilon_t}}{\text{Minimize}} \quad \sum_{t\in\mathcal{T}} w_t\epsilon_t. \tag{B.6a}$$

Note that we assign different weights to the duality gaps accrued in different time periods, in a way analogous to what we do with the absolute value of residuals in (B.3). Objective function (B.6a) is subject to

$$\sum_{b\in\mathcal{B}} a_{b,1}x_{b,1}^{m'} - p_1\sum_{b\in\mathcal{B}} x_{b,1}^{m'} + \epsilon_1 = \sum_{b\in\mathcal{B}}\left(\frac{\overline{P}_1 - \underline{P}_1}{B}\right)\overline{\psi}_{b,1} \tag{B.6b}$$

$$\sum_{b\in\mathcal{B}} a_{b,t}x_{b,t}^{m'} - p_t\sum_{b\in\mathcal{B}} x_{b,t}^{m'} + \epsilon_t = \sum_{b\in\mathcal{B}}\left(\frac{\overline{P}_t - \underline{P}_t}{B}\right)\overline{\psi}_{b,t} +$$
$$\left(r_t^u - \underline{P}_t + \underline{P}_{t-1}\right)\lambda_t^u + \left(r_t^d + \underline{P}_t - \underline{P}_{t-1}\right)\lambda_t^d \qquad t\in\mathcal{T}_{-1} \tag{B.6c}$$

$$(\text{B.2a}) - (\text{B.2c}) \tag{B.6d}$$

$$a_{b,t} \geq a_{b+1,t} \qquad\qquad t\in\mathcal{T} \tag{B.6e}$$

$$\lambda_t^u,\lambda_t^d \geq 0 \qquad\qquad t\in\mathcal{T}_{-1} \tag{B.6f}$$

$$\psi_t^{\overline{P}},\psi_t^{\underline{P}},\underline{\psi}_{b,t},\overline{\psi}_{b,t} \geq 0 \qquad\qquad t\in\mathcal{T} \tag{B.6g}$$

The set of constraints (B.6c) constitutes the relaxed strong duality conditions, which express that the objective function of the original problem at time t, previously formulated in Equation (B.1), plus the duality gap at time t, denoted by ϵ_t, must be equal to the objective function of its dual problem also at time t. Equation (B.6b) works similarly, but for $t = 1$. The constraints relative to the dual of the original problem are grouped in (B.6d). Constraint (B.6e) requires that the estimated utility be monotonically decreasing. Finally, constraints (B.6f) and (B.6g) impose the non-negative character of dual variables.

B.3.3 Statistical Learning Interpretation

The proposed method to solve the bilevel programming problem (B.4) is a heuristic in the sense that it is not guaranteed to provide the optimal solution to (B.4), that is, it may not deliver the parameters of the market bid that minimize the sum of the weighted absolute values of residuals (B.4a). However, objective function (B.4a) measures the *in-sample* prediction error and it is well known, from the theory of statistical learning [B26], that minimizing the prediction error *in-sample* is *not* equivalent to minimizing it *out-of-sample*. Consequently, the market bid that is the optimal solution to the bilevel program (B.4), i.e., that minimizes the in-sample prediction error, as given in (B.4a), is not necessarily the one performing best in practice. In fact, one could arbitrarily decrease the *in-sample* prediction error down to zero, for example, by enlarging the parameter space defining the market bid in order to overfit the data, while the *out-of-sample* prediction error would dramatically increase as a result. Our aim must be, therefore, to reduce the out-of-sample error as much as possible. In this vein, the solution method that we propose shows two major advantages over solving the bilevel program (B.4) to optimality, namely:

1. It runs swiftly as we indicate later on in Section V and can even be used for real-time trading and very short-term forecasting. In contrast, finding the optimal solution to (B.4) becomes rapidly computationally intractable for sizes of the data sample acceptable to guarantee a proper estimation of the market bid, that is, to avoid overfitting.

2. Besides its good computational properties, the relaxed problem (B.5) is parameterized on the penalty factor L, which is to be tuned by the user. Statistically speaking, this provides our solution approach with a degree of freedom that directly solving (B.4) does not have. Indeed, we can let the data decide which value of the penalty L is the best, that is, which value of L minimizes the *out-of-sample* prediction error. To compute a proxy of the out-of-sample prediction error, we conduct a thorough validation analysis [B26], which essentially consists in recreating the use of our approach in

practice for several values of L, from among which we pick up the one that returns the highest prediction performance. Furthermore, note that both the tuning of the penalty L and the consequent re-estimation of the market-bid parameters can be conducted *offline*, as new information becomes available and as soon as we perceive a statistically significant deterioration of the prediction performance of our approach.

In the case study of Section V, we demonstrate the effectiveness of the proposed solution method by evaluating its performance *out-of-sample* and comparing it against other solution approaches over the same data set from a real-life experiment.

B.4   Case Study

The proposed methodology to estimate the market bid that best captures the price-response of a pool of flexible consumers is tested using data from a real-life case study. The data relates to the Olympic Peninsula experiment, which took place in Washington and Oregon states between May 2006 and March 2007 [B20]. The electricity price was sent out every fifteen minutes to 27 households that participated in the experiment. The price-sensitive controllers and thermostats installed in each house decided when to turn on and off the appliances, based on the price and on the house owner's preferences.

For the case study, we use hourly measurements of load consumption, broadcast price, and observed weather variables, specifically, outside temperature, solar irradiance, wind speed, humidity, dew point and wind direction. Moreover, we include 0/1 feature variables to indicate the hour of the day, with one binary variable per hour (from 0 to 23), and one per day of the week (from 0 to 6). A sample of the dataset is shown in Figure B.1, where the load is plotted in the upper plot, the price in the middle plot, and the load versus the outside temperature and the dew point in the bottom plots. The lines depicted in the bottom plots represent the linear relationship between the pairs of variables, and these are negative in both cases. The high variability in the price is also noteworthy: from the 1st to the 8th of December, the standard deviation of the price is 5.6 times higher than during the rest of the month ($67.9/MWh versus $12.03/MWh).

On average, when using hourly data from the previous 3 months, i.e., 2016 samples, and a total of 37 features per sample, the time for the whole estimation process takes 287 seconds to solve, with a standard deviation of 22 seconds, on
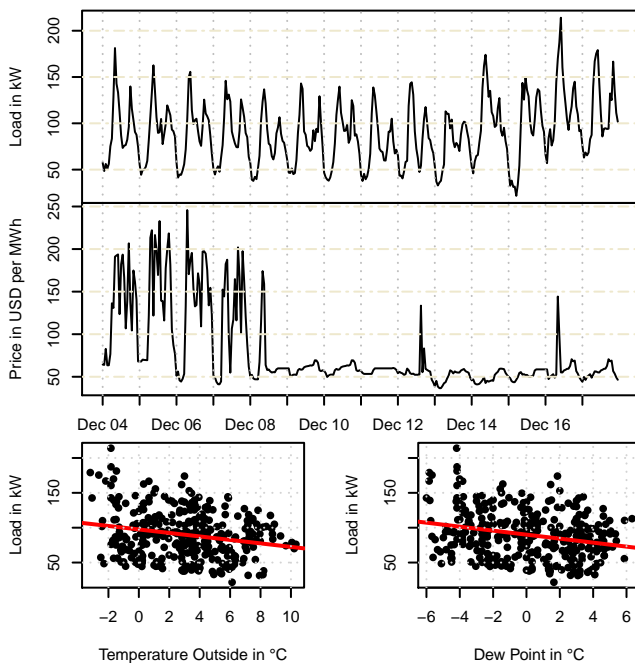
Figure B.1: The upper and the middle plot show the load and the price, respectively. The bottom plots represent the load in the vertical axis versus the outside temperature and the dew point, on the left and on the right, respectively. The data shown spans from the 4th to the 18th of December.

a personal Linux-based machine with 4 cores clocking at 2.90GHz and 6 GB of RAM. R and CPLEX 12.3 under GAMS are used to process the data and solve the optimization models. These running times depend on the number of data points and on the data itself. We conclude that the running time makes this methodology attractive for bidding on short-term electricity markets.

Furthermore, in practice, we have parallelized a great deal of the code using the High-Performance-Computing facility at the Technical University of Denmark[B27], achieving solution times in the order of seconds, so that the proposed solution algorithm can even be used to bid in current real-time/balancing markets.

Also, parallel computing proves to be specially useful to tune the penalty parameter L through cross-validation. In this regard, it is important to stress that both the value of L and the bid parameters as a function of the features can

be periodically recomputed offline (for example, every day, or every week, or every month) to capture potential changes in the pool of consumers that may eventually deteriorate the prediction performance of our method.

B.4.1 Benchmark Models

To test the quality of the market bid estimated by the proposed methodology, we quantify and assess the extent to which such a bid is able to *predict* the consumption of the cluster of price-responsive loads. For the evaluation, we compare two versions of the inverse optimization scheme proposed in this paper with the Auto-Regressive model with eXogenous inputs (ARX) described in [B5]. Note that this time series model was also applied by [B5] to the same data set of the Olympic Peninsula project. All in all, we benchmark three different models:

ARX, which stands for Auto-Regressive model with eXogenous inputs [B28]. This is the type of prediction model used in [B4] and [B5]. The consumption x_t is modeled as a linear combination of past values of consumption up to lag n, $\boldsymbol{X}_{t-n} = \{x_t, \ldots, x_{t-n}\}$, and other explanatory variables $\boldsymbol{Z}_t = \{Z_t, \ldots, Z_{t-n}\}$. In mathematical terms, an ARX model can be expressed as $x_t = \boldsymbol{\vartheta}_x \boldsymbol{X}_{t-n} + \boldsymbol{\vartheta}_z \boldsymbol{Z}_t + \epsilon_t$, with $\epsilon_t \sim \mathrm{N}(0,\sigma^2)$ and σ^2 is the variance.

Simple Inv This benchmark model consists in the utility-refining problem presented in Section B.3.2, where the parameters of maximum pick-up and drop-off rates and consumption limits are computed from past observed values of consumption in a simple manner: we set the maximum pick-up and drop-off rates to the maximum values taken on by these parameters during the last seven days of observed data. All the features are used to explain the variability in the block-wise marginal utility function of the pool of price-responsive consumers: outside temperature, solar radiation, wind speed, humidity, dew point, pressure, and hour and week-day indicators. For this model, we use B=12 blocks of utility. This benchmark is inspired from the more simplified inverse optimization scheme presented in [B16] and [B15] (note, however, that neither [B16], nor [B15] consider the possibility of leveraging auxiliary information, i.e., features, to better explain the data, unlike we do for the problem at hand).

Inv This corresponds to the inverse optimization scheme with features that we propose, which runs following the two-step estimation procedure described in Section B.3 with B=12 blocks of utility. Here we only use the outside temperature and hourly indicator variables as features. We re-parametrize weights w_t with respect to a single parameter, called forgetting factor, and

denoted as $E \geq 0$, in the following manner: $w_t = gap_t \left(\frac{t}{T}\right)^E$ for $t \in \mathcal{T}$ and T being the total number of periods. The variable *gap* indicates whether the observation was correctly measured ($gap = 1$) or not ($gap = 0$). Parameter E indicates how rapidly the weight drops (how rapidly the model forgets). When $E = 0$, the weight of the observations is either 1 or 0 depending on the variable *gap*. As E increases, the recent observations weight comparatively more than the old ones.

B.4.2 Validation of the Model and Performance in December

In this subsection we validate the benchmarked models and assess their performance during the test month of December 2006. As previously mentioned, the evaluation and comparison of the different benchmarked models is conducted in terms of *prediction errors*, and not in monetary values (e.g., in the form of market revenues). This relieves us of having to arbitrarily assume a particular market organization behind the Olympic Peninsula experiment and a particular strategy for bidding into the market based on the ARX model that we use for benchmarking. Furthermore, a well-functioning electricity market should *not* reward prediction errors, that is, energy imbalances. In fact, a number of electricity markets throughout the world explicitly penalize prediction errors through the use of a two-price balancing settlement (see, for instance, [B29] for further information on this).

To predict the aggregated consumption of the pool of flexible loads, we need good forecasts of the electricity price and the features, as these work as explanatory variables in all the considered models. For the sake of simplicity, we use the actual values of the electricity price and the features that were historically recorded as such good forecasts. Since this simplification applies to all the benchmarked models, the analysis and comparison that follow is fair.

It is worth noticing, though, that the proposed methodology need *not* a prediction of the electricity price when used for *bidding* in the market and not for predicting the aggregated consumption of a cluster of loads. This is so because the market bid expresses the desired consumption of the pool of loads for any price that clears the market. The same cannot be said, however, for prediction models of the type of ARX, which would need to be used in combination with extra tools, no matter how simple they could be, for predicting the electricity price and for optimizing under uncertainty in order to generate a market bid.

There are two parameters that need to be chosen before testing the models:
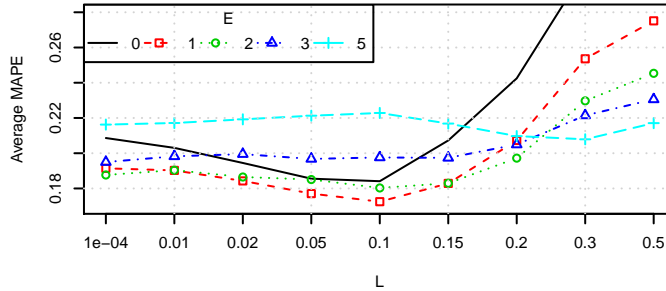
Figure B.2: Results from the validation of the input parameters L and E, to be used during December.

the penalty parameter L and the forgetting factor E. We seek a combination of parameters such that the prediction error is minimized. We achieve this by validating the models with past data, in a rolling-horizon manner, and with different combinations of the parameters L and E. The results are shown in Figure B.2. The MAPE is shown on the y-axis against the penalty L in the x-axis, with the different lines corresponding to different values of the forgetting factor E. From this plot, it can be seen that a forgetting factor of $E = 1$ or $E = 2$ yields a better performance than when there is no forgetting factor at all ($E = 0$), or when this is too high ($E \geq 3$). We conclude that selecting $L = 0.1$ and $E = 1$ results in the best performance of the model, in terms of the MAPE.

Once the different models have been validated, we proceed to test them. For this purpose, we first set the cross-validated input parameters to $L = 0.1$ and $E = 1$, and then, predict the load for the next day of operation in a rolling-horizon manner. In order to mimic a real-life usage of these models, we estimate the parameters of the bid on every day of the test period at 12:00 using historical values from three months in the past. Then, as if the market were cleared, we input the price of the day-ahead market (13 to 36 hours ahead) in the consumers' price-response model, obtaining a forecast of the consumption. Finally, we compare the predicted versus the actual realized consumption and move the rolling-horizon window to the next day repeating the process for the rest of the test period. Similarly, the parameters of the ARX model are re-estimated every day at 12:00, and predictions are made for 13 to 36 hours ahead.

Results for a sample of consecutive days, from the 10th to the 13th of December, are shown in Figure B.3. The actual load is displayed in a continuous solid line, while the load predictions from the various benchmarked models are shown with different types of markers. First, note that the *Simple Inv* model is clearly under-performing compared to the other methodologies, in terms of prediction accuracy. Recall that, in this model, the maximum and minimum
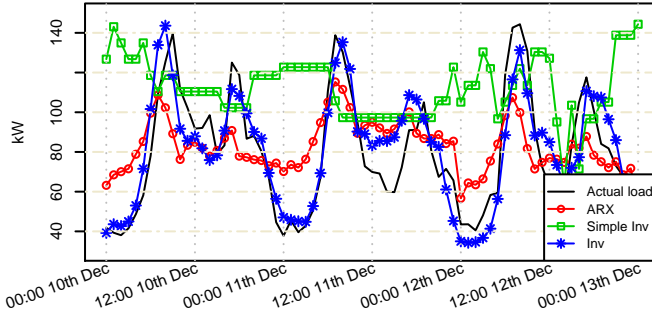
Figure B.3: Load forecasts issued by the benchmark models, and actual load, for the period between the 10th and the 13th of December.

	MAE	RMSE	MAPE
ARX	22.176	27.501	0.2750
Simple Inv	44.437	54.576	0.58581
Inv	17.318	23.026	0.1899

Table B.1: Performance measures for the three benchmarked models during December.

load consumptions, together with the maximum pick-up and drop-off rates, are estimated from historical values and assumed to remain constant along the day, independently of the external variables (the features). This basically leaves the utility alone to model the price-response of the pool of houses, which, judging from the results, is not enough. The ARX model is able to follow the load pattern to a certain extent. Nevertheless, it is not able to capture the sudden decreases in the load during the night time or during the peak hours in the morning. The proposed model (*Inv*) features a considerably much better performance. It is able to follow the consumption pattern with good accuracy.

The performance of each of the benchmarked models during the whole month of December is summarized in Table B.1. The first column shows the Mean Absolute Error (MAE), the second column provides the Root Mean Square Error (RMSE), and the third column collects the Mean Absolute Percentage Error (MAPE). The three performance metrics lead to the same conclusions: that the price-response models we propose, i.e., *Inv*, perform better than the ARX model and the *Simple Inv* model. The results collated in Table I also yield an interesting conclusion: that the electricity price is not the only driver of the consumption of the pool of houses and, therefore, is not explanatory enough to predict the latter alone. We conclude this after seeing the performance
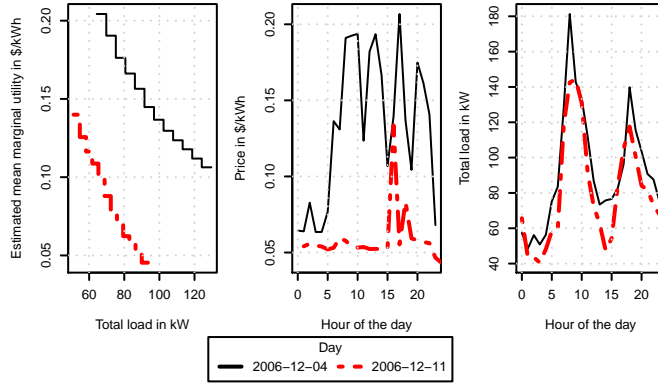
Figure B.4: Averaged estimated block-wise marginal utility function for the
Inv model (left panel), price in \$/kWh (middle panel), and load
in kW (right panel). The solid lines represent data relative to the
4th of December. Dashed lines represent data relative to the 11th
of December.

of the *Simp Inv*, which is not able to follow the load just by modeling the
price-consumption relationship by means of a marginal utility function. The
performance is remarkably enhanced when proper estimations of the maximum
pick-up and drop-off rates and the consumptions bounds as functions of the
features are employed.

The estimated block-wise marginal utility function, averaged for the 24 hours of
the day, is shown in the left plot of Figure B.4 for the *Inv* model. The solid line
corresponds to the 4th of December, when the price was relatively high (middle
plot), as was the aggregated consumption of the pool of houses (right plot). The
dashed line corresponds to the 11th of December and shows that the estimated
marginal utility is lower, as is the price on that day.

B.4.3   Performance During September and March

In this section, we summarize the performance of the benchmarked models dur-
ing September 2006 and March 2007.

In Table B.2, summary statistics for the predictions are provided for September
(left side) and March (right side). The conclusions remain similar as the ones
drawn for the month of December. The *Inv* methodology consistently achieves
the best performance during these two months as well.

	September			March		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARX	7.649	9.829	0.2350	17.439	23.395	0.2509
Simple Inv	14.263	17.8	0.4945	44.687	54.616	0.8365
Inv	5.719	8.582	0.1462	12.652	16.776	0.1952

Table B.2: Performance measures for the three benchmarked models

By means of cross-validation [B26], we find that the user-tuned parameters yielding the best performance vary over the year. For September, the best combination is $L = 0.3$, $E = 0$, while for March it is $L = 0.3$, $E = 1$.

The optimized penalization parameter L turns out to be higher in September and March than in December. This penalization parameter is highly related to the actual flexibility featured by the pool of houses. Indeed, for a high enough value of the penalty (say $L \geq 0.4$ for this case study), violating the complementarity conditions associated with the consumers' price-response model (B.1) is relatively highly penalized. Hence, at the optimum, the slacks of the complementarity constraints in the relaxed estimation problem (B.5) will be zero or close to zero. When this happens, it holds at the optimum that $r_t^u = -r_t^d$ and $\underline{P}_t = \overline{P}_t$. The resulting model is, therefore, equivalent to a linear model of the features, fit by least weighted absolute errors. When the best performance is obtained for a high value of L, it means that the pool of houses does not respond so much to changes in the price. On the other hand, as the best value for the penalization parameter L decreases towards zero, the pool becomes more price-responsive: the maximum pick-up and drop-off rates and the consumption limits leave more room for the aggregated load to change depending on the price.

Because the penalization parameter is the lowest during December, we conclude that more flexibility is observed during this month than during September or March. The reason could be that December is the coldest of the months studied, with a recorded temperature that is on average 9.4°C lower, and it is at times of cold whether when the electric water heater is used the most.

B.5   Summary and Conclusions

We consider the market-bidding problem of a pool of price-responsive consumers. These consumers are, therefore, able to react to the electricity price, e.g., by shifting their consumption from high-price hours to lower-price hours. The total amount of electricity consumed by the aggregation has to be purchased in the

electricity market, for which the aggregator or the retailer is required to place a bid into such a market. Traditionally, this bid would simply be a forecast of the load, since the load has commonly behaved inelastically. However, in this paper, we propose to capture the price-response of the pool of flexible loads through a more complex, but still quite common market bid that consists of a stepwise marginal utility function, maximum load pick-up and drop-off rates, and maximum and minimum power consumption, in a manner analogous to the energy offers made by power producers.

We propose an original approach to estimate the parameters of the bid based on inverse optimization and bi-level programming. Furthermore, we use auxiliary variables to better explain the parameters of the bid. The resulting non-linear problem is relaxed to a linear one, the solution of which depends on a penalization parameter. This parameter is chosen by cross-validation, proving to be adequate from a practical point of view.

For the case study, we used data from the Olympic Peninsula project to asses the performance of the proposed methodology. We have shown that the estimated bid successfully models the price-response of the pool of houses, in such a way that the mean absolute percentage error incurred when using the estimated market bid for predicting the consumption of the pool of houses is kept in between 14% and 22% for all the months of the test period.

We envision two possible avenues for improving the proposed methodology. The first one is to better exploit the information contained in a large dataset by allowing for non-linear dependencies between the market-bid parameters and the features. This could be achieved, for example, by the use of B-splines. The second one has to do with the development of efficient solution algorithms capable of solving the exact estimation problem within a reasonable amount of time, instead of the relaxed one. This could potentially be accomplished by decomposition and parallel computation.

Acknowledgment

Appendix B.I: Robust Constraints

Next we show how to formulate robust constraints to ensure that the estimated minimum consumption be always equal to or lower than the estimated maximum consumption. At all times, and for all plausible realizations of the external variables, we want to make sure that:

$$\underline{P}^0 + \sum_{i\in\mathcal{I}} \alpha_i^{\underline{P}} Z_{i,t} \leq \overline{P}^0 + \sum_{i\in\mathcal{I}} \alpha_i^{\overline{P}} Z_{i,t}, \quad t \in \mathcal{T}, \ \forall Z_{i,t}. \tag{B.7}$$

If (B.7) is not fulfilled, problem (B.1) is infeasible (and the market bid does not make sense). Assuming we know the range of possible values of the features, i.e., $Z_{i,t} \in [\overline{Z}_i, \underline{Z}_i]$, (B.7) can be rewritten as:

$$\underline{P}^0 - \overline{P}^0 + \underset{\substack{Z'_{i,t} \\ \text{s.t. } \underline{Z}_i \leq Z'_{i,t} \leq \overline{Z}_i \\ i\in\mathcal{I}}}{\text{Maximize}} \left\{ \sum_{i\in\mathcal{I}} (\alpha_i^{\underline{P}} - \alpha_i^{\overline{P}}) Z'_{i,t} \right\} \leq 0, \ t \in \mathcal{T}. \tag{B.8}$$

Denote the dual variables of the upper and lower bounds of $Z'_{i,t}$ by $\overline{\phi}_{i,t}$ and $\underline{\phi}_{i,t}$ respectively. Then, the robust constraint (B.8) can be equivalently reformulated as:

$$\overline{P}^0 - \underline{P}^0 + \sum_{i\in\mathcal{I}} (\overline{\phi}_{i,t}\overline{Z}_i - \underline{\phi}_{i,t}\underline{Z}_i) \leq 0 \qquad\qquad t \in \mathcal{T} \tag{B.9a}$$

$$\overline{\phi}_{i,t} - \underline{\phi}_{i,t} = \alpha_i^{\overline{P}} - \alpha_i^{\underline{P}} \qquad\qquad i \in \mathcal{I}, t \in \mathcal{T} \tag{B.9b}$$

$$\overline{\phi}_{i,t}, \underline{\phi}_{i,t} \geq 0 \qquad\qquad i \in \mathcal{I}, t \in \mathcal{T}. \tag{B.9c}$$

Following the same reasoning, one can obtain the set of constraints that guarantees the non-negativity of the lower bound and consistent maximum pick-up and drop-off rates. We leave the exposition and explanation of these constraints out of this paper for brevity.

References B

[B1]   R. Halvgaard et al. "Economic Model Predictive Control for building climate control in a Smart Grid". In: *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES* (2012), pp. 1–6.

[B2]   A. Conejo, J. Morales, and L. Baringo. "Real-Time Demand Response Model". In: *IEEE Transactions on Smart Grid* 1.3 (2010), pp. 236–242.

[B3]   A.-H. Mohsenian-Rad et al. "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid". In: *Smart Grid, IEEE Transactions on* 1.3 (2010), pp. 320–331.

[B4]   G. Dorini, P. Pinson, and H. Madsen. "Chance-Constrained Optimization of Demand Response to Price Signals". In: *IEEE Transactions on Smart Grid* 99 (2013), pp. 1–1. ISSN: 1949-3053.

[B5]   O. Corradi et al. "Controlling Electricity Consumption by Forecasting its Response to Varying Prices". In: *IEEE Transactions on Power Systems* 28.1 (2013), pp. 421–429.

[B6]   M. Zugno et al. "A bilevel model for electricity retailers' participation in a demand response market environment". In: *Energy Economics* 36 (2013), pp. 182–197.

[B7]   C. Chen, S. Kishore, and L. Snyder. "An innovative RTP-based residential power scheduling scheme for smart grids". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, pp. 5956–5959.

[B8]   F.-L. Meng and X.-J. Zeng. "A Stackelberg game-theoretic approach to optimal real-time pricing for the smart grid". In: *Soft Computing* (2013), pp. 1–16.

[B9]   G. E. Asimakopoulou, A. G. Vlachos, and N. D. Hatziargyriou. "Hierarchical Decision Making for Aggregated Energy Management of Distributed Resources". In: *Power Systems, IEEE Transactions on* 30.6 (2015), pp. 3255–3264.

[B10]   H. Mohsenian-Rad. "Optimal Demand Bidding for Time-Shiftable Loads". In: *Power Systems, IEEE Transactions on* 30.2 (Mar. 2015), pp. 939–951. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2338735.

[B11]   M. Gonzalez Vaya and G. Andersson. "Optimal Bidding Strategy of a Plug-In Electric Vehicle Aggregator in Day-Ahead Electricity Markets Under Uncertainty". In: *Power Systems, IEEE Transactions on* 30.5 (Sept. 2015), pp. 2375–2385. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2363159.

[B12]   S. Li et al. "Market-Based Coordination of Thermostatically Controlled Loads Part I: A Mechanism Design Formulation". In: *Power Systems, IEEE Transactions on* PP.99 (2015), pp. 1–9. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2015.2432057.

[B13]   L. J. Ratliff et al. "Incentive design and utility learning via energy disaggregation". In: *arXiv preprint arXiv:1312.1394* (2013).

[B14]   R. K. Ahuja and J. B. Orlin. "Inverse Optimization". In: *Operations Research* 49 (2001), pp. 771–783.

[B15]   T. C. Chan et al. "Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy". In: *Operations Research* (2014).

[B16]   A. Keshavarz, Y. Wang, and S. Boyd. "Imputing a convex objective function". In: *Intelligent Control (ISIC), 2011 IEEE International Symposium on.* Sept. 2011, pp. 613–619.

[B17]   D. Bertsimas, V. Gupta, and I. Paschalidis. "Data-driven estimation in equilibrium using inverse optimization". English. In: *Mathematical Programming* 153.2 (2015), pp. 595–633. ISSN: 0025-5610. DOI: 10.1007/s10107-014-0819-4. URL: http://dx.doi.org/10.1007/s10107-014-0819-4.

[B18]   A. Aswani, Z.-J. M. Shen, and A. Siddiq. "Inverse Optimization with Noisy Data". In: *arXiv preprint arXiv:1507.03266* (2015).

[B19]   S. Siddiqui and S. Gabriel. "An SOS1-Based Approach for Solving MPECs with a Natural Gas Market Application". English. In: *Networks and Spatial Economics* 13.2 (2013), pp. 205–227. ISSN: 1566-113X.

[B20]   D. Hammerstrom et al. *Pacific Northwest GridWise Testbed Demonstration Projects. Part I. Olympic Peninsula Project.* Tech. rep. Oct. 2007.

[B21]   D. Luenberger and Y. Ye. *Linear and Nonlinear Programming.* International Series in Operations Research & Management Science. Springer, 2008. ISBN: 9780387745022.

[B22]   A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization.* Princeton University Press, 2009.

[B23]   M. C. Ferris and T. S. Munson. "Complementarity problems in GAMS and the PATH solver". In: *Journal of Economic Dynamics and Control* 24.2 (2000), pp. 165–188.

[B24]   J. Fortuny-Amat and B. McCarl. "A representation and economic interpretation of a two-level programming problem". In: *Journal of the operational Research Society* (1981), pp. 783–792.

[B25]   E. M. L. Beale and J. A. Tomlin. "Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables". In: *Operations Research* 69 (1970), pp. 447–454.

[B26] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning.* Vol. 1. Springer series in statistics Springer, Berlin, 2001.

[B27] *DTU Computing Center.* URL: `http://www.hpc.dtu.dk/`.

[B28] H. Madsen. *Time Series Analysis.* Chapman & Hall, 2007.

[B29] A. Skajaa, K. Edlund, and J. M. Morales. "Intraday trading of wind energy". In: *Power Systems, IEEE Transactions on* 30.6 (2015), pp. 3181–3189.

PAPER C

Optimal Price-energy Demand Bids for Aggregate Price-responsive Loads

Authors:
Javier Saez-Gallego, Mahdi Kohansal, Ashkan Sadeghi-Mobarakeh and
Juan M. Morales

Optimal Price-energy Demand Bids for Aggregate Price-responsive Loads

Javier Saez-Gallego[1], Mahdi Kohansal[2], Ashkan Sadeghi-Mobarakeh[2] and Juan M. Morales[1].

Abstract

In this paper we seek to optimally operate a retailer that, on one side, aggregates a group of price-responsive loads and on the other, submits block-wise demand bids to the day-ahead and real-time markets. Such a retailer/aggregator needs to tackle uncertainty both in customer behavior and wholesale electricity markets. The goal in our design is to maximize the profit for the retailer/aggregator. We derive closed-form solutions for the risk-neutral case and also provide a stochastic optimization framework to efficiently analyze the risk-averse case. In the latter, the price-responsiveness of the load is modeled by means of a non-parametric analysis of experimental random scenarios, allowing for the response model to be non-linear. The price-responsive load models are derived based on the Olympic Peninsula experiment load elasticity data. We benchmark the proposed method using data from the California ISO wholesale electricity market.

Notation

The main notation used throughout the paper is stated below for quick reference. Other symbols are defined as required.

Indexes and sets

t Time period $t \in \{1, 2, \ldots 24\}$.

b Bidding block $b \in \{1, 2, \ldots B\}$.

w Realization of the stochastic variables, represented as scenarios $w = \{1, 2, \ldots N\}$.

Input stochastic processes

X Load.

Λ^D Day-ahead price.

Λ^R Real-time price.

Π Retail price.

Decision variables

X^D Stochastic process representing scheduled energy in the day-ahead market.

$x^D_{t,w}$ Scheduled energy in the day-ahead market for time t and scenario w.

$u_{t,b}$ Price bid for time t and block b.

Remark: a subscript t under the stochastic processes indicate the associated random variable for time t.

Parameters

ϕ_w Probability of each scenario w.

$x_{t,w}$ Load at time t and scenario w.

$\lambda^D_{t,w}$ Day-ahead price at time t and scenario w.

$\lambda^R_{t,w}$ Real-time price at time t and scenario w.

$\pi_{t,w}$ Retail price at time t and scenario w.

E_b Width of energy block b.

L Fraction of the load that must be purchased in the day-ahead market.

β Probability of occurrence of chance constraint.

C.1   Introduction

With the increasing deployment of smart grid technologies and demand response programs, more markets around the world are fostering demand bids that reflect the response of the consumers to changing electricity prices [C1, C2]. In this paper, we consider the case of a retailer who procures energy to a pool of consumers in a typical two-settlement electricity market, as for example, the California wholesale electricity market CAISO [C3]. The retailer submits price-energy demand bids to the day-ahead market, and only energy quantity bids to the real-time market in order to counterbalance the deviations from the scheduled day-ahead energy market to the actual load. The possibility of arbitrage is indirectly allowed depending on the submitted bid to the day-ahead market and the realization of the stochastic processes affecting the problem.

We assume that the load is price-responsive, in the sense that it changes depending on the price of electricity during the considered period. The proposed methodology relies on historical data of load and retail price to estimate the relationship between price and load. Because of this, we do not make any assumption on the nature of the load the retailer aggregates. Also, in the considered setup, the retail price is considered to be given exogenously. For example, the retail price can be proportional to the day-ahead price plus a fee.

The contributions of the paper are summarized as follows:

- An analytic solution to the problem of finding optimal block-wise price-energy demand bids in the day-ahead market when risk is not considered. Moreover, we propose a mixed-integer linear programming solution approach to the risk-averse case.

- The dynamic price-responsive behavior of consumers is modeled based on scenarios. The conditional probability of the load given a certain retail price trajectory is estimated using a non-parametric approach.

- We assess the practicality of the proposed methodology by using data from a real-world experiment.

The estimation of demand bids has been extensively studied in the past years [C4, ch. 7]. Several papers share the common goal of estimating price-energy bids relative to specific types of load, for example, time-shiftable loads [C5], electric vehicles [C6] and thermostatically-controlled loads [C7]. Our methodology differs with those in the fact that we do not make any assumption on the nature of the load. Methodologies based on forecasting tools [C8, C9] generally

do not make assumptions on the type of price-responsive load either, but, on the other hand, do not tackle the bidding problem.

Besides the works on forecasting, another group of papers focus on finding the optimal bid for generic loads. The work in [C10] elaborates on a robust bidding strategy against procurement costs higher than the expected one, considering uncertainty in the prices only. Uncertain prices and demand are taken into account in [C11] but minimizing imbalances and disregarding the economic side of the bidding. Our approach resembles that of [C12] with the main differences being that we use data to estimate the price-response of the dynamic load, and that we consider energy-block bidding in a one-price balancing market as the US CAISO [C3]. Authors of [C13] consider, from the theoretical point of view, the problem of allocating a deterministic load by deciding which fraction should be purchased in the day-ahead market and which in the real-time market. Finally, authors of [C14] study demand curves in an arbitrage- and risk-free situation by using a game theory.

Regarding the generation of scenarios of the stochastic processes, our methodology is inspired from [C15, C16, C17]. From the application point of view, our approach differs in the final goal, as they deal with wind energy production. To our knowledge, there is no previous work that characterizes the dynamic price-responsive load with a set of scenarios. From the methodological point of view, our approach differs with the existing literature in the estimation of the conditional distribution of the price-responsive load, taking into account the full trajectory of the day-ahead price. This enables us to capture the full dynamics of the load across the hours of the next operational day. The real-time price is modeled in an analogous manner. In both cases, we model their distributions using a non-parametric approach that allows for non-linear responses to a given day-ahead price trajectory.

The paper is structured as follows. In Section C.2 we introduce the retailer's bidding problem. Section C.3 provides the analytic solution to the risk-neutral case. In Section C.4 we formulate the stochastic optimization model for solving the bidding problem with risk constraints. Section C.5 elaborates on the scenario-generation technique. Next, in Section C.6 we analyze results from the bidding problem under the generated scenarios. Finally, in Section C.7 we draw conclusions and implications.

C.2   Problem Formulation

Consider a utility retailer/aggregator that seeks to maximize its profit based on the revenue that it collects from its loads, the payments it makes to the day-ahead market, and the payments it makes or receives in the real-time market. Mathematically speaking, we need to solve the following optimization problem:

$$\underset{X_t^D, u_{t,b}}{\text{Maximize}} \quad \mathbb{E}\left\{ \sum_{t=1}^{24} \left(\Pi_t X_t - \Lambda_t^D X_t^D - \Lambda_t^R \left(X_t - X_t^D \right) \right) \right\} \qquad \text{(C.1a)}$$

subject to

$$X_t^D = \sum_{b=1}^{B} E_b \, \mathbb{I}(u_{t,b} \geq \Lambda_t^D) \qquad\qquad \forall t,b \qquad\qquad \text{(C.1b)}$$

$$u_{t,b+1} \leq u_{t,b} \qquad\qquad \forall t, b = 1 \ldots B-1 \qquad \text{(C.1c)}$$

$$\text{P}\left(X_t^D \in [(1-L)\,X_t, (1+L)X_t] \right) \geq \beta \qquad\qquad \forall t \qquad\qquad \text{(C.1d)}$$

$$\underline{\lambda} \leq u_{t,b} \leq \overline{\lambda} \qquad\qquad \forall t,b \qquad\qquad \text{(C.1e)}$$

where $\mathbb{I}(\cdot)$ is the 0-1 indicator function.

The objective function (C.1a) is the expected total daily profit, composed of three terms. The first term represents the revenue that the retailer makes form selling energy to the consumers at the retail price. The second term represents the cost of purchasing energy from the day-ahead market. The third term accounts for the cost/revenue of purchasing/selling energy from/to the real-time market. The energy purchased or sold in the real-time market is equal to the difference between the purchased quantity at the day-ahead market and the realized load, i.e., $X_t - X_t^D$.

Constraint (C.1b) defines the scheduled energy in the day-ahead market to be equal to the sum of the width of the blocks of energy which have a price-bid higher than the market price. In other words, blocks of energy will be purchased if their price-bid is higher or equal to the day-ahead price. Note that $u_{t,b}$ is the decision variable which determines the shape of the submitted bidding curve to the day-ahead market.

Constraint (C.1d) models the risk-aversion of the retailer through two parameters. Parameter L represents the maximum fraction of the load that can be procured in the real-time market. This parameter could be defined by the retailer, but likewise could be given by the ISO as a way to ensure the stability of the system. Values of L close to 1 indicate that the full amount of the load

can potentially be bought in the real-time market. On the other hand, as L decreases, we give priority to purchasing energy in the day-ahead market. Parameter β indicates the minimum probability with which the constraint (C.1d) must be fulfilled. Values of β close to 1 indicate a hard constraint, while lower values of β indicate that the constraint is loose. The parameter β can be interpreted as the aversion of the retailer towards purchasing a certain fraction of the load in the day-ahead market. Low values of β can be interpreted as a sign that the retailer seeks to profit from arbitrage rather than from serving the load. As we show in the case study, higher values of β yield lower expected profit but also lower risk. Note that, for large L and small β, the constraint above (C.1d) becomes irrelevant, indicating the neutrality of the retailer towards risk.

Constraint (C.1c) ensures that the estimated bidding curve is monotonically decreasing which is a typical requirement in electricity markets. Finally, constraint (C.1e) set lower and upper bounds to the price bids, which are given by the market rules [C18]. All in all, the expected profit depends on the decision variable "price-bid" and also on the realization of the input stochastic variables.

The maximum number of blocks that is allowed depend on the market rules [C18] as well. The width of each block E_b must be set by the retailer depending on the magnitude of the load.

As in practice, here we assume that the retail price is given exogenously, in other words, it is not a decision variable of the retailer. The main driver for this consideration is the fact that the retail price must, to a certain extent, represent the true cost of electricity. This might not always be the case if the retail price is subject to the will of the retailer. As a consequence, the bidding curve does not directly affect the welfare of the consumers, since the welfare of the consumers depends on the retail price. Another implication is that the load, which is a random variable dependent on the retail price, is not affected by the outcome of the proposed methodology, yet only the profit of the retailer is.

C.3 Closed-Form Analytical Solution in Absence of Risk Constraints

In this subsection we elaborate on the closed-form analytic solution to problem (C.1), when the risk constraint (C.1d) is disregarded, or equivalently, when $L \to \infty$ and/or $\beta = 0$. In such a case, the bidding problem (C.1) can be decomposed by time period, so that for every time t we solve a single optimization problem.

The risk-neutral problem, for every time t, is written as follows:

$$\underset{X_t^D, u_{t,b}}{\text{Max.}} \left(\mathbb{E}\left\{ X_t \left(\Pi_t - \Lambda_t^R \right) \right\} - \mathbb{E}\left\{ X_t^D \left(\Lambda_t^D - \Lambda_t^R \right) \right\} \right) \tag{C.2}$$

subject to (C.1b), (C.1c) and (C.1e). The advantage of reformulation (C.2) is that we can perform simpler optimization problems in parallel. It is noteworthy to say that the first term of (C.2) is constant with respect to the decision variables $u_{t,b}$, however, the last terms is not.

Next we analyze the case when Λ^D and Λ^R are statistically independent. Results are presented in Theorem 1. For ease of readability, in the remaining of this section we drop the index t, which generally affects decisions and random variables.

Theorem 1   The optimal price bid u_b^* in problem (C.2), when the day-ahead and real-time prices are independent, is equal to the expected value of the real-time price.

The proof of Theorem 1 is given in Appendix C.I. Theorem 1 also shows that, given the risk-neutral setup and independent prices, we do not obtain extra benefit from bidding a curve instead of a single price-quantity bid.

The assumption of statistically independent prices is not necessarily fulfilled in practice (see, for example, [C19, Fig. 1]). For this reason, in Theorem 2 below, we provide the analytic solution to problem (C.2) when Λ^D and Λ^R are statistically dependent.

Theorem 2   A global optimum solution to problem (C.2) satisfies that the price bids for all blocks is equal to u^*. Moreover, u^* is equal to either $\underline{\lambda}$, $\overline{\lambda}$, or $\mathbb{E}\left\{ \Lambda^R | \Lambda^D = u^* \right\}$ with $\frac{d}{du} \mathbb{E}\left\{ \Lambda^R | \Lambda^D = u^* \right\} < 1$ in the latter case.

The proof of Theorem 2 is given in Appendix C.II. One could interpret the result of Theorem 2 in the following way: the optimal price bid will be the one for which price consistency holds, namely, for which the expected real-time price is equal to the day-ahead price. A second conclusion drawn from Theorem 2 is that the maximum profit is achieved with the same price-bid for each block. If there is more than one price bid that maximizes the expected profit (i.e., several global maxima), then the price bid for each block can be chosen indistinctly between them. Similarly as with Theorem 1, we do not obtain extra benefit from bidding a curve when prices are dependent.

From a practical point of view, Theorem 1 and 2 allow us to simplify the demand curve to a simple price-quantity bid. By taking into account this implication, we can obtain the optimal price bid in the case when the distributions of prices are discrete, which allow us to compute the optimal price bid when the uncertainty is modeled by scenarios. The optimal price bid can be chosen by evaluating the profit in the local maxima, which are characterized according to the following remark:

Remark 2 Given a discrete set of scenarios for the day-ahead and real-time prices, let us consider the re-ordered pair of terms $\{\lambda_w^D, \mathbb{E}\left\{\Lambda^R|\Lambda^D = \lambda_w^D\right\}\}$ such that $\lambda_w^D \leq \lambda_{w+1}^D$. Local maxima[3] are achieved at the stationary points $u^* = \lambda_w^D$ such that $\lambda_w^D \leq \mathbb{E}\left\{\Lambda^R|\Lambda^D = \lambda_w^D\right\}$ and $\lambda_{w+1}^D > \mathbb{E}\left\{\Lambda^R|\Lambda^D = \lambda_{w+1}^D\right\}$.

Note that, due to market rules, the price bid have a maximum and minimum allowed values. In practice, one needs to check also if the maximum profit is achieved when the price bid is equal to one of its bounds. Using Remark 2 one can find the optimal price-bid by just performing a finite set of simple calculations.

As a final remark, it is noteworthy to say that the results from Theorem 1 and 2 show that the solution to (C.2) does not depend on the retail price, neither on the load. From a practical point of view this means that the risk-neutral retailer acts a financial trader, making profit by selling and buying energy in both markets.

C.4 Scenario-based Solution in Presence of Risk Constraints

In this section we present a solution to problem (C.1) using a scenario-based approach. The input for every time t is a set of N scenarios, each one characterized by a realization of the retail price $\pi_{t,w}$, the day-ahead price $\lambda_{t,w}^D$, the real-time price $\lambda_{t,w}^R$, and the load $x_{t,w}$. Each scenario has a probability of occurrence of ϕ_w.

We reformulate constraint (C.1b) by adding a binary variable $y_{t,w,b}$. Then, constraint (C.1b) is replaced by:

---

[3]The proof is available upon request.

$$x_{t,w}^D = \sum_b y_{t,w,b} E_b \qquad \forall t, w$$

$$
\begin{aligned}
u_{t,b} - \lambda_{t,w}^{DA} &\le \text{M} y_{t,w,b} && \forall t, w, b \\
-u_{t,b} + \lambda_{t,w}^{DA} &\le \text{M}(1 - y_{t,w,b}) && \forall t, w, b \\
y_{t,w,b} &\in \{0, 1\} && \forall t, w, b
\end{aligned}
\tag{C.3}
$$

where M is a large enough constant. The equations above imply that $y_{t,w,b} = 1$ if $u_{t,b} \ge \lambda_{t,w}^D$ and 0 otherwise.

Next, we reformulate constraint (C.1d) by adding two extra binary variables. We first define $\underline{z}_{t,w} = 1$ if $x_{t,w}^D \le (1 - L)x_{t,w}$, and $\underline{z}_{t,w} = 0$ otherwise. Secondly, we define $\overline{z}_{t,w} = 1$ if $x_{t,w}^D \ge (1 + L)x_{t,w}$, and $\overline{z}_{t,w} = 0$ otherwise. Consequently, the chance constraint (C.1d) can be replaced by the following set of equations:

$$
\begin{aligned}
x_{t,w}^D - (1 - L)x_{t,w} &\le M(1 - \underline{z}_{t,w}) && \forall w \\
-x_{t,w}^D + (1 - L)x_{t,w} &\le M\underline{z}_{t,w} && \forall w \\
x_{t,w}^D - (1 + L)x_{t,w} &\le M\overline{z}_{t,w} && \forall w \\
-x_{t,w}^D + (1 + L)x_{t,w} &\le M(1 - \overline{z}_{t,w}) && \forall w \\
\frac{1}{N} \sum_w \left(\overline{z}_{t,w} + \underline{z}_{t,w} \right) &\le 1 - \beta.
\end{aligned}
\tag{C.4}
$$

All in all, taking into consideration the reformulations presented above, the optimal price-bid is found by maximizing, for every time t,

$$\underset{x_{t,w}^D, u_b}{\text{Maximize}} \quad \sum_w \phi_w \left(\pi_{tw} x_{tw} - \lambda_{tw}^D x_{tw}^D - \lambda_{tw}^R (x_{tw} - x_{tw}^D) \right) \tag{C.5}$$

subject to (C.1c), (C.1e), (C.3), and (C.4).

C.5   Scenario Generation

In this section we elaborate on the modeling of the stochastic variables by scenarios. The proposed approach to generate scenarios has several advantages. First, we do not need to make any assumption on the type of price-responsive load we model. The response of the load to the price is directly observed in the data and modeled by a non-parametric distribution. For this very same reason, the response of the load to the price is allowed to be non-linear. Second, it

is a fast approach, hence, big datasets can be quickly processed. Finally, the proposed approach is adequate for bidding purposes, since forecasting the load is not the main goal of the paper but rather account for its uncertainty in order to make an informed decision.

Each scenario is characterized by a 24-long sequence of day-ahead prices, real-time prices, retail prices and observed load. The proposed method to approximate their joint distribution is summarized as follows. First of all, we model the marginal distribution of the day-ahead price. Note that the day-ahead price is not dependent on the real-time price, neither on the bid of a small price-taker consumer. Second, we model the distribution of the load conditioned on the retail price using a non-parametric approach. Lastly, we model the distribution of the real-time price conditioned on the day-ahead price. The real-time price depends on the day-ahead price, but not on the load of a price-taker retailer.

The rest of this section is organized as follows. First, in Section C.5.1, we briefly elaborate on the technique to generate scenarios of day-ahead price. Then, for each scenario of day-ahead price, we generate conditional scenarios of real-time price and load in Section C.5.2.

C.5.1 Day-ahead Price Scenarios

The first step in the scenario generation procedure is to model the day-ahead price using an Autoregressive Integrated Moving Average model (ARIMA). We choose the most adequate model according to the AICc criteria [C20]. Using the estimated model, we draw scenarios using the methodology explained in [C21]. Because the scenarios are used in day-ahead trading, they are generated in a rolling horizon manner everyday at 12:00 with a lead time of 13 to 36 hours.

C.5.2 Load and Real-time Price Scenarios

In this section, we elaborate on the proposed methodology to draw scenarios from the distribution of load conditioned on the retail price. The methodology to generate conditional real-time price scenarios is analogous, hence, we omit it for brevity.

For this subsection, we consider a scenario of day-ahead prices $\tilde{\lambda}^D = \{\lambda_1^D, \ldots, \lambda_{24}^D\}$ that is generated using the methodology explained in Section C.5.1. Under the considered setup, as explained in the sections above, the retail price is given exogenously. In the case study, we assume the retail price to be proportional to

the day-ahead price, that is, $\Pi = k\lambda^D$. Therefore a scenario of retail price is directly specified from a scenario of day-ahead price.

The procedure outlined next allows us to weigh the historical trajectories, such that trajectories with a retail price "closer" to the given retail price $\tilde{\pi}$ weigh more. These weights are used later in this section to compute the conditional density function of the load, given $\tilde{\pi}$. To begin with, we define $\pi^{(j)}$ as the 24-long vectors of retail price, with each element referring to an hour of the day, and with j referring to the index of the historical day considered. Then, we compute the Euclidean distance $d^{(j)} = ||\pi^{(j)} - \tilde{\pi}||$. In this way, we "summarize" each historical price trajectory $\pi^{(j)}$ with a single value, so that trajectories "closer" to the given retail price $\tilde{\pi}$ have a lower distance. Next, we use a Gaussian kernel to weight trajectories, such that the weights are equal to $w^{(j)'} = f(d^{(j)})$, where f is the probability density function of a normal distribution with mean 0 and standard deviation σ_f. For the case study, we used $\sigma_f = K\sigma_d$, meaning that the standard deviation for f is equal to the standard deviation of the distances σ_d, multiplied by a *bandwidth* parameter K. Finally, we normalize the weight $w^{(j)} = \frac{w^{(j)'}}{\sum w^{(j)'}}$ so that their sum is equal to 1.

The effect of the bandwidth parameter K over the weights can be seen in the left plot of Fig. C.1. On its x-axis, we represent $d^{(j)}$ and on the y-axis the weights $w^{(j)}$. A smaller bandwidth penalizes price references further away. This is the reason why, when $K = 0.5$, there are few scenarios with a weight significantly greater than zero. On the other hand, when $K = 10$, all scenarios weigh similarly.

The procedure to generate each scenario is inspired from [C21] and [C16]. In short, we first transform the load data to a normal distribution using a non-parametric transformation. Then, we compute its covariance, and finally, generate random correlated Gaussian errors that are transformed back to the original distribution. The procedure consists of the following seven steps:

1. For each hour of the day, we compute a non-parametric estimation of the density of the price-responsive load [C22] conditional on a retail price trajectory $\tilde{\pi}$. We do this by computing the kernel density estimator at hour t with the weights $w^{(j)}$ in the following way:

$$\hat{f}_t(x|\tilde{\pi}) = \frac{1}{J}\sum_{j=1}^{J} w^{(j)} G_h(x - x_t^{(j)}), \qquad (C.6)$$

where $G_h(x)$ is a kernel (non-negative function that integrates to one and has zero mean), h is its bandwidth, and $x_t^{(j)}$ is the observed load at time
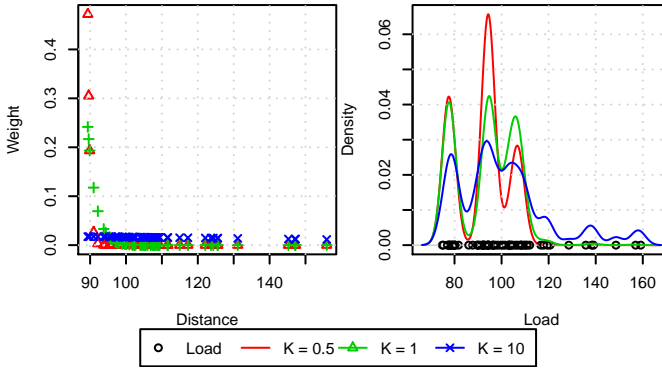
Figure C.1: On the left, the weights of the historical retail price trajectories are
shown against their distance to the price reference. On the right,
the estimated conditional distribution of the load given the retail
price is shown for different values of the bandwidth parameter K.

t and day j. An example of a estimated density using a Gaussian kernel
is shown on the right plot in Fig. C.1, for different values of K and same
h. For K close to zero ($K = 0.5$ in the case study), the weighting gives
relatively high importance to few observations, therefore, the estimated
density is more localized around them.

2. Using $\hat{f}_t(x|\tilde{\pi})$ from Step 1, we compute the cumulative density function,
   called $\hat{F}_t(x|\tilde{\pi})$.

3. The transformed load values $y_t^{(j)} = \hat{F}_t(x_t^{(j)}|\tilde{\pi})$ for every hour t follow a
   uniform distribution $U(0,1)$. Then, we normalize the load data through
   the transformation $z_t^{(j)} = \Phi^{-1}(y_t^{(j)})$, where $\Phi^{-1}(Y)$ is the probit function.
   Consequently, $(z_t^{(1)}, \ldots, z_t^{(J)}) \equiv Z_t \sim N(0,1)$.

4. We estimate the variance-covariance matrix Σ of the transformed load Z,
   relative to the 24 hours of the day. One could do it recursively as in [C21].

5. Using a multivariate Gaussian random number generator, we generate a
   realization of the Gaussian distribution $\tilde{Z} \sim N(0, \Sigma)$.

6. We use the inverse probit function to transform $\tilde{Z}$ to a uniform distribu-
   tion, that is, $\tilde{Y} = \Phi(\tilde{Z})$.

7. Finally, we obtain a scenario of load by transforming back $\tilde{Y}$ using the in-
   verse cumulative density function from step 2, that is, $\tilde{x}_t = \hat{F}_t^{-1}(\tilde{Y}_t|\tilde{\pi}), \forall t$.
   Numerically, we use a smoothing spline to interpolate $\hat{F}_t^{-1}(\tilde{Y}_t|\tilde{\pi})$.

The procedure outline above generates a scenario of load conditioned on the retail price. Steps 5 to 7 are repeated as may times as needed if more scenarios of load per retail price are desired.

C.6   Case Study

In this section we first introduce the datasets and the generated scenarios using the methodology from Section C.5. Then, in Section C.6.2, we analyze in detail the solution of the bidding model with and without considering risk. Afterwards, in Section C.6.3, we benchmark the performance of the proposed models and present the final conclusions.

C.6.1   The Data and Practical Considerations

The scenarios of day-ahead and real-time prices are generated using historical hourly values from CAISO [C3]. We use three months of training data, from August to October 2014. The test period spans over November 2014. For the retail price and price-responsive load, we use data from the Olympic Peninsula experiment [C23]. In this experiment, the electricity price was sent out every fifteen minutes to 27 households that participated in the experiment. The price-sensitive controllers and thermostats installed in each house decided when to turn on and off the appliances, based on the price and on the house owner's preferences. The training and test months are the same as for the CAISO data, but relative to year 2006.

Some practical considerations need to be addressed. Firstly, that the day-ahead price and the retail price come from two different datasets. For this reason, prices are normalized. The second practical consideration is that we assume $\Pi = k\Lambda^D$ with $k = 1$ even though it is not fulfilled in practice. However, this does not affect the comparison of the proposed models, due to the fact that we use the same set of scenarios for the benchmark and for all the models. This issue could be solved in future work when data from new experiments becomes available.

Throughout the case study, we set a total of 20 blocks, where the width is equality distributed between a maximum and a minimum bidding quantities, set to be equal to the historical range of the scenarios at every hour. They are represented by the dotted lines in Fig. C.5(b).
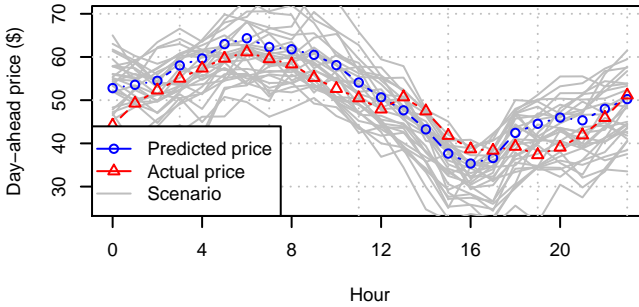
Figure C.2: Actual price, point forecast and generated scenarios for the day-ahead price.

For the case study we use a total of 150 scenarios. For the estimation of the densities, we use a Gaussian kernel with a bandwidth h given by Silverman's rule of thumb [C22]. Also, the bandwidth parameter is set to $K = 0.5$. For the model of the day-ahead price we use an ARIMA(3,1,2)(1,1,1) with a seasonal period of 24 hours. The Root Mean Square Error (RMSE) for the model of the day-ahead price (13 to 36 lead hours) is, on average, 3.22$, which is in line with the forecasting performance that other authors have achieved using similar methods [C24].

A subset of the generated scenarios of day-ahead price is given in Fig. C.2. By graphical inspection we conclude the scenarios of day-ahead price are a plausible representation of the actual day-ahead price and its uncertainty.

C.6.2 Model Analysis

To begin with, we discuss the results from the risk-neutral model (C.2). The solution to this model, for a given set of scenarios, is calculated either using Remark 2 or by solving (C.5) with $\beta = 0$. In Fig. C.3, we show the scenarios of retail price and load in dots for hour 20 of November 1st, and hour 2 of November 2nd. The estimated bidding curve is displayed as a dashed green line. In accordance with Theorem 2, the resulting bidding curve is flat.

Next, we discuss the results from the risk-averse model (C.5). We start by analyzing the effect of the risk parameters L and β on the expected profit and the feasibility of the problem. In Fig. C.4 we show, on the right axis, the feasibility frontier plot for L and β and its standard deviation in a shadowed area. We calculated it empirically, using data relative to the 1st of November.
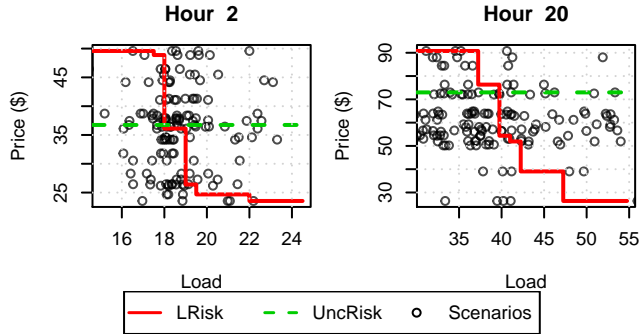
Figure C.3: The left figure is relative to the 2nd of November, while the right figure is relative to the 1st of November.

The combinations of L and β shown *below* the displayed dark line result, on average, in an infeasible solution. The frontier line is dependent on the scenarios of load: higher variability in the scenarios of load will require a greater value of L for the problem to be feasible. On the left axis of Fig. C.4, we show the expected profit for the risk-averse problem, with the combination of β and L that lay on the frontier line. Naturally, the highest profits are achieved for low values of β, that is, when the retailer is less risk averse. From now on, we set the risk parameter β to 0.8. The value of L is chosen from the frontier plot, to be as small as possible.

Fig. C.5(a) shows the scenarios of day-ahead price (continuous lines), together with the estimated optimal price bids (horizontal segments), for each hour. We observe that the magnitude of the price bid depends on the scenarios of day-ahead price. In Fig. C.5(b), we show the amount of energy bought in the day-ahead market for each scenario and the span of the bidding blocks in dashed lines. In (c), we show the scenarios of load. On average, we buy approximately in day-ahead market the expected value of the load.

The estimated price bid by the risk-averse model is represented by the continuous red line in Fig. C.3. Note that, at hour 2, the estimated price-responsiveness is much smaller than during hour 20. The reason is that, according to the scenarios of load, the load shows a lower variation during the early morning than during the early night.
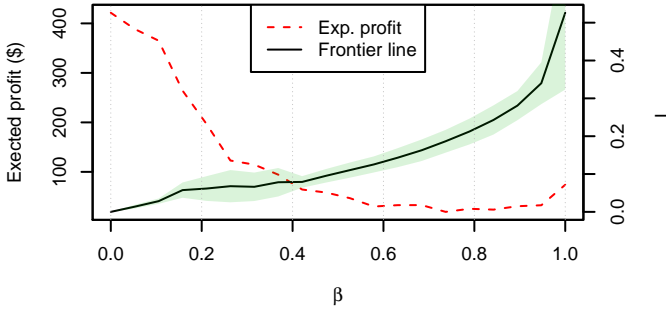
Figure C.4: On the left axis, in dashed red, the expected profit for every β, with L being in the feasibility frontier line. On the right axis, the feasibility frontier line is shown for combinations of β and L.
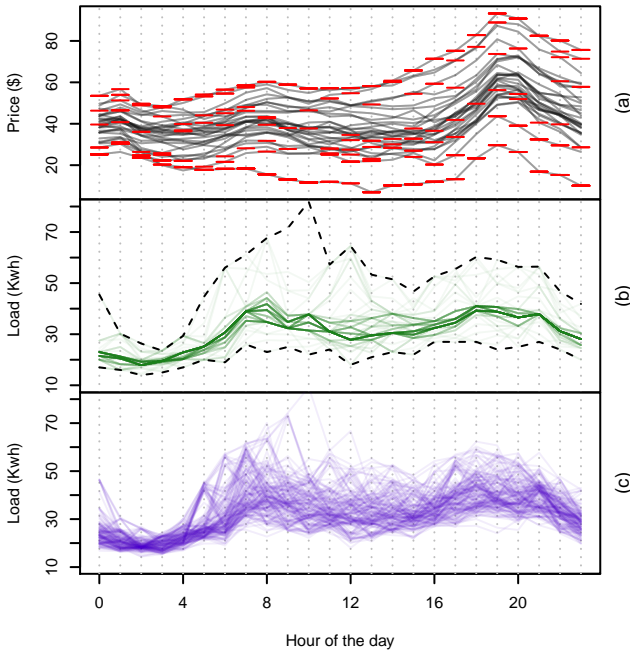
Figure C.5: In (a), the day-ahead scenarios (lines) are shown together with the estimated price-bids (horizontal segments). The day-ahead purchase for each scenario, and the load in each scenario, are shown in (b) and (c), respectively.

	Mean	Std. dev.
ExpBid	-1.78	34.52
LRisk	22.26	45.22
UncRisk	188.82	259.62

Table C.1: Mean and standard deviation of the profit for the benchmarked models during November and December.

C.6.3   Benchmark: Results in November and December

In this subsection, we benchmark the following models:

ExpBid Single block model, where E_1 is equal to the expected value of the load, and the price-bid of the single block is equal to infinity. In other words, we always buy the expected load in the day-ahead market. No optimization is needed as the solution is trivial.

LRisk Risk-averse model (C.5) with 20 bid blocks. The price-bid for each block is optimized.

UncRisk Unconstrained risk model (C.2). The solution can be obtained by using Remark 2 or by solving (C.5) with $\beta = 0$.

In order to reproduce the real-time functioning of the markets, we validate the models using a rolling horizon procedure. Everyday at 12:00, we generate scenarios for the next operational day, and afterwards obtain the optimal bidding curve for all the benchmark models. The data from the last two months is used in the scenario-generation procedure, and the process is repeated daily all over the months of November and December.

In Table C.1 we show the mean (1st column) and the standard deviation (2nd) of the profit for the three benchmark models, during November and December. We observe that the simple model *ExpBid* under-performs the rest of the models and, indeed, delivers a negative expected profit. The risk-optimized problem *LRisk* yields positive expected profit, with a variance greater than the *ExpBid* model but substantially lower than for the *UncRisk* problem. The risk-unconstrained model *UncRisk*, as anticipated, provides the highest mean returns.

C.7   Conclusion

In this paper we consider the bidding problem of a retailer that buys energy in the day-ahead market for a pool of price-responsive consumers. Under the considered setup, the deviations from the purchased day-ahead energy are traded at the real-time market. We provide an analytic solution in the case that the retailer is not risk averse. Additionally, we formulate a stochastic programming model for optimal bidding under risk aversion. The price-responsiveness of the consumers is derived from a real-life dataset, and the modeling approach is non-parametric where non-linear relationships are allowed.

The analytic results show that, in the risk-unconstrained case, the optimal bid is a single price, meaning that there is no extra benefit from bidding a curve. On the other hand, the computational results from the risk-averse case show that a block-wise bidding curve successfully mitigate the risk in terms of profit volatility. Altogether, the proposed methodology allows the retailer to optimally bid in the day-ahead market, whether it is for expected-profit maximization (by leveraging arbitrage opportunities), or for the purpose of safely procuring energy.

Appendix C.I: Proof of Theorem 1

We start by computing the expected profit, conditional on the day-ahead price (i.e., we treat $\Lambda^D = \lambda^D$ as a parameter). We disregard the first term in (C.2) since it is constant with respect to the decision variables u_b and X^D, and therefore, does not affect the solution. The expected profit (C.2) conditioned on the day-ahead price λ^D is thus given by

$$\mathbb{E}\left\{ X^D \left(\Lambda^R - \lambda^D \right) | \Lambda^D = \lambda^D \right\} = \tag{C.7}$$

$$\sum_{b=1}^{B} \mathbb{I}\left(u_{t,b} \geq \lambda_t^D \right) E_b \left(\mathbb{E}\left\{ \Lambda^R \right\} - \lambda^D \right). \tag{C.8}$$

Note that, since λ_t^D is given, X^D can be computed as $\sum_{b=1}^{B} \mathbb{I}\left(u_{t,b} \geq \lambda_t^D \right) E_b$. We distinguish three cases:

(a) When $\mathbb{E}\left\{ \Lambda^R \right\} > \lambda^D$, the second term in (C.8) is positive, hence the expected profit is maximized when $u_b \geq u_{b+1} \geq \lambda^D$, $\forall b$. This implies that $u_B \geq \lambda^D$.

(b) When $\mathbb{E}\left\{\Lambda^R\right\} < \lambda^D$, the second term in (C.8) is negative, hence the profit is maximized when $u_{b+1} \leq u_b < \lambda^D$, $\forall b$. This implies that $u_1 \leq \lambda^D$.

(c) When $\mathbb{E}\left\{\Lambda^R\right\} = \lambda^D$, any solution that satisfies $u_{b+1} \leq u_b$ is optimal.

Finally, we conclude that the expected value of the real-time price is an optimal price bid, since $u_b^* = \mathbb{E}\left\{\Lambda^R\right\}$ maximizes the retailer's expected income in the three cases above.∎

Appendix C.II: Proof of Theorem 2

Analogously as in Appendix C.I, from Equation (C.8), the expected profit conditioned on $\Lambda_t^D = \lambda^D$ is proportional to

$$\sum_{b=1}^{B} \mathbb{I}\left(u_{t,b} \geq \lambda_t^D\right) E_b\Big(\mathbb{E}\left\{\Lambda^R|\Lambda^D = \lambda^D\right\} - \lambda^D\Big). \tag{C.9}$$

Next, recall that, from the basic properties of the expected value, $\mathbb{E}_X\left\{g(X)\right\} = \int_{-\infty}^{\infty} g(x)f_X(x)dx$. We compute the expected value of (C.9) with respect to Λ^D, which is equal to:

$$\int_{\Lambda^D} g(\lambda^D)f_{\Lambda^D}(\Lambda^D = \lambda^D)d\lambda^D \tag{C.10}$$

with $g(\lambda^D)$ equal to (C.9). Arranging terms, we obtain that (C.10) is equal to

$$\sum_{b=1}^{B} E_b\Bigg(\int_{-\infty}^{u_b}\bigg(\int_{-\infty}^{\infty} \lambda^R f_{\Lambda^R}(\lambda^R|\Lambda^D = \lambda^D)d\lambda^R\bigg) \times$$
$$f_{\Lambda^D}(\lambda^D)d\lambda^D - \int_{-\infty}^{u_b} \lambda^D f_{\Lambda^D}(\lambda^D)d\lambda^D\Bigg). \tag{C.11}$$

Now we relax problem (C.2) by dropping constraint (C.1c). Then, the problem becomes decomposable by block, since (C.11) is a sum of B elements. For notational purposes, let us rename each of the B terms in the summation in (C.11) by $h_b(u_b)$. Note the functions $h_b(u_b)$ are continuous, since the integral of a continuous function is continuous. Then, for each block, the relaxed problem consists in maximizing the continuous function $h_b(u_b)$ subject to $\underline{\lambda} \leq u_b \leq \overline{\lambda}$. By the intermediate value theorem, we know that the maximum of each term in

the summation will be achieved either at $u_b^* = \underline{\lambda}$, at $u_b^* = \overline{\lambda}$, or otherwise inside the interval $(\underline{\lambda}, \overline{\lambda})$.

Considering the case when u_b^* is inside the interval, we proceed to find the u_b such that it maximizes $h_b(u_b)$. In order to achieve this, we calculate $\frac{d}{du_b} h_b(u_b) = 0$. Note that $\frac{d}{du} \int_{-\infty}^{u} \phi(x)dx = \phi(u)$. With this in mind, the derivative of $h_b(u_b)$ is equal to

$$\int_{-\infty}^{\infty} \lambda^R f_{\Lambda^R}(\lambda^R | \Lambda^D = u_b) d\lambda^R f_{\Lambda^D}(\lambda^D = u_b)$$
$$- u_b f_{\Lambda^D}(\lambda^D = u_b). \tag{C.12}$$

Assuming that $f_{\Lambda^D}(\lambda^D = u_b)$ is different than zero, and solving $\frac{d}{du_b} h_b(u_b) = 0$, we obtain the stationary point:

$$\left\{ u_b^* | u_b^* = \mathbb{E}\left\{ \Lambda^R | \Lambda^D = u_b^* \right\} \right\}. \tag{C.13}$$

Next we calculate the second derivative[4] of $\frac{d^2}{du_b^2} h_b(u_b^*)$. Its sign depends on the value of $(\frac{d}{du_b} \mathbb{E}\left\{ \Lambda^R | \Lambda^D = u_b^* \right\} - 1)$, which can be interpreted as the sensitivity of the expected real-time price to the day-ahead price at the stationary point. Depending on the sign of the second derivative, we distinguish three cases:

(a) When $\frac{d}{du_b} \mathbb{E}\left\{ \Lambda^R | \Lambda^D = u_b^* \right\} < 1$, u_b^* is a local maximum. From a practical point of view, it means that at day-ahead price $\lambda^D = u_b^*$, any marginal increase of this price will imply a comparatively lower marginal increase in the expected real-time price, hence, it becomes not profitable to buy energy from the day-ahead market at price levels greater than u_b^*.

(b) When $\frac{d}{du_b} \mathbb{E}\left\{ \Lambda^R | \Lambda_t^D = u_b^* \right\} > 1$, u_b^* is a local minimum.

(c) When $\frac{d}{du_b} \mathbb{E}\left\{ \Lambda^R | \Lambda^D = u_b^* \right\} = 1$, the solution u_b^* is an inflection point that delivers an expected profit equal to zero.

After having identified the possible candidates u_b that might maximize $h_b(u_b)$, it is easy to see that at least one global optimum to problem (C.2) satisfies that all u_b are all equal to each other, i.e., $u_b = u^*, \forall b$. This is so because functions h_b are all identical for all blocks, hence, the solution u_b^* that yields the highest expected profit for one block b will also deliver the highest expected profit for the remaining blocks.

---

[4]The calculation of $\frac{d^2}{du_b^2} h_b(u_b^*)$, where u_b^* is given by (C.13), is available upon request.

Finally, we should point out that this global solution to the relaxed problem (C.2)—without constraint (C.1c)— naturally satisfies constraint (C.1c), hence, it must also be a global solution to the original problem (C.2).∎

References C

[C1]   F. Rahimi and A. Ipakchi. "Demand response as a market resource under the smart grid paradigm". In: *Smart Grid, IEEE Transactions on* 1.1 (2010), pp. 82–88.

[C2]   N. O'Connell et al. "Benefits and challenges of electrical demand response: A critical review". In: *Renewable and Sustainable Energy Reviews* 39 (2014), pp. 686–699.

[C3]   *California ISO Open Access Same-time Information System (OASIS).* URL: http://oasis.caiso.com.

[C4]   J. M. Morales et al. *Integrating renewables in electricity markets: Operational problems.* Vol. 205. Springer Science & Business Media, 2013.

[C5]   H. Mohsenian-Rad. "Optimal Demand Bidding for Time-Shiftable Loads". In: *Power Systems, IEEE Transactions on* 30.2 (Mar. 2015), pp. 939–951. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2338735.

[C6]   M. Gonzalez Vaya and G. Andersson. "Optimal Bidding Strategy of a Plug-In Electric Vehicle Aggregator in Day-Ahead Electricity Markets Under Uncertainty". In: *Power Systems, IEEE Transactions on* 30.5 (Sept. 2015), pp. 2375–2385. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2363159.

[C7]   S. Li et al. "Market-Based Coordination of Thermostatically Controlled Loads Part I: A Mechanism Design Formulation". In: *Power Systems, IEEE Transactions on* PP.99 (2015), pp. 1–9. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2015.2432057.

[C8]   J. W. Taylor, L. M. De Menezes, and P. E. McSharry. "A comparison of univariate methods for forecasting electricity demand up to a day ahead". In: *International Journal of Forecasting* 22.1 (2006), pp. 1–16.

[C9]   J. Hosking et al. "Short-term forecasting of the daily load curve for residential electricity usage in the Smart Grid". In: *Applied Stochastic Models in Business and Industry* 29.6 (2013), pp. 604–620.

[C10] K. Zare et al. "Multi-market energy procurement for a large consumer using a risk-aversion procedure". In: *Electric Power Systems Research* 80.1 (2010), pp. 63–70.

[C11]   J. Saez-Gallego et al. "A Data-driven Bidding Model for a Cluster of Price-responsive Consumers of Electricity". In: *IEEE Transactions on Power Systems* (2016).

[C12]   S.-E. Fleten and E. Pettersen. "Constructing bidding curves for a price-taking retailer in the Norwegian electricity market". In: *Power Systems, IEEE Transactions on* 20.2 (2005), pp. 701–708.

[C13]   Y. Liu and X. Guan. "Purchase allocation and demand bidding in electric power markets". In: *Power Systems, IEEE Transactions on* 18.1 (2003), pp. 106–112.

[C14]   A. B. Philpott and E. Pettersen. "Optimizing demand-side bids in day-ahead electricity markets". In: *Power Systems, IEEE Transactions on* 21.2 (2006), pp. 488–498.

[C15]   P. Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind Energy* 12.1 (2009), pp. 51–62.

[C16]   J. M. Morales, R. Minguez, and A. J. Conejo. "A methodology to generate statistically dependent wind speed scenarios". In: *Applied Energy* 87.3 (2010), pp. 843–855.

[C17]   X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang. "Scenario generation of wind power based on statistical uncertainty and variability". In: *Sustainable Energy, IEEE Transactions on* 4.4 (2013), pp. 894–904.

[C18]   M. Kohansal and H. Mohsenian-Rad. "A Closer Look at Demand Bids in California ISO Energy Market". In: *IEEE Transactions on Power Systems* (2016).

[C19]   H. Mohsenian-Rad. "Optimal bidding, scheduling, and deployment of battery systems in california day-ahead energy market". In: *Power Systems, IEEE Transactions on* 31.1 (2016), pp. 442–453.

[C20]   K. P. Burnham and D. R. Anderson. "Multimodel inference understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2 (2004), pp. 261–304.

[C21]   P. Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind energy* 12.1 (2009), pp. 51–62.

[C22]   W. Hardle. *Applied nonparametric regression*. Cambridge Univ Press, 1990.

[C23]   D. Hammerstrom et al. *Pacific Northwest GridWise Testbed Demonstration Projects. Part I. Olympic Peninsula Project*. Tech. rep. Oct. 2007.

[C24]   J. Contreras et al. "ARIMA models to predict next-day electricity prices". In: *IEEE transactions on power systems* 18.3 (2003), pp. 1014–1020.

Paper D

Short-term Forecasting of Price-responsive Loads Using Inverse Optimization

Authors:
Javier Saez-Gallego, Juan M. Morales

Short-term Forecasting of Price-responsive Loads Using Inverse Optimization

Javier Saez-Gallego[1] and Juan M. Morales[1].

Abstract

We consider the problem of forecasting the aggregate demand of a pool of price-responsive consumers of electricity. The price-response of the aggregation is modeled by an optimization problem that is characterized by a set of marginal utility curves and minimum and maximum power consumption limits. The task of estimating these parameters is addressed using a generalized inverse optimization scheme that, in turn, requires solving a nonconvex mathematical program. We introduce a solution method that overcomes the non-convexities by solving instead two linear problems with a penalty term, which is statistically adjusted by using a cross-validation algorithm. The proposed methodology is data-driven and leverages information from regressors, such as time and weather variables, to account for changes in the parameter estimates. The power load of a group of heating, ventilation, and air conditioning systems in buildings is simulated, and the results show that the aggregate demand of the group can be successfully captured by the proposed model, making it suitable for short-term forecasting purposes.

Notation

The notation used throughout the paper is stated below for quick reference. Other symbols are defined as required.

Indexes

t Time period, ranging from 1 to T.
b Marginal utility block, ranging from 1 to B.
r Regressor, ranging from 1 to R.

Decision variables

$x_{b,t}$ Load from energy block b and time t.
$\underline{P}_t$ Lower bound for electricity consumption at time t.
$\overline{P}_t$ Upper bound for electricity consumption at time t.
$u_{b,t}$ Marginal utility of load block b at time t.
$\underline{\mu}$ Intercept for the lower load-consumption bound.
$\overline{\mu}$ Intercept for the upper load-consumption bound.
μ^u Intercept for marginal utility.
$\underline{\alpha}_r$ Coefficient relative to the affine dependence of the lower load-consumption bound on regressor r.
$\overline{\alpha}_r$ Coefficient relative to the affine dependence of the upper load-consumption bound on regressor r.
α_r^u Coefficient relative to the affine dependence of marginal utility on regressor r.
ϵ_t Duality gap at time t.
$\underline{\lambda}_t$ Dual variable associated with the lower bound for total load at time t.
$\overline{\lambda}_t$ Dual variable associated with the upper bound for total load at time t.
$\underline{\phi}_{b,t}$ Dual variable associated with the positive block-size constraint for block b at time t.
$\overline{\phi}_{b,t}$ Dual variable associated with the maximum block-size constraint for block b at time t.
$\overline{\xi}_t^+$ Feasibility slack variable linked to the upper load-consumption bound at time t.
$\underline{\xi}_t^+$ Feasibility slack variable linked to the lower load-consumption bound at time t.
$\overline{\xi}_t^-$ Infeasibility slack variable linked to the upper load-consumption bound at time t.
$\underline{\xi}_t^-$ Infeasibility slack variable linked to the lower load-consumption bound at time t.

Parameters

x_t' Measured load at time t.
$\tilde{x}_{b,t}'$ Adjusted measured load for block b at time t.

p_t Price of electricity at time t.
$Z_{r,t}$ Value of regressor r at time t.
$E_{b,t}$ Width of load block b at time t.
K   Feasibility penalty parameter.

D.1   Introduction

Demand response programs aim to alter the power consumption profile of end-users by external stimulus [D1], with the final goal of avoiding over-investing in transmission lines and generating capacities. A popular scheme amongst the numerous programs for demand-side management is *Real-time Pricing* (RTP), where the external stimulus consists of varying prices along the day reflecting the change of balance between supply and demand [D2, D3]. Consumers of electricity, equipped with a smart grid meter and an Energy Management Controller (EMC), seek the most favorable pattern of consumption according to the dynamic price. In the case of households, the EMC comprises a home automation equipment that considers both the price of electricity and the personal preferences of the users to optimally schedule their electricity demand needs and their appliances [D4, D5]. All in all, under the RTP paradigm, the consumers are *price-responsive*.

Forecasting the expected electricity demand at aggregate levels, i.e., *load forecasting*, is of utmost importance for network operators to enhance planning, for example, by mitigating grid congestion during peak-demand periods. Also, it is widely used by electric utilities to minimize the costs of over- or under-contracting power in electricity markets. The increasing penetration of smart grid technologies call for solutions able to forecast the aggregate price-responsive load as accurately as possible.

In response to these challenges, the contributions of this paper are threefold:

1. A methodology to forecast the aggregate consumption of a cluster of price-responsive power loads using inverse optimization.

2. A computationally efficient method that approximates the solution to a generalized inverse optimization problem by solving instead two linear programming problems. The proposed approach relies on cross-validation techniques to optimally tune a penalty parameter so that the out-of-sample forecasting error is minimized.

3. A comprehensive analysis of the performance of the proposed forecasting methodology. The analysis is based on a case study that considers the sim-

ulated price-response of a group of buildings equipped with heat pumps. Furthermore, we benchmark our methodology against persistence forecasting and a state-of-the-art autoregressive moving average model with exogenous inputs [D6].

The presented methodology relates to the existing literature in several aspects. First of all, its final goal is to predict a demand for electricity, hence, it fits into the realm of load forecasting. Amongst the vast load forecasting literature [D7], there are some authors that in the last years have focused on modeling the effect of the price on the load, for example, using a B-spline approach [D8], an Auto-Regressive Model With eXogeneous Inputs (ARX) [D6], neural networks [D9], or a hybrid approach with data association mining algorithms [D10]. The novelty of the proposed methodology in this paper with respect to the existing literature, lies in the characterization of the response of the load to price by an optimization problem. Indeed, to the best of our knowledge, we are the first ones to exploit inverse optimization for time series forecasting and, in particular, for load prediction.

The first formal description of the inverse linear programming problem is given by [D11], which seeks to find the minimal perturbation of the objective function cost vector that makes a given data point optimal. More recent works address the case where the observations are noisy and an exact solution of the inverse problem might not exist [D12, D13, D14, D15, D16, D17]. The proposed methodology neither makes any assumption on whether the data measurements are noisy or not, nor on the existence of a solution to the exact inverse optimization problem.

Here, as in [D12, D18], we extend the concept of inverse optimization to the case where right-hand side parameters of the forward linear programming problem are also to be estimated. Authors of [D18] assume that a feasible region for the forward problem exists, whereas we do not make any assumption in this regard. Indeed, we calculate the best feasible region, in terms of forecasting capabilities, even though it makes the observed data infeasible. The novelty of this paper with respect to [D12] is twofold. First, we propose an inverse optimization scheme that is especially tailored to one-step ahead forecasting, and not to market bidding. Second, the estimation problem we formulate is not based on relaxing the KKT conditions of the forward problem. Instead, we statistically determine the feasible region and the objective function of the forward problem that render the best out-of-sample prediction performance.

The rest of this paper is structured as follows. In Section D.2 we provide a general overview of the proposed forecasting methodology and the associated estimation problem. Then, in Section D.3, the specific load forecasting model is

provided. Section D.4 introduces the framework we have used to simulate the price-response of a group of buildings equipped with heat pumps. In Section D.5, we discuss results from a case study, and finally, in Section C.7, conclusions are duly drawn.

D.2   Inverse Optimization Methodology

Next we introduce the problem of forecasting using inverse optimization and describe the methodology that is applied later, in Section D.3, to predict price-responsive electricity load.

We start from the premise that the choices made by a certain decision-maker (e.g., an aggregation of price-responsive power loads) at a certain time t, denoted by $\boldsymbol{x}_t$, are driven by the solution to the following linear optimization problem:

$$\text{RP}_t(\rho_t|\boldsymbol{c}, \boldsymbol{b}): \qquad \underset{\boldsymbol{x}_t}{\text{Maximize}} \ (\boldsymbol{c} - \rho_t \boldsymbol{e})^T \boldsymbol{x}_t$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x}_t \leq \boldsymbol{b} \qquad\qquad (\text{D.1})$$
$$\boldsymbol{x}_t \leq \boldsymbol{u},$$

where ρ_t is a given time-varying input (e.g., the electricity price) and $\boldsymbol{e}$ is an all-ones vector of an appropriate size. In the technical literature, problem (D.1) is typically referred to as the *reconstruction problem* or the *forward problem* [D13, D15].

Now assume that the matrix of coefficients $\boldsymbol{A}$ and the right-hand side vector $\boldsymbol{u}$ are known and that we are able to observe the multivariate time series $\boldsymbol{X}' = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_T]$, which is presumed to be the solution to the reconstruction problem (D.1) at every time t. That is, $\boldsymbol{x}'_t$ represents the choices actually made by the decision-maker at time t. The basic goal of our inverse optimization approach is to infer the unknown parameter vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ from $\boldsymbol{X}'$ given $\boldsymbol{A}$, $\boldsymbol{u}$, and the series of measured inputs ρ_t. To this end, one tries to find values for the unknowns $\boldsymbol{c}$ and $\boldsymbol{b}$ such that the observed choices $\boldsymbol{X}'$ are as optimal as possible for every problem (D.1). With this aim in mind, we solve the following

generalized inverse optimization problem:

$$\text{GIOP:} \underset{\boldsymbol{b}, \boldsymbol{\phi}_t, \boldsymbol{\lambda}_t, \boldsymbol{c}}{\text{Minimize}} \sum_{t=1}^{T} \epsilon_t$$

$$\begin{aligned}
\text{subject to } & \boldsymbol{b}^T \boldsymbol{\lambda}_t + \boldsymbol{u}^T \boldsymbol{\phi}_t - \epsilon_t = (\boldsymbol{c} - \rho_t \boldsymbol{e})^T \boldsymbol{x}'_t && \forall t \\
& [\boldsymbol{A}^T \ \boldsymbol{I}][\boldsymbol{\lambda}_t^T \ \boldsymbol{\phi}_t^T]^T = (\boldsymbol{c} - \rho_t \boldsymbol{e}) && \forall t \\
& \boldsymbol{A} \boldsymbol{x}'_t \leq \boldsymbol{b} && \forall t \\
& \boldsymbol{\phi}_t, \boldsymbol{\lambda}_t, \epsilon_t \geq 0 && \forall t
\end{aligned} \tag{D.2}$$

where $\boldsymbol{I}$ is the identity matrix of an appropriate size. The objective of optimization problem (D.2) is to minimize the sum over time of the duality gaps associated with the primal-dual reformulation of problem (D.1). Thus, when the objective function of GIOP is equal to zero, namely, the accumulated duality gap is zero, $\boldsymbol{x}'_t$ is optimal in $\text{RP}_t(\rho_t | \boldsymbol{c}, \boldsymbol{b})$, $\forall t$. The first and second constraints in (D.2) are the relaxed strong duality condition and the dual problem constraints of (D.1), respectively. The third inequality represents the primal feasibility constraint involving the unknown right-hand side vector $\boldsymbol{b}$. The second primal constraint in (D.1) is, in contrast, omitted in (D.2), because it does not involve any decision variable in GIOP.

Several challenges arise when solving problem (D.2). The most noticeable one is its nonlinear, nonconvex nature, which is the result of the product of variables $\boldsymbol{b}^T \boldsymbol{\lambda}_t$ appearing in the strong duality condition. This nonlinearity makes GIOP computationally expensive and hard to solve in general. From this point of view, a method capable of obtaining a good solution to (D.2) in a reasonable amount of time is needed, even if such a solution may be suboptimal.

Once the parameter vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ have been estimated by solving (D.2), we can use the reconstruction problem (D.1) to forecast future choices of the decision maker. The estimation of $\boldsymbol{c}$ and $\boldsymbol{b}$ through GIOP is anchored in the following two assumptions about the observed choices $\boldsymbol{x}'_t$:

1. Feasibility: $\boldsymbol{x}'_t$ is feasible in $\text{RP}_t(\rho_t | \boldsymbol{c}, \boldsymbol{b})$.

2. Optimality: Given the true $\boldsymbol{c}$ and $\boldsymbol{b}$, $\boldsymbol{x}'_t$ is optimal in $\text{RP}_t(\rho_t)$ and hence, ϵ_t can be decreased to zero.

These two assumptions, however, do not usually hold in practice for a number of reasons [D16]. First, the forward problem (D.1) might be *misspecified* in the sense that it might not represent the actual optimization problem solved by the decision-maker. Therefore, there might not exist $\boldsymbol{c}$ and $\boldsymbol{b}$ such that the observed choices $\boldsymbol{x}'_t$ are both feasible and optimal for (D.1). Second, the

decision-maker might suffer from *bounded rationality* or *implementation errors*. That is, even if the forward problem (D.1) does prompt the optimal choices to be made by the decision-maker, she might be content with suboptimal choices (due to cognitive or computational limitations, for instance) or there might not be a way to implement such optimal choices without some level of error. Finally, the observed choices $\boldsymbol{X}'$ might be corrupted by *measurement noise*.

In this work, though, our intention is to use inverse optimization to *forecast* the future choices of the decision-maker by using the reconstruction problem (D.1). This has two important practical implications at least. First, we are not that concerned with the fact that the forward problem (D.1) might be misspecified (this will be indeed the case in the application problem we present later). What we demand from this problem, instead, is that it features good *predictive power* on the futures choices of the decision-maker. In other words, our aim is not to determine values for $\boldsymbol{c}$ and $\boldsymbol{b}$ that make the observed choices $\boldsymbol{X}'$ both feasible and optimal for (D.1), but to find the values of these parameters that minimize the *out-of-sample prediction error*.

Given all these practical considerations, in order to compute appropriate values for $\boldsymbol{c}$ and $\boldsymbol{b}$, we develop a two-step estimation procedure that deals with the assumptions of feasibility and optimality of $\boldsymbol{X}'$ in a *statistical* sense, i.e., with a view to minimizing the out-of-sample prediction error. Furthermore, the proposed two-step estimation procedure overcomes the nonconvexity and computational issues mentioned above regarding the solution to problem (D.2).

The first step of the estimation procedure consists in finding a "good" feasible region. Note that if $\boldsymbol{b} \to \infty$, then the second constraint in (D.1) is always satisfied. For this reason, we do not want just to find a region for which $\boldsymbol{X}'$ is feasible, but the most adequate one in terms of prediction performance. For this purpose, we solve optimization problem (D.3), which we refer to as the *feasibility problem* FP(K) and which minimizes a trade-off between the "infeasibility slack variables" $\boldsymbol{\xi}^-$, and the "feasibility slack variables" $\boldsymbol{\xi}^+$, being $0 \leq \mathrm{K} < 1$ a parameter that controls the trade-off between these two quantities.

$$\text{FP}(K): \quad \underset{\boldsymbol{b}, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-}{\text{Minimize}} \sum_{t=1}^{T} \left(\mathrm{K} \|\boldsymbol{\xi}_t^+\| + (1-\mathrm{K}) \|\boldsymbol{\xi}_t^-\| \right)$$

$$\text{subject to } \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_t' = \boldsymbol{\xi}_t^+ - \boldsymbol{\xi}_t^- \qquad \forall t$$

$$\boldsymbol{D}^1 \boldsymbol{b} \leq \boldsymbol{d}^1 \qquad\qquad\qquad (\text{D.3})$$

$$\boldsymbol{\xi}_t^+, \boldsymbol{\xi}_t^- \geq 0 \qquad\qquad \forall t.$$

The value of parameter K is computed by means of cross-validation, as explained below in Section D.2.1. Parameters $\boldsymbol{D}^1$ and $\boldsymbol{d}^1$ define constraints on $\boldsymbol{b}$, known a

priori, which might be imposed by the nature of $\boldsymbol{b}$. An example of such a-priori constraint may simply be $\boldsymbol{b} \geq 0$.

In the second step, we consider $\boldsymbol{b}$ given as the solution to FP(K), denoted by $\widehat{\boldsymbol{b}}$. Also, we adjust the observed quantity $\boldsymbol{A}\tilde{\boldsymbol{x}}'_t = \boldsymbol{A}\boldsymbol{x}'_t - \boldsymbol{\xi}_t^{-*}$, where $\boldsymbol{\xi}_t^{-*}$ is taken from the solution of FP(K). This modification makes $\tilde{\boldsymbol{x}}'_t$ feasible in $\mathrm{RP}_t(\rho_t|\boldsymbol{c}, \widehat{\boldsymbol{b}})$. As we show later in Section D.3.2, we may need to impose further constraints on the adjusted quantity $\tilde{\boldsymbol{x}}'_t$ in those cases where it is not univocally determined by $\boldsymbol{A}\tilde{\boldsymbol{x}}'_t = \boldsymbol{A}\boldsymbol{x}'_t - \boldsymbol{\xi}_t^{-*}$. Then, we solve the following linear programming problem:

$$
\mathrm{OP}(\widehat{\boldsymbol{b}}): \underset{\boldsymbol{\phi}_t, \boldsymbol{\lambda}_t, \boldsymbol{c}}{\mathrm{Minimize}} \sum_{t=1}^{T} \epsilon_t
$$

$$
\begin{aligned}
\text{subject to } & \widehat{\boldsymbol{b}}^T \boldsymbol{\lambda}_t + \boldsymbol{u}^T \boldsymbol{\phi}_t - \epsilon_t = (\boldsymbol{c} - \rho_t \boldsymbol{e})^T \tilde{\boldsymbol{x}}'_t \ \ \forall t \\
& [\boldsymbol{A}^T \ \boldsymbol{I}][\boldsymbol{\lambda}_t^T \ \boldsymbol{\phi}_t^T]^T = (\boldsymbol{c} - \rho_t \boldsymbol{e}) \qquad \forall t \\
& \boldsymbol{D}^2 \boldsymbol{c} \leq \boldsymbol{d}^2 \\
& \boldsymbol{\phi}_t, \boldsymbol{\lambda}_t, \epsilon_t \geq 0 \qquad\qquad\qquad \forall t.
\end{aligned}
\tag{D.4}
$$

The first, the second, and the last constraints in (D.4) are analogue to the ones in (D.2). The third constraint defines a-priori conditions on $\boldsymbol{c}$, specified by the parameters $\boldsymbol{D}^2$ and $\boldsymbol{d}^2$. The outcome of this problem is the estimated value of the coefficient vector $\boldsymbol{c}$, named as $\widehat{\boldsymbol{c}}$.

Finally, given $\widehat{\boldsymbol{b}}$, $\widehat{\boldsymbol{c}}$ and ρ_{T+1}, we can forecast the future decision-maker's choices by solving $\mathrm{RP}_{T+1}(\rho_{T+1}|\widehat{\boldsymbol{c}}, \widehat{\boldsymbol{b}})$.

D.2.1 Statistical Determination of K

In practice, we find the value of the penalty parameter K using cross-validation [D19, Ch. 7]. We partition $\boldsymbol{X}'$ in three subsets: the training set $\boldsymbol{X}'_{tr}$, the validation set $\boldsymbol{X}'_{val}$, and the test set. In a few words, for each given value of K, we use the training set for parameter fitting and the validation set to asses the forecasting performance. The best choice of K is, thus, the one that minimizes the out-of-sample prediction error.

The advantage of using this approach is threefold. First, by tuning the value of K for the validation set, we seek to minimize the out-of-sample prediction error (a criterion specially suited for forecasting purposes). Second, we solve three LP problems for each tested value of K, hence, the computational burden of the tuning algorithm is relatively low. Finally, the evaluations of different values of K are independent of each other so they can be executed in parallel.

D.2.2   Leveraging Auxiliary Information

When using inverse optimization for *forecasting* a time series, it is relevant to consider the case where we let the unknown parameter vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ vary over time so as to capture structural changes in the decision-making problem (D.1). To this end, we assume that we also observe a number of time-varying *regressors* $\boldsymbol{Z}_t$ that, to a lesser or greater extent, may affect the decision maker's choices. We then describe the unknown vectors $\boldsymbol{c}$ and $\boldsymbol{b}$ as functions of those regressors by letting $\boldsymbol{c}_t = f_c(\boldsymbol{Z}_t)$ and $\boldsymbol{b}_t = f_b(\boldsymbol{Z}_t)$ in problems (D.3) and (D.4), respectively. In this way, functions $f_c(\cdot)$ and $f_b(\cdot)$ become decision variables in our estimation problem. In this paper, we consider $f_c(\cdot)$ and $f_b(\cdot)$ to be affine functions. Hence, the inverse optimization problem seeks the most optimal set of intercepts and affine coefficients that relate $\boldsymbol{Z}_t$ with $\boldsymbol{c}_t$ and $\boldsymbol{b}_t$, as we exemplify below. Note that the past choices of the decision-maker, namely, $\boldsymbol{X}'$, can also be treated as regressors.

D.3   Methodology Applied to Forecast Price-responsive Loads

In this section we illustrate the use of the proposed inverse optimization approach to forecast the aggregate power load of a pool of price-responsive consumers.

We consider that the available information is the measured power consumption x'_t of the pool; the electricity price p_t, which is broadcast to every load in the pool; and the realizations of a set of explanatory variables $Z_{r,t}$ for every time period t. The aggregate response $\boldsymbol{x}_t$ of the loads to the price of electricity at time t is assumed to be the solution to the following forward/reconstruction problem:

$$\underset{\boldsymbol{x}_t}{\text{Maximize}} \sum_{b=1}^{B} x_{b,t}\left(u_{b,t} - p_t\right) \tag{D.5a}$$

$$\text{subject to } \underline{P}_t \le \sum_{b=1}^{B} x_{b,t} \le \overline{P}_t \qquad (\underline{\lambda}_t, \overline{\lambda}_t) \tag{D.5b}$$

$$0 \le x_{b,t} \le E_{b,t} \qquad (\underline{\phi}_{b,t}, \overline{\phi}_{b,t}) \ \ \forall b. \tag{D.5c}$$

Problem (D.5) takes the form of (D.1). Its objective function (D.5a), to be maximized, represents the aggregate consumers' surplus or welfare, given as the

product of the pool consumption and the difference between the marginal utility and the electricity price. We consider a step-wise marginal utility curve made up of B blocks, each of a width $E_{b,t}$, as enforced by (D.5c), and a value $u_{b,t}$. The aggregate load of the pool, given as $\sum_{b=1}^{B} x_{b,t}$, is bounded from below and above by $\underline{P}_t$ and $\overline{P}_t$, respectively, as expressed by (D.5b). Symbols within parentheses correspond to the dual variables associated with each constraint.

The goal of our inverse optimization methodology is to estimate appropriate values for $u_{b,t}$, $\underline{P}_t$ and $\overline{P}_t$, based on the observed x'_t, p_t, and $Z_{r,t}$, such that the solution to the reconstruction problem (D.5) serves as a good forecast of the future aggregate power consumption x_{t+1} of the pool of loads. For this purpose, we employ the estimation procedure outlined in Section D.2. Note that the width of each block, $E_{b,t}$, is treated as a parameter and need not to be estimated. Later, in Section D.3.2, we give a practical rule for fixing it.

Next we provide concrete formulations for the estimation of $\underline{P}_t$, $\overline{P}_t$ and $u_{b,t}$. The problem of estimating the bounds $\underline{P}_t$ and $\overline{P}_t$, which we refer to as the *bound estimation problem*, is presented in Section D.3.1. The problem of estimating the marginal utilities $u_{b,t}$, which we call the *marginal utility estimation problem*, is presented in Section D.3.2. A discussion about the proposed methodology is finally given in D.3.3.

D.3.1   Bound Estimation Problem

The bound estimation problem is derived from the feasibility problem (D.3) and consists in determining the bounds $\underline{P}_t$ and $\overline{P}_t$ by minimizing the following objective function:

$$
\underset{\underline{P},\overline{P},\xi,\mu,\alpha}{\text{Minimize}} \sum_{t=1}^{T} \left((1-\text{K}) \left(\overline{\xi}_t^+ + \underline{\xi}_t^+ \right) + \right.
$$

$$
\left. \text{K} \left(\overline{\xi}_t^- + \underline{\xi}_t^- \right) \right) \tag{D.6a}
$$

subject to

$$\overline{P}_t - x'_t = \overline{\xi}_t^+ - \overline{\xi}_t^- \qquad \forall t \qquad (D.6b)$$

$$x'_t - \underline{P}_t = \underline{\xi}_t^+ - \underline{\xi}_t^- \qquad \forall t \qquad (D.6c)$$

$$\underline{P}_t \leq \overline{P}_t \qquad \forall t \qquad (D.6d)$$

$$\underline{P}_t = \underline{\mu} + \sum_{r=1}^{R} \underline{\alpha}_i Z_{r,t} \qquad \forall t \qquad (D.6e)$$

$$\overline{P}_t = \overline{\mu} + \sum_{r=1}^{R} \overline{\alpha}_i Z_{r,t} \qquad \forall t \qquad (D.6f)$$

$$0 \leq \overline{\xi}_t^+, \overline{\xi}_t^-, \underline{\xi}_t^+, \underline{\xi}_t^- \qquad \forall t \qquad (D.6g)$$

where $\boldsymbol{\xi} = \left[\overline{\boldsymbol{\xi}}^+; \underline{\boldsymbol{\xi}}^+; \overline{\boldsymbol{\xi}}^-; \underline{\boldsymbol{\xi}}^-\right]$, $\boldsymbol{\mu} = [\overline{\boldsymbol{\mu}}; \underline{\boldsymbol{\mu}}]$, and $\boldsymbol{\alpha} = [\overline{\boldsymbol{\alpha}}; \underline{\boldsymbol{\alpha}}]$.

The objective function (D.6a) comprises two terms, weighted by the parameter K with $0 \leq \mathrm{K} < 1$. The first and second terms represent the amount of measured load that falls inside and outside the interval $[\underline{P}_t, \overline{P}_t]$, respectively.

Constraint (D.6d) ensures that the estimated lower bound is always lower than the upper bound. Constraints (D.6e) and (D.6f) impose an affine relationship between the regressors and the load bounds. We denote the estimates of the lower and upper bounds at the optimum as $\widehat{\underline{P}}_t$ and $\widehat{\overline{P}}_t$, respectively. It is worth mentioning the special case where previous load observations $x'_{t-1} \ldots x'_{t-l}$ are included as regressors, in a similar way as traditional auto-regressive models do [D20]. Also, it should be noted that we do not treat the price at time t as a regressor here, since its effect is captured through the objective function of the forward problem by solving the optimality problem (D.8).

The parameter K is computed using the cross-validation approach outlined in Section D.2.1. Values of K close to 1 yield a "wide" interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$, whereas values of K close to zero produce a narrow interval. Therefore, K can be interpreted as an indicator of the *price responsiveness* of the load, since the precise value that the load will take on within the interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$ is left to be explained by the electricity price. Notice that when K = 0 the bound estimation problem boils down to fitting an ARX model by minimizing the MAE, because in this case it holds that $\widehat{\underline{P}}_t = \widehat{\overline{P}}_t$.

Now consider the solution to (D.6). In order for the reconstruction problem (D.5) to be feasible, the estimated bounds must satisfy that $\widehat{\underline{P}}_t \leq \widehat{\overline{P}}_t$. This is enforced for the training data set by Equation (D.6d), but it is not necessarily satisfied for any *plausible* data point outside this set. Generally, when

forecasting, consistent bounds $\underline{P}_t \leq \overline{P}_t$ must be obtained for all possible future realizations of the regressors $Z_{r,t}$. We can ensure this by robustification [D21], as done in [D12].

D.3.2 Marginal Utility Estimation Problem

The marginal utility estimation problem is derived from the optimality problem (D.4) once the bounds $\widehat{\underline{P}}_t$ and $\widehat{\overline{P}}_t$ have been estimated by solving problem (D.6).

Prior to solving the marginal utility estimation problem, the measured load is adjusted so that it becomes feasible in the reconstruction problem (D.5). For this purpose, we define the adjusted load as $\tilde{x}'_t = x'_t - \overline{\xi}_t^{-*} + \underline{\xi}_t^{-*}$, where $\overline{\xi}_t^{-*}$ and $\underline{\xi}_t^{-*}$ are taken from the solution to problem (D.6). Note that this is equivalent to defining $\tilde{x}'_t$ as

$$\tilde{x}'_t = \sum_{b=1}^{B} \tilde{x}'_{b,t} = \begin{cases} \widehat{\underline{P}}_t & \text{if } x'_t < \widehat{\underline{P}}_t \\ x'_t & \text{if } \widehat{\underline{P}}_t \leq x'_t \leq \widehat{\overline{P}}_t \\ \widehat{\overline{P}}_t & \text{if } x'_t > \widehat{\overline{P}}_t \end{cases} \tag{D.7}$$

with $\tilde{x}'_{b,t} \leq E_{b,t}$. We further impose that the load blocks are to be filled in sequential order starting with $b = 1$.

To fix the width of each load block, we proceed as follows. We set the width of the first block to be equal to the lower bound, namely, $E_{1,t} = \widehat{\underline{P}}_t$. The width of the remaining blocks is computed such that the interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$ is equally divided, that is, $E_{b,t} = (\widehat{\overline{P}}_t - \widehat{\underline{P}}_t)/(B-1), \ \forall b > 1, \forall t$. In order for the lower bound to be effective, we set the marginal utility for the first block (denoted as $u_{1,t}$) to a large number. This is done in Equation (D.8f). Consequently, at the optimum, the first block of energy is always filled with $x_{1,t} = E_{1,t} = \widehat{\underline{P}}_t$, since its corresponding marginal utility is always higher than the electricity price. This practical rule allows us to enforce the lower bound through the use of $E_{1,t}$ and $u_{1,t}$.

The optimization variables in the marginal utility estimation problem (D.8) are $\boldsymbol{\Omega} = \{\boldsymbol{\epsilon}, \boldsymbol{u}, \boldsymbol{\mu}^u, \boldsymbol{\alpha}^u, \underline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\phi}}, \overline{\boldsymbol{\phi}}\}$. This problem aims to minimize the sum of duality gaps ϵ_t of the reconstruction problem (D.5), that is, to find the marginal utilities $u_{b,t}$ such that the adjusted observed load $\tilde{x}'$ is as optimal as possible.

$$\underset{\mathbf{\Omega}}{\text{Minimize}} \sum_{t=1}^{T} \epsilon_t \tag{D.8a}$$

$$\text{subject to } \widehat{\overline{P}}_t \overline{\lambda}_t - \widehat{\underline{P}}_t \underline{\lambda}_t + \sum_{b=1}^{B} E_b \overline{\phi}_{b,t} - \epsilon_t =$$

$$\sum_{b=1}^{B} \tilde{x}'_{b,t} \left(u_{b,t} - p_t \right) \qquad \forall t \tag{D.8b}$$

$$-\underline{\phi}_{b,t} + \overline{\phi}_{b,t} - \underline{\lambda}_t + \overline{\lambda}_t = u_{b,t} - p_t \quad \forall b, t \tag{D.8c}$$

$$u_{b,t} = \mu_b^u + \sum_r \alpha_r^u Z_{r,t} \qquad \forall b, t \tag{D.8d}$$

$$\mu_b^u \geq \mu_{b+1}^u \qquad \forall b < B \tag{D.8e}$$

$$\mu_1^u \geq 200 + \mu_2^u \tag{D.8f}$$

$$0 \leq \overline{\lambda}_t, \underline{\lambda}_t, \underline{\phi}_{b,t}, \overline{\phi}_{b,t} \qquad \forall b, t. \tag{D.8g}$$

Constraint (D.8b) defines the relaxed strong duality condition, with the objective function of the dual of problem (D.5) minus the duality gap on the left-hand side, and the primal objective function on the right-hand one. Equations (D.8c) are the constraints of the dual of problem (D.5). Constraint (D.8d) defines the marginal utilities as affine combinations of the regressors. Constraint (D.8e) forces the estimated marginal utility to be monotonically decreasing, and constraint (D.8f) imposes a high utility for the first block. Finally, constraint (D.8g) enforces the non-negative character of the dual variables.

D.3.3   Discussion

The proposed inverse optimization framework can be seen as a generalization of a linear time series model: the relationship between the load and the regressors is linear, but the relationship between the load and the price at time t is not.

Recall that the proposed methodology is composed by two problems that are solved sequentially: the feasibility problem (D.6) and the optimality problem (D.8). In the feasibility problem, we model the linear relationship between the load and the regressors, excluding the price at time t. The penalty parameter K is optimally chosen by cross-validation, and it affects the width of the interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$. Afterwards, in the optimality problem we model the non-linear relationship between the load that falls inside the interval $[\widehat{\underline{P}}_t, \widehat{\overline{P}}_t]$, and the price at

time t. For this reason, a narrow interval implies that the variability of the load left to be explained by the price at time t is very small. On the other hand, a wide interval indicates that the load can be explained by the price at time t to a large extent. Its non-linear relationship is estimated by the optimality problem. Unlike in the proposed scheme, in a simple linear regression model, the relationship between the load and the price is given by an affine coefficient.

D.4 Simulation of Price-responsive Buildings

We simulate the price-response behavior of a pool of buildings equipped with heat pumps. To this end, we use the work in [D22]. The heating dynamics of each building is described by a state-space model that consists of three states: indoor air temperature y_t^r, floor temperature y_t^f, and temperature of the water y_t^w inside a tank connected to a heat pump. The only input is the electricity consumption x_t. The state-space model writes as follows, where $\boldsymbol{y}_t = [y_t^r, y_t^f, y_t^w]^T$:

$$\boldsymbol{y}_t = \mathbf{A}\boldsymbol{y}_{t-1} + \mathbf{B}x_{t-1} + \mathbf{E}\boldsymbol{z}_{t-1} \qquad \forall t \tag{D.9}$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{E}$ are the matrices of the coefficients defining the state-space model. The temperature of the air outside the building z_t^a and the solar irradiance z_t^s are considered as external disturbances in $\boldsymbol{z}_t = [z_t^a, z_t^s]^T$.

The heat pump in each building schedules its consumption by solving an Economic Model Predictive Control (EMPC) problem that minimizes the cost of its consumption plus a penalty term for not complying with a comfort band:

$$\underset{\boldsymbol{y},\boldsymbol{x},\boldsymbol{v}}{\text{Minimize}} \quad \sum_{t=1}^{T} p_t x_t + \rho v_t \tag{D.10a}$$

$$\text{subject to} \quad \boldsymbol{y}_t = \mathbf{A}\boldsymbol{y}_{t-1} + \mathbf{B}x_{t-1} + \mathbf{E}\boldsymbol{z}_{t-1} \; \forall t \tag{D.10b}$$

$$0 \leq x_t \leq x^{max} \qquad\qquad \forall t \tag{D.10c}$$

$$y_{t,min}^r \leq y_t^r + v_t \qquad\qquad \forall t \tag{D.10d}$$

$$y_{t,max}^r \geq y_t^r - v_t \qquad\qquad \forall t \tag{D.10e}$$

$$v_t \geq 0 \qquad\qquad \forall t. \tag{D.10f}$$

The objective function (D.10a) minimizes the cost of purchasing x_t kWh of energy at the price p_t, with a penalization of ρv_t if the room temperature is not within the desired comfort band. Equation (D.10b) determines the time evolution of the states of the model. The maximum power consumption of the heat pump, set by Equation (D.10c), is x^{max} kW. Finally, equations (D.10d),
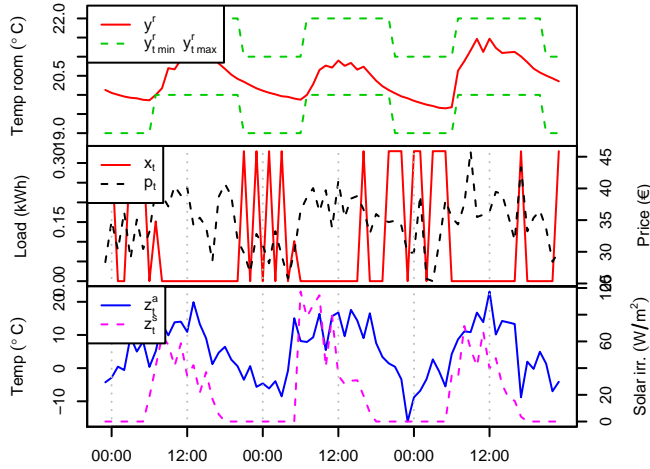
Figure D.1: Evolution of room temperature, comfort bands, price, load, and disturbances for one building during 72 hours

(D.10e) and (D.10f) define the comfort temperature band, given by $y^r_{t,min}$ and $y^r_{t,max}$, and the slack variable v_t.

The values for the coefficients of the EMPC model (D.10) are taken from [D23, Tab. 1]. The effect of the solar irradiance on the room temperature is set to be equal to $0.01\,°C/(W/m^2)$, and of $0.001\,°C/(W/m^2)$ on the floor temperature. An example of the behavior of one building during 72 hours, with hourly observations, is depicted in Fig. D.1.

A total of 100 buildings are simulated by randomly perturbing the heat-transfer coefficients that define matrix **A**. Modifying slightly these coefficients allows us to simulate the behavior of buildings with different structural characteristics. The perturbations are randomly drawn from a uniform distribution centered around zero with a variance equal to 1/50 the magnitude of the corresponding coefficient. The magnitude of the perturbations is chosen high enough so that different building structures are modeled, but not too high so that the state-space system becomes unstable. The magnitude of the perturbations has been chosen by trial-and-error, and its effectiveness is proven to be useful as explained in the remaining of this section and in the case study of Section D.5.

We simulate the behavior of two classes of buildings and aggregate the simulated information in two data sets. In the first one, called *no flex*, the comfort bands for the temperature inside the room are equal to each other. In the second case, called *flex*, the comfort bands for the temperature of the air inside the
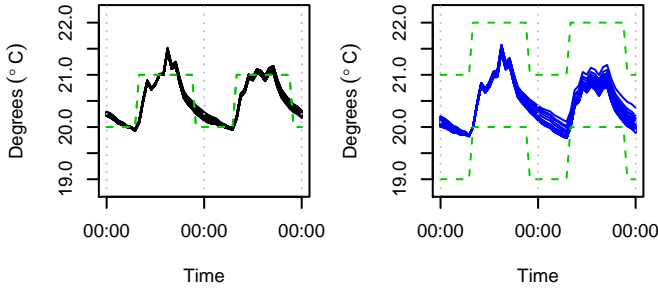
Figure D.2: Room temperature and comfort bands. Left: no-flex case; right: flex case
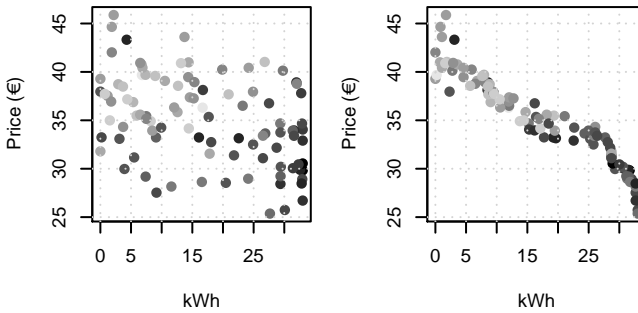
Figure D.3: Price and aggregate load for the non-flexible (left) and flexible (right) cluster of buildings. Dark tonalities indicate low ambient temperature.

room are $2\,°C$ apart from each other. A sample of the simulated comfort bands and temperatures inside the rooms is shown in Fig. D.2. On the left plot, we show the *no flex* case. Naturally, the temperature inside the room is as close as possible to the desired one. On the right plot, for the *flex* case, the temperature inside the room features a higher variation across buildings.

The effect of the electricity price on the aggregated load, for the two data sets, is displayed in Fig. D.3. On the left plot, the *no flex* data set shows barely no relationship between load and price. On the other hand, on the right plot, the *flex* data set shows a clear non-linear relationship. In both plots, black colors indicate that the temperature of the outside air is low. Naturally, the aggregated load is higher at times with low ambient temperature because of the need for heating up the water tank.

To sum up, the simulated data sets seem to be fair representations of the be-

havior of a pool of price-responsive buildings, hence, we proceed to use the simulated data sets for testing the performance of the proposed load forecasting method.

D.5   Case Study

We now asses the performance of the reconstruction problem (D.5) when forecasting the load of the pool of buildings one step ahead, that is, one hour in advance. In problem (D.5), the marginal utilities, the minimum power, and the maximum power are calculated using the methodology introduced in Section D.2 and D.3. We compare the performance of the proposed methodology using the *flex* and *no flex* data sets, which were introduced in Section D.4.

The regressors that we consider for describing the dynamics of the estimated parameters in (D.6e), (D.6f) and (D.8d), are the hour indicator, outside temperature, solar irradiance, and historical lagged price and load data. At every time period t, we assume that the regressors up to that period are known. We also assume that the price at time t is known. The training set consists of 505 data points, that is, three weeks of data. Furthermore, we set the total number of load blocks to $B = 20$.

The first step in the estimation procedure is to tune parameter K in (D.6) following the cross-validation strategy from Section D.2.1. The results, displayed in Fig. D.4, show the Root Mean Square test Error (RMSE) for the two data sets, using a validation period of one week of data. For the *flex* dataset, the optimal value of K turns out to be 0.98. Recall that values of K close to 1 indicate that the interval $[\widehat{\underline{P}}, \widehat{\overline{P}}]$ is wide. Therefore, in the considered application, this also means higher responsiveness of the load to the price. The continuous line in Fig. D.4 represents the RMSE for the *no flex* dataset, for different values of K. The best forecasting performance is achieved for K ≤ 0.5. In the *no flex* case, the load is independent of the price. Consequently, the best forecasts are achieved when the interval $[\widehat{\underline{P}}, \widehat{\overline{P}}]$ is very small, namely, when $\widehat{\underline{P}} = \widehat{\overline{P}}$. It is noteworthy to say that the solution in this case is equivalent to fitting an autoregressive linear model with exogenous inputs by minimizing the mean absolute error.

On average, when using 3 weeks of data, 20 load blocks, and 38 regressors, the time for the whole estimation process takes around 10 seconds on a personal Linux-based machine with 4 cores clocking at 2.90GHz and 12 GB of RAM. R and CPLEX 12.3 under GAMS are used to process the data and solve the optimization models. We conclude that, because of its low computational requirements, the proposed methodology is attractive for implementation in a
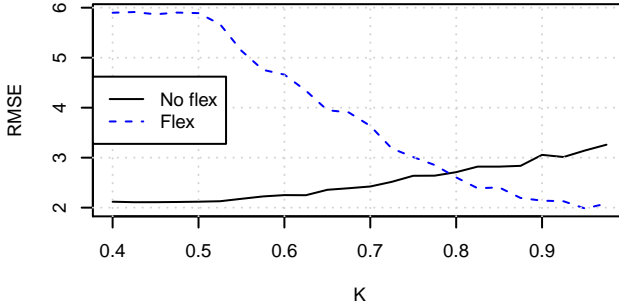
Figure D.4: The RMSE of the 1-steap ahead predictions is shown for different values of K, using the *no flex* dataset (dashed) and the *flex* dataset (continuous)

real-life setup.

D.5.1 Benchmark on a Test Period

We benchmark the forecasting capability of the proposed methodology against two other methods. The first one is a simple persistence model, where the forecast load at time t is set to be equal to the observed load at $t - 1$. The second model is an Autoregressive Moving Average Model with eXogneous inputs (ARMAX) [D20, Ch. 5], similar to the one used in [D6]. The aggregate load x_t is modeled as a linear combination of the past values of load, past errors, and regressors. In mathematical terms, the ARMAX model can be written as

$$x_t = \mu + \epsilon_t + \sum_{p=1}^{P} \varphi_p x_{t-p} + \sum_{r=1}^{R} \gamma_r Z_r + \sum_{q=1}^{Q} \theta_q \varepsilon_{t-q} \tag{D.11}$$

with $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ and σ^2 being the variance. The optimal combination of P and Q is chosen according to the AICc criteria [D24]. In order to make reasonable comparisons, the same explanatory variables are used for our inverse-optimization-based model and for the ARMAX, including the price at time t. Recall that the price at time t is considered in the optimality problem (D.8) but not in the feasibility problem (D.6).

We run 1-step ahead predictions in a rolling-horizon manner for a period of 5 days, re-estimating the parameters at every hour. The upper plot of Fig. D.5 shows the actual aggregate load together with those predicted by the proposed
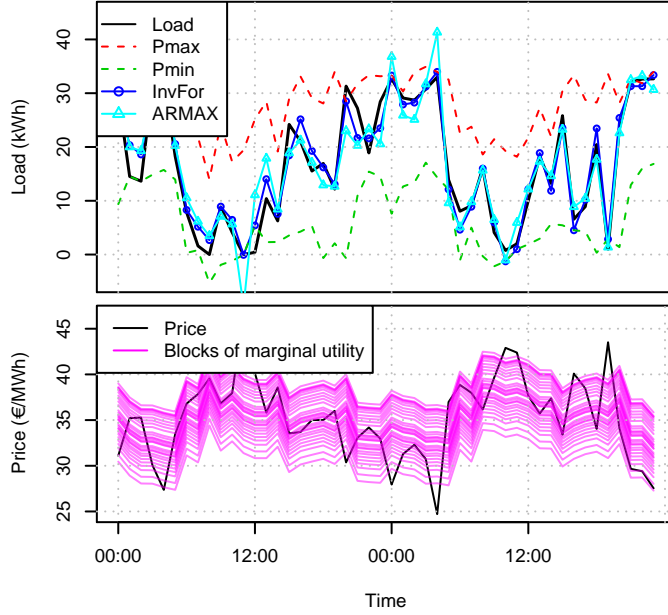
Figure D.5: On the top, the actual load is displayed together with the predictions from the inverse-optimization methodology and the AR-MAX model. On the bottom, the price is shown together with the estimated marginal utility blocks

inverse-optimization model (*InvFor*) and the ARMAX. The estimated minimum and maximum load bounds are able to explain a certain part of the variability of the load. The remaining variability is explained by the relationship between the marginal utilities and the price. The predictions made by the ARMAX are also able to anticipate the behavior of the load, but to a lesser extent. On the bottom plot of Fig. D.5, the electricity price is displayed together with the estimated marginal utility blocks, for each hour of the test period. The magnitude and distribution of the marginal utilities change with time and capture the dynamic response of the load to the price.

Performance metrics computed over the test set are summarized in Table D.1. Each row is relative to one of the three benchmark models. Columns 2 and 4 give information on the Normalized Root Mean Square Error (NRMSE), defined as

$$\text{NRMSE} = \frac{1}{x^{max} - x^{min}} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\sum_{b=1}^{B} \widehat{x}_{b,t} - x'_t \right)} \qquad \text{(D.12)}$$

and columns 3 and 5 on the Symmetric Mean Absolute Percentage Error (SMAPE)

$$\text{SMAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|\sum_{b=1}^{B} \widehat{x}_{b,t} - x'_t|}{(|\sum_{b=1}^{B} \widehat{x}_{b,t}| - |x'_t|)/2}. \tag{D.13}$$

In Table D.1 we also compare the performance of the proposed forecasting method using the two simulated data sets. On the left part, we show the performance measures relative to the *no flex* data set. The ARMAX and the *InvFor* models yield almost identical results in terms of NRMSE and SMAPE. This is indeed reasonable because, as mentioned earlier in Section D.2, the *InvFor* model with a penalty parameter of $K = 0$ is equivalent to fitting an ARX.

The differences between the ARMAX and the *InvFor* stand out when used for predicting the *flex* data set. On the right side of Table D.1, we see that our methodology outperforms the ARMAX with a NRMSE and a SMAPE 32% and 16.8% lower, respectively. The persistence model, as expected, exhibits the worst performance. We conclude that the non-linear relationship between the price and the load is well captured by the *InvFor* model.

Table D.1: Benchmark for the test set

| | No Flex | | Flex | |
	NRMSE	SMAPE	NRMSE	SMAPE
Persistence	0.1727	0.1509	0.3107	-
ARMAX	0.10086	0.08752	0.13107	0.08426
InvFor	0.10093	0.0886	0.08903	0.07003

D.6   Conclusion

This paper proposes a new method to forecast price-responsive electricity consumption. The price response is described by an optimization problem, which is characterized by a set of unknown parameters. The problem of estimating these parameters is nonlinear and nonconvex. We formulate a two-step algorithm to statistically approximate its solution, where in each step we solve a linear problem. The proposed approach is data-driven and makes use of a cross-validation scheme to minimize the out-of-sample prediction error. Moreover, a set of regressors is used to explain the variability of the price-response of the load.

A simulation framework is used to asses the performance of the proposed methodology. The simulation comprises a set of price-responsive buildings equipped

with a heat pump. The presented methodology is used for 1-step ahead predictions. Results show that the non-linear relationship between the price and the aggregate load is successfully captured and that the proposed method outperforms well-known benchmark models.

References D

[D1] N. O'Connell et al. "Benefits and challenges of electrical demand response: A critical review". In: *Renewable and Sustainable Energy Reviews* 39 (2014), pp. 686–699.

[D2] J. S. Vardakas, N. Zorba, and C. V. Verikoukis. "A survey on demand response programs in smart grids: pricing methods and optimization algorithms". In: *IEEE Communications Surveys & Tutorials* 17.1 (2015), pp. 152–178.

[D3] S. Borenstein. "The long-run efficiency of real-time electricity pricing". In: *The Energy Journal* (2005), pp. 93–116.

[D4] A.-H. Mohsenian-Rad et al. "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid". In: *Smart Grid, IEEE Transactions on* 1.3 (2010), pp. 320–331.

[D5] A. Conejo, J. Morales, and L. Baringo. "Real-Time Demand Response Model". In: *IEEE Transactions on Smart Grid* 1.3 (2010), pp. 236–242.

[D6] O. Corradi et al. "Controlling Electricity Consumption by Forecasting its Response to Varying Prices". In: *IEEE Transactions on Power Systems* 28.1 (2013), pp. 421–429.

[D7] R. Weron. *Modeling and forecasting electricity loads and prices: a statistical approach*. Vol. 403. John Wiley & Sons, 2007.

[D8] J. Hosking et al. "Short-term forecasting of the daily load curve for residential electricity usage in the Smart Grid". In: *Applied Stochastic Models in Business and Industry* 29.6 (2013), pp. 604–620.

[D9] Z. Yun et al. "RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment". In: *Power Systems, IEEE Transactions on* 23.3 (2008), pp. 853–858.

[D10] A. Motamedi, H. Zareipour, and W. D. Rosehart. "Electricity price and demand forecasting in smart grids". In: *Smart Grid, IEEE Transactions on* 3.2 (2012), pp. 664–674.

[D11] R. K. Ahuja and J. B. Orlin. "Inverse Optimization". In: *Operations Research* 49 (2001), pp. 771–783.

[D12]   J. Saez-Gallego et al. "A Data-driven Bidding Model for a Cluster of Price-responsive Consumers of Electricity". In: *IEEE Transactions on Power Systems* (2016).

[D13]   T. C. Chan et al. "Generalized Inverse Multiobjective Optimization with Application to Cancer Therapy". In: *Operations Research* (2014).

[D14]   A. Keshavarz, Y. Wang, and S. Boyd. "Imputing a convex objective function". In: *Intelligent Control (ISIC), 2011 IEEE International Symposium on*. Sept. 2011, pp. 613–619.

[D15]   A. Aswani, Z.-J. M. Shen, and A. Siddiq. "Inverse Optimization with Noisy Data". In: *arXiv preprint arXiv:1507.03266* (2015).

[D16]   P. M. Esfahani et al. "Data-driven Inverse Optimization with Incomplete Information". 2015.

[D17]   D. Bertsimas, V. Gupta, and I. C. Paschalidis. "Data-driven estimation in equilibrium using inverse optimization". In: *Mathematical Programming* 153.2 (2015), pp. 595–633.

[D18]   Z. Xu et al. "Data-Driven Pricing Strategy for Demand-Side Resource Aggregators". In: *IEEE Transactions on Smart Grid* 99 (2016).

[D19]   J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*. Vol. 1. Springer series in statistics Springer, Berlin, 2001.

[D20]   H. Madsen. *Time Series Analysis*. Chapman & Hall, 2007.

[D21]   A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University Press, 2009.

[D22]   R. Halvgaard et al. "Economic Model Predictive Control for building climate control in a Smart Grid". In: *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES* (2012), pp. 1–6.

[D23]   M. Zugno et al. "A bilevel model for electricity retailers' participation in a demand response market environment". In: *Energy Economics* 36 (2013), pp. 182–197.

[D24]   K. P. Burnham and D. R. Anderson. "Multimodel inference understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2 (2004), pp. 261–304.