UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI ECONOMIA E MANAGEMENT

DOTTORATO DI RICERCA IN MODELLI E METODI PER L'ECONOMIA E IL MANAGEMENT
(ANALYTICS FOR ECONOMICS AND MANAGEMENT - AEM)

Settore Scientifico Disciplinare
SECS–S/06 – METODI MATEMATICI DELL'ECONOMIA
E DELLE SCIENZE ATTUARIALI E FINANZIARIE

CICLO
XXXVI

OPTIMAL INCENTIVE POLICY AND EQUILIBRIUM SHARING RULES FOR RENEWABLE ENERGY COMMUNITIES IN DETERMINISTIC AND STOCHASTIC ENVIRONMENTS

Dott.ssa Ruffini Alessandra, candidata

SUPERVISORE:
Prof. Paolo Falbo

CO-SUPERVISORE:
Prof. Cristian Pelizzari

CO-SUPERVISORE:
Prof. Carlos Ruiz Mora
Universidad Carlos III de Madrid, Dipartimento di Statistica, Spagna

Contents

Abstract

Il cosiddetto “Earth Overshoot Day” (EOD) è avvenuto in anticipo negli ultimi anni. Il consumo eccessivo di risorse naturali contribuisce direttamente alle emissioni di gas serra e aggrava il problema del riscaldamento globale. In questo contesto, gli esseri umani sentono l’urgenza di limitare il cambiamento climatico. Un approccio strategico è la transizione energetica. Le Comunità Energetiche Rinnovabili (CER) sono fondamentali perché incorporano i principi di collaborazione, innovazione e difesa che le rendono indispensabili nel panorama energetico globale. Per aiutare l’implementazione pratica delle CER, è essenziale approfondire il quadro legislativo emanato dall’Unione Europea (UE). Le Direttive Europee 2018/2001 (RED II) e 2019/944 (IEM) giocano un ruolo cruciale: queste direttive mirano a creare quadri regolatori di supporto, fornire incentivi finanziari e incoraggiare la collaborazione per accelerare la transizione.
Il framework virtuale delle CER è stato introdotto in Italia dal Decreto Legge n.162/2019. In tale quadro, tutta l’elettricità prodotta viene venduta alla rete e quindi acquistata quando consumata. Il coinvolgimento efficace delle parti interessate è vitale per il successo delle CER; tuttavia, favorire una collaborazione significativa può essere difficile. Un punto distintivo della normativa italiana riguarda il meccanismo di incentivo basato sull’autoconsumo, che definisce l’autoconsumo come il minimo tra l’energia prodotta e l’energia consumata in un determinato momento. All’interno del framework virtuale delle CER, è cruciale gestire l’allocazione congiunta degli incentivi all’autoconsumo e le decisioni di investimento, data la loro influenza reciproca.
La prima parte di questa tesi è dedicata all’ottimizzazione di tale decisione con una struttura a due livelli. Il livello superiore è giocato da un “policy maker” che definisce la regola di condivisione dell’incentivo sull’autoconsumo per massimizzare l’autoconsumo. Il livello inferiore comprende le CER costituite da membri eterogenei (produttori di biogas e famiglie), che ottimizzano l’investimento in capacità per massimizzare i loro profitti. La diversità degli agenti consente di studiare quanto ognuno contribuisce all’autoconsumo e mette in evidenza le interazioni tra i membri. Inoltre, l’esistenza del livello superiore garantisce una decisione neutrale. Per chiarire l’interazione tra i membri si sviluppano due modelli di iterazione. La prima versione considera un singolo membro per tipo, e le interazioni avvengono in una singola fase. Questi risultati sono influenzati dalle dimensioni dei membri e dall’ordine di iterazione, quindi caratterizzano un equilibrio di Nash non puro. Pertanto, viene sviluppato il secondo modello, dove i due membri sono aggregazioni di individui più piccoli, le interazioni sono divise in due fasi e i risultati sono un equilibrio di Nash puro. La ricerca nei mercati dell’energia spesso richiede un approccio che gestisca in simultanea il rischio e varie fonti di incertezza, che non agiscono indipendentemente. Queste dipendenze, spesso non lineari, pongono una sfida significativa nella fase di “modelling”. Di conseguenza, le operazioni, le decisioni di investimento e la gestione del rischio diventano estremamente complesse.
La seconda parte di questa tesi introduce un metodo mirato a modellare efficacemente le dipendenze lineari, e non, tra più fonti di incertezza, basandosi sulla metodologia avanzata da Cerqueti et al. (2017a), la quale usa Catene di Markov di ordine k per bootstrappare e simulare processi stocastici multivariati; inoltre questo metodo è caratterizzato da una

selezione parsimoniosa dei parametri. L'innovazione distintiva del metodo proposto in questa tesi è la capacità di gestire diverse frequenze temporali. La qualità del metodo proposto viene valutata attraverso test statistici.

Extended Abstract

The so-called "Earth Overshoot Day" (EOD) has occurred earlier and earlier in the last years. This day marks the moment when humanity has exhausted all the natural resources that Earth can regenerate in a year. For the rest of the year, humans draw down local resource stocks. The excessive consumption of natural resources directly contributes to emissions of greenhouse gases, such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). This human behaviour, especially burning fossil fuels for energy production, transportation, and industrial production, exacerbates the problem of global warming, which, ultimately, contributes to climate change. In this context, humans feel the urge to limit climate change, because, with no action, the inauspicious consequences will be not only irreversible but also explosive. One strategic approach adopted to address this challenge is energy transition. It refers to the shift from conventional energy sources to renewable energy sources (RES), thereby reducing greenhouse gas emissions and mitigating their impact on global warming. This transition does not consist of an abrupt abandonment of fossil fuels; rather, it requires a gradual and managed shift across all sectors of the economy. This journey involves technical and infrastructural challenges because, on one side, it is crucial to prevent unstable power grids, while, on the other side, it is essential to establish a flexible grid, which guarantees the balance between demand and supply in a scenario that is partly predictable and partly unplannable due to weather and climate conditions. Contrary to conventional fuels, RES generate electricity without emitting greenhouse gases and are virtually inexhaustible. While the need for an energy transition has gained prominence in recent years, it is notable that certain green technologies have been used in the last centuries. A brief description of anaerobic digestion and solar photovoltaic, two technologies that will be used in this thesis, follows. First, anaerobic digestion is the oldest form of renewable energy technology, with the first plant that dates back to 1859 in India. Also called biodigestion, it involves the breakdown of organic matter by microorganisms in the absence of oxygen, producing biogas primarily composed of methane and carbon dioxide. This process not only efficiently manages organic waste, but also produces biogas, a valuable source of renewable energy. However, despite its potential, the adoption of anaerobic digestion greatly varies among countries due to differing policies and incentives. Germany has embraced it as a key component of its renewable energy strategy, promoting widespread adoption and investment, while other countries lag behind, hindered by regulatory barriers and lack of financial support. Second, solar photovoltaic is more recent: until the end of the 20th century, it played a marginal role. However, thanks to technological innovations that have made it economically competitive concerning fossil fuels, solar photovoltaic is experiencing rocketing growth.

Renewable Energy Communities (RECs) are fundamental to energy transition for a multitude of reasons. First, they decentralize energy production, allowing communities to generate their own renewable energy locally, which in turn enhances energy security and resilience, particularly in times of disruptions or natural disasters. Second, RECs play a pivotal role in promoting the adoption of clean energy technologies, such as solar, wind, and hydro-power, thus significantly contributing to the reduction of greenhouse gas emis-

sions and fighting climate change. Third, beyond environmental benefits, RECs empower local communities by granting them a stake in their energy production and consumption. This empowerment leads to economic advantages through job creation, revenue generation, and local investment opportunities. Fourth, collaborative efforts within RECs foster social cohesion and community engagement as individuals come together to make decisions about energy practices, instilling a sense of ownership and responsibility towards sustainable actions. Fifth, RECs offer flexibility and adaptability to changing energy needs and technological advancements, allowing communities to tailor their renewable energy projects accordingly. Sixth, RECs can advocate for supportive policies at various levels of governance to facilitate energy transition, influencing decision-makers to implement measures that promote clean energy adoption and support community-led initiatives. In essence, RECs embody the principles of empowerment, collaboration, innovation, and advocacy that make them indispensable agents of change in the global energy landscape.
To help the practical implementation of RECs, it is essential to delve into the legislative framework enacted by the European Union (EU) to facilitate the energy transition of its Member States. The European Directives 2018/2001 (RED II) and 2019/944 (IEM) play a crucial role in shaping the scope of RECs. Directive RED II focuses on promoting renewable energy and empowering consumers, facilitating community-owned renewable energy projects and ensuring fair access to the grid. Meanwhile, Directive IEM emphasizes the importance of energy self-sufficiency and community engagement in decision-making processes. Together, these directives aim to create supportive regulatory frameworks, provide financial incentives, and encourage collaboration among stakeholders to accelerate the transition towards a sustainable and decentralized energy system in Europe.
Italy has implemented European laws. Among the others, Law Decree No.162/2019 introduced the virtual framework of RECs. In such a framework, all the electricity produced by RECs' plants is sold to the grid and then purchased back when consumed. The framework confronts various challenges. Navigating through the complex regulatory environment, shaped by a multitude of laws and regulations at both national and regional levels, often leads to bureaucratic hurdles and delays for community initiatives. Furthermore, the integration of virtual energy sharing within existing energy markets presents challenges regarding pricing mechanisms, market rules, and regulatory oversight, necessitating careful coordination to ensure fair compensation for energy transactions while maintaining market stability. Moreover, ensuring data privacy, protection, and cybersecurity is crucial to fostering trust among participants and safeguarding against potential breaches or misuse of data. Effective stakeholder engagement, as mandated by Law Decree No. 162/2019, is vital for RECs success; however, fostering meaningful participation and collaboration among diverse stakeholders may be challenging within virtual frameworks. A distinctive point of the Italian regulation is concerned with the incentive mechanism based on self-consumption, as outlined in Law Decree No. 162/2019, which defines self-consumption as the minimum between the energy produced and the energy consumed at a given time. Within the virtual framework of RECs, it is crucial to manage the joint allocation of self-consumption incentives and investment decisions, due to their mutual influence. The quantities produced are influenced by the installed capacities. However, these capacities are installed based on the offered incentive, which affects investment profitability. Therefore, this interplay must be jointly managed.
The first part of this thesis is devoted to optimizing such a decision. In particular, an optimization problem in a bi-level structure is advanced. The upper level is played by an agent who mixes the roles of a REC administrator and a policy maker. The agent defines the self-consumption incentive-sharing rule to maximize energy self-consumption. The lower level comprises a REC consisting of two types of members, namely biogas producers and households, that optimally size their investments in capacity plants to maximize their

profits. This kind of REC characterizes both urban and agricultural contexts because photovoltaic plants will be the most common technology for urban households, and biodigestors will be a widespread technology in agricultural scenarios. Moreover, the agent diversity allows for the examination of each member's contribution to self-consumption and highlights the interactions among the members. The legislative framework does not prevent the REC from defining the sharing rule internally. However, the presence of an upper-level agent, responsible for allocating the incentive, ensures a neutral decision. To elucidate the interaction among RECs' members, which is a key aspect of this study, two iteration models are developed. The first version considers a single member per type, and interactions occur in a single round. While the results of this model characterize a Nash equilibrium, they are influenced by the iteration order and the size of the members, precluding the attainment of a pure Nash equilibrium. Therefore, the definition of the second model becomes necessary. In this second version, the two members (biogas producer and household) are understood as aggregations of smaller individuals. Additionally, the interactions are divided into two rounds: in the first round, intentions to install are collected, but actual installations do not occur until confirmed in the subsequent round. The previous interaction is iterated until an individual confirms the decision made in the previous round. The results of this second model characterize a pure Nash equilibrium because they are not influenced by the iteration order.

In the realm of RES technologies, such as photovoltaic systems, understanding the relevance of uncertainty sources - like sales prices, solar radiation, demand levels, and supply costs - is paramount in operations and investment decisions. Indeed, it is crucial to recognize the interdependence among these stochastic factors: supply costs, weather conditions, sale prices, and demand levels do not act independently; rather, they are intricately connected. These dependencies, often non-linear, require an approach to manage risk that simultaneously considers these various sources of uncertainty, posing a significant challenge in modelling. Consequently, operations, investment decisions, and risk management become exceedingly complex. Given the interdependencies and complexity of the stochastic processes that are considered in this thesis, the chosen analytical instrument is Markov chains. Markov chains manage the interdependence of stochastic processes by modelling systems where future states depend only on the present state and not on the sequence of events that preceded it (this is the Markov property, which simplifies the analysis and prediction of complex, non-independent sequences). Furthermore, transition probabilities between states are implemented to facilitate the understanding of long-term behaviour and steady-state distributions. Moreover, Markov chains provide a structured approach to simulate and optimize systems influenced by randomness, enhancing decision-making and strategic planning in uncertain environments. The simplicity and versatility of Markov chains make them invaluable tools for capturing and analyzing the dynamics of interdependent stochastic processes.

The second part of this thesis introduces a method aimed at effectively modelling linear and non-linear dependencies among multiple sources of uncertainty, building upon the methodology advanced by Cerqueti et al. (2017a). This method revolves around bootstrapping and simulating multivariate stochastic processes using Markov Chains of order k. The focus lies in modelling the quadrivariate stochastic process encompassing natural gas prices, electricity demand (load), electricity prices, and solar radiation. A distinguishing feature of this method, compared to previous applications, is its ability to accommodate the varying temporal frequencies of the four components - natural gas prices exhibit a daily frequency, while the others possess an hourly frequency. This approach enables the bootstrapping and simulation of the quadrivariate stochastic process leveraging a parsimonious selection of parameters relative to alternative methods. The applied method is now briefly described: first, each stochastic process follows an align-

ment of the historical data downloaded using the "forward filling method" (i.e. missing dates, such as weekends and holidays, are filled with the last previous available element). To simplify the simulation steps that follow, non-recurrent dates (such as February 29th in the leap years) have been excluded from the historical dataset and from the simulation step. Furthermore, series with an hourly frequency required an additional step to align every day to 24 hours. This is required because on days of time switches from daylight saving time to standard time, 25 values are encountered (therefore, the 2 : 00 am price is cut), while only 23 are met with the switch in the opposite sense (therefore, 2 : 00 am price is duplicated). Second, detrendization and deseasonalization are applied to the adjusted series to remove the trend and daily, weekly, and yearly seasonalities. Detrendization is applied to the logarithms of the adjusted series to avoid negative values in the simulation step. Solar radiation does not face detrendization because it is considered stationary in the analyzed period; the adjusted series is transformed into daily logarithms before the deseasonalization and given its deterministic cycle, its deseasonalization process removes only the yearly seasonality. After these procedures, load and electricity prices are transformed from hourly to daily to be aligned with the gas prices and the solar radiation. Third, the series of quadruplets of daily residuals face the cluster analysis using "Ward clustering" to obtain the pairing series of states. Fourth, based on the series of states, the transition probability matrix is estimated to fit a Markov Chain of order 2. Fifth, once the residuals are bootstrapped, trend and seasonalities are added back, the exponentiation removes the logarithm and the simulated series of gas prices, load, electricity prices, and radiation are obtained. The quality of the method is assessed through statistical tests that evaluate the similarity between the distributions of statistics calculated on scenarios and historical data.

Chapter 1

Literature Review

This chapter offers a comprehensive literature review and meticulous examination of research studies and practical applications of RECs and Markov Chain Bootstrapping (MCB).

1.1 Renewable Energy Communities

RECs are a novelty over the last few years. "RECs are a legal entity: which, in accordance with the applicable national law, is based on open and voluntary participation, is autonomous, and is effectively controlled by shareholders or members that are located in the proximity of the renewable energy projects that are owned and developed by that legal entity; the shareholders or members of which are natural persons, SMEs or local authorities, including municipalities; the primary purpose of which is to provide environmental, economic or social community benefits for its shareholders or members or for the local areas where it operates, rather than financial profits"[1]. The literature is still scattered.

Some authors have studied governance models of RECs. Lowitzsch et al. (2020) tests 67 best-practice cases of consumer co-ownership from 18 countries: the authors define renewable energy clusters to take into consideration the complementarity of different energy sources, flexibility, inter-connectivity of actors, and bi-directionality of energy flows. They also discuss the regulatory frameworks, explore the policy implications under the European Clean Energy Package, and argue that renewable energy clusters and RECs are socio-technical mirrors of the same concept; thus, both are necessary for the energy transition. However, they lack empirical data and case studies that illustrate the practical implementation of consumer co-ownership discussed. Inês F.G. Reis et al. (2021) provides a comprehensive view of the emergent and prevailing business models applied to RECs across Europe. It offers valuable insights about key issues and emerging trends in the field but does not deeply investigate each business model's specifics.

Other authors have focused their attention on the role and relevance of consumers. Gjorgievski et al. (2021) enhances how social arrangements and technical designs impact RECs providing a holistic perspective on RECs' multifaceted nature. However, the authors limit their work to a theoretical discussion and avoid empirical analysis that would substantiate their discussion. Pons-Seres de Brauwer and Cohen (2020) quantifies citizens' potential in financing and participating in RECs across Europe and provides a contribution to decentralized energy systems discussions but, it does not analyze financial mechanisms or investment frameworks related to citizen-financed renewable energy projects. Li et al. (2022) analyzes the current heating price models, which do not sup-

[1]DIRECTIVE (EU) 2018/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2018 on the promotion of the use of energy from renewable sources

port the reverse heat supply from prosumers[2] to the central district heating system. The authors optimize heat prosumers' economic performance offering practical solutions to enhance the efficiency of water tank thermal energy storage but, they focus only on heat prosumers, lacking a broader view that integrates different technologies.
Other contributions have focused on how current distribution network pricing can be changed to enable the transition to a smart grid in a low-carbon economy (Faerber et al., 2018), on electricity flows and the related costs (Fina et al., 2022), and on photovoltaic cells material properties (Zhang et al., 2022). These works provide valuable contributions but lack discussions on implications beyond their focus, leaving space for further comparative analysis. Sousa et al. (2023) proposes an optimization model to support REC investment decisions in RES plants and operational electricity sharing management. In particular, it discusses that incremental investments in RES plants are justified if the marginal realized price exceeds the levelized cost of electricity. Additionally, the results demonstrate that within a REC the optimal investment is higher than in individual self-consumption configurations. However, the authors do not consider policies, which are crucial for RECs' implementation.
Furthermore, Pera et al. (2022) describes the circular bioeconomy model of a plant currently working in South Italy. This plant applies the dry anaerobic digestion and the subsequent digestate composting process to the organic fraction of municipal solid waste. The system is powered by photovoltaic energy generated by the photovoltaic plant on the roof. Although this approach offers valuable insights into circular bioeconomy strategies, its focus on the plant currently working in South Italy limits the generalization to other contexts. Also Zhang et al. (2021) includes biogas: the authors propose to use biogas digesting thermodynamics to compensate for the fluctuating outputs of other RES like hydro, wind or solar. Authors apply multi-energy management strategies, and offer holistic solutions, to integrate energy systems in sustainable biogas-dominated hubs but, the technology focus on biogas-related plants is very specific and does not comprehend other RES.
The interactions and strategic decisions within RECs can be studied with the game theory, which helps to understand the incentives and behaviours of REC members and contributes to more resilient energy systems. Lilliu et al. (2023) proposes a market design to discourage energy production curtailment, minimize congestion, promote self-consumption, and ensure a common strategy among selfish prosumers while considering the formation of coalitions. Then, Lilliu and Reforgiato Recupero (2024) proposes a cooperative game theory approach for incentive systems in arbitrary-size coalitions, analyzes the proposed mechanism for large coalitions and creates a new incentive mechanism with new selling functions that optimize peak shaving. Similarly, Gomes and Vale (2024) develops a cost-less renewable energy distribution model of electricity among members' buildings, based on cooperative game theory. This model implements the Shapley value to distribute electricity according to each member's contribution, which receives the energy without costs. Moreover, Maldet et al. (2022) discusses the evolving trends in local electricity market design, emphasizing the regulatory barriers and the role of grid tariffs, which are crucial considerations in implementing game-theoretic solutions for renewable energy integration at the community level.
Finally, Gallego-Castillo et al. (2021) provides a regional analysis of optimal self-consumption under the Spanish new legal framework, while Di Silvestre et al. (2021) reviews existing legal frameworks and discusses the regulation about self-consumption but limits its analysis

[2]EU legislation does not have a definition of "prosumers" but rather uses the term "self-consumption (-generators)": Article 2 of Directive 2018/2001 defines a "renewables self-consumer" as a "final customer who generates renewable electricity for its consumption, and who may store or sell self-generated renewable electricity, provided that, for a non-household renewables self-consumer, those activities do not constitute its primary commercial or professional activity".

to the interaction between the REC and the grid. Both these papers limit their analysis to a single context (Spain or Italy legislative frameworks) and consequently hinder the relevance of the analysis to broader international discussions on RECs.

1.2 Markov chain bootstrapping

An important extension of the RECs model, presented in the first part of the thesis, will regard its development in a stochastic context. To develop such a model, the treatment of the stochasticity of variables and their interdependencies is crucial. Indeed, the generation of multivariate scenarios is pivotal to extend the work of this thesis. The chosen methodology to generate multivariate scenarios is based on resampling techniques that apply Markov Chain Theory. Therefore, a comprehensive literature review on MCB follows.

There are various strands of literature on resampling procedures based on Markov Chain theory. In general, contributions can be distinguished between those concerned with Markov processes that are not necessarily Markov Chains and those related to Markov Chains (discrete-time Markov processes with finite states).

The first strand of literature includes the so-called "block bootstrap method", where consecutive observations of a set of stationary data are divided into blocks of equal length; some blocks are then joined to generate bootstrap samples. This method first appeared in Hall (1985) and was born to solve the problem of dependence disruption. Hall (1985) is trailblazing in the field of resampling methods, providing foundational insight into the theory and the application of resampling techniques. However, being the first one he does not address certain complexities. Indeed, it has been developed in many variants: the non-overlapping block bootstrap (Carlstein, 1986), the blocks-of-blocks bootstrap (Politis and Romano, 1992), the stationary bootstrap (Politis and Romano, 1994), and the tapered block bootstrap (Paparoditis and Politis, 2001a, 2002b). These methods reach the goal of avoiding dependence disruption only partially since they still face the loss of dependency among blocks. An interesting extension of this literature includes the approach of Hounyo et al. (2017), where the blocks-of-blocks bootstrap (Politis and Romano, 1992) is linked with the wild bootstrap in a new method, called "wild block-of-blocks bootstrap", advanced to successfully handle both the dependence and heterogeneity of the squared pre-averaged returns. Furthermore, Hounyo et al. (2017) offers a comprehensive methodological framework to deal with the market micro-structure noise, which is a significant issue in financial econometrics. However, the complexity of the proposed method limits its applicability to specialized agents in the field.

The second strand of literature includes contributions related to Markov Chains. The papers face the problem of maintaining the original data dependency. The "sieve (Markov) bootstrap method" was first advanced by Bühlmann (1997): it fits Markovian models to data series and resamples randomly from the residuals. The author further developed this method and proposed a novel approach: the "variable length Markov Chain sieve bootstrap method" (Bühlmann, 2002). The authors deal with categorical times series that have changing dependencies but, the practical implementation, without appropriate software tools, is challenging. Rajarshi (1990) and Horowitz (2003) estimate a Markov process transition density function through kernel probability estimates. The idea of kernels has been further developed by Paparoditis and Politis (2001b) and Paparoditis and Politis (2002a) with the so-called "local bootstrap method": they assume that similar trajectories will tend to show similar transition probabilities, even though common empirical observations contradict this hypothesis (the existence of structural breaks in the historical series is sufficient to demonstrate the underlying inconsistency of this hypothesis, Bauwens et al. (2015)). Furthermore, Anatolyev and Vasnev (2002) proposes

the "Markov Chain bootstrap method", which is based on a finite state discrete Markov Chain: states are formed by uniformly distributing the values in some percentiles, without being further grouped, and an arbitrary number of time lags bound the relevant path length. This method deals with auto-regressive processes within the bootstrap framework and addresses a common problem in time series analysis. Another approach called "regenerative (Markov Chain) block bootstrap" was first developed by Athreya and Fuh (1992) and Datta and McCormick (1993), and further extended by Bertail and Clémençon (2006) and Bertail and Clémençon (2007). This method uses an atom (a chosen recurring state) and observed cycles, or blocks (consecutive observations between departure from and return to the atom), and the bootstrap proceeds by sampling at random from the observed cycles. The method closes the gap between Markov Chain bootstrapping procedures and block bootstrapping because the cutting points are data-driven and not randomly chosen anymore. Unfortunately, the regenerative block bootstrap is heavily dependent on the identification of the atom, which is unknown. In the same strand of literature, it is worth mentioning Cerqueti et al. (2017b), where the authors apply Markov Chain theory to bootstrap continuous processes. The authors discretize the support of the process and suggest Markov Chains of order k to model the evolution of the time series. This approach results in a too-big transition probability matrix, even for small k. Therefore, they propose a methodology to reduce the number of rows by clustering similar ones using Mixed Integer Linear Programming. They partition the state space of a continuous-valued process into a finite number of intervals and use the contiguity constraint developed in Cerqueti et al. (2015). The explosion of the number of alternative partitions is addressed with a Tabu Search algorithm. Their result is an aggregated transition probability matrix that does not affect the bootstrapping procedure because the typical features of the original series are maintained in the resampled ones, thus confirming the good consistency properties of their method. Despite offering a practical approach to reduce the computational complexity of MCB, the methodology is computationally intensive and the adaptation to different datasets and scenarios raises practical challenges.
A third strand of literature collects the contributions on the estimation of the Markov Chain order, or memory. A certain number of papers (Merhav et al., 1989; Finesso, 1992; Kieffer, 1993; Liu and Narayan, 1994; Csiszár and Shields, 2000; Csiszár, 2002; Peres and Shields, 2005; Chambaz et al., 2009) face the order estimation problem by assuming that all states at all time lags, up to the estimated order, are relevant. The work Cerqueti et al. (2017a) advances a method to estimate the order of a Markov Chain and to identify its relevant states. The resampled series obtained by applying this method share the structural features of the original series and show a controlled diversification. In the context of categorical series representing paths in a network, Petrović and Scholtes (2022) advances a multi-order modelling framework based on a Bayesian learning technique. The method is robust to partial knowledge of the underlying constraints; moreover, it is more correct in estimating the Markov order of paths and less prone to overfitting it than a competing method based on the likelihood ratio test. However, the method is specific to path-based graph analysis and does not have direct applicability to other types of data.
The fourth strand of literature includes works related to information theory and data compression that focus on both the relevant states and the order of a Markov Chain. Among these works, an important method advanced in Bühlmann and Wyner (1999) is the "Variable Length Markov Chain". It is characterized by a variable order that depends on the state that occurred at past time lags. In particular, starting from time lag 1, states are distinguished if they contribute to differentiate future evolution, otherwise, they are grouped. This method identifies a Markov model whose memory changes according to the trajectory followed by the process. Cerqueti et al. (2017a) is strongly related to Bühlmann and Wyner (1999), given that both works aim at reducing the state space of a Markov

Chain as much as possible. However, Cerqueti et al. (2017a) allows for a non-hierarchical selection of the relevant time lags, that is, the relevance of farther time lags is not conditioned to depend on that of the closer ones.

This thesis contributes to the existing literature in two ways. First, given the absence of technical and economic analysis related to interaction dynamics among REC members, incentive distribution mechanisms and integration of legislative frameworks, this work provides a new perspective on RECs by analyzing the interaction among heterogeneous REC members through competitive games, by finding the optimal sharing rule of the incentive that maximizes self-consumption, and by integrating the Italian legislative framework in the mathematical model. Second, the literature does not consider the different time frequencies and interdependencies of stochastic processes in energy markets and focuses mainly on homogeneous time-scale stochastic processes and linear dependencies. Therefore, the model proposed in Chapter 3 builds a transition probability matrix and fits a Markov chain starting by four stochastic processes with heterogeneous time frequencies and with strong interdependencies among them.

Chapter 2

Efficient Incentive Policies of Renewable Energy Communities

2.1 Introduction

Climate change poses significant global risks that, without a decrease in greenhouse gas emissions, will have impacts, not only irreversible but with further escalation. To fight it, energy transition is necessary and decarbonization is the key driver. In this context, a virtuous example of how people can actively contribute to the energy transition and combat climate change is represented by Renewable Energy Communities (RECs). A REC is a legal entity which is based on the open and voluntary participation of members that are in the proximity of renewable energy projects; members have full control of the legal entity and their purpose is to promote self-production and self-consumption of electricity[1]. Thus, through RECs, it is possible to reduce CO_2 emissions, energy waste, and costs related to energy consumption.
Concerning the European legislative framework for RECs, the EU has issued the "Clean energy for all Europeans package". It is a set of directives that aims at reaching carbon neutrality (net-zero emissions) by 2050. Among the other directives, EU Directive 2018/2001 (RED II), and EU Directive 2019/944 are worth mentioning. The former sets binding renewable energy targets for 2030 and lays down financial support rules on self-consumption and RES-based electricity production. The latter sets new rules for the active participation of consumers and sets common guidelines for the internal market of electricity.
Italy has integrated the European Directives with Law Decree 162/2019 (enacted into Law 08/2020), which authorized the creation of RECs and collective self-consumption schemes. The most relevant issue related to Italian regulation on RECs is the choice between adopting a virtual or physical framework. As shown in Figure 2.1, in the virtual framework, all the electricity produced by REC plants is sold to the grid and then purchased back when consumed. In the physical framework, the REC members consume the electricity produced and sell to the grid only the residual electricity. In Italy, the virtual framework is applied.
The Ministry of the Environment and Energy Security, with Decree 162/2019[2] (updated with Decree 414/2023), introduced the definition of self-consumption and the incentive

[1] DIRECTIVE (EU) 2018/2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2018 on the promotion of the use of energy from renewable sources

[2] Law Decree 30 December 2019, n. 162 "Disposizioni urgenti in materia di proroga di termini legislative, di organizzazione delle pubbliche amministrazioni, nonchè di innovazione technological", art 42-bis introduces also two major constraints to enter a REC: first plants must be powered by renewable sources with a total power not exceeding 200 KW; second, the withdrawal points and the injection points must be located on low voltage electric networks at the same medium voltage/low voltage transformer substation.

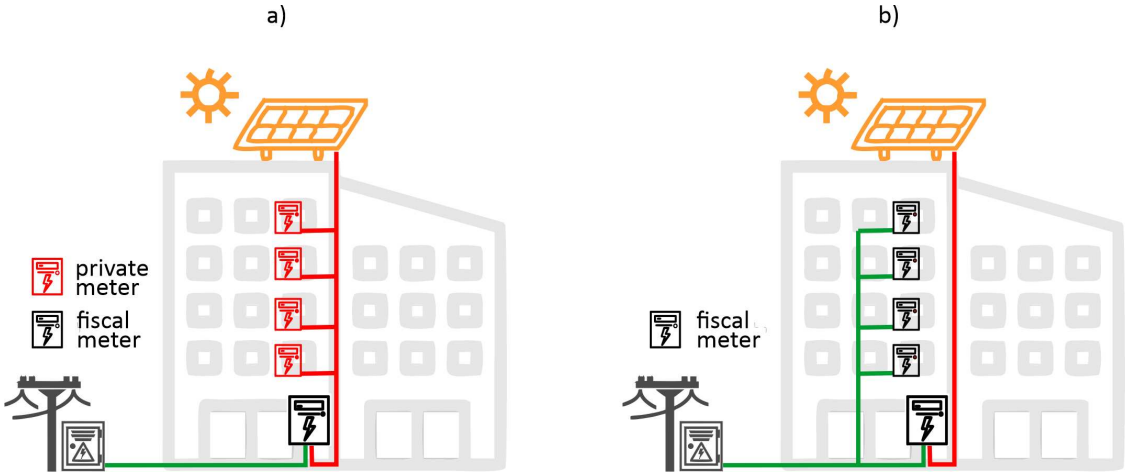

Figure 2.1: Difference between physical (a) and virtual (b) REC frameworks.

on self-consumption (from now on referred to as "incentive"). Self-consumption is defined as the minimum between the consumption and production of electricity by the REC members, for every hour t[3]. Mathematically speaking it can be written as:

$$\text{self-consumption}_t = \min(\text{total load}_t, \text{total production}_t).$$

Self-consumption is multiplied for the incentive[4] and generates revenues that effectively decrease the bills of REC members. Before proceeding with the discussion, the following mechanism must be specified: typically, individuals consume energy by purchasing it from the market paying the price set by their supplier. If they install a RES-based plant, the energy produced is sold on the market at a price that is usually lower than the purchase cost. However, this situation can be improved if one is a REC member, as an additional source of revenue is obtained through the incentive. In addition, given that among members there can be consumers (i.e. individuals who do not have a plant, do not produce electricity but only consume it) RECs answer to the demand for solutions for "energy poverty" because consumers will enjoy the incentive by being REC members. While there is a regulatory framework in place and recognized benefits associated with RECs, the practical implementation of these measures presents hurdles, such as the optimal sizing of plant capacities and how RECs members can share the incentive. First, the optimal sizing of capacities to install is an issue that must be addressed to avoid over-investments. Second, the incentive brings forward the issue of how REC members will interact and allocate it among themselves. The solutions proposed in this thesis aim to find the optimal size of capacity installations and the optimal sharing rule of the incentive that maximizes self-consumption to have a stable and well-managed REC. Due to their mutual influence, the incentive allocation and the investment decisions require joint management. Therefore, the model is structured as a bi-level problem, where, at the upper level, there is a policy maker (or an administrator), while, at the lower level, there are RECs characterized by heterogeneous agents (households and biogas producers). The upper level defines the sharing rule of the incentive that maximizes the self-consumption of RECs at the lower level. At the lower level, RECs size their investment in RES plants to

[3]Law Decree 30 December 2019, n. 162 "Disposizioni urgenti in materia di proroga di termini legislativi, di organizzazione delle pubbliche amministrazioni, nonchè di innovazione technological", art 42-bis.

[4]Established in 100 €/MWh for collective self-consumption schemes and in 110 €/MWh for RECs by Ministry of the Environment and Energy Security Decree of 16 September 2020, "Individuazione della tariffa incentivante per la remunerazione degli impianti a fonti rinnovabili inseriti nelle configurazioni sperimentali di autoconsumo collettivo e comunità energetiche rinnovabili, in attuazione dell'articolo 42-bis, comma 9, del decreto-legge n. 162/2019, convertito dalla legge n. 8/2020."

maximize their profits, taking into consideration the sharing rule and the self-consumption generated. The legislative framework does not prevent the REC from defining the sharing rule internally. However, the introduction of the upper level as the external entity that sets the sharing rule is necessary to avoid conflicts among REC members and to ensure an effective REC life.
Most of the literature discussed in Chapter 1 is theoretical, focuses on photovoltaic plants only, implements one-level problems, and limits the analyses to RECs' external influences. Through the research efforts of this thesis, the objective is to address existing gaps in the literature and contribute to its advancement. First, following current knowledge, there exists a gap in technical and economic analyses regarding interaction dynamics among REC members and incentive distribution mechanisms. The management of this interaction is crucial to allow a correct and effective REC life. Consequently, a bi-level problem is set, where, the entity at the upper level (it can be a central authority, a policy-maker, or an administrator) defines the sharing rule of the incentive that maximizes self-consumption of the REC at the lower level. The introduction of the upper level as the external individual who defines the sharing rule is necessary to avoid conflicts among REC members and to ensure an effective REC life. Second, according to the literature studied in Chapter 1, except for Gallego-Castillo et al. (2021) and Di Silvestre et al. (2021), many models are abstracted from legal constraints, incentive policies, and current regulatory frameworks. Therefore, the present thesis proposes a model which includes the relevant features of the virtual framework for RECs (referring to the Italian regulatory framework). Third, despite the capillary presence of farms and agriculture-related activities, literature related to biogas and biodigestor dynamics is still scattered. Thus, the present work focuses on RECs with heterogeneous members to integrate the urban context (photovoltaic households) with the agricultural one (farms with biogas digestors). This work wants to collect the opportunity of joining in the same REC two different technologies, the photovoltaic plant and the biodigestors. The ownership of photovoltaic plants, whether households, enterprises or farmers, is irrelevant for the purposes of this work. However, from an empirical point of view, the household sector relevance is substantial: in Italy it has grown from $756,799$ plants in 2020 to $1,355,687$ plants in 2023 (capacity has more than doubled, from $3,458$ MW in 2020 to $7,032$ MW in 2023) while the photovoltaic from the agricultural sector has increased from $38,115$ plants ($2,497$ MW installed) in 2020 to $45,560$ plants ($2,877$ MW installed) in 2023[5]. Fourth, to avoid over-investments, sizing capacities problem is taken into consideration.
The outline of this chapter is as follows: in Section 2.2 the problem and the bi-level model structure are described, in Section 2.3 the sensitivity analysis and results discussions are presented. Finally, Section 2.4 illustrates the conclusions.

2.2 Model

The model is concerned with a bi-level optimization problem. The upper level (which can be a central authority, a policy maker, or an administrator) decides the sharing rule of the incentive that maximizes the self-consumption of the REC at the lower level. At the lower level, there is a REC with heterogeneous members. REC members maximize their profit function by optimally sizing the installation of RES plant capacities.
Let us proceed to introduce some notation. First, the 24-hour of a day are indexed by $t \in T$, with $t = \{1, \dots, \bar{T}\}$. At the lower level, heterogeneous agents characterize each REC, so different letters to identify them are introduced: agents are biogas producers, who are labelled with the letter b, and households, who are labelled with the letter h.

[5]Data are available at GSE in "Photovoltaic - Statistical Report 2020" and "Photovoltaic - Statistical Report 2023".

A different number of agents can form the REC, therefore, it is necessary to introduce the index $i \in I$, with $i = \{1, \dots, \bar{I}\}$, to identify the different biogas producers, and index $j \in J$,with $j = \{1, \dots, \bar{J}\}$, to identify the different households.
Each agent maximizes his profit, which is labelled with π. Consequently, π_i^b is the profit of the i-th biogas producer, and π_j^h is the profit function of the j-th household. Profits are influenced by many factors, among them the installed capacities and the consequent produced quantities are worth mentioning. Given that agents can install different types of RES plants, it is necessary to introduce labels for each kind of technology: in this problem, biogas producers install a biodigestor, labelled with the letter g, and a gas-to-electricity turbine, labelled with the letter e, while households install a photovoltaic plant, labelled with the initials pv.
Furthermore, capacities to install are labelled with capital Q, while production quantities are labelled with small q. Therefore, installed capacities of the i-th biogas producer are defined as Q_i^g and Q_i^e, for the biodigestor and gas-to-electricity turbine, respectively. Q_j^{pv} identifies the photovoltaic plant of the j-th household. Consequently, the quantities produced by each plant are $q_{t,i}^g$, $q_{t,i}^e$, and $q_{t,j}^{pv}$ for the biodigestor, the turbine, and the photovoltaic plant, respectively, at each hour $t \in T$. Finally, load is labelled with the letter d: $d_{t,i}^b$ and $d_{t,j}^h$ are the load of the i-th biogas producer and the j-th household, with $t \in T$.
In this section is extremely important to introduce three key concepts: the self-consumption that the upper level aims to maximize, the sharing rule that the upper level defines to share the incentive among REC members, and the two model versions at the lower level.

First, self-consumption is defined as the minimum between the electricity produced and consumed by REC members. Self-consumption is denoted as q_t^{REC} because it is computed for every hour t, with $t \in T$. Mathematically speaking, self-consumption depends on the quantity of electricity produced, which, itself depends on the capacities installed by REC members. Therefore, its mathematical definition is the following:

$$q_t^{REC}(q_{t,i}^e, q_{t,j}^{pv}) = \min \left\{ \sum_{i \in I} d_{t,i}^b + \sum_{j \in J} d_{t,j}^h \ , \ \sum_{i \in I} q_{t,i}^e + \sum_{j \in J} q_{t,j}^{pv} \right\}, t \in T, \tag{2.1}$$

with

$$q_{t,i}^e \leq Q_i^e,$$
$$q_{t,j}^{pv} \leq Q_j^{pv} \gamma_t^{pv},$$

where the first term sums the electricity loads of all biogas producers and households, while the second term sums the electricity productions of all biogas producers and households participating in the REC, and productions are constrained by the installed capacities.[6] In this equation the quantity of biogas $q_{t,i}^g$ does not appear because self-consumption considers the production, and demand, of electricity. The production of electricity $q_{t,i}^e$ influences the production of biogas, as is shown later. According to national regulation, each REC benefits from an incentive prizing self-consumption within the community. In particular, in Italy, the "virtual" framework is applied (as anticipated in Figure 2.1): each member sells all the energy self-generated to the market and purchases, from the grid,

[6]This is a nonlinear condition but it can be linearized. In particular, $x = \min\{a, b\}$ can be rewritten as:

$$\begin{aligned} &a - M(1-u) \leq x \leq a \\ &b - Mu \leq x \leq b \\ &uM > b - a \\ &(1-u)M > a - b \\ &u \in \{0, 1\} \end{aligned}$$

where M is a sufficiently large constant.

the electricity to satisfy his load. Consequently, self-consumption occurs only virtually, it is recorded through smart meters punctually distributed on the grid, accounted for by the central authority and finally compensated through the incentive. This incentive must be distributed among the members and here comes into play the sharing rule.

Second, as a consequence of the virtual framework, the administrator (upper level) defines a sharing rule to distribute the incentive among REC members and to maximize self-consumption q_t^{REC}. The sharing rule is labelled with δ and is defined by the upper level to avoid conflicts among RECs members and ensure an effective REC life.

Third, at the lower level, in each REC, self-consumption allows to highlight the interaction among REC members because it considers the produced quantities and the loads of all agents. However, this interaction presents challenges that lead to the development of two versions. In the first version, agents are sufficiently large, they have great negotiation power and consequently, depending on who acts first, the final outcomes change in terms of optimized profits, installed capacities and generated self-consumption. This version is explained in Section 2.2.2 "Incremental commitment hypothesis". In particular, the results of this version are influenced by the iteration order. Therefore, a second version of the model characterizing the lower level has become necessary. In this second version, there are still two types of players but they are not aggregated in two large entities, i.e. each agent is considered as numerous individual entities. Let us provide an example. Suppose that in the first version there is 1 biogas producer whose load is equal to x MWh, while in the second version, there are n biogas producers whose individual load is y MWh, with $x > y$. Consequently, in the second version, $n = \frac{x}{y}$ biogas producers are needed to reach the load level of the single biogas producer in the first version.
Furthermore, the two models differ because in the first version, each agent can either confirm or increase his previously stated capacity, he can not withdraw from it and the interaction is depicted in one round. In the second version, more rounds are introduced: the interaction in the first round works as in the previous version, while, in the second round, negative capacities are allowed to represent the possibility of rethinking the previous choice. Further details on these interactions are provided in Section 2.2.2 Rethink hypothesis (pure Nash equilibrium).
Figures 2.2 and 2.3 represent the structure of the two models implemented: the first model, characterized by only one agent per member (i.e. one biogas producer and one household) is shown in Figure 2.2, while the second model characterized by multiple agents per each type of member is shown in Figure 2.3.

2.2.1 Upper level - Policy maker

The main objective of the policymaker is to maximize the self-consumption of RECs, q_t^{REC}, with $t \in T$, acting on the sharing rule δ of incentive z.
To maximize self-consumption, the policymaker promotes the installation of renewable resources plants and the participation of neighbourhoods to RECs by prizing self-consumption with an incentive z. This incentive is then allocated to each REC member through the sharing rule δ. The policymaker will choose the δ that maximizes self-consumption. The role of the policymaker is essential for the efficient functioning of RECs since his neutrality towards the involved parties makes him impartial, enabling him to make decisions objectively. The optimization is defined as in the following:

$$\max_{\delta} \sum_{t=1}^{T} q_t^{REC}(q_{t,i}^{e}, q_{t,j}^{pv}) \tag{2.2a}$$

s.t.

$$0 \leq \delta \leq 1 \tag{2.2b}$$

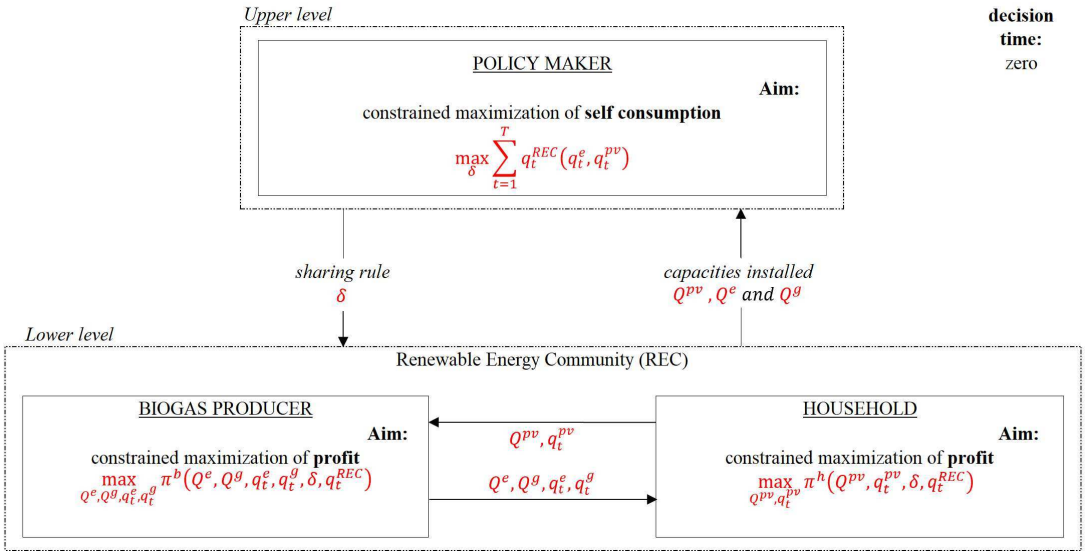

Figure 2.2: Structure of the bilevel problem in the Incremental commitment hypothesis model.

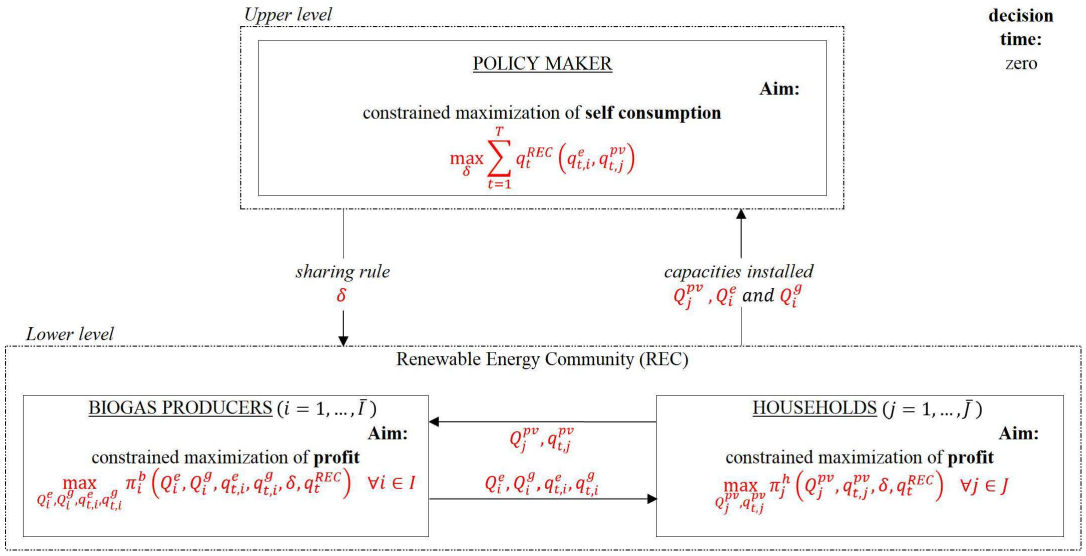

Figure 2.3: Structure of the bilevel problem in the Rethink hypothesis (pure Nash equilibrium) model.

where Constraint (2.2b) fixes the sharing rule δ between 0 and 1 (i.e. how much incentive is given to each member).

2.2.2 Lower level - Renewable Energy Community

At time zero and within the same REC, biogas and household must size their power generation plants to maximize their profits, which are influenced by the energy community regulation and by the sharing rule δ on incentive.
In particular, in a given REC the i-th biogas producer must decide the capacity of a gas-to-power turbine, Q_i^e (in MW), and the j-th household must size the capacity of a photovoltaic plant, Q_j^{pv} (also in MW). Figure 2.4 shows how the sharing rule interacts with the capacity decision of both members, in a simplified framework where profit is

linear with respect to the installed capacities. For ease of notation, the indexes i and j related to agents are momentarily dropped, and the capacities are identified as Q^e and Q^{pv}. For clarity, in this example, the generation is assumed equal to the corresponding capacity installed, therefore $q_t^{pv} = Q^{pv}$ and $q_t^e = Q^e$.
Letting d the energy load of the community, the possible values of capacity installed (Q^{pv} and Q^e) are distinguished in two intervals, a and b.
In each panel of Figure 2.4 (left for the household, right for the biogas producer), it is supposed that only the corresponding member installs some generation capacity (while the other installs nothing) so that he only can satisfy the load of the community.
Given the regulation, as long as $Q^{(\cdot)} \leq d$ (interval a), with $(\cdot) = e, pv$, the community cashes in both the self-consumption incentive and the revenues of selling the energy to the market. As soon as $Q^{(\cdot)} > d$ (interval b), the incentive stops and the extra capacity (beyond d) only generates revenues from the sale to the market. So, three main outcomes can be identified:

i. member x has positive marginal profit whatever is δ, so the optimal capacity size Q^x is bounded by the budget or technical constraints, whatever binds first;

ii. member x has positive marginal profit as long as $Q^x \leq d$. In this case, losing the incentive for the generation resulting from the extra capacity installed over d changes the sign of the marginal profit from positive to negative. Optimal investment occurs at $Q^x = d$;

iii. member x has negative marginal profit whatever is δ, so it is optimal to him to install no capacity;

In the spirit of the EU directive, an outcome of type ii. can be taken as the benchmark, since it represents a solution where self-generation is sized to match load. In this way, self-consumption is maximized efficiently, i.e. avoiding any over-investment.
However, the actual solution of this optimization problem is going to be more complicated

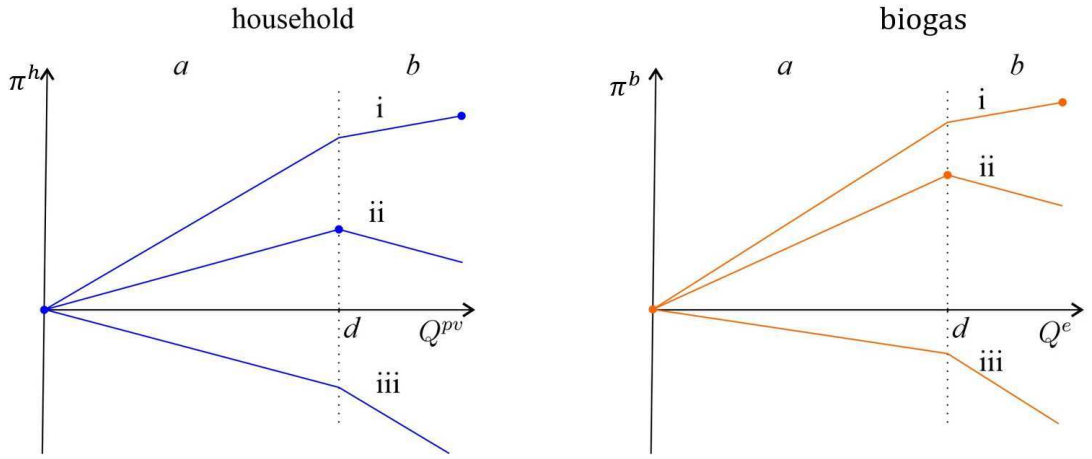

Figure 2.4: Profit for a household and biogas as a function of the capacity installed and depending on the sign of the marginal profit occurring in the intervals $a = [0, d)$ and $b = [d, \infty)$.

than what is represented in Figure 2.4 for several reasons.
First of all, the members of the energy community are free to install their generation capacity, so their joint capacity must be compared with the joint load of the community. This freedom may cause over-investments in capacities.
Second, different values of δ have an impact on the optimal capacity decision, so these

decision variables interact. The analysis of how adjustments in the optimal capacity decision respond to variations in the values of δ must be carried out.
Finally, in these settings, the biogas producers can compare, on an hourly basis, if selling the biogas as biogas on the gas market is more profitable than directing it to the gas-to-power turbine (i.e. to meet the electricity load of the energy community). See Section 2.2.2 for a more comprehensive explanation. So, this model encompasses both initial investment and everyday management decisions given the market prices of both electricity and gas.

Incremental commitment hypothesis

In this framework, 1 biogas producer and 1 household are considered. Given a REC, the biogas producer must size a biodigestor plant (Q^g) and a gas-to-electricity turbine (Q^e) to maximize his profits (π^b), while the household must size a photovoltaic plant (Q^{pv}) to maximize his profits (π^h). Each profit is influenced by the capacities, productions, and load of the agent and the capacities, productions, and load of the other agent (through self-consumption q_t^{REC}), and the sharing rule δ.
The interaction among players is now explained. Each player acts knowing the decision of the previous player. For example, suppose that a REC whose members have no plants is being built, and suppose that the biogas producer acts first (thus, the capacity of the household is set to zero). The biogas producer maximizes his profit. Then, it is the household turn. The household knows the biogas capacities and maximizes his profit, accordingly. After household maximization, the biogas is aware of household decision and can increase or confirm his previous decision. This iteration stops when both agents confirm their choices. In the following paragraphs, the mathematical formulations of biogas producer and household optimization problems are presented.

Biogas producer Given a REC, let us start with the biogas producer problem: he aims to maximize his profit π^b. Revenues are earned from three possible sources:

- the sale of biogas q_t^g (m^3) at the market price p_t^g (€/m^3),
- the sale of electricity on the market by converting biogas through the gas-to-power turbine q_t^e (MWh) at the spot market price p_t^{me} (€/MWh),
- and the share δ (%) of incentive z (€/MWh) obtained on shared electricity q_t^{REC} (MWh).

Costs comprehend bills and investment costs. Bills are represented by the purchase of electricity to satisfy load d_t^b. This electricity is purchased at the market price p_t^{me} increased by k (€), which takes into account the retailer profit margin[7]. Investment costs include fixed costs C^g (i.e., to pay for licenses and authorizations, €), and variable costs p^{qg} (€/m^3) and p^{qe} (€/MW), which are the unit costs for a cubic meter of anaerobic biodigestor and an MW of a gas-to-power turbine, respectively. These unit costs are multiplied by the installed capacities of the biogas digestor Q^g (m^3) and turbine Q^e (MW). Investment costs are multiplied by the parameter a which represents the depreciation rate over the time horizon of operations (i.e. each cost is depreciated over the entire lifespan of the plant, with the unit of time being the operational time horizon). This optimization is constrained by technical and economic constraints. Constraint 2.3b sets the physical condition where the production of electricity can not exceed the installed turbine capacity. Furthermore, for simplicity, the generation of electricity from the turbine (if installed) is

[7]It is assumed that the electricity price to buy back the electricity from the market differs from the selling price by a constant k, which is not proportional to the price itself, p_t^{me}.

assumed equal to the capacity installed in each time interval $t \in T$.
Constraint 2.3c defines the quantity sold of biogas as the residual biogas, that is what remains after a portion of it is converted into electricity. In this model, the transformation from biogas (in m^3) to electricity (in MWh) is performed using the parameter m^e, which measures the amount of MWh of electricity produced by a cubic meter of biomethane (MWh/m^3). Similarly, the quantity of electricity (measured in MWh) generated by the turbine can be converted into m^3 by multiplying it by $\frac{1}{m^e}$. These two constraints describe the choice that the biogas producer has to make between producing biogas or transforming it into electricity. This choice is taken by confronting the prices on the spot market of biogas (expressed in €/MWh), $\frac{p_t^g}{m^e}$, with the electricity price p_t^{me} (€/MWh). The choice is represented with the indicator functions $\mathbb{1}$: namely, if $p_t^{me} \geq \frac{p_t^g}{m^e}$, then it is more convenient to transform biogas into electricity and sell it on the market; otherwise, the transformation is not convenient and the biogas producer sells biogas directly on the spot market.
Constraint 2.3d sets the economic constraint where the investment costs must be covered by the personal budget of the biogas producer, B^b.
Last, Constraint 2.3e explains the influence of household on biogas choices because self-consumption is impacted by the production and load of the household. This equation has key relevance in the model because it is used to represent the interaction among the heterogeneous agents.
The mathematical definition is the following:

$$\max_{Q^e, Q^g, q_t^g, q_t^e} \pi^b = \sum_{t=1}^{\bar{T}} \left(p_t^g q_t^g + p_t^{me} q_t^e + z\delta q_t^{REC} - (p_t^{me} + k) d_t^b \right) - a \left(C^g + p^{qg} Q^g + p^{qe} Q^e \right) \tag{2.3a}$$

s.t.

$$0 \leq q_t^e = Q^e \cdot \mathbb{1}_{\{p_t^{me} > p_t^g m^e\}} \leq Q^e \quad \forall t \tag{2.3b}$$

$$0 \leq q_t^g = \left(Q^g - \frac{q_t^e}{m^e} \right) \cdot \mathbb{1}_{\{p_t^{me} \leq p_t^g m^e\}} \quad \forall t \tag{2.3c}$$

$$0 \leq C^g + p^{qg} Q^g + p^{qe} Q^e \leq B^b \tag{2.3d}$$

$$q_t^{REC} = (q_t^e + q_t^{pv}) \cdot \mathbb{1}_{\{q_t^e + q_t^{pv} \leq d_t^b + d_t^h\}} + (d_t^b + d_t^h) \cdot \mathbb{1}_{\{q_t^e + q_t^{pv} > d_t^b + d_t^h\}} \tag{2.3e}$$

Household Given a REC, the household maximizes his profits π^h. Revenues are generated from selling photovoltaic electricity q_t^{pv} (MWh) at the spot market price p_t^{me} (€/MWh), and from the residual share $1 - \delta$ of the incentive z (€) on self-consumption q_t^{REC} (MWh). Costs take into account bills and investments. Bills are related to the electricity purchased to cover the load d_t^h (MWh) at the price applied by the retailer ($p_t^{me} + k$, in €/MWh). The investment in the PV plant is the product between the unitary cost of a PV plant p^{pv} (€/MW) and the installed capacity Q^{pv} (MW). The investment cost is then multiplied by the parameter b, which is the PV plant depreciation rate over the time horizon of operations (i.e. each cost is depreciated over the entire lifespan of the plant, with the unit of time being the operational time horizon).
This optimization is subject to two constraints. Constraint 2.4b dictates that the investment cost cannot exceed the individual budget B^h (€). Constraint 2.4c defines the quantity of electricity produced q_t^{pv} as the product between the capacity installed Q^{pv} and the parameter γ_t^{pv} (%), which measures the efficiency of the PV plant taking into account the geographic position of the plant.[8] Furthermore, constraint 2.4c sets the quan-

[8] For example, in northern Italian regions, the global solar radiation is different, and on average lower, than the solar radiation of the southern regions.

tity of PV electricity produced lower than the maximum capacity installed. Given that Constraint 2.4d describes the self-consumption as before, the mathematical formulation is the following:

$$\max_{Q_r^{pv}, q_t^{pv}} \pi_r^h = \sum_{t=1}^{\bar{T}} \left(p_t^{me} \gamma_t^{pv} Q^{pv} + z(1-\delta) q_t^{REC} - (p_t^{me} + k) d_t^h \right) - b p^{pv} Q_r^{pv} \tag{2.4a}$$

s.t.

$$0 \leq p^{pv} Q_j^{pv} \leq B^h \tag{2.4b}$$

$$q_t^{pv} = \gamma_t^{pv} Q^{pv} \leq Q^{pv} \quad \forall t \tag{2.4c}$$

$$q_t^{REC} = (q_t^e + q_t^{pv}) \cdot \mathbb{1}_{\{q_t^e + q_t^{pv} \leq d_t^b + d_t^h\}} + (d_t^b + d_t^h) \cdot \mathbb{1}_{\{q_t^e + q_t^{pv} > d_t^b + d_t^h\}} \tag{2.4d}$$

Derivatives analysis

Let us proceed to analyze the derivatives of biogas and household profits with respect to the installed capacities. This analysis is limited to the derivatives of the first version from Section "Incremental commitment hypothesis" but the same approach holds for the second one that will be explained in Section "Rethink hypothesis (pure Nash equilibrium)".

Biogas producer derivatives To analyze the choice of the biogas producer between installing only the biodigestor or installing the turbine, the market prices of biogas and electricity must be compared. In particular, two cases are introduced:

$$\text{case 1} : \frac{p_t^g}{m^e} < p_t^{me} \longrightarrow q_t^e = Q^e \text{ and } q_t^g = Q^g - \frac{Q_r^e}{m^e},$$

$$\text{case 2} : \frac{p_t^g}{m^e} > p_t^{me} \longrightarrow q_t^e = 0 \text{ and } q_t^g = Q_r^g.$$

The biogas price p_t^g is expressed in m^3 and is transformed into MWh by dividing it by the coefficient m^e, which computes how many MWh are produced by a m^3 of biomethane ($\mathrm{MWh/m}^3$). This conversion allows the comparison with the electricity market price p_t^{me}. In case 1, if the price of the biogas p_t^g, expressed in €/MWh, is lower than the electricity market price, then it is convenient to transform the biogas into electricity. Therefore, the biogas producer in a given REC, at each time t, produces all possible q_t^e and q_t^g is only residual. Thus, $q_t^e = Q^e$ for every $t \in T$. The biogas profit function, of a given energy community r, becomes:

$$\begin{aligned} \max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi_r^b \quad &= \quad \sum_{t=1}^{T} \left(p_t^g \left(Q^g - \frac{Q_r^e}{m^e} \right) + p_t^{me} Q^e + z\delta q_t^{REC} - (p_t^{me} + k) d_t^b \right) \\ &- \quad a \left(C^g + p^{qg} Q^g + p^{qe} Q^e \right). \end{aligned}$$

Now, it is necessary to distinguish between the two possible values that q_t^{REC} can assume. If the REC load is lower than the electricity production, then $q_t^{REC} = d_t^b + d_t^h$. The objective function becomes:

$$\begin{aligned} \text{case 1.1} \quad \max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi_r^b \quad &= \quad \sum_{t=1}^{\bar{T}} \left(p_t^g \left(Q^g - \frac{Q_r^e}{m^e} \right) + p_t^{me} Q^e + z\delta \left(d_t^b + d_t^h \right) - (p_t^{me} + k) d_t^b \right) + \\ &- \quad a \left(C^g + p^{qg} Q^g + p^{qe} Q_r^e \right), \end{aligned} \tag{2.5}$$

and the derivatives concerning the installed capacities are the following:

$$\frac{\partial \pi^b}{\partial Q_r^e} = -\frac{p_t^g}{m^e} + p_t^{me} - a p^{qe} \quad \forall t, \tag{2.6}$$

$$\frac{\partial \pi^b}{\partial Q_r^g} = p_t^g - a p^{qg} \quad \forall t.$$

Otherwise, if the load is greater than the electricity production, then it is possible to assume $q_t^e = Q^e$ and $q_t^{REC} = Q^e + \gamma_t^{pv} Q^{pv}$, where $\gamma_t^{pv} Q^{pv} = q_t^{pv}$. The objective function becomes:

$$\begin{aligned} \text{case 1.2} \quad \max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi^b &= \sum_{t=1}^{\bar{T}} \left(p_t^g \left(Q_r^g - \frac{Q_r^e}{m^e} \right) + p_t^{me} Q^e + z\delta \left(q_t^e + \gamma_t^{pv} Q^{pv} \right) - (p_t^{me} + k) d_t^b \right) + \\ &- a \left(C^g + p^{qg} Q^g + p^{qe} Q^e \right), \end{aligned} \tag{2.7}$$

and the derivatives are:

$$\begin{aligned} \frac{\partial \pi^b}{\partial Q_r^e} &= -\frac{p_t^g}{m^e} + p_t^{me} + z\delta - a p^{qe} \quad \forall t, \\ \frac{\partial \pi^b}{\partial Q_r^g} &= p_t^g - a p^{qg} \quad \forall t. \end{aligned} \tag{2.8}$$

In case 2, if the price of the biogas (in €/MWh) is greater than the electricity market price, then it is more convenient to sell the biogas rather than transforming it into electricity. Therefore, $q_t^e = 0$ and $q_t^g = Q^g$. The biogas profit function, for a given energy community r, becomes:

$$\max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi^b = \sum_{t=1}^{\bar{T}} \left(p_t^g Q^g + z\delta q_t^{REC} - (p_t^{me} + k) d_t^b \right) - a \left(C^g + p^{qg} Q^g + p^{qe} Q_r^e \right).$$

Now, let us distinguish between the two possible values that q_t^{REC} can assume.
If the overall load is lower than the electricity production, $q_t^{REC} = d_t^b + d_t^h$. The objective function is:

$$\begin{aligned} \text{case 2.1} \quad \max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi^b &= \sum_{t=1}^{\bar{T}} \left(p_t^g Q^g + z\delta \left(d_t^b + d_t^h \right) - (p_t^{me} + k) d_t^b \right) + \\ &- a \left(C^g + p^{qg} Q^g + p^{qe} Q_r^e \right), \end{aligned}$$

and the derivatives are the following:

$$\begin{aligned} \frac{\partial \pi^b}{\partial Q_r^e} &= -a p^{qe}, \\ \frac{\partial \pi^b}{\partial Q_r^g} &= p_t^g - a p^{qg} \quad \forall t. \end{aligned}$$

If instead, the load is greater than the electricity production, $q_t^{REC} = Q^e + \gamma_t^{pv} Q^{pv}$. The objective function is:

$$\begin{aligned} \text{case 2.2} \quad \max_{Q_r^e, Q_r^g, q_t^g, q_t^e} \pi^b &= \sum_{t=1}^{\bar{T}} \left(p_t^g Q^g + z\delta \left(q_t^e + \gamma_t^{pv} Q^{pv} \right) - (p_t^{me} + k) d_t^b \right) + \\ &- a \left(C^g + p^{qg} Q^g + p^{qe} Q_r^e \right), \end{aligned}$$

and the derivatives are:

$$\begin{aligned} \frac{\partial \pi^b}{\partial Q_r^e} &= -a p^{qe}, \\ \frac{\partial \pi^b}{\partial Q_r^g} &= p_t^g - a p^{qg} \quad \forall t. \end{aligned}$$

Now, let us explain in more detail the mechanisms that define the three possible results anticipated in Figure 2.4. There can be three different outcomes according to the value of the derivatives. If the biogas producer installs a quantity greater than the demand, then one must look at the right side of the graph, in the area labelled with b. Here, equations (2.5) and (2.6) hold. Otherwise, if the biogas producer installs a quantity lower than the demand, one is in the area labelled with a and the Equations (2.7) and (2.8) hold.

Household derivatives Since the household has only one technology (the photovoltaic plant), the comparison of different prices is not necessary and the analysis is limited to the two possible values of q_t^{REC}.
If the overall load is lower than the electricity production, $q_t^{REC} = d_t^b + d_t^h$ and the derivative is:

$$\frac{\partial \pi^h}{\partial Q_r^{pv}} = p_t^{me}\gamma_t^{pv} - bp^{pv} \quad \forall t. \tag{2.9}$$

If instead, the load is greater than the electricity production, $q_t^{REC} = Q^e + \gamma_t^{pv}Q^{pv}$ and the derivative is:

$$\frac{\partial \pi^h}{\partial Q_r^{pv}} = p_t^{me}\gamma_t^{pv} + z(1-\delta)\gamma_t^{pv} - bp^{pv} \quad \forall t. \tag{2.10}$$

By looking at Figure 2.4, the approach for the household is the same: by looking at area a of the graph, then Equation (2.9) holds; otherwise, in area b, Equation (2.10) holds.

Rethink hypothesis (pure Nash equilibrium)

This version is characterized by two main differences with respect to the version explained in Section 2.2.2 "Incremental commitment hypothesis". First, one biogas producer and one household are not considered anymore but a group of biogas producers and a group of households are taken into account. Each agent within a group is smaller than the single agents studied before (i.e. lower budget and lower load). In this way, there is a differentiation from the model in Section 2.2.2 because by iterating over smaller agents, the iteration order and the size of each agent do not influence the results anymore. Furthermore, the order of agent interviews is alternated: following each biogas producer, an individual household is interviewed, and so forth.
Second, the iteration is split into two rounds to respect the process of REC creation that is effectively implemented in Italy. In the first round, biogas producers and households are asked about their willingness to enter the REC, their interest in installation, and the potential quantity they intend to install. In the first round, there is only an intent collection and the actual installation does not take place. The first round ends when a predefined threshold is satisfied. The threshold can be chosen among a handful of alternatives: for example, the first round is complete when the REC load is covered by the REC member's plant production. Once the first round is completed, the second round starts. At the beginning of the second round, each agent is aware of all the choices taken in the first round. Given the information set, each agent can confirm or change his previous choice. For simplicity, the agent order is assumed equal in both rounds (i.e. if you were the first one in the first round, you are the first one in the second round too, and so on). If an agent changes his previous decision, the following agent has an information set updated with the updated choice. If all agents change their decision taken in the previous round, the second round ends and the third one begins. This process repeats until an agent confirms his choice from the previous round because no further iterations are required, given that the optimization problem is the same for the following agents. Once the round ends, the effective installations occur.
The description of the mathematical model follows but, before that, it is necessary to introduce some notation to ease the explanation. First, in this version a plurality of agents is considered, therefore, the indexes $i \in I$ and $j \in J$ are applied to identify the different biogas producers and households, respectively. Second, the label pre is introduced to identify the capacities and the quantities decided by the *pre*vious agents. Third, the numbers 1 and 2 are used to distinguish the rounds where the choices were taken. The mathematical explanation of the optimization problems in the two rounds by describing only the equations and constraints that differ from the model defined previously in Section 2.2.2 "Incremental commitment hypothesis" follows.

Round 1 Each agent is asked to state his willingness to enter the REC, his intention to install and the capacity he is willing to install. In this round, there is no effective installation because choices must be first confirmed in the following round.
The first round is completed when the load threshold is achieved, i.e. the production of electricity from the REC plants is sufficient to cover the load of REC members.
The iteration alternates one agent per type: suppose that one biogas producer acts first, then the second agent will be an household, while the third one will be a biogas producer again, and so forth.

Biogas Each biogas producer maximizes his profits by taking into account what has been decided by agents (both biogas and households) before him.
Profits are labelled with π_i^b to identify the i-th biogas producer in a given REC. The profit function structure is the same as previously defined in Problem (2.3).
Now, the notation that differentiates the quantities and the capacities chosen by the previous agents from the quantities and the capacities chosen by the i-th agent is introduced.
The biodigestor and gas-to-electricity turbine capacities chosen in round 1, by the previous agents up to the i-th one (excluded), are labelled as $Q_i^{g1.pre}$, and $Q_i^{e1.pre}$, respectively. Indeed, $Q_i^{g1.pre} = \sum_{c=1}^{i-1} Q_c^{g1}$ and $Q_i^{e1.pre} = \sum_{c=1}^{i-1} Q_c^{e1}$ where all the capacities chosen by $i-1$ agents are considered (see Constraints (2.11b) and (2.11c), respectively). The i-th one is excluded from this notation because his choices are depicted by Q_i^{g1} and by Q_i^{e1}, respectively.
Similarly, the quantities of biogas and of electricity, respectively, sold on the spot market by agents before the i-th one in round 1 are labelled as $q_{t,i}^{g1.pre}$ and $q_{t,i}^{e1.pre}$. These quantities are the sum of the quantities produced (and sold) by $i-1$ agents, namely: $q_{t,i}^{g1.pre} = \sum_{c=1}^{i-1} q_{t,c}^{g1}$ (see Constraint (2.11d), and $q_{t,i}^{e1.pre} = \sum_{c=1}^{i-1} q_{t,c}^{e1}$ (see Constraint (2.11e).
The quantities of biogas and electricity produced by the i-th biogas producer are labelled with $q_{t,i}^{g1}$ and $q_{t,i}^{e1}$, respectively.
Likewise, $q_{t,i}^{REC1.pre}$ is the self-consumption from previous agents up to the i-th one excluded, in the first round (as defined in Constraint (2.11i), and $q_{t,i}^{REC1}$ is the self-consumption of the i-th agent in round 1 (its definition holds as in Section 2.2.2).
Furthermore, the load of previous agents is represented by $d_{t,i}^{b1.pre}$ (see Constraint (2.11f) while the load of the i-th agent is labelled with $d_{t,i}^{b1}$. Other constraints work as explained in Section 2.2.2 "Incremental commitment hypothesis". The mathematical formulation, with $\Xi = Q_i^{g1}, Q_i^{e1}, q_{t,i}^{g1}, q_{t,i}^{e1}$ and with only the constraints that differ from Problem (2.3),

is the following:

$$\max_{\Xi} \pi^b = \sum_{t=1}^{\bar{T}} \left(p_t^g \left(q_{t,i}^{g1.pre} + q_{t,i}^{g1} \right) + p_t^{me} \left(q_{t,i}^{e1.pre} + q_{t,i}^{e1} \right) + z\delta \left(q_{t,i}^{REC1.pre} + q_{t,i}^{REC1} \right) + \right.$$
$$\left. - (p_t^{me} + k) \left(d_{t,i}^{b1.pre} + d_{t,i}^{b1} \right) \right) - a \left(C^g + p^{qg} \left(Q_i^{g1.pre} + Q_i^{g1} \right) + p^{qe} \left(Q_i^{e1.pre} + Q_i^{e1} \right) \right) \tag{2.11a}$$

s.t.

$$Q_i^{g1.pre} = \sum_{c=1}^{i-1} Q_c^{g1} \quad \forall i \tag{2.11b}$$

$$Q_i^{e1.pre} = \sum_{c=1}^{i-1} Q_c^{e1} \quad \forall i \tag{2.11c}$$

$$q_{t,i}^{g1.pre} = \sum_{c=1}^{i-1} q_{t,i}^{g1} \quad \forall i,t \tag{2.11d}$$

$$q_{t,i}^{e1.pre} = \sum_{c=1}^{i-1} q_{t,c}^{e1} \quad \forall i,t \tag{2.11e}$$

$$d_{t,i}^{b1.pre} = \sum_{c=1}^{i-1} d_{t,c}^{b1} \quad \forall i,t \tag{2.11f}$$

$$x = \sum_{c=1}^{i-1} q_{t,c}^{e1.pre} + \sum_{c=1}^{j-1} q_{t,c}^{pv1.pre} \quad \forall i,j,t \tag{2.11g}$$

$$y = \sum_{c=1}^{i-1} d_{t,c}^{b1.pre} + \sum_{c=1}^{j-1} d_{t,c}^{h1.pre} \quad \forall i,j,t \tag{2.11h}$$

$$q_{t,i}^{REC1.pre} = x \cdot \mathbb{1}_{\{x \leq y\}} + y \cdot \mathbb{1}_{\{x > y\}} \tag{2.11i}$$

$$v = q_{t,i}^{e1} + q_{t,j}^{pv1} \quad \forall i,j,t \tag{2.11j}$$

$$w = d_{t,i}^{b1} + d_{t,j}^{h1} \quad \forall i,j,t \tag{2.11k}$$

$$q_{t,i}^{REC1} = v \cdot \mathbb{1}_{\{v \leq w\}} + w \cdot \mathbb{1}_{\{v > w\}} \tag{2.11l}$$

In Constraint (2.11i) the influence of household choices is evident: self-consumption is influenced not only by the loads and productions of other biogas producers but also by the loads and productions of households. This is the key factor that is used to represent the interaction among heterogeneous agents.

Household After the biogas optimization, it is the household's turn to optimize his profits given what has been decided by previous agents (both biogas producers and households). The household takes into consideration the installation of only one technology, the photovoltaic, and his profits, labelled with π_j^h, are defined as in Problem (2.4).
Similarly to the biogas producer problem, the variables $Q_j^{pv.pre}$ and $q_{t,j}^{pv1.pre}$ are the capacity of PV plant installed and the quantity of electricity sold on the spot market, respectively, up to the j-th agent excluded, in the first round (as defined in Constraints (2.12b and (2.12c, respectively). Consequently Q_j^{pv1} and $q_{t,j}^{pv1}$ are the capacity and the quantity sold, respectively, of the j-th agent in round 1.
Variables $q_{t,j}^{REC1.pre}$, $q_{t,j}^{REC1}$ and the loads, $d_{t,j}^{h1}$ and $d_{t,j}^{h1.pre}$, work as in Problem (2.11) but index i is replaced by j, and b is replaced by h, to represent the households. The mathematical formulation, with only constraints that differ from Problem (2.4) are stated, is

the following:

$$\max_{Q_j^{pv1}, q_{t,j}^{pv1}} \pi^h = \sum_{t=1}^{\bar{T}} \Bigg(p_t^{me} \left(q_{t,j}^{pv1.pre} + q_{t,j}^{pv1} \right) + z(1-\delta) \left(q_{t,j}^{REC1.pre} + q_{t,j}^{REC1} \right) + \\ - (p_t^{me} + k) \left(d_{t,j}^{h1.pre} + d_{t,j}^{h1} \right) \Bigg) - b \left(p^{pv} \left(Q_j^{pv1.pre} + Q_j^{pv1} \right) \right) \tag{2.12a}$$

s.t.

$$Q_j^{pv1.pre} = \sum_{c=1}^{j-1} Q_c^{pv1} \quad \forall j \tag{2.12b}$$

$$q_{t,j}^{pv1.pre} = \sum_{c=1}^{j-1} q_{t,c}^{pv1} \quad \forall j, t \tag{2.12c}$$

Round 2 At the beginning of the second round, each agent possesses full knowledge of the decisions made by all agents in the first round. Within this context, each agent, equipped with the information set, has the opportunity to either validate or revise his previous choice. If an agent opts to alter his prior decision, the round advances, and the information set available to the subsequent agent is adjusted to reflect the latest choice. The second round completes when an agent confirms his initial decision from the first round, as no further iterations are needed, given the invariability of the optimization problem for ensuing agents. It is worth noting that the iteration may extend beyond the second round because if all agents change their previous choice, the second round ends and a third one begins, and so forth. However, for the sake of simplicity, this analysis confines itself to the second round, as any subsequent rounds would adhere to the same structural framework. Upon completion of the decision-making process, effective installations ensue. The problem's structure is coherent with the explanations in Problem (2.11) and (2.12) but some adjustments are necessary. First, the number 2 identifies the capacities and quantities related to the second round. Second, the quantities labelled with *pre* are influenced by the iterations that occurred in the first round and by the choices taken in the second round before the current agent.

Biogas This description is limited to constraints that differ from Problem (2.11). In Constraint (2.13b) the value of $Q_i^{e2.pre}$ is defined as the sum between all the capacities chosen in round one ($\sum_{i=1}^{\bar{I}} Q_i^{e1}$) and all the capacities confirmed in the iterations of the second round, up to the i-th biogas producer excluded ($\sum_{c=1}^{i-1} Q_c^{e2.pre}$). The same holds for $Q_i^{g2.pre}$ as explained in Constraint (2.13c).
Furthermore, Q_i^{e2} and Q_i^{g2} are influenced by the quantities defined in Problem (2.11). This relationship is explained in Constraints (2.13g) and (2.13h) because Q_i^{e1} and Q_i^{g1} are the lower bounds of Q_i^{e2} and Q_i^{g2}, respectively (i.e. agent i can not divest more than the optimal quantity chosen before). The same holds for the quantity produced $q_{t,i}^{e2}$ which is constrained by $-q_{t,i}^{e1}$, as the lower physical limit to respect what was stated before, and by Q_i^{e2} as the physical upper limit.
The mathematical formulation is the following:

$$\max_{Q_i^{e2},Q_i^{g2},q_{t,i}^{e2},q_{t,i}^{g2}} \pi^b = \sum_{t=1}^{\bar{T}} \left(p_t^g \left(q_{t,i}^{g2.pre} + q_{t,i}^{g2} \right) + p_t^{me} \left(q_{t,i}^{e2.pre} + q_{t,i}^{e2} \right) + z\delta \left(q_{t,i}^{REC2.pre} + q_{t,i}^{REC2} \right) + \right.$$
$$\left. - \left(p_t^{me} + k \right) d_t^{b2} \right) - a \left(C^g + p^{qg} \left(Q_i^{g2.pre} + Q_i^{g2} \right) + p^{qe} \left(Q_i^{e2.pre} + Q_i^{e2} \right) \right) \tag{2.13a}$$

s.t.

$$Q_i^{e2.pre} = \sum_{i=1}^{\bar{I}} Q_i^{e1} + \sum_{c=1}^{i-1} Q_c^{e2.pre} \tag{2.13b}$$

$$Q_i^{g2.pre} = \sum_{i=1}^{\bar{I}} Q_i^{g1} + \sum_{c=1}^{i-1} Q_c^{g2.pre} \tag{2.13c}$$

$$q_{t,i}^{e2.pre} = \sum_{i=1}^{\bar{I}} q_i^{e1} + \sum_{c=1}^{i-1} q_{t,c}^{e2.pre} \quad \forall t \tag{2.13d}$$

$$q_{t,i}^{g2.pre} = \sum_{i=1}^{\bar{I}} q_i^{g1} + \sum_{c=1}^{i-1} q_{t,c}^{g2.pre} \quad \forall t \tag{2.13e}$$

$$d_t^{b2} = \sum_{i=1}^{\bar{I}} d_i^{b1} \tag{2.13f}$$

$$- Q_i^{e1} \leq Q_i^{e2} \leq Q_i^{g2} m^e \quad \forall i \tag{2.13g}$$

$$- Q_i^{g1} \leq Q_i^{g2} \quad \forall i \tag{2.13h}$$

$$- q_{t,i}^{e1} \leq q_{t,i}^{e2} = Q_i^{e2} \mathbb{1}_{p_t^{me} > p_t^g m^e} \leq Q_i^{e2} \quad \forall t, i \tag{2.13i}$$

$$q_{t,i}^{g2} = \left(Q_i^{g2} - q_{t,i}^{e2}/m^e \right) \mathbb{1}_{p_t^{me} \leq p_t^g m^e} \quad \forall t, i \tag{2.13j}$$

Household The following description is limited to constraints that differ from Problem (2.12). In constraint (2.14b) the value of $Q_j^{pv2.pre}$ is defined as the sum between all the capacities chosen in round one ($\sum_{j=1}^{\bar{J}} Q_j^{pv1}$) and the capacities confirmed in previous iterations in this round ($\sum_{c=1}^{j-1} Q_c^{pv2}$). Furthermore, as explained in (2.14e), Q_j^{pv2} is strongly related to the quantity Q_j^{pv1} defined in Round 1 in Problem (2.12) because Q_j^{pv1} is the lower bound of Q_j^{pv2}. Indeed, agent j cannot divest more than the optimal quantity stated before. The same holds for the quantity produced $q_{t,j}^{pv2}$ which is constrained by $-q_{t,j}^{pv1}$, to respect what was stated before, and by Q_j^{pv2} as the physical upper limit. Other constraints work as before in Problem (2.12). The mathematical formulation is the following:

$$\max_{Q_j^{pv2}, q_{t,j}^{pv2}} \pi^h = \sum_{t=1}^{\bar{T}} \left(p_t^{me} \left(q_{t,j}^{pv2.pre} + q_{t,j}^{pv2} \right) + z(1-\delta) \left(q_{t,j}^{REC2.pre} + q_{t,j}^{REC2} \right) - (p_t^{me} + k)\, d_t^{h2} \right) +$$
$$- b \left(p^{pv} \left(Q_j^{pv2.pre} + Q_j^{pv2} \right) \right) \tag{2.14a}$$

s.t.

$$Q_j^{pv2.pre} = \sum_{j=1}^{\bar{J}} Q_j^{pv1} + \sum_{c=1}^{j-1} Q_c^{pv2} \tag{2.14b}$$

$$q_{t,j}^{pv2.pre} = \sum_{j=1}^{\bar{J}} q_{t,j}^{pv1} + \sum_{c=1}^{j-1} q_{t,c}^{pv2} \quad \forall t, j \tag{2.14c}$$

$$d_{t,j}^{h2} = \sum_{J}^{j-1} d_{t,j}^{h1} \quad \forall t \tag{2.14d}$$

$$- Q_j^{pv1} \leq Q_j^{pv2} \quad \forall j \tag{2.14e}$$

$$- q_{t,j}^{pv1} \leq q_{t,j}^{pv2} = Q_j^{pv2} \gamma_t^{pv} \leq Q_j^{pv2} \quad \forall t, j \tag{2.14f}$$

2.3 Application and results

2.3.1 Application

The two versions of the model, "Incremental commitment hypothesis" and "Rethink hypothesis (pure Nash equilibrium)" are applied separately. Furthermore, each problem is run with different iteration orders. First, the "Incremental commitment hypothesis" with the biogas producer acting first is executed. Then, the iteration is repeated but with the household starting first. In "Rethink hypothesis (pure Nash equilibrium)" the iteration order alternates the two kinds of agents: for example, if the first one is a biogas producer, then the second one will be a household, and so forth; vice versa, if the first one is a household, then the second one will be a biogas producer, and so forth.
In both versions, the optimization is performed on 24 hours ($T = 24$) because this time frame aligns with the natural daily cycles of processes, making it highly relevant to real-world scenarios. By concentrating on a single day, the challenges faced within this time frame are directly addressed, such as energy usage patterns, load and solar radiation cycle. Moreover, a 24-hour horizon ensures that the solutions generated are both feasible and realistic. Daily optimization also enhances efficiency by making the optimization of a REC life more manageable while still capturing the essential dynamics of the system.
To apply the first version, explained in Section "Incremental commitment hypothesis", 1 REC composed of 1 household and 1 biogas producer is considered. In this version, the household load is set equal to 0.04 MWh. This value is then distributed over the 24 hours by multiplying it for the following load pattern: the electricity load is set equal to 10% from midnight to 5:00 am (i.e., load related to home appliances), 70% from 5:00 am to 7:00 am (i.e. people start to get up), 100% from 7:00 am to 9:00 pm (i.e., normal living to take into account households, elderly people, smart workers and use of electricity to cook for lunches and dinners), 80% from 9:00 pm to 11:00 pm (i.e. TVs, lights, computers, washing machines) and 20% from 11:00 pm to midnight (i.e. only few people are still awake, home appliances). Biogas load is set to 0.05 MWh and is assumed constant over the 24 hours.
For the second version, explained in Section "Rethink hypothesis (pure Nash equilibrium)",

the individual household electricity load is set equal to 0.008 MWh and an individual biogas electricity load is assumed equal to 0.01 MWh. The individual household load is distributed over 24 hours with the same pattern explained above. Furthermore, in this model, the threshold chosen to close the first round is load: the first round is complete when the REC load is covered by REC members' plant production.
Table 2.1 lists all parameters used in both versions and their values: in particular the turbine, biodigestor and fixed installation costs are already discounted by a rate of 6% over a lifetime of 15 years, while the price of the photovoltaic plant is discounted by a rate of 8% over 20 years lifetime.
Figures 2.5 and 2.6 collect the parameters that range over the time horizon of 24 hours. In Figure 2.5 the price of electricity (M€/MWh)[9], the solar radiation of an average day in Italy (that is the photovoltaic efficiency, in %), and the biogas price on the spot market (M€/m^3)[10] are represented. Figure 2.6 shows the average loads of households and biogas producers for both versions of the model.

Table 2.1: This table lists the parameters, their values, the unit of measure, their definition and source. This list is common to both versions of the model, any differences between the two versions are highlighted.

Parameter	Unit of measure	Value	Definition and source
T	hours	24	Optimization time horizon
z	€/KWh	0.110	Incentive on self-consumption if member of a REC[11]
p^{pv}	M€/MW	0.7	Price of photovoltaic plant
B^h	M€	0.30	Household budget
B^h_j	M€	0.05	Individual household budget
p^{qe}	M€/MW	0.685	Price of the gas-to-electricity turbine
p^{qg}	M€/m^3	13.699	Price of the biodigestor
C^g	M€	0.001	Fixed installation cost
B^b	M€	0.36	Biogas producer budget
B^b_i	M€	0.13	Individual biogas producer budget
m^e	MWh/m^3	0.01055	Ratio that converts cubic metres into MW[12]

2.3.2 "Incremental commitment hypothesis" results

The results of the first version of the model are presented: the discussion is divided into two parts to distinguish the two iteration orders: when "Biogas producer acts first" and when "Household acts first". Table 2.2 lists the sensitivity analyses common to both iterations.

Biogas producer acts first Let us start by analyzing the impact of δ on the decisions of biogas producer and household (see Figure 2.7). For $\delta \leq 0.5$, the biogas producer receives an incentive that is not sufficient to make Q^e convenient and he installs only

[9]Gestore del Mercato Elettrico (GME), `https://www.mercatoelettrico.org/It/download/DatiStorici.aspx`, day-ahead market single national price (PUN, Prezzo Unico Nazionale)

[10]Gestore del Mercato Elettrico (GME), `https://www.mercatoelettrico.org/It/download/DatiStoriciGas.aspx`, Single National Price established in the Virtual Exchange Point (PSV, Punto di Scambio Virtuale)

[11]Ministry of the Environment and Energy Security Decree of 16 September 2020, "Individuazione della tariffa incentivante per la remunerazione degli impianti a fonti rinnovabili inseriti nelle configurazioni sperimentali di autoconsumo collettivo e comunità energetiche rinnovabili, in attuazione dell'articolo 42-bis, comma 9, del decreto-legge n. 162/2019, convertito dalla legge n. 8/2020."

[12]The conversion chart is available at `https://learnmetrics.com/m3-gas-to-kwh/`.

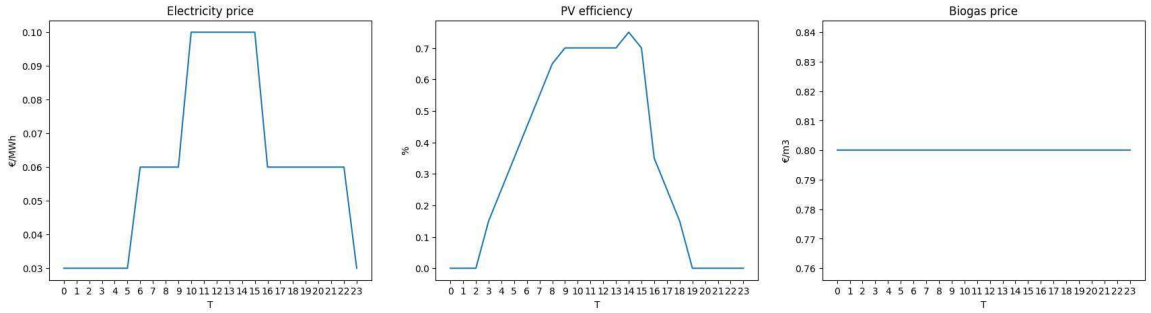

Figure 2.5: Plots, over 24 hours, of electricity prices p_t^{me} (in €/MWh), solar radiation efficiency γ_t^{pv} (in %), and biogas prices p_t^g (in €/m^3).

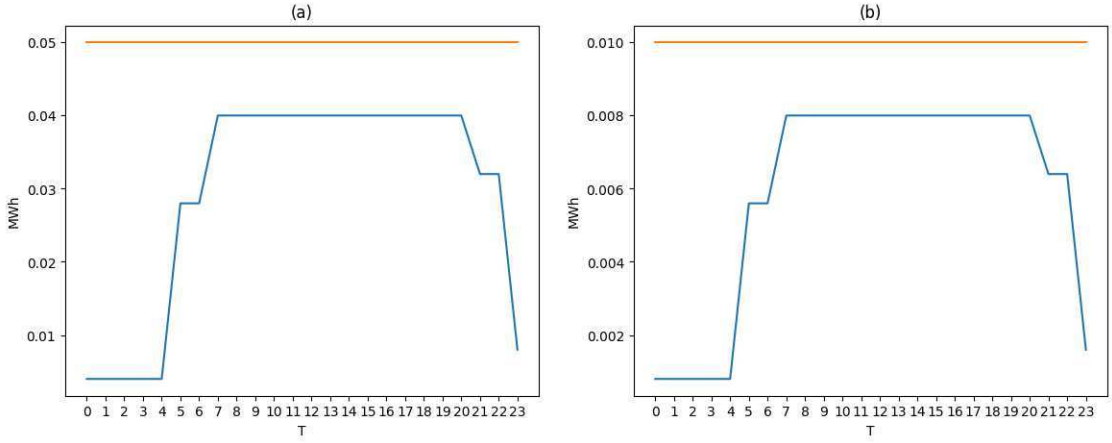

Figure 2.6: Plots of household (blue) and biogas (orange) loads in Incremental commitment hypothesis (panel (a)) and in Rethink hypothesis (pure Nash equilibrium) (panel (b)).

Parameter		Reference value
z	$0.055 - 0.1705$	0.110 €/KWh
p^{qg}	$6.8493 - 21.2329$	13.698630 M€/m^3
p^{qe}	$0.3425 - 1.0616$	0.684932 M€/MW
p^{pv}	$0.35 - 1.085$	0.7 M€/MW

Table 2.2: This table summarizes the sensitivity analysis carried out on the results in the "Incremental commitment hypothesis".

Q^g. At the same time, when $\delta \leq 0.3$ the household installs Q^{pv} without spending all his budget (the investment cost is 0.18M€ while the available budget is 0.30M€). He does not spend all his budget because with $Q^{pv} = 0.257143$ he produces more than the demand level for 12 hours (from 04:00 am to 04:00 pm) but, during the other 12 hours (from midnight to 04:00 am and from 04:00 pm to midnight) the production is lower than or equal to the demand. Therefore, further expenses in Q^{pv} would not be compensated by the revenues, obtained from selling the electricity, because the incentive is received on the quantity up to the demand level for half of the day (12 hours over 24). When $\delta \in [0.4, 0.5]$ the household decreases the PV installation because he obtains a lower share of incentive. However, the level of installation is still higher than the demand but he produces more than the demand level only for 10 hours per day, improving his position with respect to the previous situation. Indeed, the household is forced to install Q^{pv} to gain the incentive because the biogas producer does not install Q^e. When

$\delta = 0.6$ the situation changes. The biogas producer finds the installation of Q^e convenient. Given that the budget constraint is active (i.e. he is spending all his budget for Q^g) there is a substitution effect: the biogas producer has to decrease Q^g to have money available for Q^e. However, he does not install the maximum Q^e because the incentive is still high enough for the household to install Q^{pv}. The investment in Q^e is coherent with self-consumption: the production from the turbine (assumed to be set equal to the capacity installed, i.e. the plant works at maximum capacity every hour) exactly covers the demand of REC members. This is the first proof that the method allows to avoid over-investments. At this level of δ the profit functions show a non-monotonic trend. In particular, the biogas producer has a fall because the huge investment in Q^e and the consequent revenues, together with the incentive, do not compensate for the lower revenues from biogas sales. Conversely, the household has a peak because the lower investment costs, together with the incentive, compensate for the lower revenues from selling electricity. With $\delta \geq 0.7$ it is convenient for the biogas producer to install the maximum possible Q^e capacity. Therefore, Q^g is maintained to allow a sufficient level of Q^e that maximizes self-consumption. Consequently, the household does not find the installation of Q^{pv} convenient and it falls to zero. For any level of δ, household's profits are always negative. However, he installs the photovoltaic plant because the revenues from selling electricity and the incentive generated from self-consumption allow the household to partially counterbalance the installation costs. Furthermore, the household installation favours the generation of the self-consumption incentive, which is gained by both agents.

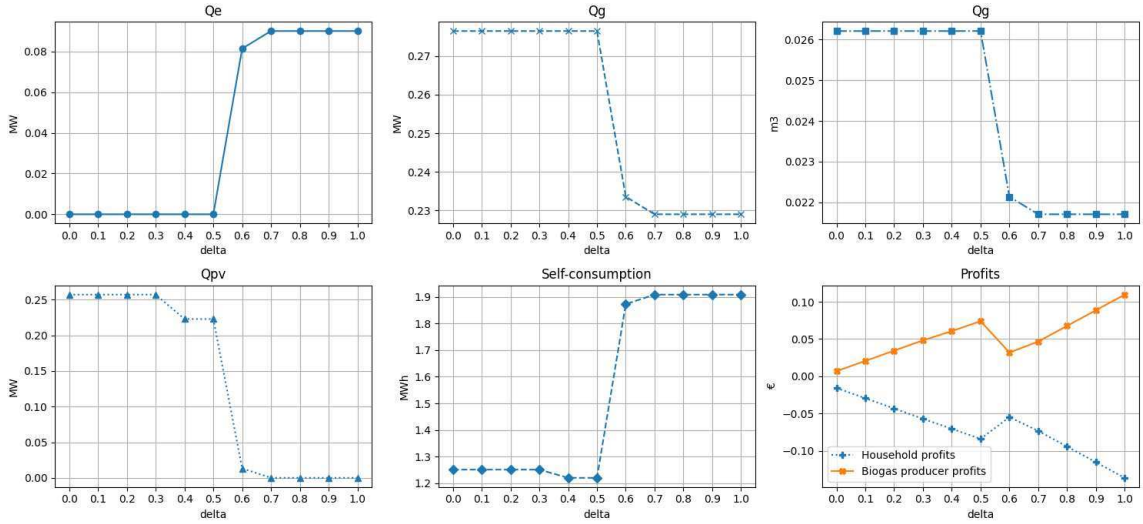

Figure 2.7: Q^e (in MW), Q^g (in m^3 and MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of δ when the biogas producer acts first in "Incremental commitment hypothesis".

The sensitivity analysis of the intricate relationships among variables by analyzing the sensitivity analysis on some key parameters, in particular the installation costs of the technologies and the incentive, follows.

Sensitivity analysis on incentive, z Figure 2.8 collects the results for different values of z. As a general comment, all graphs follow a rational trend but, three exceptions are worth mentioning. First, with $z = 0.055$ and $\delta = 0.9$ the self-consumption falls to zero (labelled as point 1.). This is due to two reasons: the household does not install Q^{pv} because, with a so high δ, the installation is not convenient anymore; and the biogas producer receives a too low z which makes Q^e not convenient, therefore he installs only

Q^g. Second, with $z = 0.17$ and with $\delta = 0.3$ the household profit has a peak (labelled as point 2.) due to the non-investment in Q^{pv} (i.e. no installation costs), while Q^e is maximized causing a maximization in self-consumption levels. As a consequence, as before, the lower installation costs, together with the incentive, exceed the lost revenues. Third, the falls in biogas profits (labelled as point 3.) are coherent with the different moments where Q^e occurs (and the simultaneous decrease of Q^g). Q^e generates revenues that, together with the incentive, do not compensate for the losses from selling biogas.

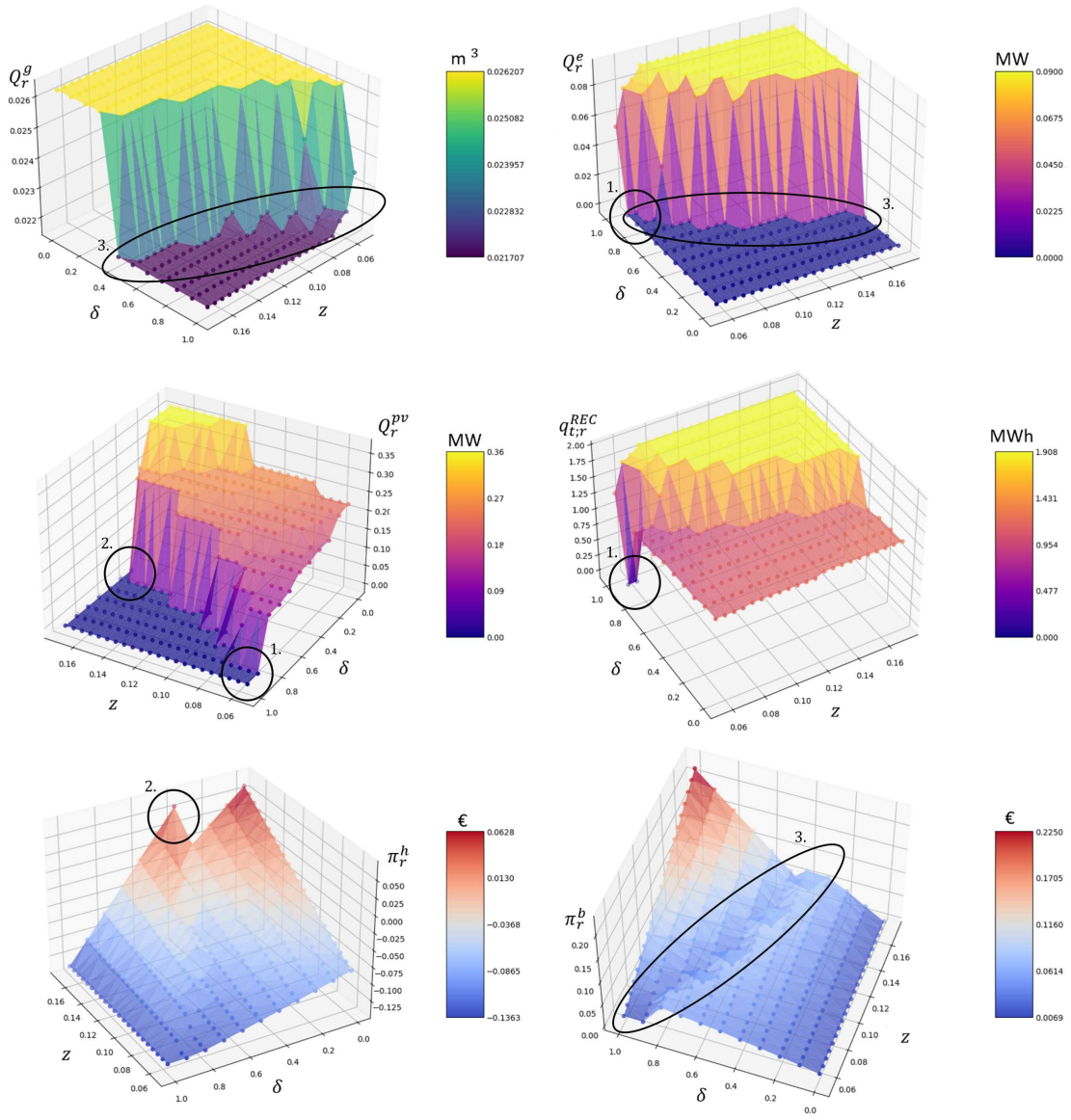

Figure 2.8: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of z (and δ) when the biogas producer acts first in "Incremental commitment hypothesis".

Sensitivity analysis on biodigestor installation cost, p^{qg} Figure 2.9 shows the sensitivity analysis on p^{qg}. It highlights the strong relationship and the consequent substitution effect between Q^g and Q^e. In particular, when $p^{qg} \geq 19.863$ and $\delta \geq 0.7$, (labelled as point 1.) the biogas producer installs enough Q^g to install Q^e: the biogas producer finds convenient Q^g, despite the high p^{qg}, because it is compensated by the revenues obtained from selling electricity and by the incentive. Furthermore, with $p^{qg} \geq 18.4932$ and $\delta = 0.5$

there is a peak in π^h (labelled as point 2.). This is due to the incentive and null installation costs that compensate for the lost revenues from selling electricity.

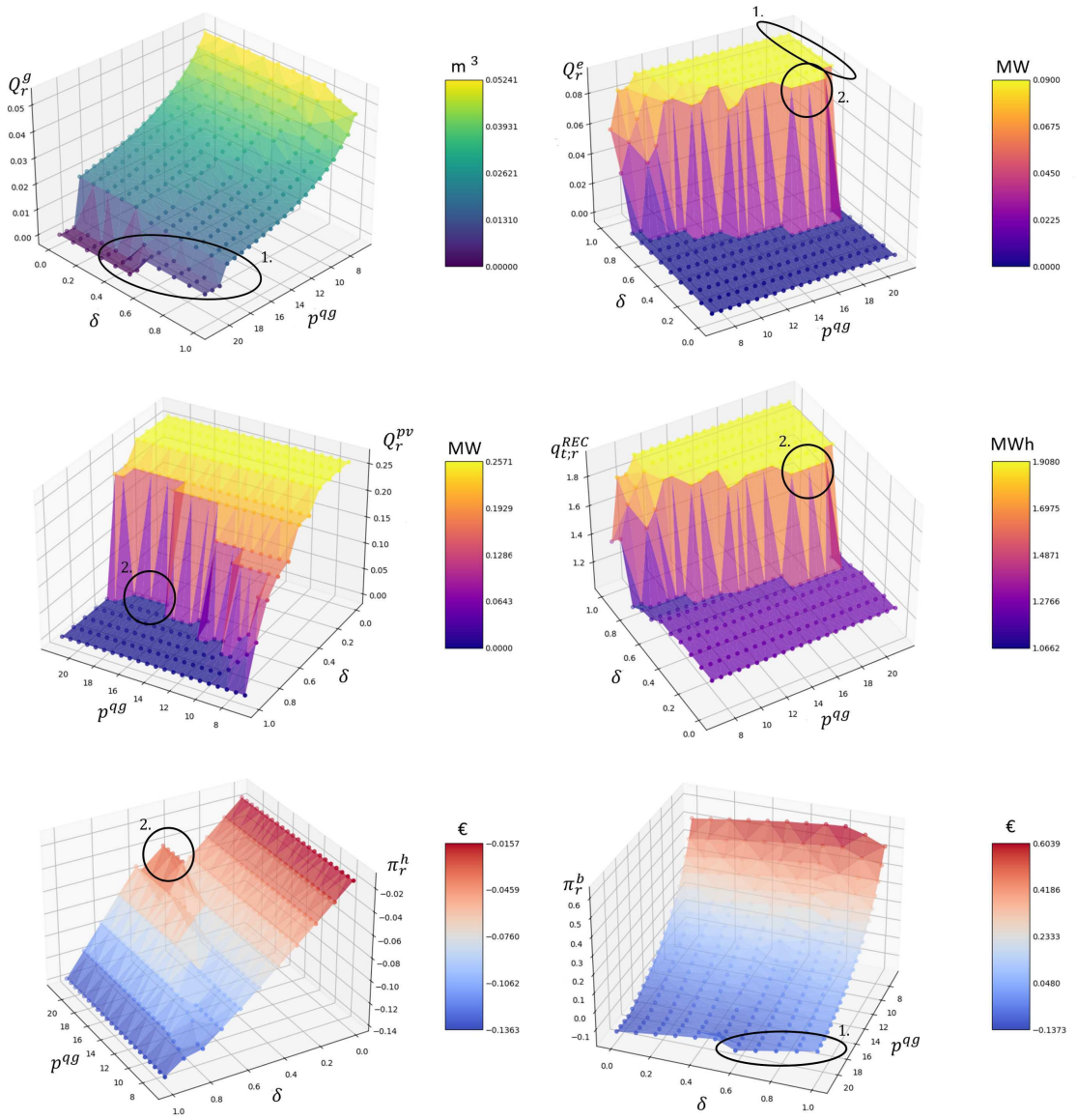

Figure 2.9: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qg} (and δ) when the biogas producer acts first in "Incremental commitment hypothesis".

Sensitivity analysis on turbine installation cost, p^{qe} Figure 2.10 shows the impact of p^{qe} on Q^g, Q^e, Q^{pv}, self-consumption and both profit functions. As a general comment, as p^{qe} increases, higher levels of δ are required to make Q^e convenient. Consequently, Q^g and Q^{pv} decrease accordingly. Q^e has a major impact on self-consumption. In particular, when $p^{qe} = 0.3425$ and $\delta = 0.3$ there is a non-monotonic trend in self-consumption (labelled as point 1.). This non-monotonic trend is due to the maximized Q^e, with the consequent non-installation of Q^{pv}. However, Q^e does not produce for all 24 hours: up to 06:00 am and at 11:00 pm the biogas producer finds more convenient the production and sale of biogas. This time frame coincides with low demand levels that impact on self-consumption computation. Conversely, from 06:00 am to 11:00 pm the biogas producer finds it convenient to transform biogas into electricity, which accrues to self-consumption

and results in a total lower level. In this context, the biogas producer faces a drop in his profits because the investment in Q^e and the following revenues from selling electricity, together with the incentive, do not compensate for the lower revenues from selling biogas. Conversely, the household has a peak in its profits because, as before, the lower investment costs, together with the incentive, exceed the lost revenues. Furthermore, as p^{qe} increases, higher values of delta postpone the shift in biogas producer profits but do not eliminate it (labelled as point 2.).

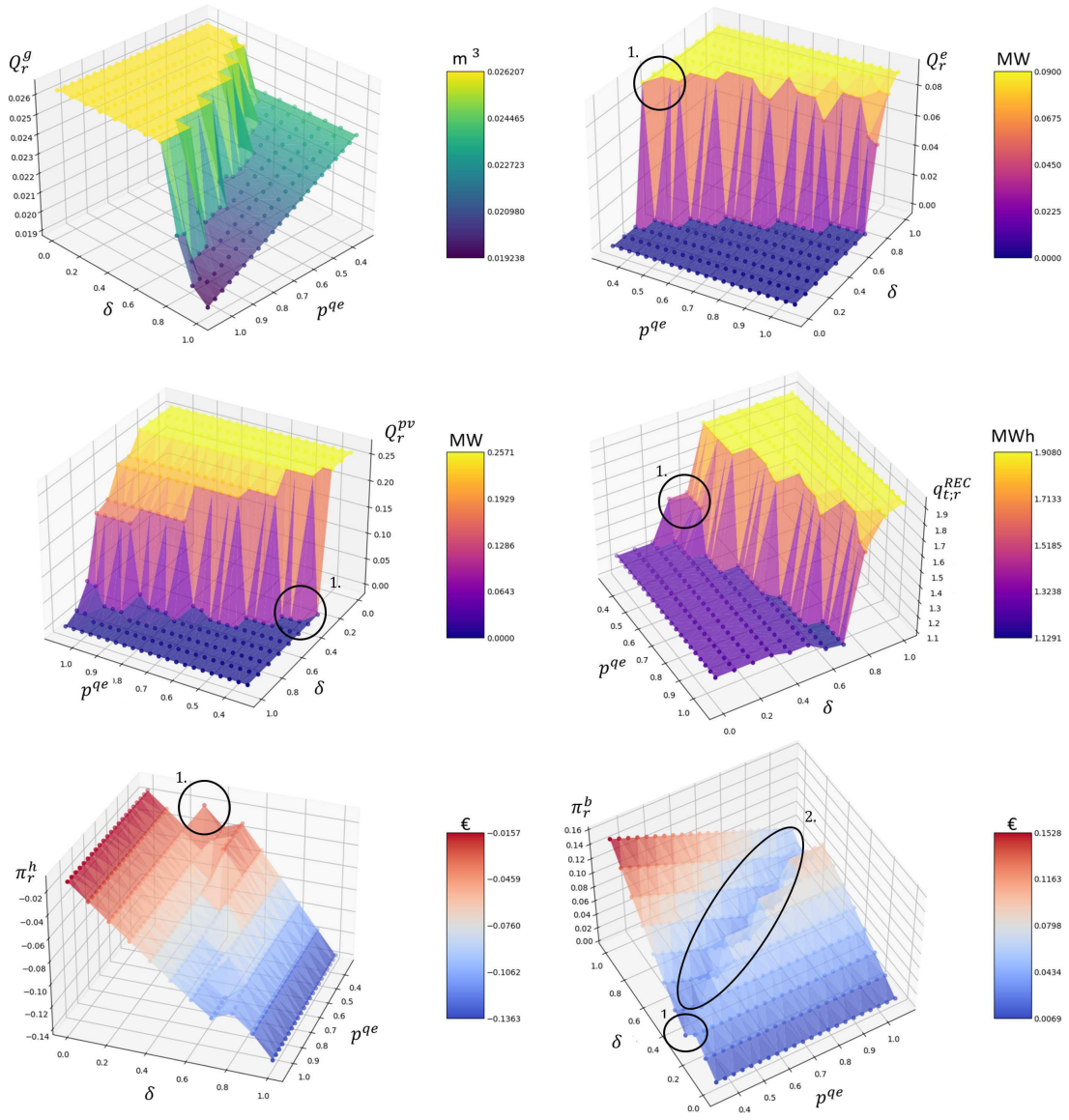

Figure 2.10: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qe} (and δ) when the biogas producer acts first in "Incremental commitment hypothesis".

Sensitivity analysis on photovoltaic installation cost, p^{pv} Figure 2.11 collects the impacts of p^{pv} for every value of δ. First, when $p^{pv} \leq 0.63$ Q^{pv} is convenient for every δ(labelled as point 1.). Coherently, $Q^e = 0$ (only Q^g is installed) and self-consumption reaches intermediate levels. However, when $p^{pv} > 0.63$ a distinction is necessary: for $\delta \leq 0.5$ (labelled as point 2.) Q^{pv} is convenient (Q^e is not convenient and is not installed) but self-consumption decreases, causing a decrease in household profits; for $\delta > 0.5$ (labelled as point 3.) Q^{pv}

is not convenient anymore (it is not installed) and Q^g decreases for the substitution effect with Q^e, which is installed and maximizes self-consumption. Consequently, household profits increase slightly but they decrease due to the small δ. Coherently with Q^g and Q^e, biogas profits are not harmed when $p^{pv} \leq 0.63 \; \forall \; \delta$ or when $p^{pv} > 0.63$ (for $\delta \leq 0.5$). However, when Q^e occurs, there is a shift in biogas profits because the investment and the lost revenues from biogas are not covered by electricity revenues and incentive.

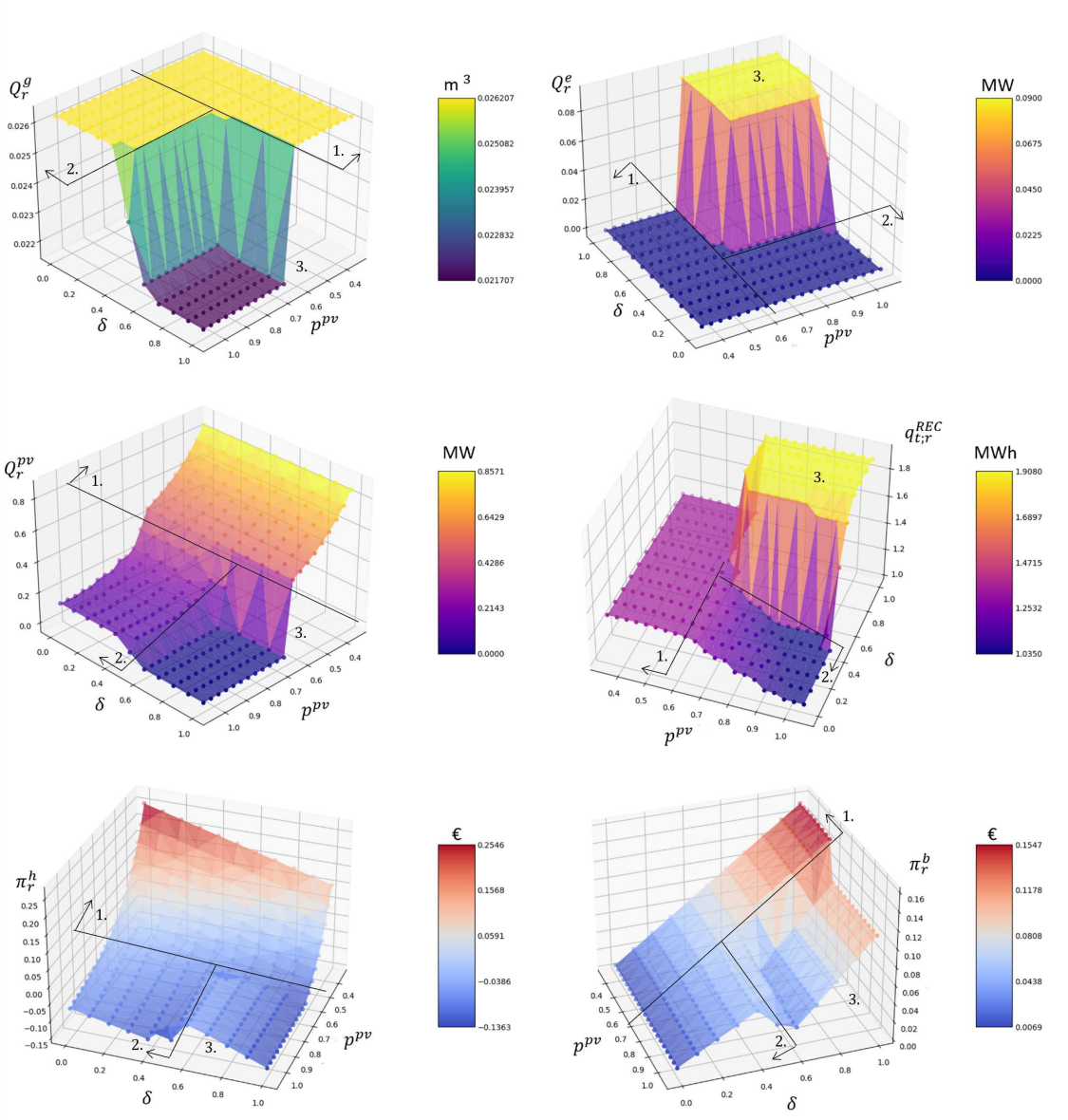

Figure 2.11: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{pv} (and δ) when the biogas producer acts first in "Incremental commitment hypothesis".

Household acts first Let us start by analyzing the impact of δ on the decisions of biogas producer and household (see Figure 2.12). With $\delta \leq 0.5$ the situation mirrors the results of the previous paragraph, when the biogas is acting first. With $\delta \in [0.6, 0.7]$, given that the household receives $1 - \delta$, he decreases Q^{pv}. Consequently, the self-consumption decreases. However, the biogas producer finds it not convenient to install Q^e, therefore, only Q^g is installed. The situation changes when $\delta \geq 0.8$: the household further decreases Q^{pv}. However, Q^e is now convenient: the biogas producer starts to install it at $\delta = 0.8$ and increases it, up to $\delta = 1$. Simultaneously, he decreases Q^g because of the substitution effect between the technologies. Q^e increases the self-consumption. This substitution effect among the three technologies has an impact on both profits: with $\delta \in [0.8, 0.9]$, after a monotonic decrease, there is a less negative marginal profit for the household; while the biogas, after a monotonic increase, faces a less positive marginal profit. The household faces lower investment costs that, together with the incentive, compensate for the lower revenues. Conversely, the biogas producer faces an increase in revenues for selling electricity and for the incentive that compensates almost exactly the lower revenues from selling biogas. Furthermore, when $\delta = 1$ the lost revenues and the absence of incentive cause a more negative marginal profit in household profit, while the biogas producer has a higher marginal positive profit because revenues from selling electricity and the incentive exceed the lost revenues from biogas.For any level of δ, household's profits are always negative. However, he installs the photovoltaic plant because the revenues from selling electricity and the incentive generated from self-consumption allow the household to partially counterbalance the installation costs. Furthermore, the household installation favours the generation of the self-consumption incentive, which is gained by both agents.

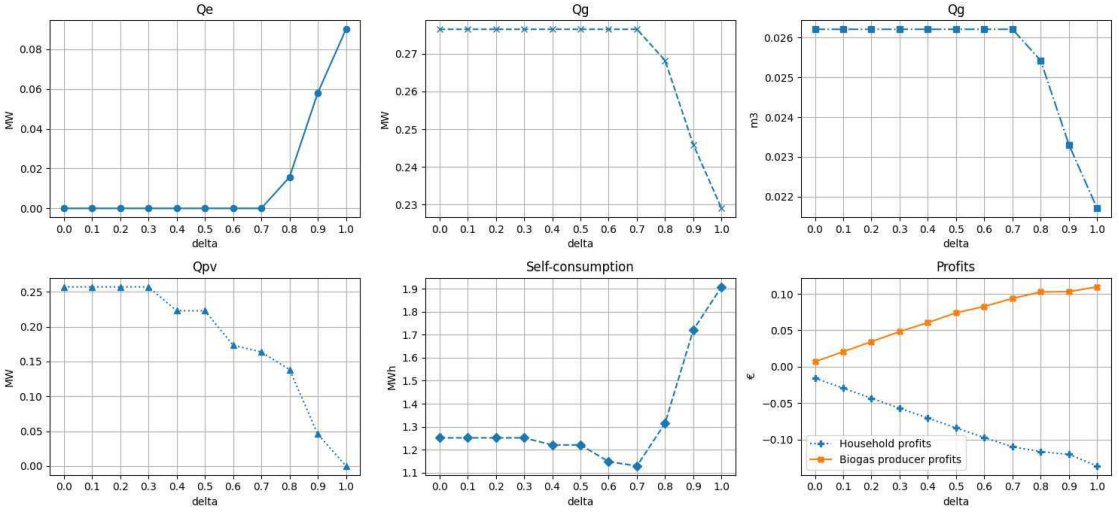

Figure 2.12: Q^e (in MW), Q^g (in m^3 and MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of δ when the household acts first in "Incremental commitment hypothesis".

Sensitivity analysis on incentive, z Figure 2.13 shows the impact of changing z. In particular, when $\delta = 0.9$ and $z = 0.055$ (labelled as point 1.) q_t^{REC} is zero because, given the low incentive, both Q^{pv} and Q^e are not convenient. Consequently, both profits show a decrease. Furthermore, with $\delta = 0.9$, when $z \in [0.0825, 0.1265]$ (labelled as point 2.) there is an inversion of trend in the technologies installation: the household finds convenient to install Q^{pv}, consequently Q^e decreases and the biogas producer invests in Q^g.

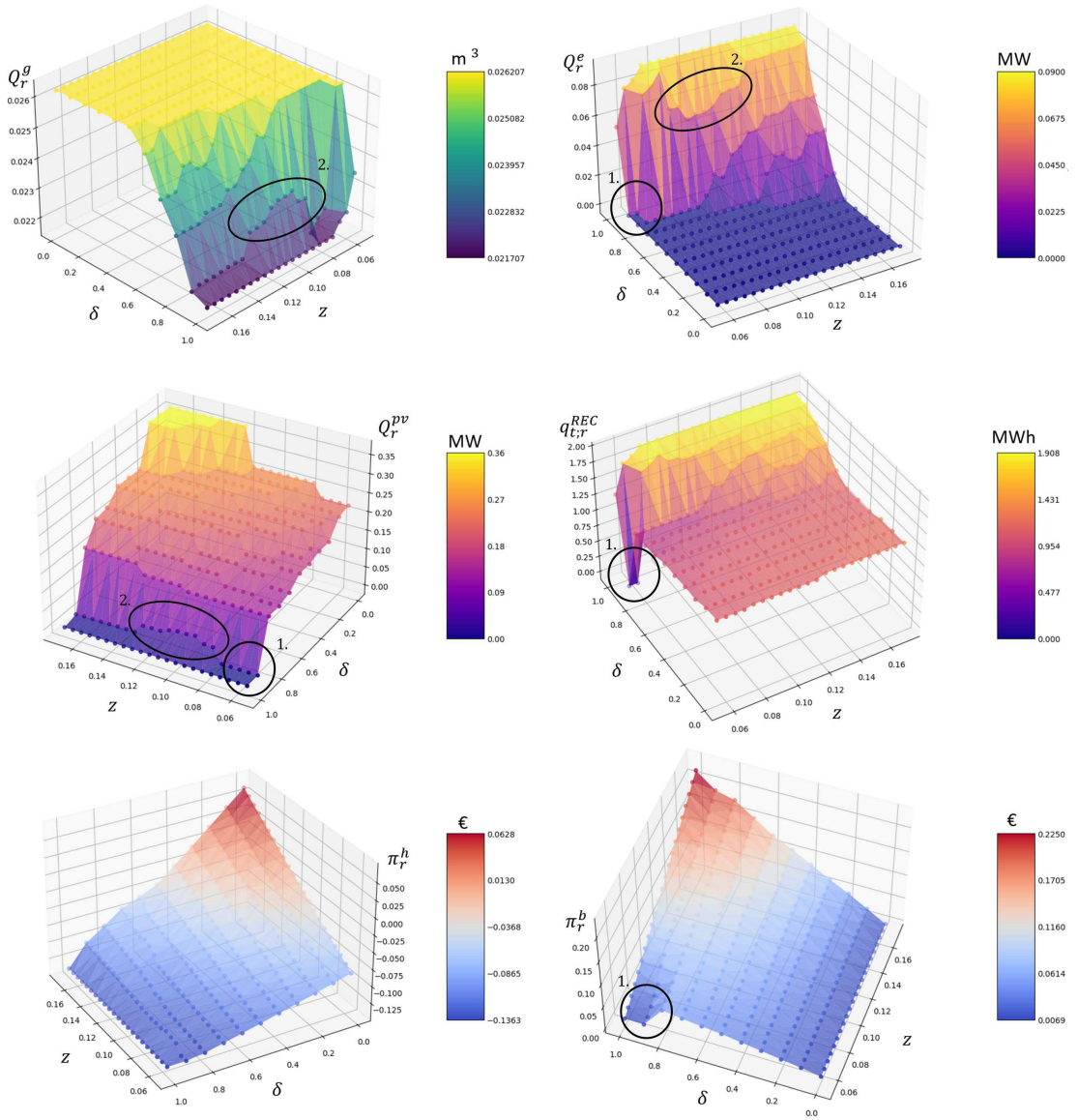

Figure 2.13: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of z (and δ) when the household acts first in "Incremental commitment hypothesis".

Sensitivity analysis on biodigestor installation cost, p^{qg} For this analysis, by looking at Figure 2.14, it is worth mentioning that, for high values of p^{qg} and $\delta \geq 0.7$ (labelled as point 1.), Q^g is installed to allow the installation of Q^e. In other words, the cost of installing Q^g is compensated by the benefits from installing Q^e, i.e. the incentive from self-consumption and the revenues from selling electricity. Furthermore, with $\delta = 0.9$, p^{qg} affects Q^{pv} (labelled as point 2.): when $p^{qg} \leq 9.5890$ the installation of Q^g is convenient (Q^e is not installed) and Q^{pv} is installed to gain a minimum quantity of self-consumption; for higher values of p^{qg}, Q^g and Q^{pv} decrease proportionally to increases of Q^e. However, with $p^{qg} \geq 20.5479$, on one hand Q^g results too expensive and is divested, with Q^e being divested accordingly, but on the other hand, Q^{pv} increases to compensate for the turbine in self-consumption generation. This behaviour of Q^{pv} is reflected in household higher profits.

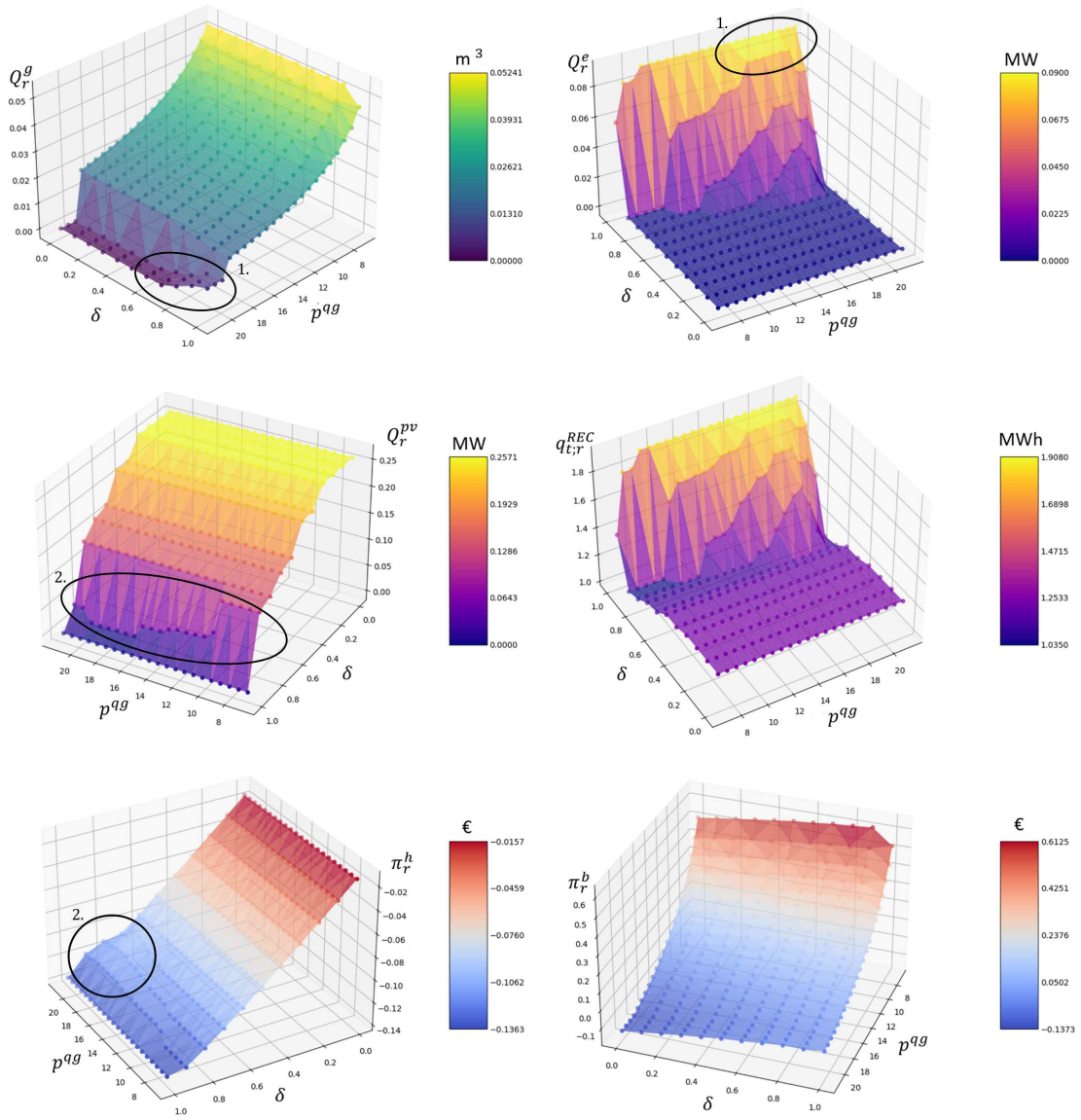

Figure 2.14: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qg} (and δ) when the household acts first in "Incremental commitment hypothesis".

Sensitivity analysis on turbine installation cost, p^{qe} Figure 2.15 plots the sensitivity analysis carried out for the different values of p^{qe} and δ. As a general comment, as p^{qe} increases, higher values of δ are necessary to make Q^e convenient. Consequently, as Q^e occurs, Q^g is decreased, accordingly. In particular, when $p^{qe} \geq 0.9932$ and $\delta = 0.9$ (labelled as point 1.) Q^{pv} is more convenient than Q^e, therefore Q^g is installed; consequently, the household profits face a decrease, while the biogas profits face a peak. The self-consumption plot depicts very clearly the influences of turbine and PV plant installations: as Q^{pv} decreases, self-consumption decreases; conversely, as Q^e increases, self-consumption increases.

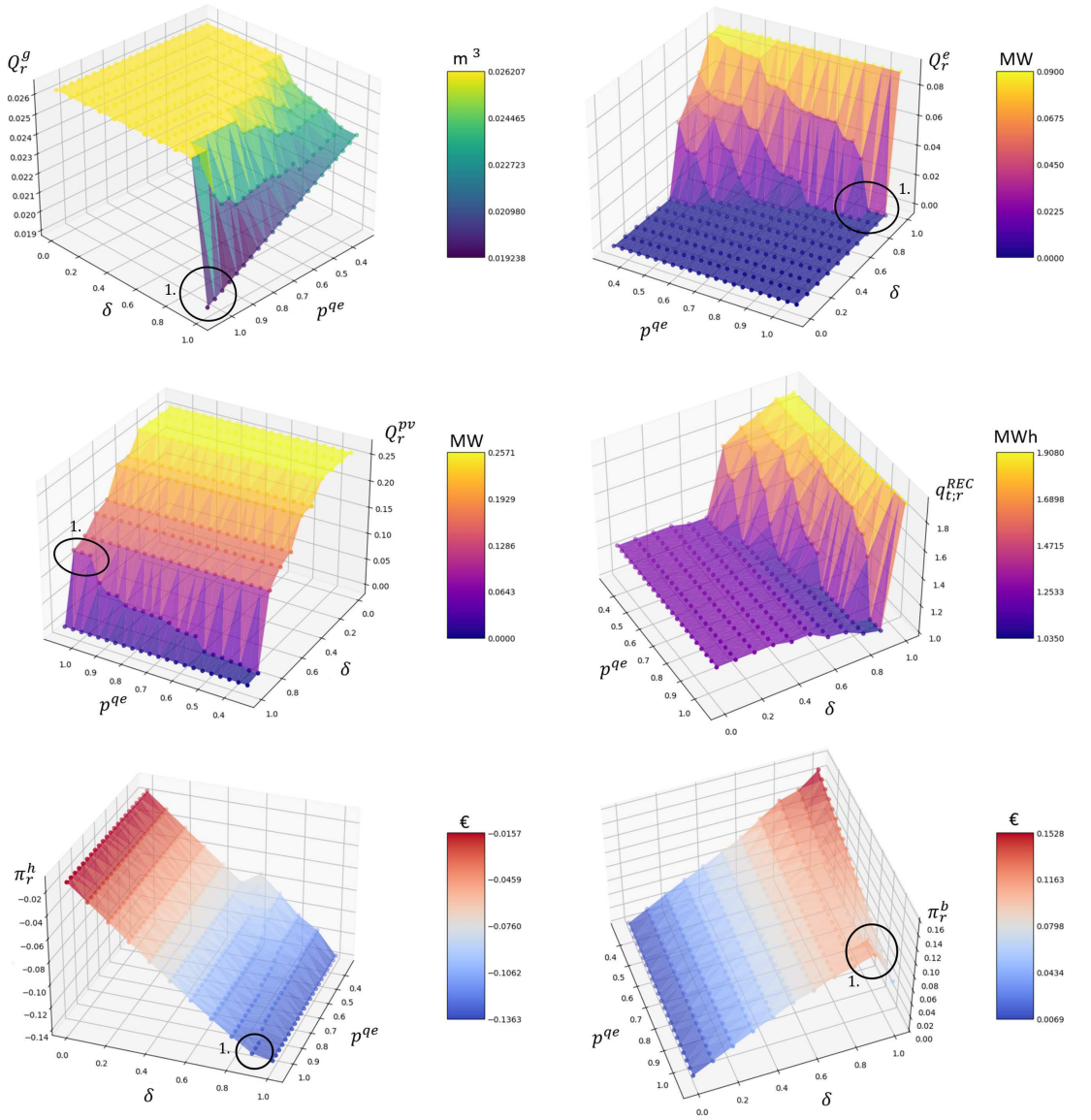

Figure 2.15: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qe} (and δ) when the household acts first in "Incremental commitment hypothesis".

Sensitivity analysis on PV installation cost, p^{pv} As shown in Figure 2.16, when p^{pv} increases, Q^{pv} decreases. However, with $\delta \leq 0.5$, it is convenient to keep Q^{pv} installed to gain the incentive from self-consumption; in this context, the biogas producer installs only Q^g. With $\delta \geq 0.6$ and $p^{pv} \geq 0.7$ the biogas producer installs Q^e, Q^{pv} falls to zero but self-consumption increases. Therefore, the lower installation costs for the household compensate for the lower revenues from selling electricity and profits increase, while biogas producer profits decrease because the investment in Q^e, together with the consequent revenues from selling electricity, and the incentive from self-consumption, do not compensate for the lower revenues from selling biogas.

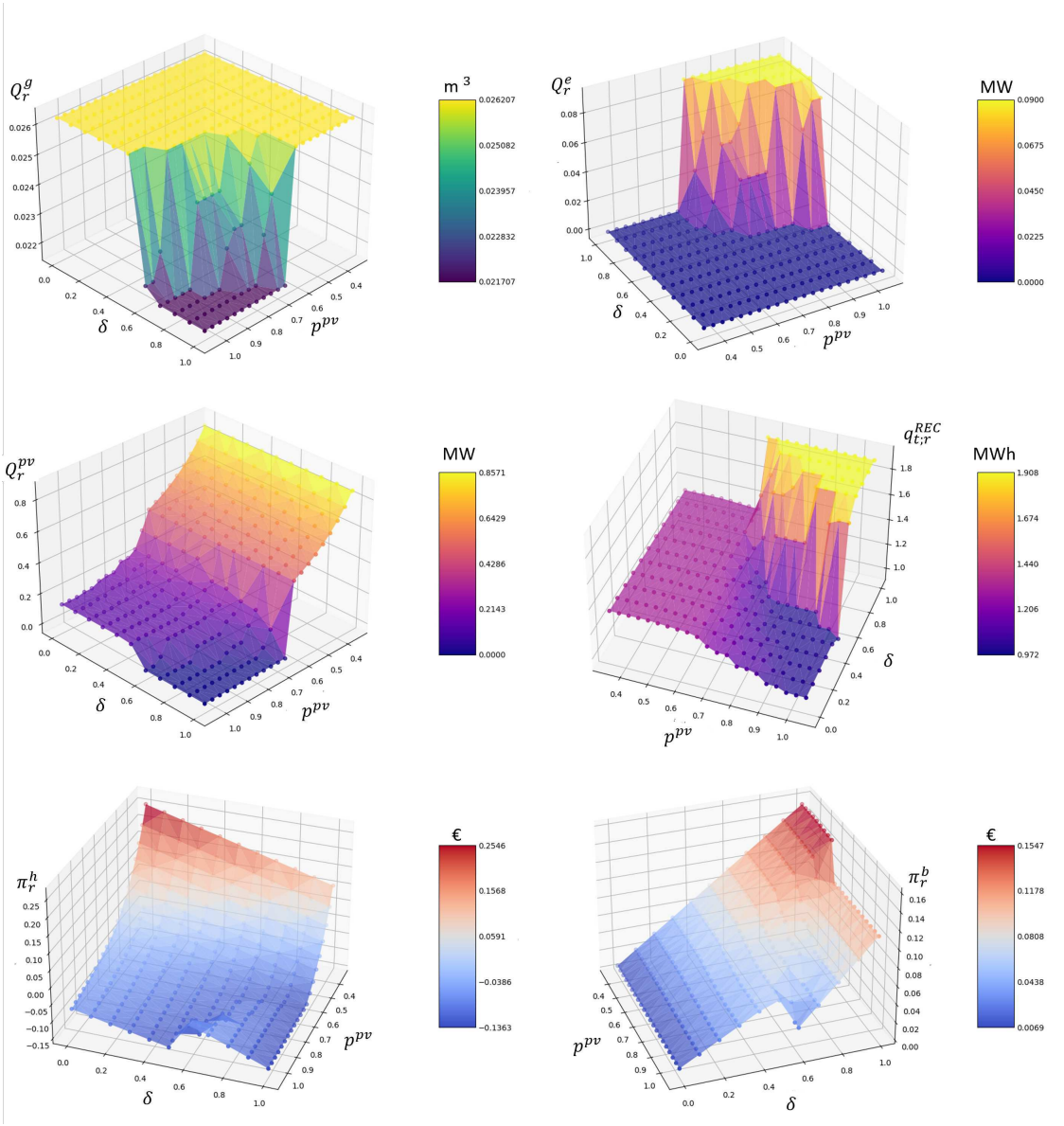

Figure 2.16: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{pv} (and δ) when the household acts first in "Incremental commitment hypothesis".

2.3.3 "Rethink hypothesis (pure Nash equilibrium)" results

The following discussion does not differentiate according to the iteration order, because it is not relevant anymore. Furthermore, the results discussed are obtained at the end of the second round, with the confirmed choice. Every quantity has to be intended as the cumulative value after all iterations (to simplify the discussion the indexes $i \in I$ and $j \in J$, that identify the agents. and the numbers 1 and 2, that distinguish the rounds, are dropped). Table 2.3 lists the sensitivity analyses carried out in this version.

By analyzing the results obtained for different values of δ (see Figure 2.17) it is immediately noticeable that the maximum level of self-consumption is reached with $\delta = 1$. In particular, self-consumption does not follow a monotonic trend. With $\delta \leq 0.6$ self-consumption follows Q^{pv2}. With $\delta \geq 0.6$, Q^{pv2} decreases but the biogas producer installs Q^{e2}, which compensates for Q^{pv2}, therefore self-consumption increases. Furthermore, Q^{g2}

Parameter		Reference value
z	$0.055 - 0.1705$	0.110 €/MWh
p^{qg}	$6.8493 - 21.2329$	13.698630 M€/m^3
p^{qe}	$0.3425 - 1.0616$	0.684932 M€/MW
p^{pv}	$0.35 - 1.085$	0.7 M€/MW

Table 2.3: This table summarizes the sensitivity analysis carried out on the results in the "Rethink hypothesis (pure Nash equilibrium)".

is coherent with the turbine installation because of the substitution effect between the two technologies and the binding budget constraint (i.e. when Q^{e2} becomes convenient, Q^{g2} decreases to release money that can be spent in Q^{e2}). It is worth noticing that with $\delta = 0.8$ there is an increase in the household profits and a decrease in the biogas profits. This happens because the high Q^{e2} generates revenues that, together with the incentive, do not compensate for the lost sales from Q^{g2}. On the other hand, the high installation Q^{e2} allows the household to decrease Q^{pv2}, and the consequent lower investment costs compensate for the lower revenues from selling electricity.

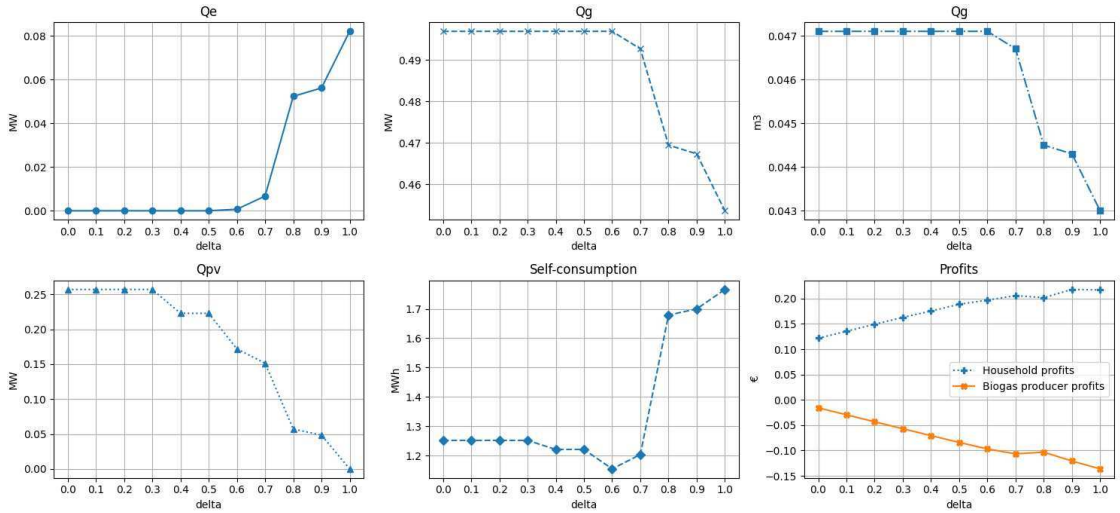

Figure 2.17: Q^{e2} (in MW), Q^{g2} (in m^3 and MW), Q^{pv2} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of δ in "Rethink hypothesis (pure Nash equilibrium)."

Sensitivity analysis on incentive, z Figure 2.18 clearly represents how self-consumption is influenced by Q^{pv2} and Q^{e2}. With $\delta \leq 0.8$ self-consumption is defined by Q^{pv2}; with $\delta = 0.9$, $Q_j^{pv} = Q_i^e = 0$, and also self-consumption falls to zero: this happens because the low level of incentive z makes the installations not convenient. With $\delta = 1$, Q^{e2} is convenient. Consequently, as happened before, this substitution effect between the Q^{e2} and Q^{pv2} impacts on self-consumption and profits. This substitution effect can be observed also with $\delta = 0.6$ where Q^{pv2} decreases while Q^{e2} increases. The lower installation of Q^{pv2} (i.e. lower investment costs), together with the self-consumption obtained through the installation of Q^{e2} compensates the lost revenues from selling electricity and the household profit has a peak while the biogas producer profit has a fall.

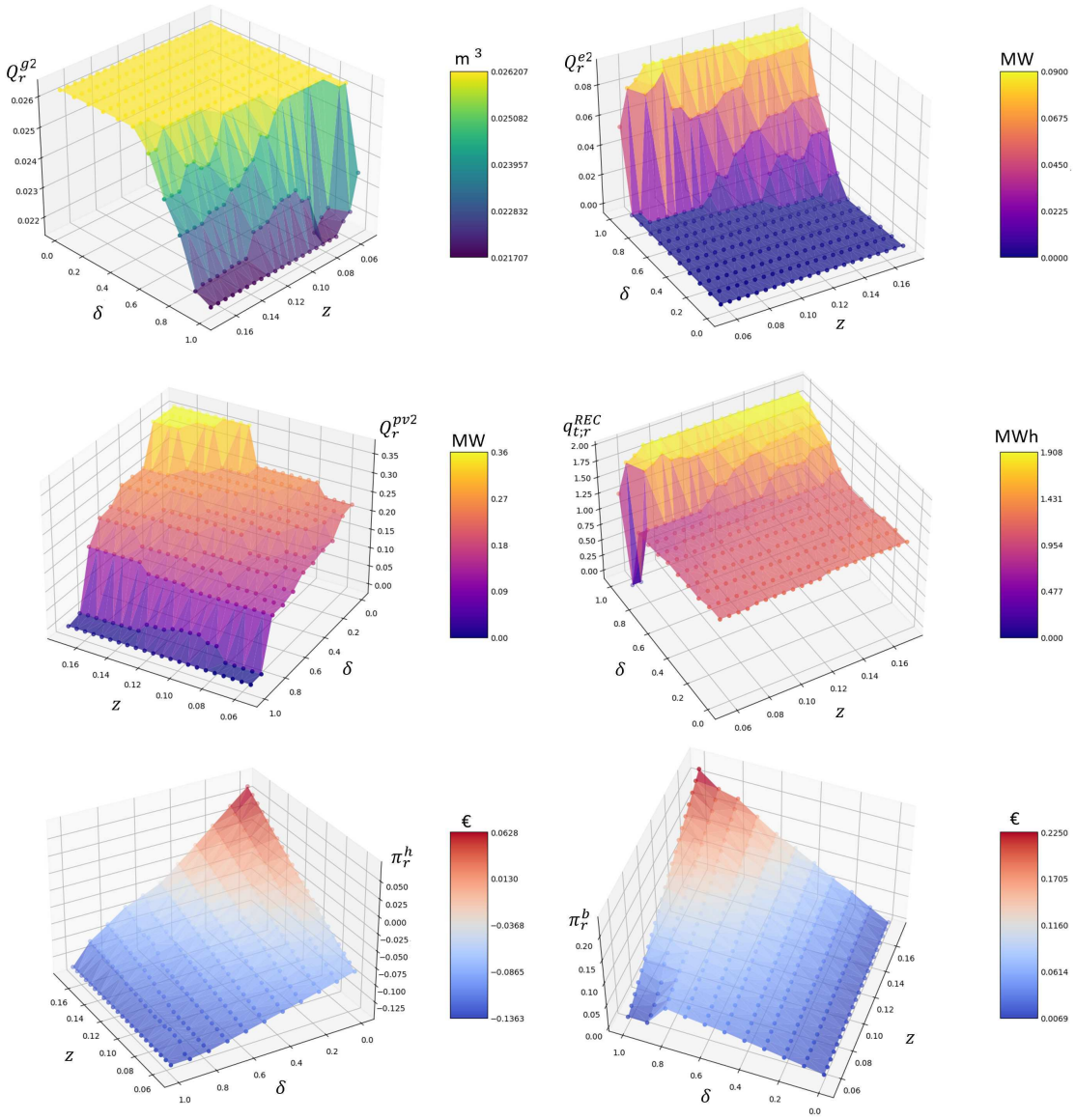

Figure 2.18: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of z (and δ) in "Rethink hypothesis (pure Nash equilibrium)".

Sensitivity analysis on biodigestor installation cost, p^{qg} From Figure 2.19 one can see that self-consumption is maximized when p^{qg} and δ are maximum (labelled as point 1.). This is due to the biogas producer installing a small Q^{g2}, enough for installing the maximum Q^{e2}. Furthermore, with $\delta = 0.7$ (labelled as point 2.), Q^{g2} is more convenient than Q^{e2} and the substitution effect with Q^{pv2} is evident.
Indeed, as p^{qg} increases, Q^{g2} is divested and Q^{pv2} is installed. Furthermore, Q^{pv2} is substituted with Q^{e2} when p^{qg} becomes too expensive.

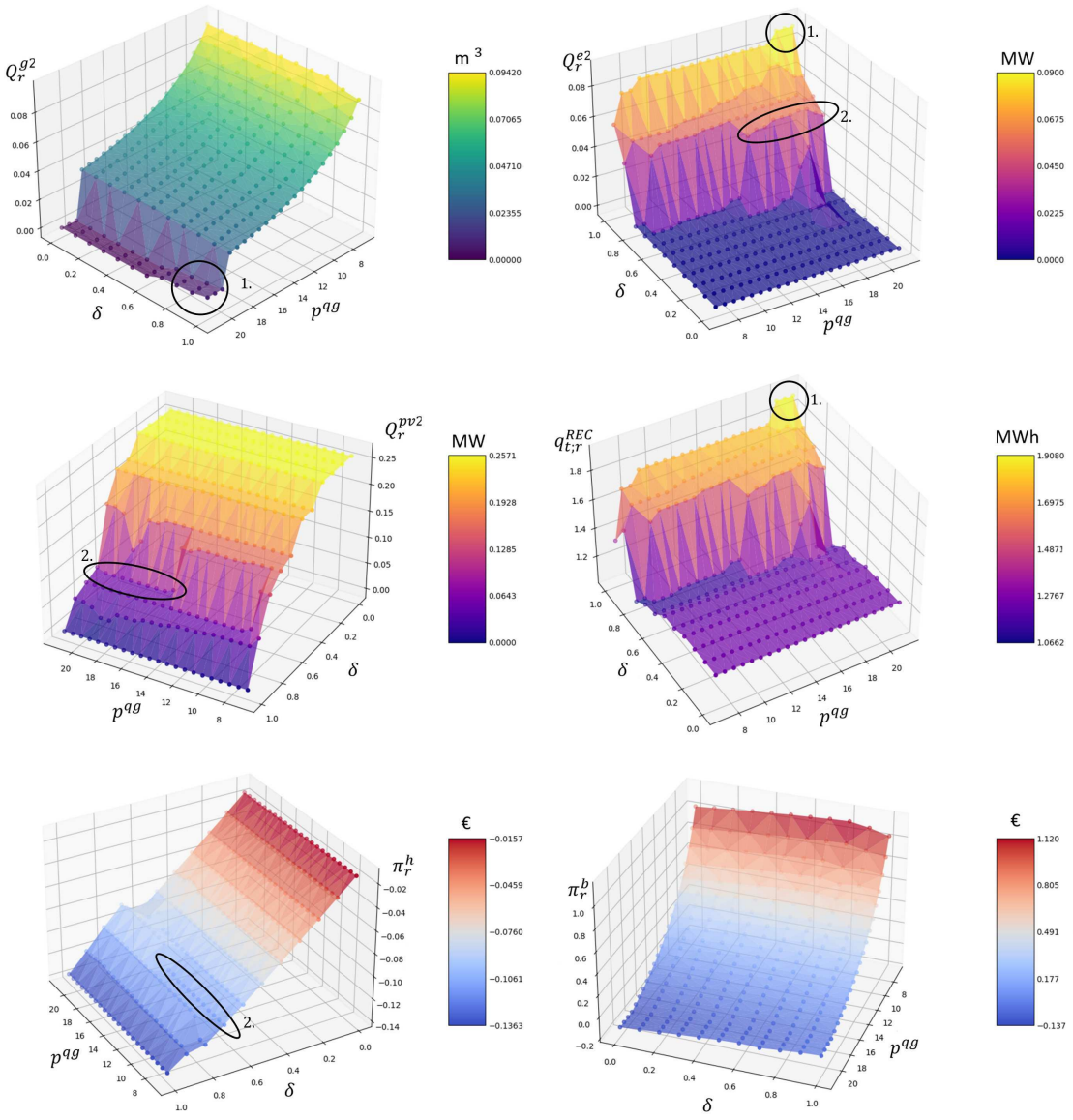

Figure 2.19: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qg} (and δ) in "Rethink hypothesis (pure Nash equilibrium)".

Sensitivity analysis on turbine installation cost, p^{qe} Figure 2.20 shows how p^{qe} impacts on decision variables and profits. As p^{qe} increases, higher values of δ are required to make Q^{e2} convenient. However, with respect to p^{qe}, Q^{e2} does not follow a monotonic trend: with $p^{qe} \geq 0.5479$ and $\delta = 0.6$, $Q_i^e = 0$ and only Q^{pv2} is installed. The behaviour of Q^{e2} and Q^{pv2} is reflected on self-consumption: in particular, with $\delta = 0.6$ self-consumption decreases because Q^{pv2} has an efficiency (represented by γ_t^{pv}) lower than Q^{e2}. Furthermore, when $\delta = 1$ and p^{qe} is maximum, $Q^{pv2} = 0$ but Q^{e2} is maximized (and Q^{g2} decreases accordingly). This investment causes a decrease in the biogas profit because the revenues from selling electricity, together with the incentive, do not compensate for the lost revenues from biogas.

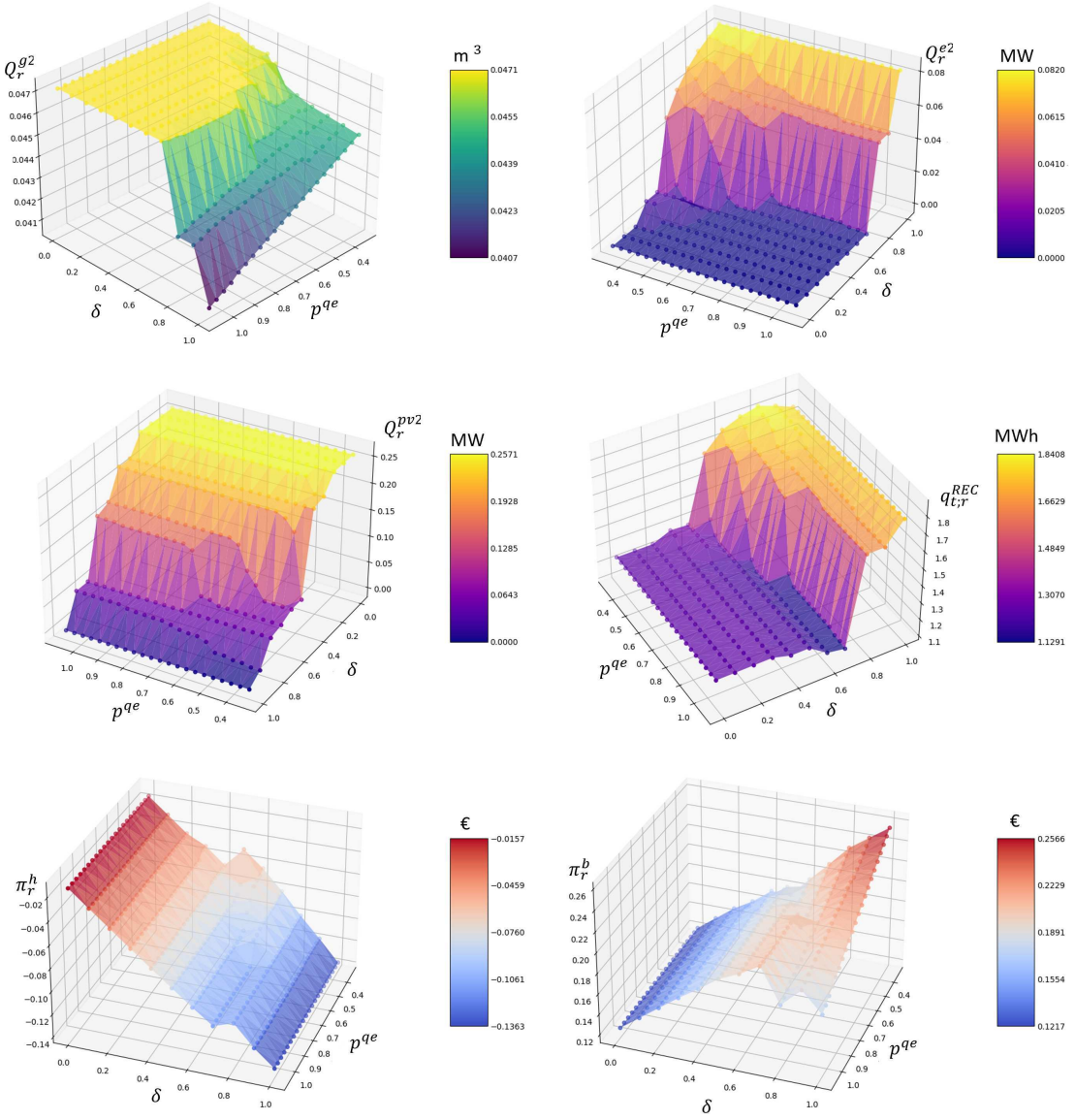

Figure 2.20: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{qe} (and δ) in "Rethink hypothesis (pure Nash equilibrium)".

Sensitivity analysis on PV installation cost, p^{pv} By looking at Figure 2.21 one can differentiate 2 situations: with $p^{pv} \leq 0.63$ and with $p^{pv} > 0.63$. For the latter, another distinction is necessary: $\delta \leq 0.5$ and $\delta > 0.5$. In the first situation, as p^{pv} increases, Q^{pv2} decreases. This does not have an impact on self-consumption (in this interval, it is defined by the demand), Q^{e2}, or Q^{g2}. In the second situation, Q^{pv2} decreases even further, causing a decrease in self-consumption. In the third situation, Q^{pv2} is not installed anymore, Q^{e2} is maximized (Q^{g2} decreases accordingly) and it maximizes self-consumption. Consequently, household profits face a rise due to lower investment costs, while the biogas producer profits have a shift due to the high investment costs.

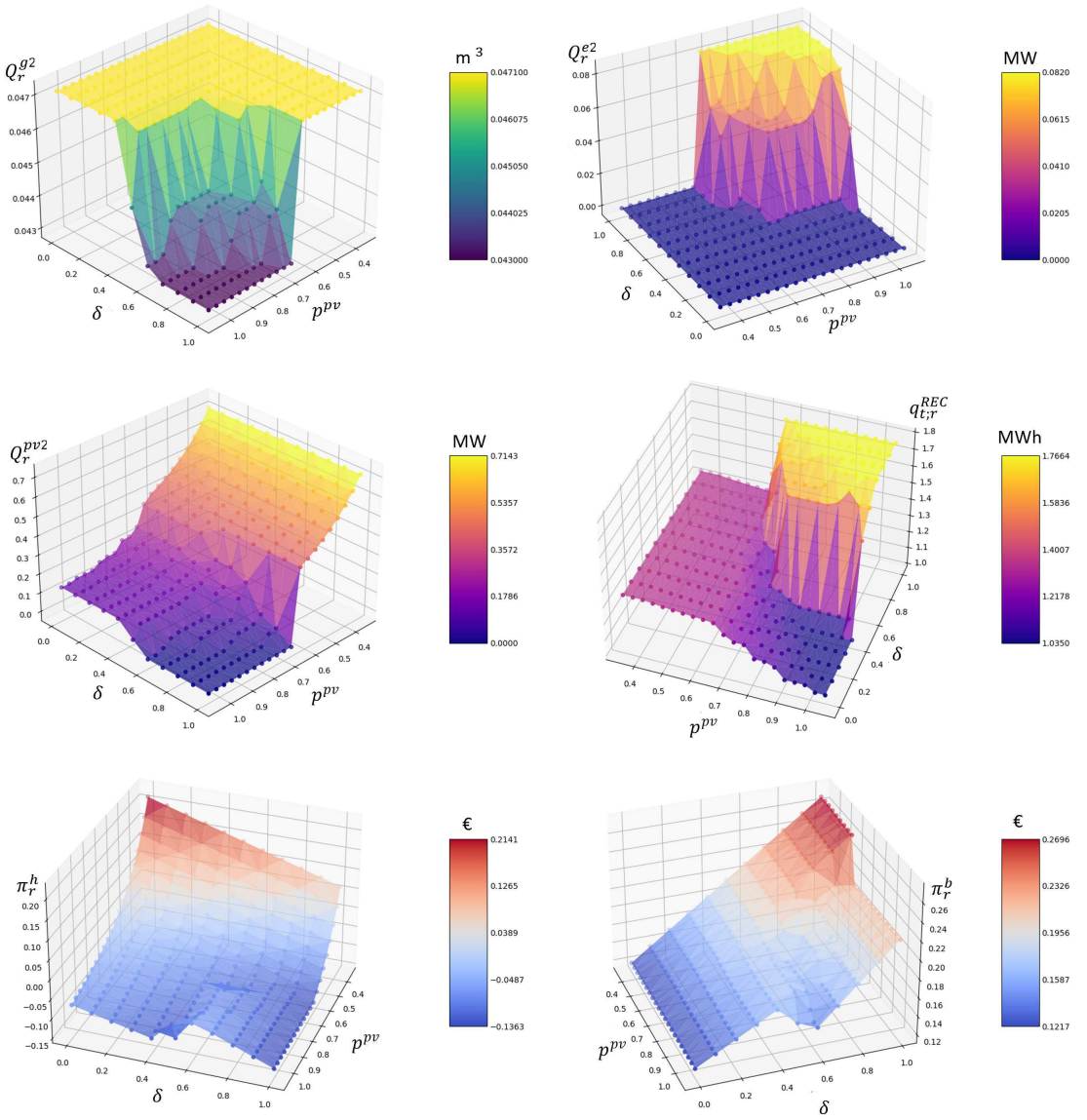

Figure 2.21: Sensitivity analysis on Q^g (in m^3), Q^e (in MW), Q^{pv} (in MW), q_t^{REC} (in MWh), π^b (in M€), and π^h (in M€), for different values of p^{pv} (and δ) in "Rethink hypothesis (pure Nash equilibrium)".

2.4 Conclusions

The models presented in this study allow the examination of the intricate dynamics within RECs, with a particular focus on the interplay among REC members and the regulatory influence on decision-making processes.

First, the version explained in Section "Incremental commitment hypothesis" enables the exploration of how size, and iteration order, impact final decisions. The results characterize a non-pure Nash equilibrium because it is influenced by the iteration sequencing due to agents' different market powers.

Second, in consideration of real-world complexities such as diverse market powers and potential conflicts, the introduction of an upper-level administrator is necessary to ensure REC's effectiveness. This administrator, acting as an external decision-maker determining sharing rules to maximize self-consumption, aligns REC members' interests and promotes decision equity.

Third, in the second version,"Rethink hypothesis (pure Nash equilibrium)", where smaller individual agents are considered (still categorized into two main groups), iteration order becomes irrelevant, and results characterize a pure Nash Equilibrium.
Fourth, the incorporation of at least two rounds aligns with real-world REC implementations and Italian legislative frameworks. This realistic approach enhances the applicability of the model to practical scenarios.
Fifth, the model introduces a novel aspect by addressing interactions among heterogeneous agents. This approach expands the analysis beyond capacity sizing for single-technology RECs by incorporating multiple agent types and technology choices, thereby elucidating internal REC dynamics influenced by both external (regulation, prices, incentives) and internal (peer decisions) factors.
Sixth, by optimizing capacity sizing, over-investments are mitigated, while the joint optimization of capacities and sharing rules ensures effective REC operations.
In conclusion, this research provides a fresh perspective on REC implementation strategies and offers insights into navigating challenges inherent in the implementation process.

Chapter 3

Markov Chain Bootstrapping and Simulation for a Quadrivariate Stochastic Process in Energy Market Scenario Generation

3.1 Introduction

A huge amount of interconnected and stochastic factors profoundly influence decision-making processes in energy markets. Fluctuations in the cost of energy production, influenced by resource availability, geopolitical tensions, and technological advancements, significantly impact the profitability of energy market participants. Weather conditions, including temperature variations, precipitation patterns, and extreme events, exert a pronounced influence on energy demand and supply. Sale prices are influenced, among others, by geopolitical events, regulatory changes, and market speculation. Finally, demand levels are influenced by economic activity, population growth, technological advancements, and consumer behaviour. In addition to the stochastic nature of the previous factors, it is essential to recognize the intricate connections among them. These connections are often nonlinear, which makes them challenging to accurately model. It is of crucial importance to be able to deal with the previous factors when making investment and operations decisions. One step in this direction is the availability of (risk management) models and methods capable of capturing the stochasticity and connection of the decision variables and parameters.

Cerqueti et al. (2017a) addresses the previous challenges in advancing a methodological framework that has proven effective in modelling nonlinear dependencies among multiple sources of uncertainty. This approach, based on the approximation of n-variate stochastic processes using Markov Chains of order k, offers an accurate and robust framework for capturing the intricacies of factors relevant to decision-making in energy markets.

According to the works examined in Chapter 1, the literature on resampling procedures based on multivariate Markov Chains is mainly focused on Markov Chains whose components share the same temporal frequency. The present work contributes to the literature in advancing a method to deal with heterogeneous temporal frequencies in bootstrapping and simulating n-variate stochastic processes approximated by n-variate Markov Chains. Indeed, this research applies the bootstrapping and simulation approach of Cerqueti et al. (2017a) to the quadrivariate stochastic process of gas prices, load, electricity prices, and solar radiation. A key advantage of leveraging the bootstrapping and simulation approach of Cerqueti et al. (2017a) lies in its parsimonious parameterization compared to

alternative approaches. By keeping to a minimum the number of parameters required for modelling, the bootstrapping and simulation offer computational efficiency without compromising accuracy in the generation of scenarios for the quadrivariate stochastic process under scrutiny.
In particular, while gas prices exhibit a daily frequency, the other three components have an hourly basis. This temporal frequency heterogeneity adds another layer of complexity to the bootstrapping and simulation process, requiring tailored approaches to effectively capture and analyze the stochasticity and connection among variables operating at different temporal frequencies. Indeed, to the goal of using the theory of Markov Chains to jointly model stochastic processes with heterogeneous temporal frequencies, their time frequencies must be made homogeneous. Special treatments are advanced to deal with the daily and hourly frequencies of the four components and make all of them into daily stochastic processes. The resulting daily quadrivariate stochastic process of gas prices, load, electricity prices, and solar radiation is simulated based on fitting a Markov Chain of order 2 to the discretized version of such a stochastic process. Order 2 is chosen to verify if it contributes to modelling the temporal dependence in the series. The goodness-of-fit of the present method is tested by calculating several statistics on the simulated series and by comparing them with the same statistics calculated on the historical series. In particular, the statistics calculated on the historical series are meant to capture the autodependencies and interdependencies of the quadrivariate stochastic process. The comparison of such statistics with the ones calculated on the simulated series is meant to test the robustness and accuracy of the method in reproducing these intricate connections.
The chapter outline is the following: in Section 3.2 the model is presented; it is followed by the methodological issues that are described in Section 3.3; in Section 3.4 all the transformations and computations are applied; last, the goodness-of-fit results and the conclusions are presented in Sections 3.5 and 3.6, respectively.

3.2 Model

Let us introduce two types of stochastic processes. Let $\{\mathbf{X}_d^{(N)}\}_{d\in D} = (X_d^1, X_d^2, \dots, X_d^N)_{d\in D}$ be an N-variate stochastic process taking values in $\mathbb{R}^N$ with time index d and $D = \{0, \dots, \bar{D}\}$. Let $\{\mathbf{X}_i^{(M)}\}_{i\in I} = (X_i^1, X_i^2, \dots, X_i^M)_{i\in I}$ be an M-variate stochastic process taking values in $\mathbb{R}^M$ with time index i and $I = \{0, \dots, \bar{I}\}$. Let $\bar{D} < \bar{I}$, that is, the cardinality of D is smaller than the cardinality of I, but let $\bar{D}$ and $\bar{I}$ refer to the same point in time measured on an absolute scale of times with a given starting point related to a historical event. Therefore, the N-variate stochastic process is said to have a lower time frequency than the M-variate stochastic process. Without loss of generality, the cardinality of I is assumed as a multiple of the cardinality of D, that is $\frac{|I|}{|D|} \in \mathbb{N} \setminus \{0, 1\}$. Let us introduce the parameter $J = \frac{|I|}{|D|}$, that is, the number of times that index i is related to the same value of index d.
As an example, let us consider a stochastic process with daily gas prices and the stochastic process of hourly electricity prices and load. Focus on December 2022, $D = \{0, \dots, 30\}$ for gas prices and $I = \{0, \dots, 743\}$ for the electricity prices and load. $\bar{D} = 30$ is the end of December 31, 2022 and $\bar{I} = 743$ is the end of the 24th hour of December 31, 2022. Consequently, $J = \frac{744}{31} = 24$, which means that 24 hours (24 values of index i) are related to each day (each value of index d).
To the goal of using the theory of Markov Chains to jointly model both types of stochastic processes, their time frequencies must be homogeneous. The necessary step is to modify the M-variate stochastic process into a stochastic process with the same time-frequency

of the N-variate stochastic process. To do so, the following transformation is performed:

$$(X_i^1, X_i^2, \ldots, X_i^M)_{i\in I} \longrightarrow (Y_d^1, Y_d^2, \ldots, Y_d^M)_{d\in D},$$

where $Y_d^1 = f_1(d, X_i^1; J)$, $Y_d^2 = f_2(d, X_i^2; J), \ldots, Y_d^M = f_M(d, X_i^M; J)$. The previous M functions transform the M-variate stochastic process with index i into a corresponding stochastic process with index d. As an example, one could adopt the following functions:

$$f_1(d, X_i^1; J) = \sum_{j=Jd}^{J(d+1)-1} X_j^1,$$

$$f_2(d, X_i^2; J) = \frac{1}{J} \sum_{j=Jd}^{J(d+1)-1} X_j^2,$$

and so on. The first function calculates the sum of values with index i related to each value of index d, the second function calculates the average of values with index i related to the same value of index d, and so on. Choosing f_1 for the electricity prices and f_2 for the load, the computations become, respectively:

$$f_1(d, X_i^1; 24) = \sum_{j=24d}^{24(d+1)-1} X_j^1$$

and

$$f_2(d, X_i^2; 24) = \frac{1}{24} \sum_{j=24d}^{24(d+1)-1} X_j^2.$$

The first function transforms the hourly electricity prices into their daily sums and the second function transforms the hourly load into its daily average. Other functions could be used for this frequency transformation. After performing this step, the following $(N+M)$-variate stochastic process with index d, $(X_d^1, \ldots, X_d^N, Y_d^1, \ldots, Y_d^M)_{d\in D} \in \mathbb{R}^{N+M}$ are obtained. For ease of notation, the stochastic process $(X_d^1, \ldots, X_d^N, Y_d^1, \ldots, Y_d^M)_{d\in D}$ is labelled as $\{\mathbf{Z}_d\}_{d\in D}$.

To model $\{\mathbf{Z}_d\}_{d\in D}$ as a (discrete state) Markov Chain, such a process must be discretized, that is, the support $\mathbb{R}$ of each component of the stochastic process is partitioned. Suppose to deal with finite partitions. For example, the support of $\{Z_d^1\}_{d\in D}$ could be partitioned into $k \geq 1$ intervals. For ease of notation, each interval of this partition is labelled, and its members are listed, as $s^{1,1}, s^{1,2}, \ldots, s^{1,k}$. The previous partitioning step into k intervals is, then, repeated for the other components of $\{\mathbf{Z}_d\}_{d\in D}$.[1] A given list of $N+M$ intervals is called state, each of them chosen from the partition of $\mathbb{R}$ adopted for the corresponding component in $\{\mathbf{Z}_d\}_{d\in D}$. Based on $N+M$ partitions of $\mathbb{R}$ into k intervals, the overall number of states is $(N+M)^k$. For example, the state $s^1 = (s^{1,1}, \ldots, s^{N,1}, s^{N+1,1}, \ldots, s^{N+M,1})$ collects the first intervals taken from the $N+M$ partitions of the support of $\{\mathbf{Z}_d\}_{d\in D}$. Analogously, the state $s^{(N+M)^k} = (s^{1,k}, \ldots, s^{N,k}, s^{N+1,k}, \ldots, s^{N+M,k})$ collects the k-th intervals of the same partitions. The previous two states could be labelled as "extreme" states. The other possible states combine the k intervals of each partition in ways other than those of the extreme states. The index ν is introduced to label such states. Therefore, they can be listed in the set $\mathbf{s}^\nu = (s^1, \ldots, s^\nu, \ldots, s^{(N+M)^k})$. Finally, given $\{\mathbf{Z}_d\}_{d\in D}$, its discretized stochastic process $\{\mathbf{S}_d\}_{d\in D}$ whose realizations are

[1] It is implicitly assumed that all the partitions have the same number, k, of elements. However, the partitioning step could be generalized to partitions with different cardinalities across the components of $\{\mathbf{Z}_d\}_{d\in D}$.

$\mathbf{s}_d = (s_d^1, \ldots, s_d^N, s_d^{N+1}, \ldots, s_d^{N+M})$, with $d \in D$, can be built. Each realization is one of the previous $(N+M)^k$ states.
Now, the transition probability matrix for the discretized stochastic process $\{\mathbf{S}_d\}_{d\in D}$ can be introduced. Such a matrix is composed of conditional probabilities that, as is usual for a Markov Chain, measure the likelihood of a state given a state, with the understanding that the conditioning state is observed one time before the conditioned state. For example, to estimate the conditional probability $p(s^2|s^1)$, one counts how many times the couple (s^1, s^2) was observed in two *consecutive* realizations of $\{\mathbf{S}_d\}_{d\in D}$, with s^2 following s^1. The estimated conditional probability is obtained by dividing the previous count by the number of times when state s^1 was a realization $\mathbf{s}_d$, with $d \in D$.
A generalization of such probabilities can be introduced, and speak of conditional probabilities of a state given more than one state. For example, one could estimate $p(s^1|s^1, s^2)$, understanding that the conditioned state is s^1 and the conditioning states are s^1 and s^2. The conditioning states are observed two times (s^1) and one time (s^2) before the time of observation of the conditioned state, respectively. These kinds of probabilities are collected in a "rectangular" transition probability matrix and model a Markov Chain of order 2, given that two consecutive states are needed to determine the following state. The generalization of this example leads to Markov Chains of order $k \geq 1$.
It is important to highlight that the degree of determinism/stochasticity associated with a Markov Chain of order k depends on the number of transition probabilities equal to 1. The transition probability matrix is said deterministic if, given any conditioning event (k-tuples of consecutive states observed in a temporal sequence), there is only one state that can be observed at the following time. In other words, each row has only one probability equal to 1 and all the other probabilities are null. The lower the number of transition probabilities equal to 1, the higher the level of stochasticity associated with the Markov Chain of order k. To obtain the desired level of determinism/stochasticity, one can act on the number of intervals that partition $\mathbb{R}$ when building the discretized stochastic process $\{\mathbf{S}_d\}_{d\in D}$.
In this work, the quadrivariate stochastic process of gas prices, electricity load, electricity prices, and solar radiation is modelled. The processes are labelled as g, ℓ, e, and r, respectively. In particular, gas prices have a daily frequency and the other processes have an hourly frequency. This causes the introduction of two indexes, $i \in I$ and $d \in D$, to represent the hourly and daily frequencies, respectively. The quadrivariate stochastic process has to undergo a transformation that makes hourly frequencies into daily frequencies. After this step, the quadrivariate stochastic process is discretized and, finally, modelled as a Markov Chain of order 2. More details on these steps are included in Section 3.3.

3.3 Methodological issues

3.3.1 Intraday distributions and seasonalities

The estimation of models and their bootstrapping and simulation based on multivariate series whose components display different frequencies (some with an hourly frequency and others with a daily frequency) is an aspect of no secondary importance and must be managed.
There are two points in the bootstrapping and simulation method where this aspect must be considered. The first point is concerned with the data used for fitting the Markov Chain of order 2. The second point is concerned with the simulated series.
Regarding the first point, among the possible solutions, hourly series are transformed into daily ones, thus aligning the series with higher frequency to the series with lower frequency. This inevitably bring to a loss of information when summarizing the hourly data,

but allows for the fitting of a Markov Chain to the quadrivariate time series of daily data. The transformation of hourly data into daily data is carried out after having detrended (Section 3.3.2) and deseasonalized (Section 3.3.3) the hourly data. This transformation is explained in Section 3.3.3 and affects the three hourly series of residuals corresponding to load, electricity prices, and solar radiation.
Concerning the second point, after having fit a Markov Chain of order 2 to the quadrivariate series of residuals obtained after detrendization and deseasonalization of the historical data (Section 3.3.4), the bootstrapping of these residuals and their transformation from daily into hourly, is performed. This step is needed for the simulation of the 3 time series with hourly frequency.
The nature of the problem arising when transforming bootstrapped daily residuals into hourly ones is due to the fact that the clustering of daily residuals into states of the set $\mathbf{s}^\nu$ was performed to fit the Markov Chain of order 2 and was based on the "distance" among the daily residuals; such a distance is not necessarily preserved between the hourly residuals corresponding to two daily residuals belonging to the same state $s^\nu \in \mathbf{s}^\nu$, even more so in the case of hourly residuals corresponding to daily residuals that do not belong to the same state. In the end, hourly residuals may exhibit different yearly and/or weekly seasonalities even in the presence of similar daily residuals (i.e., clustered in the same state $s^\nu \in \mathbf{s}^\nu$.)
To provide an example, suppose that the daily load residual of a certain Sunday of April was bootstrapped and used to simulate the load of a certain Thursday in April. If the distributions of load residuals of the two days differ, there may be a problem in simulating one day by using the bootstrapped residual belonging to the other. The problem is graphically captured in Figure 3.1, which plots the frequency distributions of daily load residuals of all Thursdays and all Sundays of April in the historical series. It is easy to see that the two distributions differ. By focusing on their medians, it is easily noticeable that the median of all Thursdays of April is around 0.9137 and the median of all Sundays of April is around -0.1054. Therefore, to simulate a Thursday of April by bootstrapping a Sunday of April, small values are extracted from the frequency distribution of all Sundays of April with a higher probability than it would happen if the frequency distribution of all Thursdays of April was used. The opposite would happen when extracting high values. In a few words, the simulated Thursday values, with respect to the typical values observed for a Thursday, would be distorted. The problem affecting daily residuals may be magnified when using the hourly residuals of the bootstrapped day to obtain the simulated day values. The graphical display of this problem is in Figure 3.2, where the frequency distributions of hourly load residuals of 03:00 am of all Sundays and all Thursdays of April are plotted. The medians of the two distributions are different, the one of 03:00 am of all Sundays of April is around 0.1304, whereas that of the same hour of all Thursdays of April is around 0.0250. To simulate 03:00 am of a Thursday of April by bootstrapping 03:00 am of a Sunday of April, one extracts small values from the frequency distribution of all Sundays of April with a lower probability than it would happen if one used the frequency distribution of all Thursdays of April. The opposite would happen when extracting high values. In addition, the two frequency distributions show another distorting feature: the hourly residuals of Sundays show a minimum that is higher than that of the hourly residuals of Thursdays, and vice versa for the maximum. This means that the simulation of Sundays based on bootstrapped Thursdays can display minimum values never observed on Sundays and maximum values always lower than those observed on Sundays.
To address the previous problem, an adjustment in the simulation step is advanced: it focuses on a specific feature of the bootstrapped daily residual to obtain the simulated daily residual. This feature is the percentile of the bootstrapped daily residual with respect

to the frequency distribution of daily residuals of the bootstrapped day. The percentile is, then, used to extract the simulated daily residual with the same percentile from the frequency distribution of daily residuals of the simulated day. The two frequency distributions, that of residuals for the bootstrapped day and that of residuals for the simulated day, include all daily residuals in the historical series that share the same bootstrapped day (and month) and the same simulated day (and month), respectively. The adjustment does not assign to the simulated day residuals belonging to the bootstrapped day, but chooses, among the residuals of the simulated day, the element that shares the same percentile that the bootstrapped residual scores with respect to the frequency distribution of the bootstrapped day. Thus, the bootstrapping is not completely disregarded when adopting the previous adjustment, since the percentile of the bootstrapped daily residual, not its very value, is used to choose the simulated daily residual.
All the previous explanations of the present adjustment could be referred to as hourly residuals of a bootstrapped day with respect to the same hour of a simulated day.
Returning to the example of daily residuals, the distribution of bootstrapped residuals includes the daily residuals of all Sundays of April of the years considered in the historical series. Similarly, the distribution of simulated residuals includes the daily residuals of all Thursdays of April.
A detailed description of the adjustment is included in Section 3.4.4, where it is directly explained in the context of the application covered by this chapter.

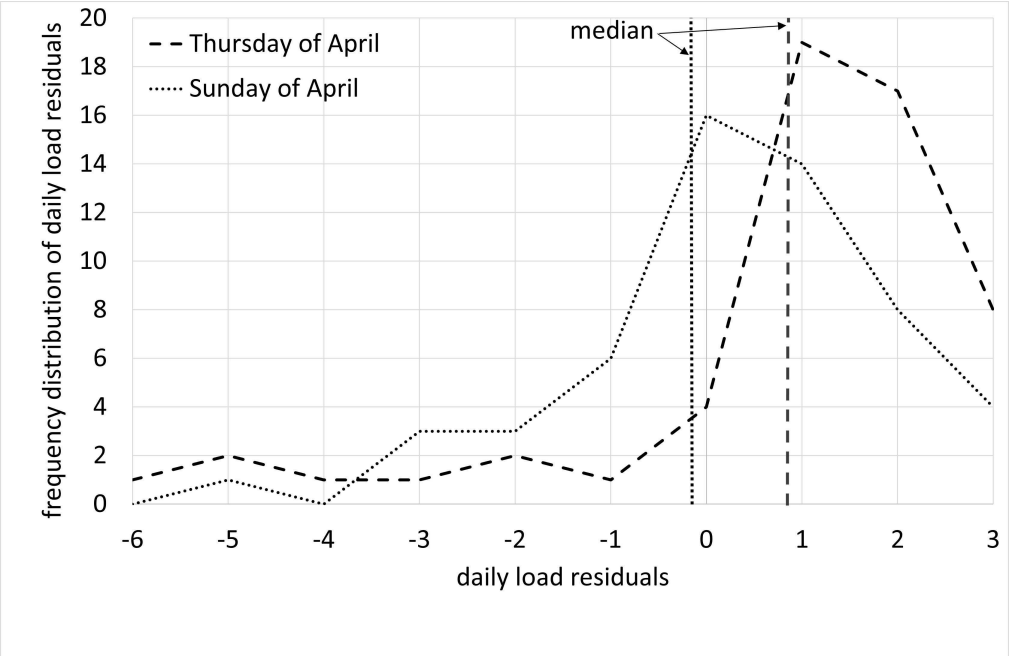

Figure 3.1: Frequency distributions of daily load residuals of all Thursdays and all Sundays of April in the historical series.

3.3.2 Detrendization

The detrendization allows to remove the trend from historical data. It is not applied to solar radiation due to the physical characteristics that make it to be considered stationary: it may have changed over very long periods, but, in the period analyzed, there are no obvious alterations such as to highlight a trend, which should, therefore, be estimated and removed. A linear detrendization model for gas prices, electricity prices, and load is applied. Historical series are labelled as w (see Section 3.4.1 for details about the

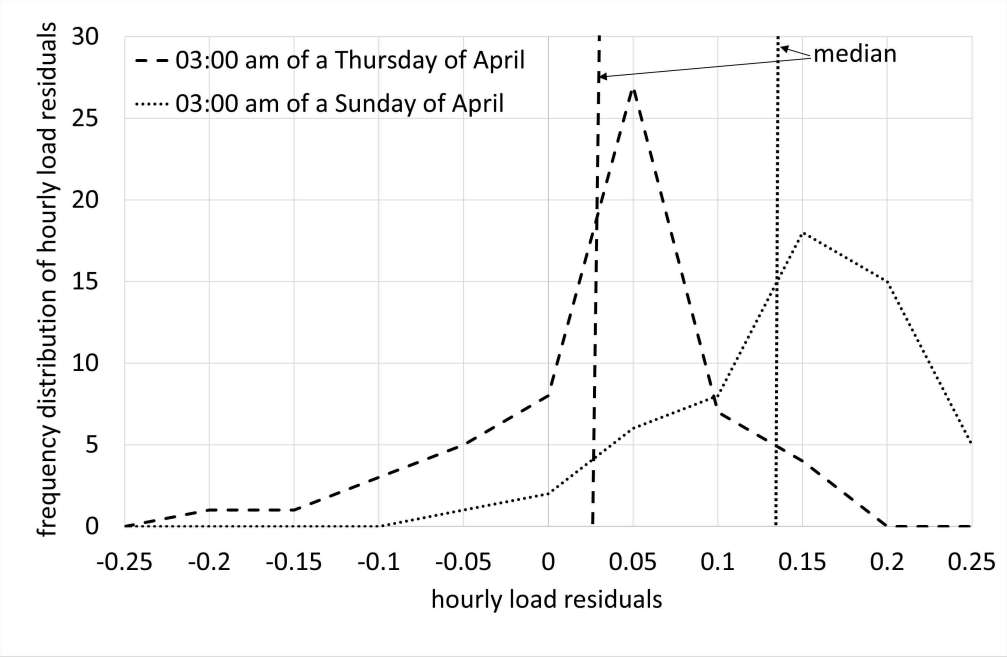

Figure 3.2: Frequency distributions of hourly load residuals of 03:00 am of all Thursdays and all Sundays of April in the historical series.

sources of such series and their preliminary treatment), whereas the detrended residuals are labelled as η. The linear detrendization model includes coefficients, the intercept α_0 and the slope α_1. These coefficients are estimated through Ordinary Least Squares (OLS) and their numerical values are shown in Section 3.4.2. The models for the detrendization are:

$$\ln(w_d^g) = \alpha_0^g + \alpha_1^g \, d + \eta_d^g, \text{ for gas prices,} \tag{3.1}$$

$$\ln(w_i^\ell) = \alpha_0^\ell + \alpha_1^\ell \, i + \eta_i^\ell, \text{ for load,} \tag{3.2}$$

$$\ln(w_i^e) = \alpha_0^e + \alpha_1^e \, i + \eta_i^e, \text{ for electricity prices.} \tag{3.3}$$

Indexes $i \in I$ and $d \in D$ count the hours and the days covered by the historical series, respectively. To avoid obtaining negative values in the simulated series (Section 3.4.5), the detrendization is applied to the logarithmic version of the original values.

3.3.3 Deseasonalization

The deseasonalization removes the seasonality from historical data. The deseasonalized residuals are labelled as ε. In particular, deseasonalization is applied to detrended historical series. Given the specificity of solar radiation, which has a deterministic cycle and does not have weekly seasonality, it receives a simplified treatment: because of its deterministic cycle, only the yearly seasonality is removed. In addition, the residuals of the hourly series are transformed into daily residuals to be used in the cluster analysis for the estimation of the Markov Chain transition probability matrix (Section 3.3.4). The deseasonalization is performed via a linear regression model whose parameters are estimated through OLS and their numerical values are shown in Section 3.4.2. The daily, weekly and yearly seasonalities are estimated. Before going ahead, the following sets must be introduced:

- $H = \{\textit{00:00:00}, \ldots, \textit{23:00:00}\} \equiv \{0, \ldots, 23\}$,
- $W = \{\textit{Monday}, \ldots, \textit{Sunday}\} \equiv \{0, \ldots, 6\}$,

- $C = \{\textit{01-01}, \ldots, \textit{12-31}\} \equiv \{0, \ldots, 364\}$,
- $M = \{\textit{January}, \ldots, \textit{December}\} \equiv \{0, \ldots, 11\}$.

H includes the hours of a day, W includes the weekdays, C includes the calendar days, and M includes the months of a year.
The linear regression model includes an intercept to deal with multicollinearity. Multicollinearity is a typical phenomenon when using dummy explanatory variables and is related to the fact that explanatory variables can be expressed as a linear combination of other explanatory variables. Multicollinearity corresponds to a singular/non-invertible correlation matrix of explanatory variables. If the correlation matrix is not invertible, the OLS estimates are not unique. This makes the interpretation of the coefficients problematic and can lead to unstable and unreliable predictions. Among the existing methods to deal with multicollinearity in the case of dummy explanatory variables, the "dummy variable drop" method is applied. It is also known as the "reference category" method, which consists in excluding an appropriate number of dummy explanatory variables so that their influence remains captured by the model intercept.
In this case, given the estimation of daily, weekly and yearly seasonalities, the first hour, the first weekday, the first calendar day, and the first month has been removed from each of the previously defined sets, H, W, C, and M, respectively. The sets without their first element are labelled as H_{-1}, W_{-1}, C_{-1}, and M_{-1}, respectively. The intercept of each linear regression model is labelled as α_2. It compensates the effects of the "dummy variable drop".
In the following, the linear regression models used for the deseasonalization of the detrended historical series are advanced.

Gas price residuals The detrended residuals of gas prices, η_d^g, $d \in D$, have a daily frequency, therefore the deseasonalization process focuses on removing weekly seasonalities and yearly seasonalities on a monthly basis. The linear regression model is:

$$\eta_d^g = \alpha_2^g + \sum_{j \in W_{-1}} \gamma_j^g \chi_{(wd(d)=j)} + \sum_{j \in M_{-1}} \delta_j^g \chi_{(m(d)=j)} + \varepsilon_d^g,\ d \in D. \tag{3.4}$$

The functions $wd(d)$ and $m(d)$ transform the index $d \in D$ into elements of W and M, respectively: $wd(d)$ transforms the day $d \in D$ into the corresponding weekday in W; $m(d)$ transforms the day $d \in D$ in the corresponding month in M. The dummy explanatory variable for weekday $j \in W_{-1}$ is $\chi_{(wd(d)=j)}$, $d \in D$. Analogously, the dummy explanatory variable for month $j \in M_{-1}$ is $\chi_{(m(d)=j)}$, $d \in D$. The regression coefficients γ_j^g and δ_j^g capture the weekly seasonalities and the yearly seasonalities on a monthly basis, respectively. Finally, the intercept is α_2^g. At the outset of such a deseasonalization, the series of gas price residuals ε_d^g, $d \in D$ is obtained.

Load residuals The detrended residuals of load, η_i^ℓ, $i \in I$, have an hourly frequency, therefore the deseasonalization process focuses on removing daily and weekly seasonalities, as well as yearly seasonalities on a monthly basis. The linear regression model is:

$$\eta_i^\ell = \alpha_2^\ell + \sum_{j \in H_{-1}} \beta_j^\ell \chi_{(hr(i)=j)} + \sum_{j \in W_{-1}} \gamma_j^\ell \chi_{(wd(i)=j)} + \sum_{j \in M_{-1}} \delta_j^\ell \chi_{(m(i)=j)} + \varepsilon_i^\ell,\ i \in I. \tag{3.5}$$

The functions $hr(i)$, $wd(i)$, and $m(i)$ transform the index $i \in I$ into elements of H, W, and M, respectively: $hr(i)$ is the hour in a day in H corresponding to hour $i \in I$; $wd(i)$ is the weekday in W corresponding to hour $i \in I$; $m(i)$ is the month in M corresponding to hour $i \in I$. The dummy explanatory variable for hour $j \in H_{-1}$ is $\chi_{(hr(i)=j)}$, $i \in I$.

The dummy explanatory variable for weekday $j \in W_{-1}$ is $\chi_{(wd(i)=j)}$, $i \in I$. The dummy explanatory variable for month $j \in M_{-1}$ is $\chi_{(m(i)=j)}$, $i \in I$. The regression coefficients β_j^ℓ, γ_j^ℓ, and δ_j^ℓ capture the daily and weekly seasonalities and the yearly seasonalities on a monthly basis, respectively. Finally, the intercept is α_2^ℓ. At the outset of such a deseasonalization, the series of load residuals ε_i^ℓ, $i \in I$, is obtained.
Furthermore, the hourly residuals are transformed into daily residuals in the following way:

$$\varepsilon_d^\ell = \sum_{d(i)=d} \varepsilon_i^\ell,\ d \in D,\ i \in I, \tag{3.6}$$

where the summation condition $d(i) = d$ collects the hourly residuals ε_i^ℓ such that $d(i) = d$, that is, the day of hour i is equal to d. In a few words, the 24 hourly load residuals of each day in the historical series are summed up.

Electricity price residuals The deseasonalization of electricity prices follows an analogous treatment to that of load. Here, the linear regression model is:

$$\eta_i^e = \alpha_2^e + \sum_{j \in H_{-1}} \beta_j^e \chi_{(hr(i)=j)} + \sum_{j \in W_{-1}} \gamma_j^e \chi_{(wd(i)=j)} + \sum_{j \in M_{-1}} \delta_j^e \chi_{(m(i)=j)} + \varepsilon_i^e,\ i \in I. \tag{3.7}$$

At the outset of such a deseasonalization, the series of electricity price residuals ε_i^e, $i \in I$, is obtained.
Furthermore, the hourly residuals are transformed into daily residuals in the following way:

$$\varepsilon_d^e = \frac{1}{24} \sum_{d(i)=d} \varepsilon_i^e,\ d \in D,\ i \in I, \tag{3.8}$$

where the summation condition $d(i) = d$ collects the hourly residuals ε_i^e such that $d(i) = d$, that is, the day of hour i is equal to d. In a few words, the 24 hourly electricity price residuals of each day in the historical series are averaged.

Solar radiation As anticipated, solar radiation faces a simplified treatment, which, first of all, does not include any detrendization. As for the deseasonalization, first daily values of solar radiation based on the hourly series are computed:

$$w_d^r = \left(\sum_{d(i)=d} w_i^r \right),\ d \in D,\ i \in I, \tag{3.9}$$

where the summation condition $d(i) = d$ collects the hourly solar radiations w_i^r such that $d(i) = d$, that is, the day of hour i is equal to d. In a few words, the 24 hourly solar radiations of each day in the historical series are added up. Next, the natural logarithm to the daily solar radiations is applied:

$$\eta_d^r = \ln(w_d^r),\ d \in D. \tag{3.10}$$

The natural logarithm is calculated to avoid obtaining negative values in the simulated series (Section 3.4.5). It is applied to the daily solar radiations instead of the hourly solar radiations to avoid the natural logarithm of null values during the night hours.
Finally, a yearly seasonality based on calendar days is estimated, therefore weekly seasonalities and yearly seasonalities on a monthly basis are estimated jointly. The linear regression model is:

$$\eta_d^r = \alpha_2^r + \sum_{j \in C_{-1}} \phi_j^r \chi_{(cd(d)=j)} + \varepsilon_d^r,\ d \in D. \tag{3.11}$$

The function $cd(d)$ transforms the index $d \in D$ into elements of C: $cd(d)$ is the calendar day in C corresponding to day $d \in D$. The dummy explanatory variable for calendar day $j \in C_{-1}$ is $\chi_{(cd(d)=j)}$, $d \in D$. The regression coefficients ϕ_j^r capture the yearly seasonality based on calendar days. Finally, the intercept is α_2^r. At the outset of such a deseasonalization, the resulting series of solar radiation residuals is ε_d^r, $d \in D$.

3.3.4 Cluster analysis and transition probability matrix

The series of quadruplets of residuals of gas prices, load, electricity prices, and solar radiation obtained from the corresponding adjusted series w_d^g, w_i^ℓ, w_i^e, and w_i^r after the steps explained in Sections 3.3.2 and 3.3.3 is labelled as $\boldsymbol{\varepsilon}_d = (\varepsilon_d^g, \varepsilon_d^\ell, \varepsilon_d^e, \varepsilon_d^r)$, $d \in D$. Such a series serves as the input of a cluster analysis aimed at partitioning the values of each component into k adjacent intervals. The output of such a procedure pairs the series of residuals $\boldsymbol{\varepsilon}_d$ with the series of corresponding states $\mathbf{s}_d = (s_d^g, s_d^\ell, s_d^e, s_d^r)$, $d \in D$. The series of states is, therefore, the realization of the discretized stochastic process corresponding to the stochastic process of residuals (whose realization is the series of residuals). The two stochastic processes were labelled as $\{\mathbf{S}_d\}_{d\in D}$ and $\{\mathbf{Z}_d\}_{d\in D}$, respectively, in Section 2.2. Based on the series of states, a Markov Chain of order 2 is fit by estimating the corresponding transition probability matrix. The cluster analysis and the estimation of the transition probability matrix are detailed in Section 3.4.3.

3.4 Application

3.4.1 Data sources and treatment

The historical series of gas prices, electricity prices, electricity load, and solar radiation are collected over 13 years, from January 1, 2010 to December 31, 2022. Given such a historical period, the two sets I and D of Section 3.3.2 become, respectively:

- $I = \{\textit{01-01-2010 00:00}, \ldots, \textit{12-31-2022 23:00}\} \equiv \{0, \ldots, 113,879\}$,
- $D = \{\textit{01-01-2010}, \ldots, \textit{12-31-2022}\} \equiv \{0, \ldots, 4,744\}$.

These sets do not include values of February 29 for leap years 2012, 2016, and 2020. Such an exclusion allows the application of homogeneous treatments, especially to simplify the simulation step by dealing with a standard yearly length of 365 days. Furthermore, the set of historical years T is introduced:

- $T = \{\textit{2010}, \ldots, \textit{2022}\} \equiv \{0, \ldots, 12\}$.

Since each series comes from a different source, a different treatment is required before detrendization.

Gas prices Gas prices are expressed in €/MWh and have daily frequency. They are related to the EEX Natural Gas Trading Hub Europe Index recorded by the London Stock Exchange Group (LSEG).[2] These historical data are labelled as u_d^g, $d \in D$. Because gas prices are not quoted at some weekends and some holidays and considering that the other 3 series are quoted all the days of a year, the missing dates are filled in with the last available gas price (forward filling method; for example, Friday, December 23 was copied on December 24, 25, and 26). The adjusted series is labelled as w_d^g.

[2]Source of data: LSEG Workspace, EEX Natural Gas Trading Hub Europe Index (EEX-NCG-D).

Load Load is expressed in MWh and has an Italian hourly indexing. It is related to the Italian day-ahead market total purchases (Italian hourly load) recorded by Gestore del Mercato Elettrico (GME).[3] These historical data are labelled as u_i^ℓ, $i \in I$. In the case of time switches from daylight saving time to standard time, 25 values were encountered; in the case of time switches in the opposite sense, only 23 values were downloaded. The series has been adjusted by cutting one hour for days with 25 prices and adding one hour for days with 23 prices. In particular, the 2:00 am price when switching from daylight saving time to standard time (last Sunday of October) was removed and the 2:00 am price when switching from standard time to daylight saving time (last Sunday of March) was copied to the next hour. Therefore, based on the Italian hourly load available at the GME website,[3] the Italian hourly average load was computed in the following way:

$$\bar{u}^\ell = \frac{\sum_{i \in I} u_i^\ell}{113{,}880} \cdot 1{,}000 = 33{,}521{,}714 \quad \text{(kWh)},$$

where $|I| = 113{,}880$. The hypothesis that, on average, a household consumes 2,400 kWh per year (365 days, or 8,760 hours) is introduced to compute the hourly average load per household in the following way:

$$\bar{o}^\ell = \frac{2{,}400}{8{,}760} = 0.2739726 \quad \text{(kWh)}.$$

The ratio

$$f^\ell = \frac{\bar{o}^\ell}{\bar{u}^\ell} = 8.1729889\text{E-}09$$

is a pure number and links the hourly average load per household ($\bar{o}^\ell$) with the Italian hourly average load ($\bar{u}^\ell$). The Italian hourly load u_i^ℓ, $i \in I$, is multiplied by f^ℓ and by 1,000 to obtain

$$w_i^\ell = u_i^\ell \cdot f^\ell \cdot 1{,}000 \quad \text{(MWh)},$$

which is the hourly load per household.

Electricity prices Electricity prices are expressed in €/MWh and have Italian hourly indexing. They are related to the Italian day-ahead market single national price recorded by GME.[4] These historical data are labelled as u_i^e, $i \in I$. Similarly to load, the series was adjusted by cutting one hour for days with 25 prices and adding one hour for days with 23 prices. The adjusted series is labelled as w_i^e.

Solar radiation Solar radiation is expressed in Wh/m^2 and has Universal local time reference (UT). It is related to the global horizontal all sky irradiation recorded in Brescia, Italy, by Copernicus Atmosphere Monitoring Service (CAMS).[5] These historical data are labelled as u_i^r, $i \in I$. The UT time of the historical series was transformed into the corresponding Italian hourly indexing, and the adjustments for the shifts from daylight saving time to standard time, and vice versa, described above for load and electricity prices, were applied. The adjusted series are labelled as w_i^r.

[3]Source of data: Gestore del Mercato Elettrico (GME), `https://www.mercatoelettrico.org/It/download/DatiStorici.aspx`, day-ahead market total purchases.

[4]Source of data: Gestore del Mercato Elettrico (GME), `https://www.mercatoelettrico.org/It/download/DatiStorici.aspx`, day-ahead market single national price (PUN, Prezzo Unico Nazionale).

[5]Source of data: Copernicus Atmosphere Monitoring Service (CAMS), `https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-solar-radiation-timeseries?tab=overview` global horizontal all sky irradiation (GHI), that is, the surface solar downward irradiation integrated over the whole spectrum available at ground level, on a horizontal surface in Brescia, Italy, at the following coordinates: latitude 45.5400 (positive North, ISO 19115), longitude 10.2100 (positive East, ISO 19115), altitude 147.00 (m).

3.4.2 Estimation of the models for trend and seasonalities

A linear regression detrendization model for gas prices, electricity prices, and load was estimated (Section 3.3.2), and, subsequently, a linear regression deseasonalization model for all four series (Section 3.3.3). The coefficients of both models were estimated through OLS and are included in Table 3.1. For illustrative purposes, Figures 3.3, 3.4, and 3.5 show historical logarithmic gas prices, load, and electricity prices with their estimated trend.
Table 3.1 suggests that the majority of the estimated coefficients are statistically significant: indeed, a substantial majority (88%) is significant at the 1% level of confidence, accompanied by a small percentage (2%) of estimated coefficients significant at the 5% level of confidence. This implies that the linear regression models used for detrending and deseasonalizing the data effectively capture significant relationships among the data.
However, a non-negligible percentage (10%) of estimated coefficients is not significant at any of the previous two levels of confidence. In particular, the non-significant coefficients are related to the deseasonalization of solar radiation (7%) and to the deseasonalization of gas prices (3%). The previous results indicate that some aspects of the data are not adequately captured by the model.
At the outset of the previous estimations, the inadequacy of some estimated coefficients is not worrying: indeed, a further estimation is to be performed, which consists in fitting a Markov Chain of order 2 (Section 3.3.4). Such a fit takes care of the remaining unexplained relationships among detrended and deseasonalized residuals, especially of the intricate autodependencies and interdependencies that inevitably characterize the quadrivariate stochastic process under scrutiny.

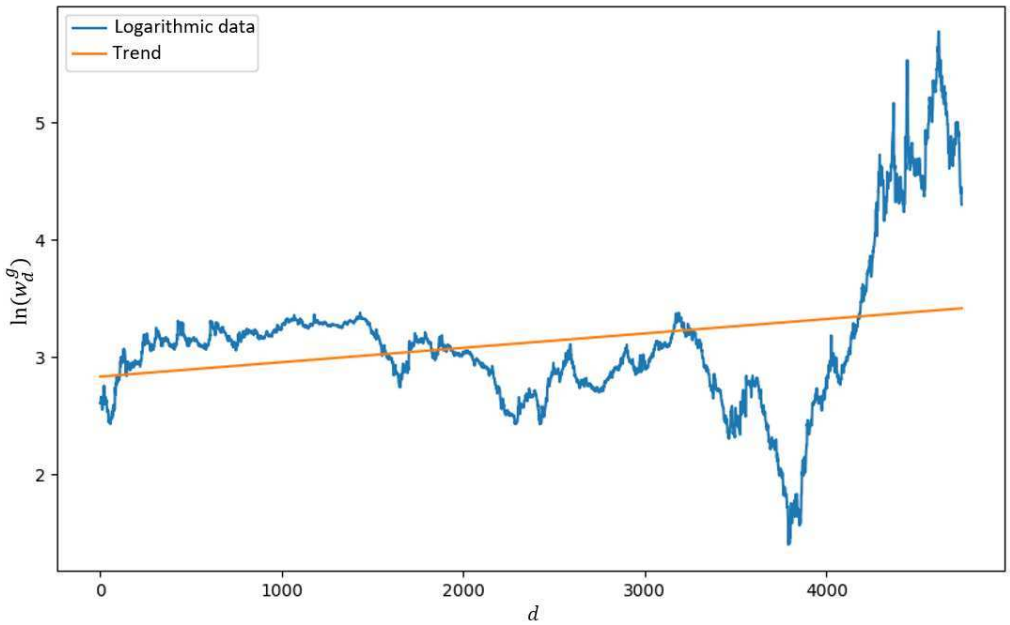

Figure 3.3: Gas prices, in logarithmic form ($ln(w_d^g)$), and their trend over $d \in D$.

3.4.3 Definition of the transition probability matrix

As anticipated in Section 3.3.4, a hierarchical clustering procedure on the residuals of each series based on Euclidean metric and Ward linkage was applied.
The main steps of this procedure are illustrated, by focusing only on the series of daily gas price residuals $\{\varepsilon_0^g, \ldots, \varepsilon_{4744}^g\}$ for the sake of simplicity. Suppose that the hierarchical

Table 3.1: Detrendization and deseasonalization coefficients.

	Gas prices (g)		Load (ℓ)			Electricity prices (e)			Solar radiation (r)
$\alpha_0^{(\cdot)}$ (†)		2.8321**			−1.2797**			3.8608**	
$\alpha_1^{(\cdot)}$ (‡)		0.0001**			−6.523E-07**			5.2579E-06**	
$\alpha_2^{(\cdot)}$ (§)		−0.0284			−0.1235**			−0.0249*	7.2922**
j	γ_j^g (¶)	δ_j^g (¶)	β_j^ℓ (††)	γ_j^ℓ (††)	δ_j^ℓ (††)	β_j^e (‡‡)	γ_j^e (‡‡)	δ_j^e (‡‡)	ϕ_c^r (§§)
1	0.0005	−0.0624	−0.0542**	0.0480**	0.0296**	−0.0950**	0.0261**	−0.0653**	−0.3401**
2	0.0034	−0.0452	−0.0854**	0.0541**	−0.0264**	−0.1640**	0.0327**	−0.1089**	−0.2206**
3	0.0039	−0.0661	−0.0998**	0.0526**	−0.1064**	−0.2152**	0.0352**	−0.1853**	−0.3440**
4	0.0019	−0.0820	−0.0970**	0.0426**	−0.0879**	−0.2185**	0.0269**	−0.1930**	−0.2463**
5	0.0032	−0.0948*	−0.0628**	−0.1024**	−0.0136**	−0.1477**	−0.0539**	−0.1200**	−0.0198
6	0.0004	−0.0437	0.0347**	−0.2147**	0.0785**	−0.0053	−0.1735**	0.0490**	−0.0136
7		0.0336	0.1589**		−0.0737**	0.1067**		0.0588**	−0.1724**
8		0.1397**	0.2615**		−0.0124**	0.1910**		0.0957**	−0.2900**
9		0.1841**	0.3002**		−0.0448**	0.1867**		0.0545**	−0.1837**
10		0.1741**	0.3040**		−0.0243**	0.1371**		0.0560**	−0.1370**
11		0.1741**	0.3020**		−0.0333**	0.0965**		0.1045**	−0.0343
12			0.2670**			0.0105			0.0070
13			0.2475**			−0.0459**			−0.2341**
14			0.2554**			−0.0188			−0.0354
15			0.2622**			0.0396**			0.0322
16			0.2769**			0.1153**			−0.0189
17			0.2987**			0.1978**			0.1219**
18			0.3147**			0.2623**			0.0525
19			0.3239**			0.3177**			0.0312
20			0.3037**			0.3040**			−0.0087
21			0.2535**			0.2260**			0.2310**
22			0.1689**			0.1344**			0.2899**
23			0.0782**			0.0431**			0.2775**
⋮									⋮
364									−0.0366(¶¶)

* Significant at 5% level of confidence.

** Significant at 1% level of confidence.

(†) The coefficient is the intercept of detrendization as estimated in Eqs. (3.1), (3.2), and (3.3).

(‡) The coefficient is the slope of detrendization as estimated in Eqs. (3.1), (3.2), and (3.3).

(§) The coefficient is the intercept of deseasonalization as estimated in Eqs. (3.4), (3.5), (3.7), and (3.11).

(¶) The coefficients are the slopes of deseasonalization as estimated in Eq. (3.4).

(††) The coefficients are the slopes of deseasonalization as estimated in Eq. (3.5).

(‡‡) The coefficients are the slopes of deseasonalization as estimated in Eq. (3.7).

(§§) The coefficients are the slopes of deseasonalization as estimated in Eq. (3.11).

(¶¶) The coefficients for solar radiation from $j = 24$ to $j = 363$ are available upon request

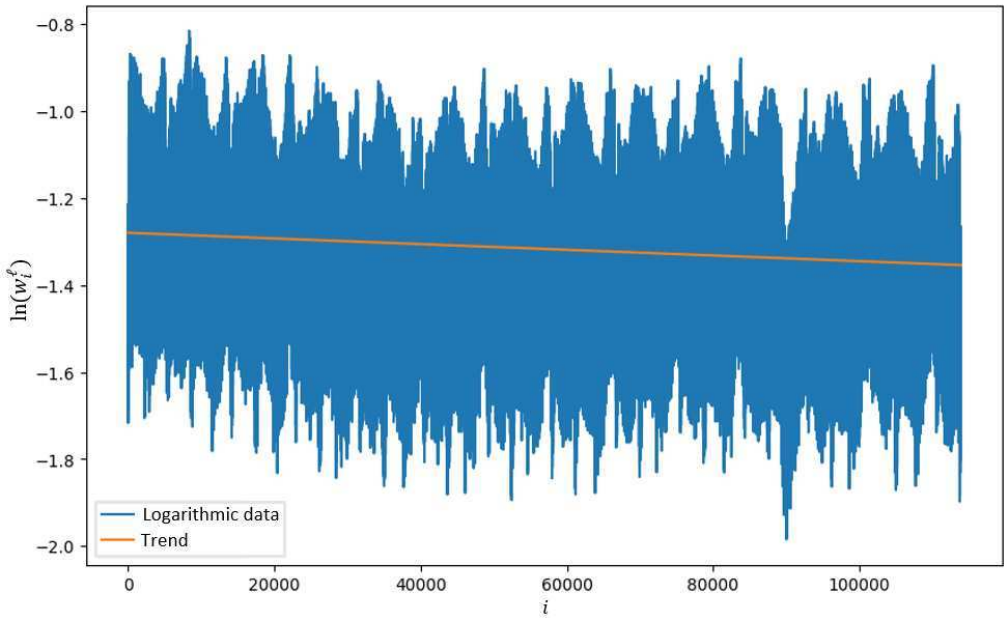

Figure 3.4: Load, in logarithmic form ($ln(w_i^\ell)$), and its trend over $i \in I$.

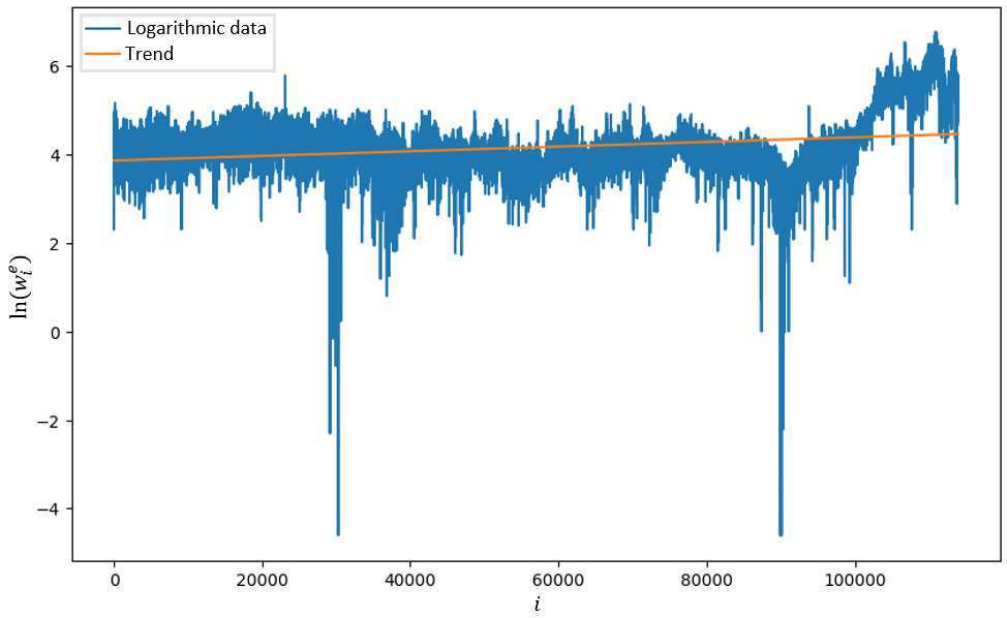

Figure 3.5: Electricity prices, in logarithmic form ($ln(w_i^e)$), and their trend over $i \in I$.

clustering procedure determines the following k adjacent intervals of the domain:

$$[min_{d\in D}\ \varepsilon_d^g, \varepsilon^{g,1}), [\varepsilon^{g,1}, \varepsilon^{g,2}), \ldots, [\varepsilon^{g,k}, max_{d\in D}\ \varepsilon_d^g].$$

The previous k intervals are called states and are labelled, respectively, as:

$$s^{g,1}, s^{g,2}, \ldots, s^{g,k}.$$

Therefore, the series $\{s_0^g, \ldots, s_{4744}^g\}$ can be built, where each element in the list can be one of the possible k labels. The other three series (electricity prices, load, and solar radiation) can be treated similarly.
After completing such an assignment of states, the series of residuals $\boldsymbol{\varepsilon}_d = (\varepsilon_d^g, \varepsilon_d^\ell, \varepsilon_d^e, \varepsilon_d^r)$ is paired with the series of corresponding states $\mathbf{s}_d = (s_d^g, s_d^\ell, s_d^e, s_d^r),\ d \in D$.

The hierarchical clustering procedure can be visualized through the dendrogram. It is a representation of the clusters of residuals obtained from a hierarchical partitioning starting from the singleton partition (where all residuals are separated, i.e., each cluster includes only one residual) and ending with the all-comprehensive partition (where all residuals populate a unique cluster).
Each of the four dendrograms is cut at a common level of $k = \bar{k}$ intervals, thus considering the partitions of each series with the same number of intervals. As a result, the number of (distinct) states that can be observed in the series $\mathbf{s}_d$, $d \in D$, is equal to $\bar{k}^4$. Such (distinct) states can be listed in the set $\mathbf{s}^\nu = (s^1, \ldots, s^\nu, \ldots, s^{\bar{k}^4})$.
Finally, a Markov Chain of order 2 is fit to the series $\mathbf{s}_d$, $d \in D$. In particular, let us count how many times any couple of states in $\mathbf{s}^\nu$ evolved in any state s^ν in the same set, based on the triplets of consecutive states observed in set $\mathbf{s}_d$, with $d \in D$, with the couple of states in the first two positions in the triplet. The transition probability of any state $s^\nu \in \mathbf{s}^\nu$, conditioned on any couple of states in the same set, is obtained by dividing the previous count by the number of times when that couple of states was observed in two consecutive states in $\mathbf{s}_d$, with $d \in D$ (see Section 2.2). For example, suppose that the couple of states

$$(s^1, s^2),\ s^1, s^2 \in \mathbf{s}^\nu$$

with $s^1 = (s^{g,1}, s^{l,2}, s^{e,1}, s^{r,5})$ and $s^2 = (s^{g,3}, s^{l,1}, s^{e,4}, s^{r,2})$, was observed seven times, based on the couples of consecutive states in $\mathbf{s}_d$, $d \in D$. In addition, suppose that the state following the previous couple was observed to be:

- $s^1 = (s^{g,1}, s^{l,2}, s^{e,1}, s^{r,5})$ one time,
- $s^2 = (s^{g,3}, s^{l,1}, s^{e,4}, s^{r,2})$ four times,
- $s^3 = (s^{g,1}, s^{l,2}, s^{e,1}, s^{r,6})$ one time,
- $s^4 = (s^{g,1}, s^{l,3}, s^{e,1}, s^{r,6})$ one time.

The transition probabilities of the previous four states, based on the couple of states (s^1, s^2) at time lags 2 and 1, were, respectively:

- $P(s^1|s^1, s^2) = 0.1429$,
- $P(s^2|s^1, s^2) = 0.5714$,
- $P(s^3|s^1, s^2) = 0.1429$,
- $P(s^4|s^1, s^2) = 0.1429$.

Based on the estimated transition probability matrix, 365 consecutive states were bootstrapped. The chosen length was meant to reproduce a whole year of daily data. Such a series is labelled as $\tilde{\mathbf{s}}_c$, $c \in C$.[6] For ease of understanding, consider time $\bar{c} \in C$. Also. suppose that the bootstrapped states at times $\bar{c} - 1$ and $\bar{c} - 2$ are $\tilde{\mathbf{s}}_{\bar{c}-1}$ and $\tilde{\mathbf{s}}_{\bar{c}-2}$, respectively. The choice of the bootstrapped state $\tilde{\mathbf{s}}_{\bar{c}}$ is based on the transition probability of the bootstrapped couple of consecutive states $(\tilde{\mathbf{s}}_{\bar{c}-2}, \tilde{\mathbf{s}}_{\bar{c}-1})$. The choice of the bootstrapped residuals, $\tilde{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}_c^g, \tilde{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \tilde{\varepsilon}_c^r)$, $c \in C$, corresponding to $\tilde{\mathbf{s}}_c$, $c \in C$, is now detailed. Consider the bootstrapped state $\tilde{\mathbf{s}}_{\bar{c}} = \left(\tilde{s}_{\bar{c}}^g, \tilde{s}_{\bar{c}}^\ell, \tilde{s}_{\bar{c}}^e, \tilde{s}_{\bar{c}}^r\right)$, as obtained through the transition probability matrix. Suppose that $\tilde{s}_{\bar{c}}^g = s^{g,1}$, that is the bootstrapped interval for gas price residuals at time $\bar{c}$ is the first interval in the partition of this series, $s^{g,1}$. Analogously, suppose that

[6] Each bootstrapped series is initialized by picking up the states observed on January 1 and January 2 of each year in the historical sample; those states were assigned to times $c = 0$ and $c = 1$, respectively. The following states of the bootstrapped series, that is those at time $c = 2, \ldots, 364$, were obtained based on the transition probability matrix.

the bootstrapped intervals for the other three components are $\tilde{s}^{\ell}_{\bar{c}} = s^{\ell,3}$, $\tilde{s}^{e}_{\bar{c}} = s^{e,2}$, and $\tilde{s}^{r}_{\bar{c}} = s^{r,1}$, respectively the third, the second, and the first interval of the corresponding partitions. Therefore, the bootstrapped state at time $\bar{c}$ is $(s^{g,1}, s^{\ell,3}, s^{e,2}, s^{r,1})$. Such a state includes "days" of the series of residuals characterized by values of the four components jointly belonging to the previous respective intervals. The choice of the quadruplet of residuals to be associated with the bootstrapped state is based on a discrete uniform probability distribution applied to all quadruplets of residuals belonging to that state. The extracted quadruplet is the element at time $\bar{c}$ of the bootstrapped series of residuals and is labeled as $\tilde{\boldsymbol{\varepsilon}}_{\bar{c}} = (\tilde{\varepsilon}^{g}_{\bar{c}}, \tilde{\varepsilon}^{\ell}_{\bar{c}}, \tilde{\varepsilon}^{e}_{\bar{c}}, \tilde{\varepsilon}^{r}_{\bar{c}})$.
At the end of Section 2.2, the fact that the degree of determinism/stochasticity associated with a Markov Chain of order k depends on the number of transition probabilities equal to 1 was highlighted. In particular, the lower the number of rows populated by transition probabilities equal to 1, the higher the level of stochasticity associated with the Markov Chain of order k. Such a level can be reached by acting on the number of intervals that partition $\mathbb{R}$ for each component when building the discretized stochastic process $\{\mathbf{S}_d\}_{d \in D}$. In this case, the support of each component (gas prices, load, electricity prices, and solar radiation) was partitioned into 5 intervals, which resulted in an estimated transition probability matrix with $2,012$ rows and 293 columns.[7] Such a matrix displays 63.17% of its $2,012$ rows only populated by a unique transition probability (equal to 1).

3.4.4 Intraday distributions and seasonality

The bootstrapped series of residuals, obtained through the steps explained in Sections 3.4.1, 3.4.2, and 3.4.3, are now manipulated to obtain the simulated series. The manipulation consists in adding back the seasonality and the trend to such residuals. These steps are illustrated in Section 3.4.5. However, an important issue needs to be faced before adding back seasonality and trend: the adjustment of daily series of bootstrapped residuals to make them become hourly series of residuals in the case of three components of the quadrivariate stochastic process: load, electricity prices, and solar radiation. Indeed, the bootstrapped residuals are daily, but the frequency of the simulated series is hourly in the previous three cases. Formulas for the transformation of hourly residuals into daily residuals were advanced in the deseasonalization step in Section 3.3.3 (Eqs. (3.6), (3.8), and (3.9)). The transformation allowed to fit a Markov Chain of order 2 to the daily residuals of all components (Section 3.3.4). Now, the hourly residuals for each bootstrapped daily residual of load, solar radiation, and electricity prices must be chosen (Section 3.3.1 motivates why such a choice was performed). In the following, the choice for the three aforementioned components is detailed.

Load Let us begin by considering load and let us focus on the bootstrapped load residual for day $c \in C$ of simulation, $\tilde{\varepsilon}^{\ell}_{c}$. This residual belongs to the bootstrapped state of day c, $\tilde{s}^{\ell}_{c}$, and was chosen based on a discrete uniform probability distribution applied to all residuals in that state (Section 3.4.3). The bootstrapped residual $\tilde{\varepsilon}^{\ell}_{c}$ is one of the residuals of load ε^{ℓ}_{d}, $d \in D$, obtained through Eq. (3.6). First, all residuals of the historical series sharing the same weekday, $wd(d)$, and month, $m(d)$, of $\varepsilon^{\ell}_{d} = \tilde{\varepsilon}^{\ell}_{c}$, are collected in a set, which is labelled as $\boldsymbol{\varepsilon}^{\ell}_{d,all}$. Recalling the example in Section 3.3.1, the previous set collects all Sundays of April in the historical series. Given that 13 years of load are

[7] The rows and columns of the theoretical transition probability matrix are much more than those of the estimated one: indeed, if one counts the possible number of states by combining the 5 intervals partitioning the support of the 4 components, he obtains $5^4 = 625$ states; moreover, if he combines these states with themselves, he gets $625^2 = 390,625$ couples of consecutive states, that is, the conditioning events of the Markov Chain of order 2; finally, the theoretical transition probability matrix includes $390,625$ rows and 625 columns.

being considered, the previous set includes a minimum of $4 \times 13 = 52$ to a maximum of $5 \times 13 = 65$ Sundays of April (April can include 4 or 5 Sundays). Second, the percentile of ε_d^ℓ with respect to the elements of $\boldsymbol{\varepsilon}_{d,all}^\ell$ is computed and such a number is called $\varepsilon_{d,perc}^\ell$. Third, another set of load residuals by collecting all residuals of the historical series sharing the same weekday and month of day $c \in C$ that is being simulated, that is, $wd(d) = wd(c)$, $m(d) = m(c)$, $d \in D$ is built and is labelled as $\boldsymbol{\varepsilon}_{c,all}^\ell$. Recalling the example in Section 3.3.1, the previous set collects the load residuals of all Thursdays of April in the historical series. Finally, the element of $\boldsymbol{\varepsilon}_{c,all}^\ell$ with percentile $\varepsilon_{d,perc}^\ell$ is computed and is called $\breve{\varepsilon}_c^\ell$. After this treatment, $\tilde{\varepsilon}_c^\ell$ was transformed into $\breve{\varepsilon}_c^\ell$ by keeping the percentile of $\tilde{\varepsilon}_c^\ell = \varepsilon_d^\ell$ in set $\boldsymbol{\varepsilon}_{d,all}^\ell$ equal to the percentile of $\breve{\varepsilon}_c^\ell$ in set $\boldsymbol{\varepsilon}_{c,all}^\ell$.

Solar radiation Let us focus on the bootstrapped solar radiation residual for calendar day $c \in C$ of the simulation, $\tilde{\varepsilon}_c^r$. This residual belongs to the bootstrapped state of day c, $\tilde{s}_c^r$, and was chosen based on a discrete uniform probability distribution applied to all residuals in that state (Section 3.4.3). The bootstrapped residual $\tilde{\varepsilon}_c^r$ is one of the residuals of solar radiation ε_d^r, $d \in D$, obtained through Eq. (3.9). First, all residuals of the historical series sharing the same calendar day, $cd(d)$, of $\varepsilon_d^r = \tilde{\varepsilon}_c^r$ are collected in a set labelled as $\boldsymbol{\varepsilon}_{d,all}^r$. For example, suppose that one wants to simulate December 20 based on a bootstrapped residual of June 15. Thus, the previous set collects all June 15 in the historical series. Given that 13 years of solar radiation are being considered, the previous set includes 13 such days. Second, the percentile of ε_d^r with respect to the elements of $\boldsymbol{\varepsilon}_{d,all}^r$ is calculated and is called $\varepsilon_{d,perc}^r$. Third, another set of solar radiation residuals is built, by collecting all residuals of the historical series sharing the same day $c \in C$ that is being simulated, that is, $cd(d) = c$, $d \in D$; such a set is labelled as $\boldsymbol{\varepsilon}_{c,all}^r$. To proceed with the example, the previous set collects the solar radiation residuals of all December 20 in the historical series. Finally, the element of $\boldsymbol{\varepsilon}_{c,all}^r$ with percentile $\varepsilon_{d,perc}^r$ was calculated and is called $\breve{\varepsilon}_c^r$.

Electricity prices Finally, the third component, electricity prices, did not undergo the aforementioned treatments for load and solar radiation.

After the previous treatments, the bootstrapped series of residuals, that is, $\tilde{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}_c^g, \tilde{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \tilde{\varepsilon}_c^r)$, becomes $\dot{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}_c^g, \breve{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \breve{\varepsilon}_c^r)$, $c \in C$.

3.4.5 Reseasonalization and retrendization

Given the residuals $\dot{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}_c^g, \breve{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \breve{\varepsilon}_c^r)$, $c \in C$, obtained in the previous section, now one can simulate all four series with a reseasonalization and a retrendization of such residuals. Remember that simulations are going to be one year long. Each simulated year is comprised of 365 days, which are identified by the elements of set $C = \{0, \ldots, 364\}$.

Simulated gas prices The simulation step takes the series of gas price residuals $\tilde{\varepsilon}_c^g$, $c \in C$, and adds back the seasonalities and the trend based on the estimated coefficients in Table 3.1, finally exponentiates the reseasonalized and retrended residuals. Seasonalities and trends are added back following Eq. (3.4) and (3.1), respectively. The reseasonalization proceeds as in the following:

$$\tilde{\eta}_c^g = \alpha_2^g + \sum_{j \in W_{-1}} \gamma_j^g \chi_{(wd(c)=j)} + \sum_{j \in M_{-1}} \delta_j^g \chi_{(m(c)=j)} + \tilde{\varepsilon}_c^g, \; c \in C,$$

where $wd(c)$ is the weekday in W_{-1} corresponding to calendar day c and $m(c)$ is the month in M_{-1} corresponding to calendar day c.
The retrendization and the exponentiation proceed as in the following:

$$\tilde{w}_c^g = e^{\alpha_0^g + \alpha_1^g \, d_c + \tilde{\eta}_c^g}, \; c \in C,$$

where

$$d_c = |D| + y \cdot 365 + c,$$

- $|D| = 4,745$ is the cardinality of set D,
- $y =$ simulated year $- 2023$.

The previous equation transforms $c \in C$ into the corresponding $d \in D$. This "renumbering" of the simulated day is needed to take into account the fact that the simulation is one year of data long, but the detrendization was performed on the whole sample of 13 years.
To give an example of such a transformation, let us assume that the simulated year is 2010. The first simulated day is identified with $c = 0$. Then, the calculation of d_0 needed for the retrendization gives $d_0 = 4,745 + (2010 - 2023) \cdot 365 + 0 = 0$, which is the value that the simulated day had in D when detrending. Let us make another example: if the year to simulate is 2022, then January 1 is associated with $c = 0$ and corresponds to $d_0 = 4,745 + (2022 - 2023) \cdot 365 + 0 = 4,745 - 365 + 0 = 4,380$ in D; likewise, the last day of 2022 is associated with $c = 364$ and corresponds to $d_{364} = 4,380 + 364 = 4,744$ in D.
Even more, the computation of y allows to simulate both years in the original sample (2010 to 2022) and years outside this sample. For example, if one wants to simulate year 2023, which is the first year after the last year of the original sample, then the first day of 2023, associated with $c = 0$, corresponds to $d_0 = 4,745 + (2023 - 2023) \cdot 365 + 0 = 4,745$; this number identifies January 1, 2023 and, although it does not belong to D, it immediately follows the last number in D, 4,745, that corresponds to December 31, 2022.

Simulated load Let us now proceed to simulate hourly load. First of all, the simulation step replaces each daily load residual $\breve{\varepsilon}_c^\ell$ with its 24 hourly residuals $\breve{\varepsilon}_{h,c}^\ell$, $h \in H$, $c \in C$. Then, the seasonalities and trend based on the estimated coefficients in Table 3.1 are added back. Finally, the resulting values are exponentiated. Seasonalities and trend are added back following Eqs. (3.5) and (3.2), respectively. The reseasonalization proceeds as in the following:

$$\tilde{\eta}_{h,c}^\ell = \alpha_2^\ell + \sum_{j \in H_{-1}} \beta_j^\ell \chi_{(hr(q)=j)} \sum_{j \in W_{-1}} \gamma_j^\ell \chi_{(wd(q)=j)} + \sum_{j \in M_{-1}} \delta_j^\ell \chi_{(m(q)=j)} + \breve{\varepsilon}_{h,c}^\ell, \; h \in H, c \in C,$$

where $q = c \cdot 24 + h$ counts the hour to be simulated. As an example, if one wants to simulate the last hour of the third calendar day, he refers to $h = 23$ and $c = 2$. Therefore, 11:00 pm of the third calendar day is the 71st hour of that simulated year and is computed as $q = c \cdot 24 + h = 2 \cdot 24 + 23 = 71$. Consequently, the functions $hr(q)$, $wd(q)$, and $m(q)$ determine the hour in H_{-1}, the weekday in W_{-1}, and the month in M_{-1} of the simulated year, respectively. The retrendization and the exponentiation proceed as in the following:

$$\tilde{w}_{h,c}^\ell = e^{\alpha_0^\ell + \alpha_1^\ell \, i_{h,c} + \tilde{\eta}_{h,c}^\ell}, \; h \in H, \; c \in C,$$

where

$$i_{h,c} = |I| + y \cdot 8,760 + c \cdot 24 + h,$$

- $|I| = 113,880$ is the cardinality of set I,
- $y =$ simulated year $- 2023$.

Similarly to gas prices simulation, the previous equation transforms $c \in C$ into the corresponding $i \in I$. To give an example of such a transformation, let us assume that the simulated year is 2010. The second simulated hour ($h = 1$) of the first day ($c = 0$) corresponds to $i = 1$ in I. Then, the calculation of $i_{1,0}$ needed for the retrendization gives $i_{1,0} = 113,880 + (2010 - 2023) \cdot 8,760 + 1 = 1$, which is the value that the simulated hour had in I when detrending. Let us provide another example: if the chosen simulated year is 2022, then 10:00 am ($h = 11$) of January 1 ($c = 0$) is associated with $i_{11,0} = 113,880 + (2022 - 2023) \cdot 8,760 + 11 = 113,880 - 8,760 + 11 = 105,131$ in I; likewise, the last hour ($h = 23$) of the last day ($c = 364$) of 2022 is $i_{23,364} = 113,880 - 8,760 + 364 \cdot 24 + 23 = 113,879$ in I.

Simulated electricity prices The simulation of electricity prices is similar to that of load. To simulate hourly values, each daily residual $\tilde{\varepsilon}^e_c$ is replaced with its 24 hourly residuals, $\tilde{\varepsilon}^e_{h,c}$, $h \in H$, $c \in C$. Then, the seasonalities and trend based on the estimated coefficients in Table 3.1 are added back. Finally, the resulting values are exponentiated. Seasonalities and trends are added back based on Eqs. (3.7) and (3.3), respectively. The reseasonalization proceeds as in the following:

$$\tilde{\eta}^e_{h,c} = \alpha^e_2 + \sum_{j \in H_{-1}} \beta^e_j \chi_{(hr(q)=j)} \sum_{j \in W_{-1}} \gamma^e_j \chi_{(wd(q)=j)} + \sum_{j \in M_{-1}} \delta^e_j \chi_{(m(q)=j)} + \tilde{\varepsilon}^e_{h,c},\ h \in H,\ c \in C,$$

with $q = c \cdot 24 + h$.
The retrendization proceeds as in the following:

$$\tilde{w}^e_{h,c} = e^{\alpha^e_0 + \alpha^e_1\, i_{h,c} + \tilde{\eta}^e_{h,c}},\ h \in H,\ c \in C,$$

where

$$i_{h,c} = |I| + y \cdot 8760 + c \cdot 24 + h.$$

Simulated solar radiation The simulation of hourly solar radiation only includes the reseasonalization of daily residuals based on the coefficients in Table 3.1. Section 3.3.2 explains the reason why a detrendization was not performed and, consequently, a retrendization is not needed now. The reseasonalization proceeds as in the following:

$$\tilde{\eta}^r_c = \alpha^r_2 + \sum_{j \in C_{-1}} \phi^r_j \chi_{(cd(c)=j)} + \tilde{\varepsilon}^r_c,\ c \in C,$$

where $cd(c)$ is the calendar day in C_{-1} corresponding to calendar day c. To the purpose of simulating hourly values, daily simulated residuals $\tilde{\eta}^r_c$ are exponentiated in the following way:

$$\tilde{w}^r_{h,c} = \frac{\bar{w}^r_{h,c}}{\sum_{h \in H} \bar{w}^r_{h,c}}\, e^{\tilde{\eta}^r_c},\ h \in H,\ c \in C, \tag{3.12}$$

where

$$\bar{w}^r_{h,c} = \frac{1}{|T|} \sum_{i \in I} w^r_i \chi_{(hr(i)=h, cd(i)=c)},\ h \in H,\ c \in C. \tag{3.13}$$

The computations $hr(i) = h$ and $cd(i) = c$ transform $i \in I$ into the corresponding hour $h \in H$ and calendar day $c \in C$, respectively. The cardinality of set T, $|T|$, is equal to 13. The exponentiation in Eq. (3.12) also includes the multiplication of each value by weight $\bar{w}^r_{h,c}$, $h \in H$, $c \in C$, which is the average of solar radiations historically observed for hour h in calendar day c. The reason for such weighting is to avoid the use of exponentiated (positive) values for simulating solar radiation at night hours when typically null values are realised. Each weight, based on the historical series, is normalized by the sum of

weights per calendar day $c \in C$.
For example, if $h = 2$ and $c = 89$, then one is considering 02:00 am of March 30. Eq. (3.13) calculates $\bar{w}^r_{2,89}$ as the sum of historical solar radiations at that hour of that day; given the 13 years of data, there are 13 such observations ($\chi_{(hr(i)=2,cd(i)=89)} = 1$ for 13 values of i and is equal to 0 for the other values of i). These observations are averaged by dividing them by $|T|= 13$.

3.5 Results

Following the approach adopted in Cerqueti et al. (2017a) to appreciate the quality of the bootstrapping and simulation method, several statistics on the simulated series are calculated and compared with the same statistics calculated on the historical series. In particular, the historical series w^g_d, w^ℓ_i, w^e_i, w^r_i, $d \in D$, $i \in I$ (Section 3.4.1) and the simulated series $\tilde{w}^g_c$, $\tilde{w}^\ell_{h,c}$, $\tilde{w}^e_{h,c}$, $\tilde{w}^r_{h,c}$, $c \in C$, $h \in H$ (Section 3.4.5) were considered. The univariate statistics of mean, standard deviation, skewness, kurtosis, minimum, and maximum to minimally describe such series were computed (Section 3.5.1). In addition, univariate auto-regressive models to the previous series were fit. The univariate autoregression coefficients of such models to highlight auto-dependencies in the series were calculated (Section 3.5.2). The previous univariate statistics and autoregression coefficients allowed to appreciate the goodness-of-fit of this method for each component of the quadrivariate stochastic process under scrutiny. Since this method is powerful in capturing interdependencies among series, a test aimed at detecting such interdependencies in the simulated series and comparing them with the corresponding interdependencies in the historical series (Section 3.5.3) was devised. In particular, a quadrivariate auto-regressive model to the series of simulated residuals, $\dot{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}^g_c, \breve{\varepsilon}^\ell_c, \tilde{\varepsilon}^e_c, \breve{\varepsilon}^r_c)$, $c \in C$, as opposed to the series of historical residuals, $\boldsymbol{\varepsilon}_d = (\varepsilon^g_d, \varepsilon^\ell_d, \varepsilon^e_d, \varepsilon^r_d)$, $d \in D$, was fit. Since vector auto-regressive models require series to be stationary, the previous test was based on the series of residuals, which do not include trend and seasonality.
20 annual series of 365 days per each year of the observation period (2010-2022) were simulated. Since the observation period includes 13 years, $13 \times 20 = 260$ series were obtained. Each series starts on January 1 of each year in the observation period (footnote 6 at page 67 provides a brief explanation of the initialization values needed for the bootstrapping step). In addition, the set D of time indexes of historical days was partitioned into 13 elements: $D_1, \ldots, D_{13}$. Each set includes 365 consecutive time index values: $D_1 = \{0, \ldots, 364\}, D_2 = \{365, \ldots, 729\}, \ldots, D_{13} = \{4{,}380, \ldots, 4{,}744\}$. The previous partition corresponds to cutting the daily historical series into 13 annual historical series, each beginning on January 1.
Given the 260 simulated series, on the one side, and the 13 annual historical series, on the other side, two frequency distributions for each univariate statistic and autoregression coefficient could be built: a distribution of values calculated on the simulated series (simulated distribution) and a distribution of values calculated on the annual historical series (historical distribution). It should be expected that an acceptable simulation comes with distributions of the aforementioned univariate statistics and autoregression coefficients that very much resemble the companion distributions based on the annual historical series. In the case of the quadrivariate auto-regressive model, the distributions of autoregression coefficients estimated on simulated residuals were compared with the analogous distributions estimated on historical residuals.
The previous presentation is suitable for the daily series of historical and simulated residuals of all 4 components and for the daily series of gas prices. In the cases of load, electricity prices, and solar radiation, the statistics on hourly series covering 365 days were calculated. Consequently, the set I of time indexes of historical hours was par-

titioned into 13 components: $I_1 = \{0,\ldots,8{,}759\}, I_2 = \{8{,}760,\ldots,17{,}519\},\ldots,I_{13} = \{105{,}120,\ldots,113{,}879\}$. The previous partition corresponds to cutting the hourly historical series into 13 annual historical series, each beginning at 00:00 on January 1.

3.5.1 Univariate statistics: mean, standard deviation, skewness, kurtosis, minimum, and maximum

Figures 3.6, 3.7, 3.8, and 3.9 show the relative frequency distributions of mean, standard deviation, skewness, kurtosis, minimum, and maximum calculated on the 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right). The simulations were based on a Markov Chain of order 2 characterized by 5 states for all components of the quadrivariate stochastic process.
As a general comment for all 4 components, the simulated distributions are centred to the historical distributions, with almost all simulated values included in the range of historical values. In addition, is important to observe that such a result was obtained based on a Markov Chain of order 2 with 5 states for each component, which corresponds to a level of determinism of 63.17% (end of Section 3.4.3). It may seem that such a high level of determinism could produce the centered distributions observed in the aforementioned figures but, on the other side, the level of stochasticity, which is the complement to 100% of the previous level of determinism, is not negligible at all. Such a high level of stochasticity would be a tough obstacle to many simulation methods, except for this one. The power of this method is tested not only through the previous univariate statistics, but also with statistics that capture further aspects of the simulated series, especially their intricate autodependencies and interdependencies (Sections 3.5.2 and 3.5.3).

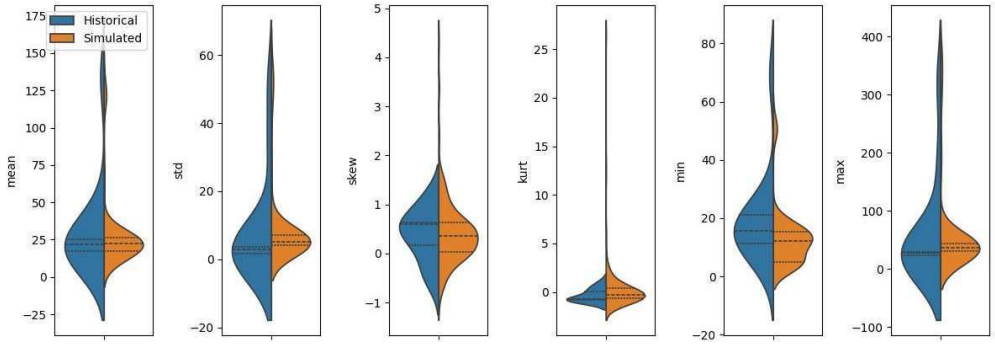

Figure 3.6: Gas prices - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right).

Figure 3.10 shows the univariate statistics of load before applying the adjustment discussed in Sections 3.3.1 and 3.4.4. Without this adjustment, it is clear that there is a misalignment between the historical and simulated distributions of skewness, kurtosis, and minimum. In particular, the simulated series show a negative skewness, opposite to the annual historical series. This is reflected in the minimum values reached by the simulated series, which are lower than the minima reached by the historical series. To reinforce the previous observation, the simulated distribution is less platykurtic compared to the historical one: indeed, the former has thinner but longer tails than the latter, thus allowing for lower minima in the former distribution compared to the minima in the latter distribution.
Similar observations can be carried out for solar radiation, which is the other component of the quadrivariate stochastic process undergoing the aforementioned adjustment.

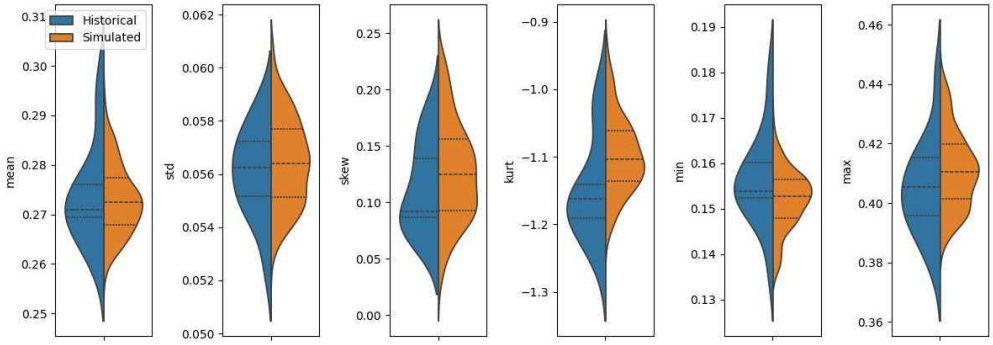

Figure 3.7: Load - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right).

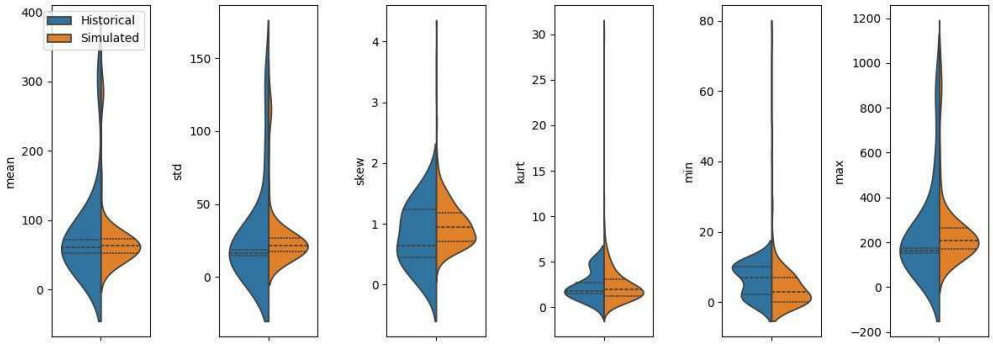

Figure 3.8: Electricity prices - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right).

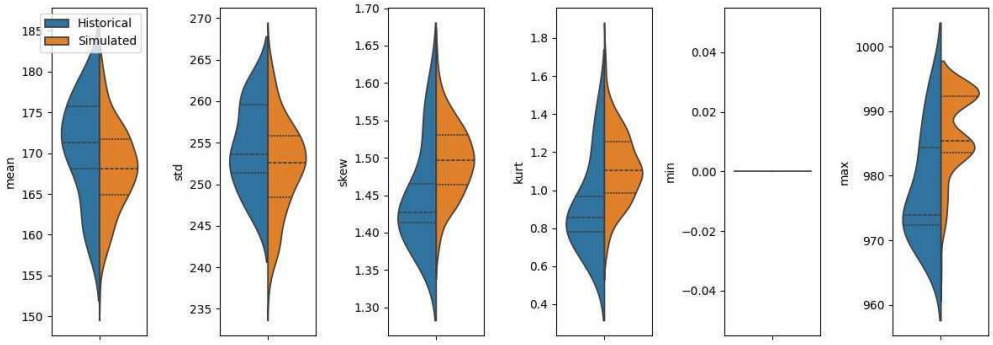

Figure 3.9: Solar radiation - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right).

Figure 3.11 shows that, without this adjustment, there is a misalignment between the historical and simulated distributions of skewness, kurtosis, and maximum. In particular, the simulated series show more positive skewness and kurtosis than the annual historical series. This misalignment causes the simulated series to reach maximum values that are well beyond the maxima reached by the historical series.
The method advanced in this work takes into account the previous distortions through the

adjustments explained in Sections 3.3.1 and 3.4.4. Figures 3.10 and 3.11 witness the serious flaw of this method in the absence of such an adjustment. It should also be mentioned that Figures 3.10 and 3.11 show the results of simulations based on a Markov Chain of order 2 characterized by a number of states for each component well above the 5 states adopted for the simulation including the adjustment. Indeed, 30 states for gas prices, 25 states for load, and 15 states for both electricity prices and solar radiation were adopted, which meant a much higher level of determinism (99.58%) in the transition probability matrix than the one associated with using 5 states per each component (63.17%). The power of this adjustment is, thus, reinforced by the fact that the misalignment shown in Figures 3.10 and 3.11, based on a high level of determinism (99.58%), is solved even for a much lower level of determinism (63.17%), as it is shown in Figures 3.7 and 3.9.

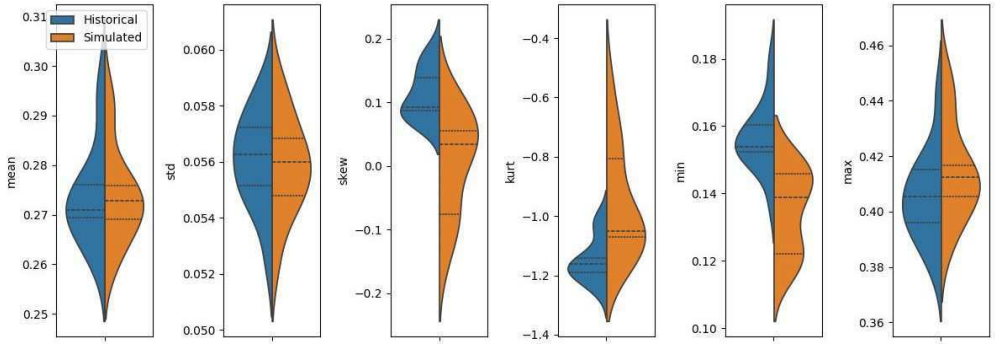

Figure 3.10: Load - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right), with simulations based on a Markov Chain of order 2 characterized by 30 states for gas prices, 25 states for load, and 15 states for both electricity prices and solar radiation. These results were obtained *before* applying the adjustment discussed in Sections 3.3.1 and 3.4.4.

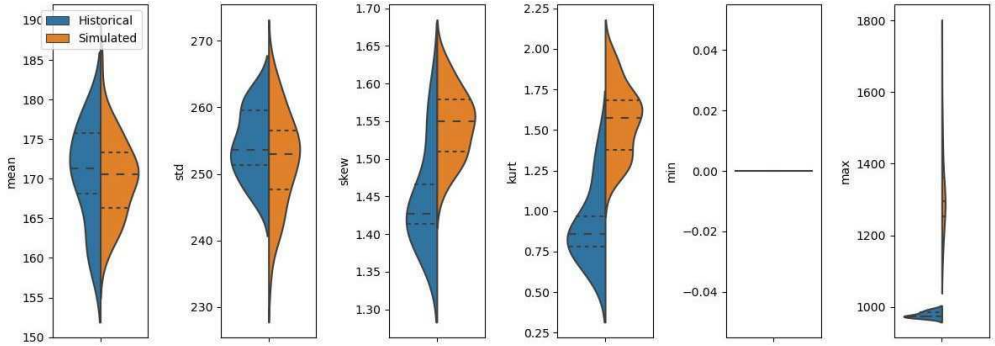

Figure 3.11: Solar radiation - Relative frequency distributions of mean, standard deviation (std), skewness (skew), kurtosis (kurt), minimum (min), and maximum (max) calculated on 13 annual historical series (historical distribution, on the left) and on the 260 simulated series (simulated distribution, on the right), with simulations based on a Markov Chain of order 2 characterized by 30 states for gas prices, 25 states for load, and 15 states for both electricity prices and solar radiation. These results were obtained *before* applying the adjustment discussed in Sections 3.3.1 and 3.4.4.

3.5.2 Univariate autoregression coefficients

The present section deals with the estimates of univariate autoregression coefficients. Another test of the goodness-of-fit of this bootstrapping and simulation consists in comparing the historical and simulated distributions of estimated coefficients obtained by fitting a univariate auto-regressive model to both sets of series. The auto-regressive model is fit only to the hourly series of load, electricity prices, and solar radiation (the daily gas prices are excluded from this analysis, although they are considered in the daily quadrivariate auto-regressive model of Section 3.5.3).
Each univariate auto-regressive model includes 5 lags: 1 hour, 2 hours, 1 day, 2 days, and 7 days. More explicitly, the three models are specified by the following equations:

- load:

$$\begin{aligned} \tilde{w}^{\ell}_{h,c} &= a^{\ell}_0 + a^{\ell}_1\, \tilde{w}^{\ell}_{h-1,c} + a^{\ell}_2\, \tilde{w}^{\ell}_{h-2,c} \\ &+ a^{\ell}_3\, \tilde{w}^{\ell}_{h,c-1} + a^{\ell}_4\, \tilde{w}^{\ell}_{h,c-2} + a^{\ell}_5\, \tilde{w}^{\ell}_{h,c-7} + \varepsilon^{\ell}_{h,c}; \end{aligned} \tag{3.14}$$

- electricity prices:

$$\begin{aligned} \tilde{w}^{e}_{h,c} &= a^{e}_0 + a^{e}_1\, \tilde{w}^{e}_{h-1,c} + a^{e}_2\, \tilde{w}^{e}_{h-2,c} \\ &+ a^{e}_3\, \tilde{w}^{e}_{h,c-1} + a^{e}_4\, \tilde{w}^{e}_{h,c-2} + a^{e}_5\, \tilde{w}^{e}_{h,c-7} + \varepsilon^{e}_{h,c}; \end{aligned} \tag{3.15}$$

- solar radiation:

$$\begin{aligned} \tilde{w}^{r}_{h,c} &= a^{r}_0 + a^{r}_1\, \tilde{w}^{r}_{h-1,c} + a^{r}_2\, \tilde{w}^{r}_{h-2,c} \\ &+ a^{r}_3\, \tilde{w}^{r}_{h,c-1} + a^{r}_4\, \tilde{w}^{r}_{h,c-2} + a^{r}_5\, \tilde{w}^{r}_{h,c-7} + \varepsilon^{r}_{h,c}. \end{aligned} \tag{3.16}$$

The model implies the estimation of 6 univariate autoregression coefficients for each equation, a total of 18 coefficients: $a^{(\cdot)}_0, a^{(\cdot)}_1, a^{(\cdot)}_2, a^{(\cdot)}_3, a^{(\cdot)}_4, a^{(\cdot)}_5$, with $(\cdot) = \ell, e, r$.[8]
The previous equations refer to simulated series and were applied to estimate the coefficients for the 260 simulated series of this test. The notation used to refer to lagged values of a simulated series is easily understood: given the value of a simulated series indexed by the pair "h, c", $c \in C$, $h \in H$, the values of that series at time lags of 1 hour, 2 hours, 1 day, 2 days, and 7 days are indexed, respectively, by "$h-1, c$", "$h-2, c$", "$h, c-1$", "$h, c-2$", and "$h, c-7$".
Mutatis mutandis, the previous equations could be written also for the 13 annual historical series used as a benchmark. In this case, $i \in I$ instead of $c \in C$ and $h \in H$ should be used. Moreover, keep in mind that the time index i should cover one year of hourly data, similarly to the hours covered by the combined time indexes c and h used for the simulated series. The series covering the observation period of 13 years is cut into 13 annual historical series based on the partition of I listed at the beginning of the present section. Each resulting series begins on January 1. Finally, a univariate auto-regressive model is fit to each annual historical series. It is observed that, since the annual historical series are indexed by i, $i \in I_j$, $j = 1, \ldots, 13$, the time lags of 1 hour, 2 hours, 1 day, 2 days, and 7 days are, respectively, "$i-1$", "$i-2$", "$i-24$", "$i-48$", and "$i-168$".
Based on Figures 3.12, 3.13, and 3.14, is easily noticeable that the simulated distributions and the historical distributions share similar ranges of values for almost all univariate coefficients in Eqs. (3.14), (3.15), and (3.16). The method successfully captures the intricate autodependencies affecting the scrutinized series even with a high level of stochasticity (36.83%, end of Section 3.4.3) characterizing the fit Markov Chain of order 2 with 5 states for each component.

[8] The coefficients are available upon request.

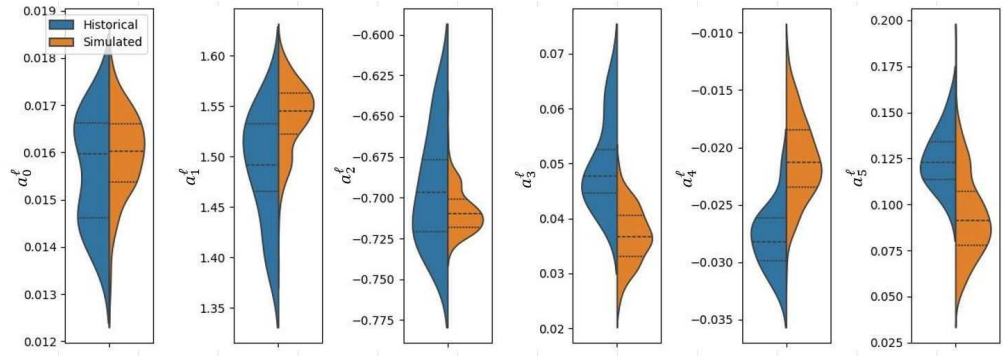

Figure 3.12: Load - Relative frequency distributions of autoregression coefficients a_0^ℓ, a_1^ℓ, a_2^ℓ, a_3^ℓ, a_4^ℓ, and a_5^ℓ in Eq. (3.14) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

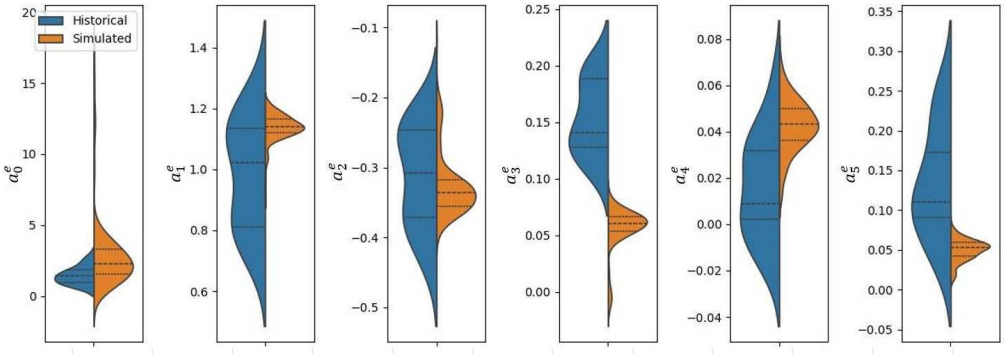

Figure 3.13: Electricity prices - Relative frequency distributions of autoregression coefficients a_0^e, a_1^e, a_2^e, a_3^e, a_4^e, and a_5^e in Eq. (3.15) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

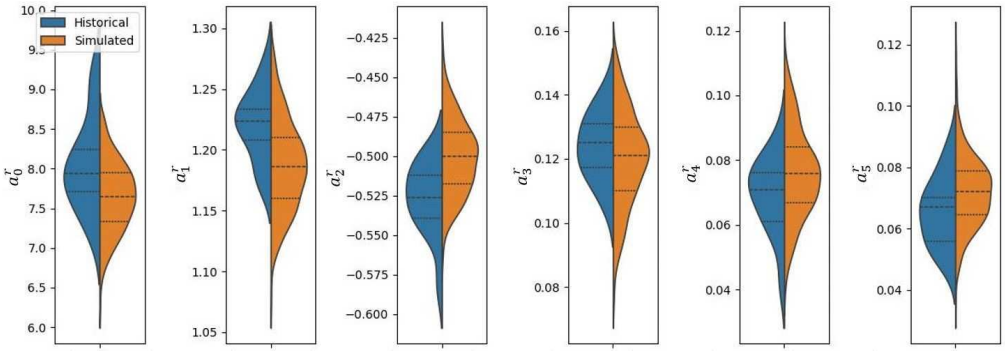

Figure 3.14: Solar radiation - Relative frequency distributions of autoregression coefficients a_0^r, a_1^r, a_2^r, a_3^r, a_4^r, and a_5^r in Eq. (3.16) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

3.5.3 Coefficients of a quadrivariate auto-regressive model

The quadrivariate auto-regressive model tested for interdependencies among residuals includes their lagged values at 1 up to 7 days. The model was fit to the daily series of historical residuals $\boldsymbol{\varepsilon}_d = (\varepsilon_d^g, \varepsilon_d^\ell, \varepsilon_d^e, \varepsilon_d^r)$, $d \in D$, and to the daily series of simulated residuals $\dot{\boldsymbol{\varepsilon}}_c = (\tilde{\varepsilon}_c^g, \breve{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \breve{\varepsilon}_c^r)$, $c \in C$. Since the auto-regressive model assumes that the

time series are stationary, it was applied to the series of residuals, that is, the series obtained after detrendization and deseasonalization. The four equations included in the model, referred to as simulated residuals, are:

- gas prices:

$$\begin{aligned}
\tilde{\varepsilon}^g_c &= b^g_0 \\
&+ b^g_1\,\tilde{\varepsilon}^g_{c-1} + b^g_2\,\breve{\varepsilon}^\ell_{c-1} + b^g_3\,\tilde{\varepsilon}^e_{c-1} + b^g_4\,\breve{\varepsilon}^r_{c-1} \\
&+ b^g_5\,\tilde{\varepsilon}^g_{c-2} + b^g_6\,\breve{\varepsilon}^\ell_{c-2} + b^g_7\,\tilde{\varepsilon}^e_{c-2} + b^g_8\,\breve{\varepsilon}^r_{c-2} \\
&+ \ldots \\
&+ b^g_{25}\,\tilde{\varepsilon}^g_{c-7} + b^g_{26}\,\breve{\varepsilon}^\ell_{c-7} + b^g_{27}\,\tilde{\varepsilon}^e_{c-7} + b^g_{28}\,\breve{\varepsilon}^r_{c-7} \\
&+ \psi^g_c, \qquad (3.17)
\end{aligned}$$

- load:

$$\begin{aligned}
\breve{\varepsilon}^\ell_c &= b^\ell_0 \\
&+ b^\ell_1\,\tilde{\varepsilon}^g_{c-1} + b^\ell_2\,\breve{\varepsilon}^\ell_{c-1} + b^\ell_3\,\tilde{\varepsilon}^e_{c-1} + b^\ell_4\,\breve{\varepsilon}^r_{c-1} \\
&+ b^\ell_5\,\tilde{\varepsilon}^g_{c-2} + b^\ell_5\,\breve{\varepsilon}^\ell_{c-2} + b^\ell_7\,\tilde{\varepsilon}^e_{c-2} + b^\ell_8\,\breve{\varepsilon}^r_{c-2} \\
&+ \ldots \\
&+ b^\ell_{25}\,\tilde{\varepsilon}^g_{c-7} + b^\ell_{26}\,\breve{\varepsilon}^\ell_{c-7} + b^\ell_{27}\,\tilde{\varepsilon}^e_{c-7} + b^\ell_{28}\,\breve{\varepsilon}^r_{c-7} \\
&+ \psi^\ell_c, \qquad (3.18)
\end{aligned}$$

- electricity prices:

$$\begin{aligned}
\tilde{\varepsilon}^e_c &= b^e_0 \\
&+ b^e_1\,\tilde{\varepsilon}^g_{c-1} + b^e_2\,\breve{\varepsilon}^\ell_{c-1} + b^e_3\,\tilde{\varepsilon}^e_{c-1} + b^e_4\,\breve{\varepsilon}^r_{c-1} \\
&+ b^e_5\,\tilde{\varepsilon}^g_{c-2} + b^e_6\,\breve{\varepsilon}^\ell_{c-2} + b^e_7\,\tilde{\varepsilon}^e_{c-2} + b^e_8\,\breve{\varepsilon}^r_{c-2} \\
&+ \ldots \\
&+ b^e_{25}\,\tilde{\varepsilon}^g_{c-7} + b^e_{26}\,\breve{\varepsilon}^\ell_{c-7} + b^e_{27}\,\tilde{\varepsilon}^e_{c-7} + b^e_{28}\,\breve{\varepsilon}^r_{c-7} \\
&+ \psi^e_c, \qquad (3.19)
\end{aligned}$$

- solar radiation:

$$\begin{aligned}
\breve{\varepsilon}^r_c &= b^r_0 \\
&+ b^r_1\,\tilde{\varepsilon}^g_{c-1} + b^r_2\,\breve{\varepsilon}^\ell_{c-1} + b^r_3\,\tilde{\varepsilon}^e_{c-1} + b^r_4\,\breve{\varepsilon}^r_{c-1} \\
&+ b^r_5\,\tilde{\varepsilon}^g_{c-2} + b^r_6\,\breve{\varepsilon}^\ell_{c-2} + b^r_7\,\tilde{\varepsilon}^e_{c-2} + b^r_8\,\breve{\varepsilon}^r_{c-2} \\
&+ \ldots \\
&+ b^r_{25}\,\tilde{\varepsilon}^g_{c-7} + b^r_{26}\,\breve{\varepsilon}^\ell_{c-7} + b^r_{27}\,\tilde{\varepsilon}^e_{c-7} + b^r_{28}\,\breve{\varepsilon}^r_{c-7} \\
&+ \psi^r_c. \qquad (3.20)
\end{aligned}$$

Eqs. (3.17), (3.18), (3.19), and (3.20) show that the each time series at calendar day c are regressed against all series at time lags from 1 (the calendar day before c, $c-1$) up to 7 (seven calendar days before c, $c-7$). The model also includes intercepts for each equation. The autoregression coefficients to be estimated are $b^g_0,\ldots,b^g_{28}$, $b^\ell_0,\ldots,b^\ell_{28}$, $b^e_0,\ldots,b^e_{28}$, and $b^r_0,\ldots,b^r_{28}$, a total of 116 coefficients. The quadrivariate auto-regressive model matrix equation is:

$$\mathbf{v}_c \;=\; \mathbf{b}_0 + \mathbf{B}_1\,\mathbf{v}_{c-1} + \mathbf{B}_2\,\mathbf{v}_{c-2} + \ldots + \mathbf{B}_7\,\mathbf{v}_{c-7} + \boldsymbol{\psi}_c,$$

where

$$\mathbf{v}_c = \begin{bmatrix} \tilde{\varepsilon}_c^g \\ \breve{\varepsilon}_c^\ell \\ \tilde{\varepsilon}_c^e \\ \breve{\varepsilon}_c^r \end{bmatrix}, \quad \mathbf{v}_{c-1} = \begin{bmatrix} \tilde{\varepsilon}_{c-1}^g \\ \breve{\varepsilon}_{c-1}^\ell \\ \tilde{\varepsilon}_{c-1}^e \\ \breve{\varepsilon}_{c-1}^r \end{bmatrix}, \quad \mathbf{v}_{c-2} = \begin{bmatrix} \tilde{\varepsilon}_{c-2}^g \\ \breve{\varepsilon}_{c-2}^\ell \\ \tilde{\varepsilon}_{c-2}^e \\ \breve{\varepsilon}_{c-2}^r \end{bmatrix}, \quad \dots \quad , \mathbf{v}_{c-7} = \begin{bmatrix} \tilde{\varepsilon}_{c-7}^g \\ \breve{\varepsilon}_{c-7}^\ell \\ \tilde{\varepsilon}_{c-7}^e \\ \breve{\varepsilon}_{c-7}^r \end{bmatrix},$$

$$\mathbf{b}_0 = \begin{bmatrix} b_0^g \\ b_0^\ell \\ b_0^e \\ b_0^r \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} b_1^g & b_2^g & b_3^g & b_4^g \\ b_1^\ell & b_2^\ell & b_3^\ell & b_4^\ell \\ b_1^e & b_2^e & b_3^e & b_4^e \\ b_1^r & b_2^r & b_3^r & b_4^r \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} b_5^g & b_6^g & b_7^g & b_8^g \\ b_5^\ell & b_6^\ell & b_7^\ell & b_8^\ell \\ b_5^e & b_6^e & b_7^e & b_8^e \\ b_5^r & b_6^r & b_7^r & b_8^r \end{bmatrix}, \quad \dots \text{ , and}$$

$$\mathbf{B}_7 = \begin{bmatrix} b_{25}^g & b_{26}^g & b_{27}^g & b_{28}^g \\ b_{25}^\ell & b_{26}^\ell & b_{27}^\ell & b_{28}^\ell \\ b_{25}^e & b_{26}^e & b_{27}^e & b_{28}^e \\ b_{25}^r & b_{26}^r & b_{27}^r & b_{28}^r \end{bmatrix}.$$

Mutatis mutandis, the previous equations could be written also for the 13 annual historical series of residuals used as a benchmark and obtained by partitioning the historical series of residuals into 13 segments according to the partition of D listed at the beginning of the present section.

Some coefficients to show the performance of this method were selected, precisely, those in matrices $\mathbf{b}_0$, $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_7$, to test the dependence between today's values and the values of yesterday, two days ago, and one week ago. The historical and simulated distributions associated with the previous coefficients are shown in Figures 3.15, 3.16, 3.17, and 3.18. Each figure shows the distributions of 13 coefficients, precisely the coefficients that share the same row in the previous matrices. For example, Figure 3.15 shows the distributions associated with $b_0^g, b_1^g, \dots, b_4^g, b_5^g, \dots, b_8^g, b_{25}^g, \dots b_{28}^g$, that is, the 13 coefficients that are shown in the first row of $\mathbf{b}_0$, $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_7$.

Figures 3.15, 3.16, 3.17, and 3.18, show that, for almost all coefficients, the simulated distributions and the historical distributions of daily residuals share similar ranges of values. The method successfully captures the intricate autodependencies and interdependencies affecting the scrutinized series of daily residuals even with a high level of stochasticity (36.83%, end of Section 3.4.3) characterizing the fit Markov Chain of order 2 with 5 states for each component.

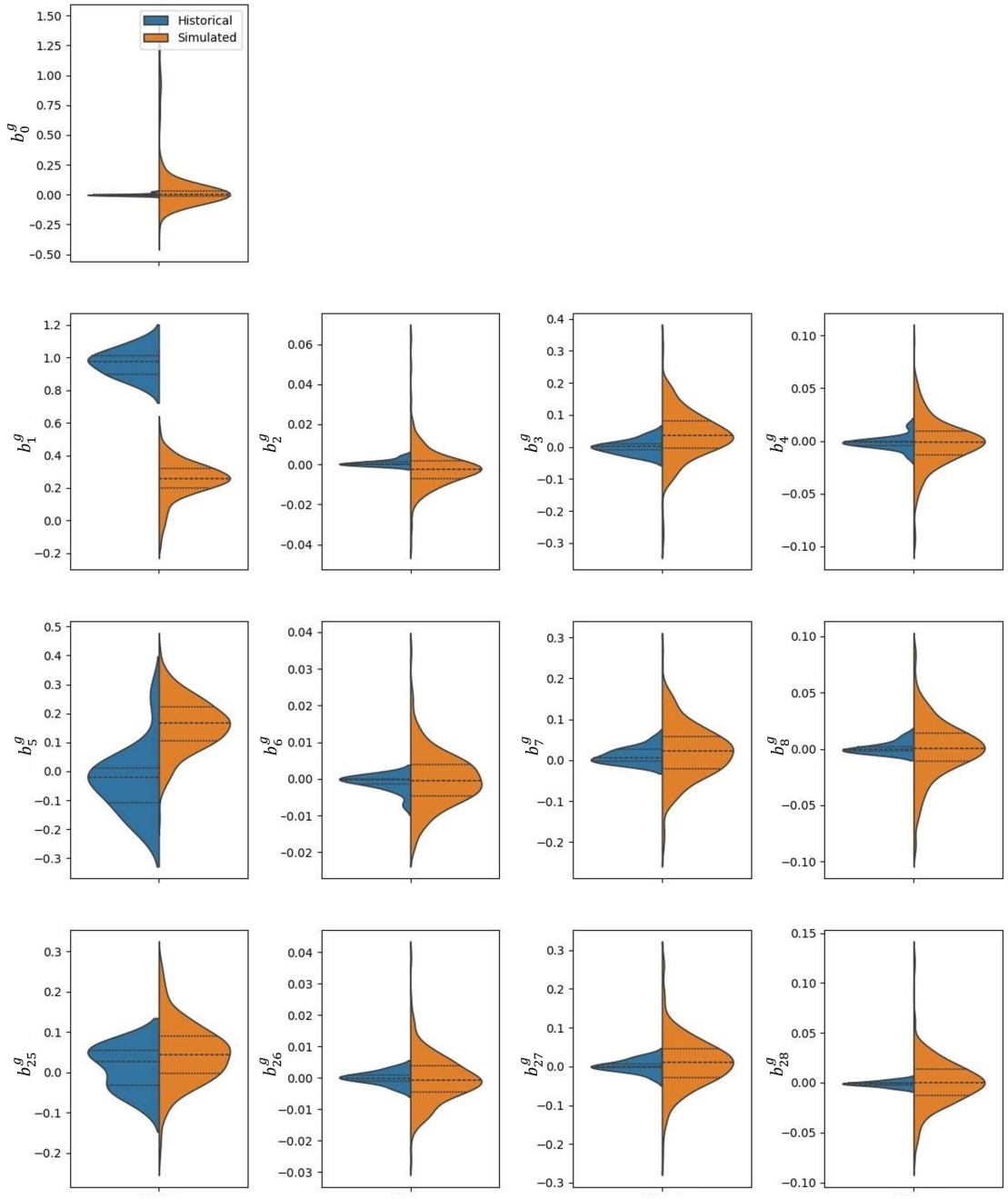

Figure 3.15: Gas prices - Relative frequency distributions of the intercept and the 12 coefficients at lags 1, 2, and 7 in Eq. (3.17) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

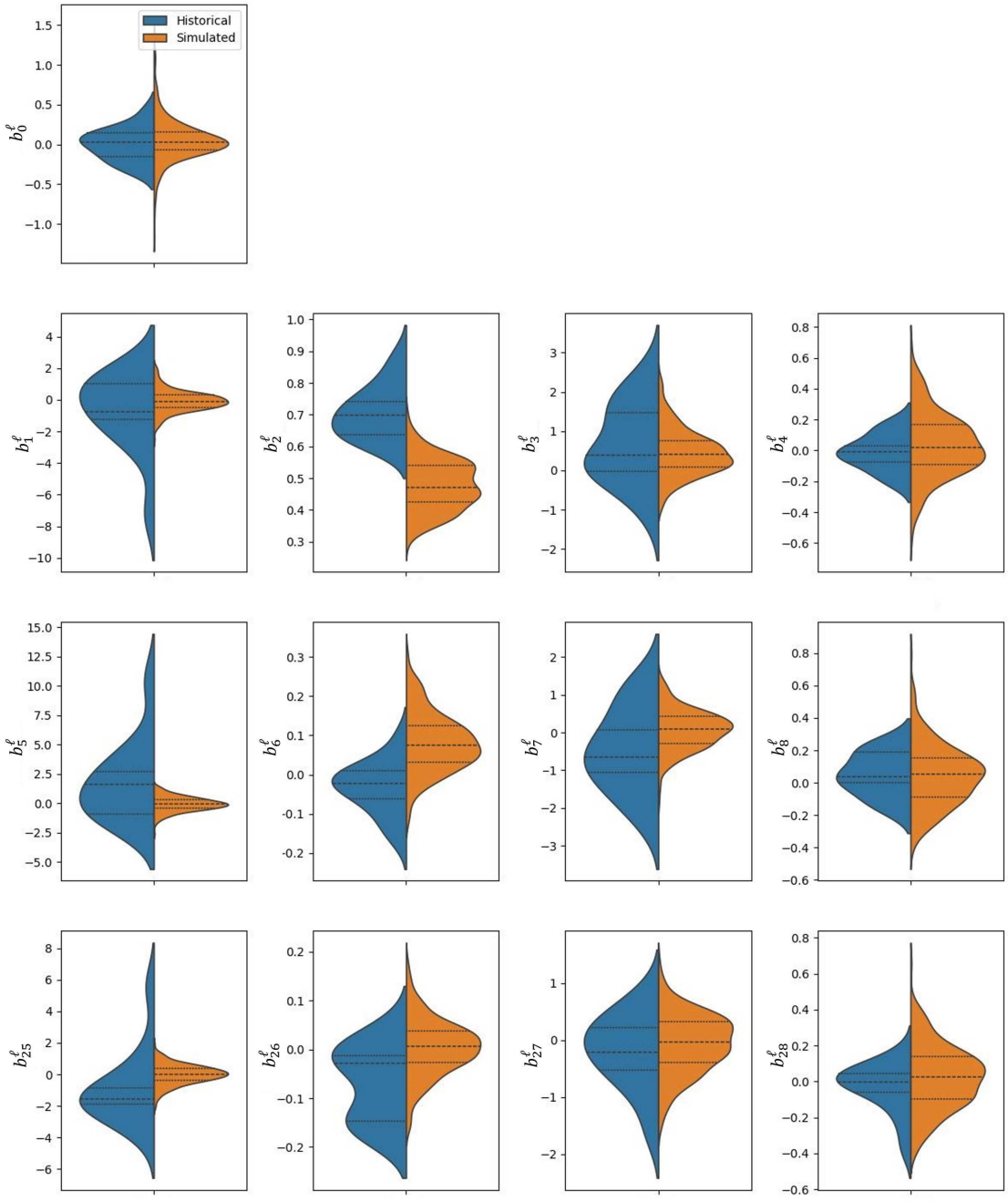

Figure 3.16: Load - Relative frequency distributions of the intercept and the 12 coefficients at lags 1, 2, and 7 in Eq. (3.18) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

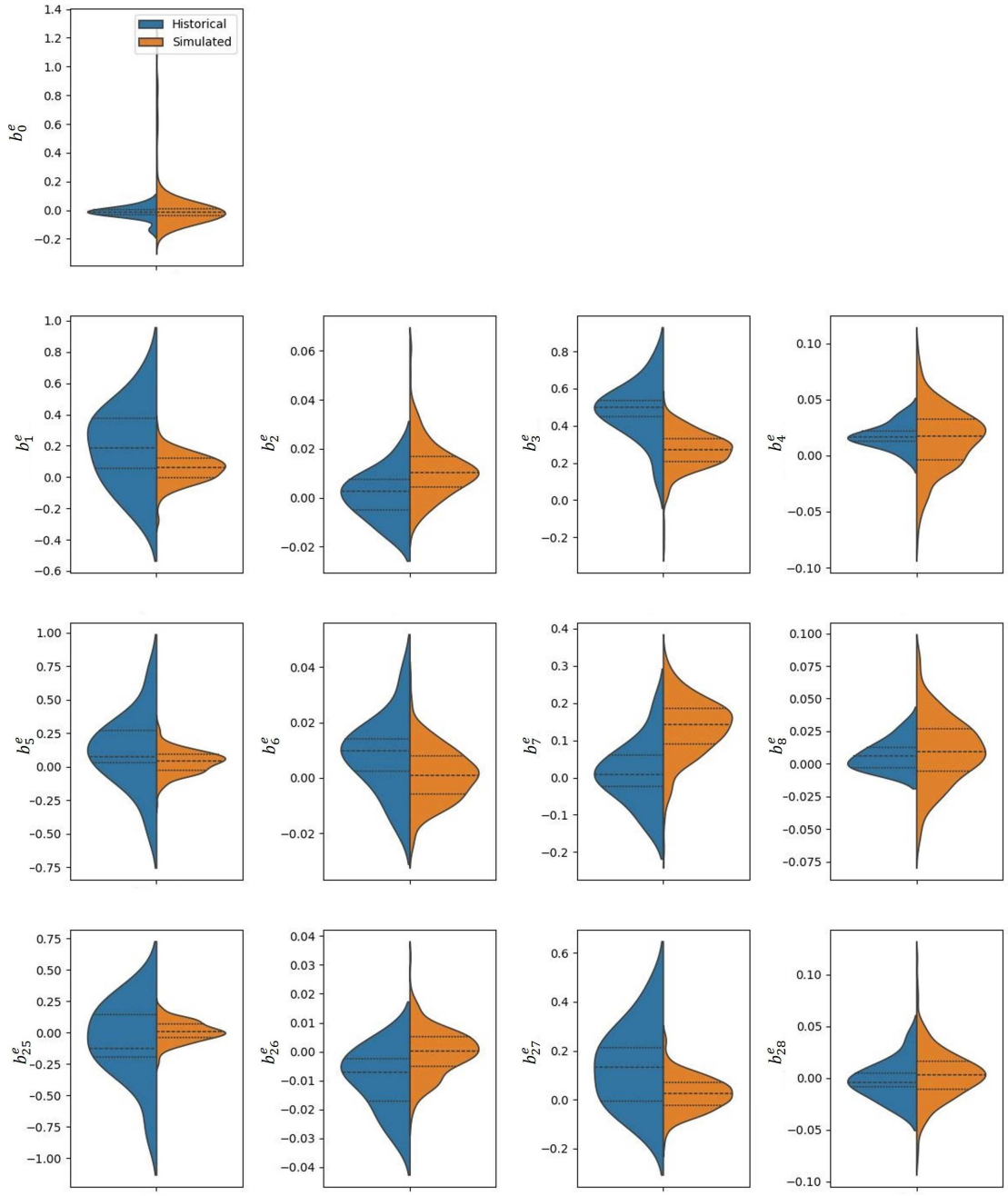

Figure 3.17: Electricity prices - Relative frequency distributions of the intercept and the 12 coefficients at lags 1, 2, and 7 in Eq. (3.19) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

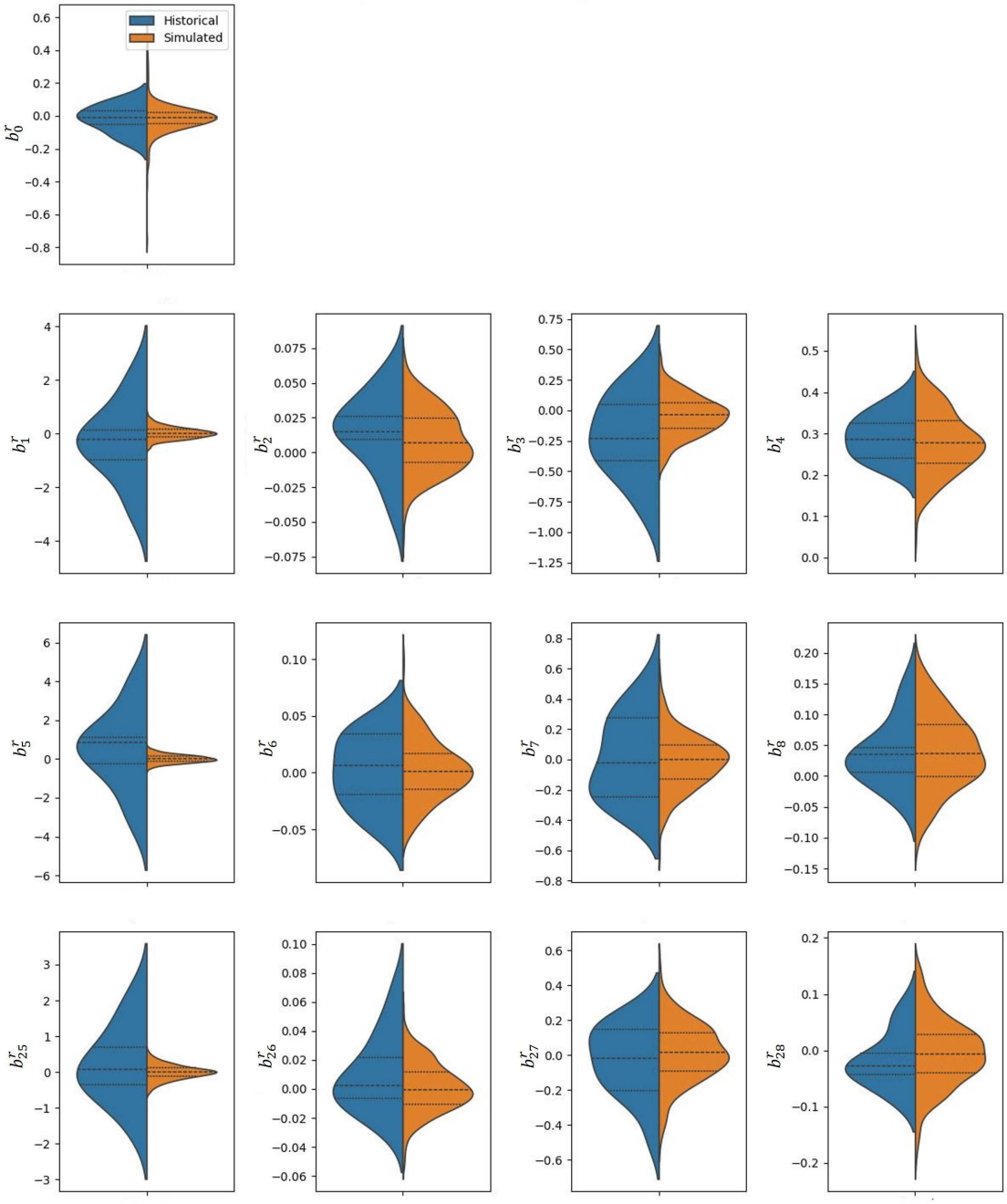

Figure 3.18: Solar radiation - Relative frequency distributions of the intercept and the 12 coefficients at lags 1, 2, and 7 in Eq. (3.20) based on 13 annual historical years (historical distribution, on the left) and on 260 simulated series (simulated distribution, on the right).

3.5.4 Kupiec test

In addition to the previous tests, results have been further analyzed by implementing the Kupiec "Proportion of Failures" (POF) test. The POF test evaluates the accuracy of a "Value-at-Risk" (VaR) model by computing the "Likelihood Ratio" (LR). The LR equation

$$LR = -2\Bigg[\ln\left((1-p)^{n-x}p^x\right) - \ln\left((1-\hat{p})^{n-x}\hat{p}^x\right)\Bigg], \tag{3.21}$$

with

- $p = 1-$ VaR confidence level,
- $n =$ number of observations,
- $x =$ number of VaR failures,
- $\hat{p} = \frac{x}{n}$,

measures if the observed frequency of VaR failures ($\hat{p}$) is consistent with the expected rate (p). This test assumes the Null Hypothesis (H_0) that the expected and observed rates of failures coincide,

$$H_0 : p = \hat{p}$$

The LR is distributed following a χ^2 distribution with 1 degree of freedom. By setting a value α, it is possible to determine the critical value of this distribution.
The comparison between the critical value of a χ^2 distribution and the LR allows to reject or accept the null hypothesis according to the following rule:

$$\text{Decision: } \begin{cases} \text{reject } H_0 & \text{if } LR > \text{critical value,} \\ \text{accept } H_0 & \text{if } LR \leq \text{critical value.} \end{cases}$$

The POF test has been applied to the daily series of historical residuals $\boldsymbol{\varepsilon}_d = (\varepsilon_d^g, \varepsilon_d^\ell, \varepsilon_d^e, \varepsilon_d^r)$, $d \in D$, and to the daily series of simulated residuals $\dot{\boldsymbol{\varepsilon}}_c = (\dot{\varepsilon}_c^g, \dot{\varepsilon}_c^\ell, \dot{\varepsilon}_c^e, \dot{\varepsilon}_c^r)$, $c \in C$. The VaR confidence level and the α have been set to 98% and 0.05, respectively. The results are collected in Table 3.2. With $\alpha = 0.05$ the critical value of the LR χ^2 distribution with 1 degree of freedom is 3.8414. By comparing the LR values with this critical value (3.8414) only load and radiation accept the null hypothesis H_0, while gas prices and electricity prices reject the H_0.

3.6 Conclusions

After a meticulous analysis and a rigorous comparison between historical data and simulated series, several significant conclusions regarding the modelling of energy market dynamics can be drawn.
First, the method proposed by Cerqueti et al. (2017a) proves instrumental in capturing nonlinear dependencies among multiple sources of uncertainty in energy markets. Through a comprehensive statistical analysis, the power of this approach in bootstrapping and simulating series that closely align with historical data across various key statistics, including mean, standard deviation, skewness, kurtosis, minimum, and maximum was proved. Moreover, the analysis of univariate autoregression coefficients and interdependencies among residuals further corroborates the validity of the present approach in capturing complex autodependencies and interrelationships among energy market factors.

Series	Variable	Value
Gas prices	p	0.020
	$\hat{p}$	0.034
	n	4,745
	x	163
	LR	41.14
Load	p	0.020
	$\hat{p}$	0.018
	n	4,745
	x	86
	LR	0.879
Electricity prices	p	0.020
	$\hat{p}$	0.032
	n	4,745
	x	152
	LR	29.70
Radiation	p	0.020
	$\hat{p}$	0.019
	n	4,745
	x	92
	LR	0.091

Table 3.2: Kupiec test results with VaR confidence level = 98%

Furthermore, this study highlights the importance of carefully managing the heterogeneous temporal frequencies among the components of the quadrivariate stochastic process. By implementing tailored adjustments, which are being advanced in this study, and extending it with respect to Cerqueti et al. (2017a), the method proves able to successfully manage different temporal frequencies, effectively capturing daily and hourly dynamics with high fidelity. The lack of adjustments would lead to a misalignment between historical and simulated series, which was particularly evident for load and solar radiation. Importantly, these adjustments prove effective even with a lower level of determinism in the bootstrapping step, reinforcing the versatility and reliability of the present approach across different levels of stochasticity.

In conclusion, the present study underscores the robustness and accuracy of the proposed methodology in bootstrapping and simulating energy market dynamics, offering valuable insights for stakeholders in navigating uncertainties and connections of multiple factors under scrutiny in decision-making processes. The choice of 5 states for each component of the quadrivariate stochastic process and of 2 time lags does not result from an optimization process but it is a choice done after several trials. Therefore, it is necessary to change the combination of time lags and states to improve the misalignments. Moving forward, further refinements and extensions of this approach could enhance its applicability across diverse energy market contexts, facilitating more informed and effective strategies for risk management of investment and operations decisions.

Bibliography

Anatolyev, S. and Vasnev, A. (2002). Markov chain approximation in bootstrapping autoregressions. *Economics Bulletin*, 3(19):1–8.

Athreya, K. B. and Fuh, C. D. (1992). Bootstrapping Markov chains: countable case. *Journal of Statistical Planning and Inference*, 33(3):311–331.

Bauwens, L., Koop, G., Korobilis, D., and Rombouts, J. V. (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, 30(4):596–620.

Bertail, P. and Clémençon, S. (2006). Regenerative block bootstrap for Markov chains. *Bernoulli*, 12(4):689–712.

Bertail, P. and Clémençon, S. (2007). Second-order properties of regeneration-based bootstrap for atomic Markov chains. *TEST*, 16(1):109–122.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148.

Bühlmann, P. (2002). Sieve bootstrap with variable-length Markov chains for stationary categorical time series. *Journal of the American Statistical Association*, 97(458):443–456.

Bühlmann, P. and Wyner, A. J. (1999). Variable length Markov chains. *The Annals of Statistics*, 27(2):480–513.

Carlstein, E. (1986). The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence. *The Annals of Statistics*, 14(3):1171–1179.

Cerqueti, R., Falbo, P., Guastaroba, G., and Pelizzari, C. (2015). Approximating multivariate Markov chains for bootstrapping through contiguous partitions. *OR Spectrum*, 37(3):803–841.

Cerqueti, R., Falbo, P., and Pelizzari, C. (2017a). Relevant states and memory in Markov chain bootstrapping and simulation. *European Journal of Operational Research*, 256(1):163–177.

Cerqueti, R., Falbo, P., Pelizzari, C., Ricca, F., and Scozzari, A. (2017b). A mixed integer linear program to compress transition probability matrices in Markov chain bootstrapping. *Annals of Operations Research*, 248(1):163–187.

Chambaz, A., Garivier, A., and Gassiat, E. (2009). A minimum description length approach to hidden Markov models with Poisson and Gaussian emissions. application to order identification. *Journal of Statistical Planning and Inference*, 139(3):962–977.

Csiszár, I. (2002). Large-scale typicality of Markov sample paths and consistency of MDL order estimators. *IEEE Transactions on Information Theory*, 48(6):1616–1628.

Csiszár, I. and Shields, P. C. (2000). The consistency of the BIC Markov order estimator. *The Annals of Statistics*, 28(6):1601–1619.

Datta, S. and McCormick, W. P. (1993). Regeneration-based bootstrap for Markov chains. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 21(2):181–193.

Di Silvestre, M. L., Ippolito, M. G., Sanseverino, E. R., Sciumè, G., and Vasile, A. (2021). Energy self-consumers and renewable energy communities in Italy: New actors of the electric power systems. *Renewable and Sustainable Energy Reviews*, 151:111565.

Faerber, L. A., Balta-Ozkan, N., and Connor, P. M. (2018). Innovative network pricing to support the transition to a smart grid in a low-carbon economy. *Energy Policy*, 116:210–219.

Fina, B., Monsberger, C., and Auer, H. (2022). Simulation or estimation? — two approaches to calculate financial benefits of energy communities. *Journal of Cleaner Production*, 330:129733.

Finesso, L. (1992). Estimation of the order of a finite Markov chain. In Kimura, H. and Kodama, S., editors, *Recent Advances in Mathematical Theory of Systems, Control, Networks, and Signal Processing: Proceedings of the International Symposium MTNS-91*, pages 643–645, Tokyo, Japan. Mita Press.

Gallego-Castillo, C., Heleno, M., and Victoria, M. (2021). Self-consumption for energy communities in Spain: A regional analysis under the new legal framework. *Energy Policy*, 150:112144.

Gjorgievski, V. Z., Cundeva, S., and Georghiou, G. E. (2021). Social arrangements, technical designs and impacts of energy communities: A review. *Renewable Energy*, 169:1138–1156.

Gomes, L. and Vale, Z. (2024). Costless renewable energy distribution model based on cooperative game theory for energy communities considering its members' active contributions. *Sustainable Cities and Society*, 101:105060.

Hall, P. (1985). Resampling a coverage pattern. *Stochastic Processes and their Applications*, 20(2):231–246.

Horowitz, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica*, 71(4):1049–1082.

Hounyo, U., Gonçalves, S., and Meddahi, N. (2017). Bootstrapping pre-averaged realized volatility under market microstructure noise. *Econometric Theory*, 33(4):791–838.

Inês F.G. Reis, Gonçalves, I., Lopes, M. A., and Antunes, C. H. (2021). Business models for energy communities: A review of key issues and trends. *Renewable and Sustainable Energy Reviews*, 144:111013.

Kieffer, J. C. (1993). Strongly consistent code-based identification and order estimation for constrained finite-state model classes. *IEEE Transactions on Information Theory*, 39(3):893–902.

Li, H., Hou, J., Tian, Z., Hong, T., Nord, N., and Rohde, D. (2022). Optimize heat prosumers' economic performance under current heating price models by using water tank thermal energy storage. *Energy*, 239:122103.

Lilliu, F. and Reforgiato Recupero, D. (2024). A cooperative game-theory approach for incentive systems in local energy communities. *Sustainable Energy, Grids and Networks*, 38:101391.

Lilliu, F., Reforgiato Recupero, D., Vinyals, M., and Denysiuk, R. (2023). Incentive mechanisms for the secure integration of renewable energy in local communities: A game-theoretic approach. *Sustainable Energy, Grids and Networks*, 36:101166.

Liu, C.-C. and Narayan, P. (1994). Order estimation and sequential universal data compression of a hidden Markov source by the method of mixtures. *IEEE Transactions on Information Theory*, 40(4):1167–1180.

Lowitzsch, J., Hoicka, C., and Tulder, F. V. (2020). Renewable energy communities under the 2019 European Clean Energy Package – Governance model for the energy clusters of the future? *Renewable and Sustainable Energy Reviews*, 122:109489.

Maldet, M., Revheim, F. H., Schwabeneder, D., Lettner, G., del Granado, P. C., Saif, A., Löschenbrand, M., and Khadem, S. (2022). Trends in local electricity market design: Regulatory barriers and the role of grid tariffs. *Journal of Cleaner Production*, 358:131805.

Merhav, N., Gutman, M., and Ziv, J. (1989). On the estimation of the order of a Markov chain and universal data compression. *IEEE Transactions on Information Theory*, 35(5):1014–1019.

Paparoditis, E. and Politis, D. N. (2001a). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory*, 17(3):540–566.

Paparoditis, E. and Politis, D. N. (2001b). Tapered block bootstrap. *Biometrika*, 88(4):1105–1119.

Paparoditis, E. and Politis, D. N. (2002a). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference*, 108(1):301–328.

Paparoditis, E. and Politis, D. N. (2002b). The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal*, 5(1):131–148.

Pera, A. L., Sellaro, M., Bencivenni, E., and D'Amico, F. (2022). Environmental sustainability of an integrate anaerobic digestion-composting treatment of food waste: Analysis of an italian plant in the circular bioeconomy strategy. *Waste Management*, 139:341–351.

Peres, Y. and Shields, P. (2005). Two new Markov order estimators.

Petrović, L. V. and Scholtes, I. (2022). Learning the Markov order of paths in graphs. In *Proceedings of the ACM Web Conference 2022*, page 1559–1569, New York, NY, USA. Association for Computing Machinery.

Politis, D. N. and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *The Annals of Statistics*, 20(4):1985–2007.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Pons-Seres de Brauwer, C. and Cohen, J. (2020). Analysing the potential of citizen-financed community renewable energy to drive europe's low-carbon energy transition. *Renewable and Sustainable Energy Reviews*, 133:110300.

Rajarshi, M. B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Annals of the Institute of Statistical Mathematics*, 42(2):253–268.

Sousa, J., Lagarto, J., Camus, C., Viveiros, C., Barata, F., Silva, P., Alegria, R., and Paraíba, O. (2023). Renewable energy communities optimal design supported by an optimization model for investment in pv/wind capacity and renewable electricity sharing. *Energy*, 283:128464.

Zhang, F., Wang, X., Wu, M., Hou, X., Han, C., and Liu, Z. (2022). Optimization design of uncertain parameters for improving the stability of photovoltaic system. *Journal of Power Sources*, 521:230959.

Zhang, K., Zhou, B., Li, C., Voropai, N., Li, J., Huang, W., and Wang, T. (2021). Dynamic modeling and coordinated multi-energy management for a sustainable biogas-dominated energy hub. *Energy*, 220:119640.

Appendix A

List of symbols

This appendix provides a comprehensive list of all the symbols used throughout the thesis, along with their descriptions and units where applicable.

Table A.1: Symbols Used in Chapter 2

Symbol	Description	Unit
i	Index to count the number of biogas producers	$= \{1, \ldots, \bar{I}\}$
I	Total number of biogas producers	$\mathbb{N}$
j	Index to count the number of households	$= \{1, \ldots, \bar{J}\}$
J	Total number of households	$\mathbb{N}$
t	Time index $\in T$	$= \{1, \ldots, \bar{T}\}$
δ	Incentive sharing rule	%
γ_t^{pv}	Solar radiation efficiency	%
π_i^b	Profit function of i-th biogas producer	€
π_j^h	Profit function of j-th household	€
d_t^b	Biogas producer load	MWh
d_t^h	Household load	MWh
$d_{t,i}^{b1}$	Load of the i-th biogas producer in round 1	MWh
$d_{t,j}^{h1}$	Load of the j-th household in round 1	MWh
$d_{t,i}^{b1.pre}$	Load before the i-th biogas producer in round 1	MWh
$d_{t,j}^{h1.pre}$	Load before the j-th household in round 1	MWh
$d_{t,i}^{b2}$	Load of the i-th biogas producer in round 2	MWh
$d_{t,j}^{h2}$	Load of the j-th household in round 2	MWh
$d_{t,i}^{b2.pre}$	Load before the i-th biogas producer in round 2	MWh
$d_{t,j}^{h2.pre}$	Load before the j-th household in round 2	MWh
B^b	Biogas producer budget	M€
B_i^b	Individual i-th biogas producer budget	M€
B^h	Household budget	M€

Continued on next page

Table A.1: Symbols Used in Chapter 2 (Continued)

Symbol	Description	Unit
B_j^h	Individual j-th household budget	M€
C^g	Fixedd installation cost	M€
k	Retailer profit margin	€/MWh
m^e	Ratio that converts m^3 into MWh	MWh/m^3
p_t^{me}	Electricity price on the spot market	€/MWh
p_t^g	Gas market price	€/m^3
p^{qe}	Unit cost of 1 MW gas-to-electricity turbine	M€/MW
p^{qg}	Unit cost of 1 m^3 biogidestor plant	M€/m^3
p^{pv}	Unit cost of 1 MW PV plant	M€/MW
Q_i^e	Turbine capacity installed by the i-th biogas producer	MW
Q_i^{e1}	Turbine capacity installed by the i-th biogas producer in round 1	MW
$Q_i^{e1.pre}$	Turbine capacity installed before the i-th biogas producer in round 1	MW
Q_i^{e2}	Turbine capacity installed by the i-th biogas producer in round 2	MW
$Q_i^{e2.pre}$	Turbine capacity installed before the i-th biogas producer in round 2	MW
$q_{t,i}^e$	Electricity produced from the turbine of the i-th biogas producer	MWh
$q_{t,i}^{e1}$	Electricity produced from the turbine of the i-th biogas producer in round 1	MWh
$q_{t,i}^{e1.pre}$	Electricity produced from the turbine before the i-th biogas producer in round 1	MWh
$q_{t,i}^{e2}$	Electricity produced from the turbine of the i-th biogas producer in round 2	MWh
$q_{t,i}^{e2.pre}$	Electricity produced from the turbine before the i-th biogas producer in round 2	MWh
Q_i^g	Biodigestor capacity installed by the i-th biogas producer	m^3
Q_i^{g1}	Biodigestor capacity installed by the i-th biogas producer in round 1	m^3
$Q_i^{g1.pre}$	Biodigestor capacity installed before the i-th biogas producer in round 1	m^3
Q_i^{g2}	Biodigestor capacity installed by the i-th biogas producer in round 2	m^3
$Q_i^{g2.pre}$	Biodigestor capacity installed before the i-th biogas producer in round 2	m^3
$q_{t,i}^g$	Biogas produced by the biodigestor of the i-th biogas producer	m^3

Continued on next page

Table A.1: Symbols Used in Chapter 2 (Continued)

Symbol	Description	Unit
$q_{t,i}^{g1}$	Biogas produced by the biodigestor of the i-th biogas producer in round 1	m^3
$q_{t,i}^{g1.pre}$	Biogas produced by the biodigestor before the i-th biogas producer in round 1	m^3
$q_{t,i}^{g2}$	Biogas produced by the biodigestor of the i-th biogas producer in round 2	m^3
$q_{t,i}^{g2.pre}$	Biogas produced by the biodigestor before the i-th biogas producer in round 2	m^3
Q_j^{pv}	PV plant capacity installed by the j-th household	MW
Q_j^{pv1}	PV plant capacity installed by the j-th household in round 1	MW
$Q_j^{pv1.pre}$	PV plant capacity installed before the j-th household in round 1	MW
Q_j^{pv2}	PV plant capacity installed by the j-th household in round 2	MW
$Q_j^{pv2.pre}$	PV plant capacity installed before the j-th household in round 2	MW
$q_{t,j}^{pv}$	Electricity produced by the PV plant of the j-th household	MWh
$q_{t,j}^{pv1}$	Electricity produced by the PV plant of the j-th household in round 1	MWh
$q_{t,j}^{pv1.pre}$	Electricity produced by the PV plant before the j-th household in round 1	MWh
$q_{t,j}^{pv2}$	Electricity produced by the PV plant of the j-th household in round 2	MWh
$q_{t,j}^{pv2.pre}$	Electricity produced by the PV plant before the j-th household in round 2	MWh
q_t^{REC}	Self-consumption	MWh
$q_{t,i}^{REC1}$	Self-consumption from the i-th biogas producer in round 1	MWh
$q_{t,j}^{REC1}$	Self-consumption from the j-th household in round 1	MWh
$q_{t,i}^{REC1.pre}$	Self-consumption before the i-th biogas producer in round 1	MWh
$q_{t,j}^{REC1.pre}$	Self-consumption before the j-th household in round 1	MWh
$q_{t,i}^{REC2}$	Self-consumption from the i-th biogas producer in round 2	MWh

Continued on next page

Table A.1: Symbols Used in Chapter 2 (Continued)

Symbol	Description	Unit
$q_{t,j}^{REC2}$	Self-consumption from the j-th household in round 2	MWh
$q_{t,i}^{REC2.pre}$	Self-consumption before the i-th biogas producer in round 2	MWh
$q_{t,j}^{REC2.pre}$	Self-consumption before the j-th household in round 2	MWh
z	Incentive on self-consumption	€/KWh

Table A.2: Symbols Used in Chapter 3

Symbol	Description	Unit
$\{\mathbf{X}_d^{(N)}\}_{d\in D}$	N-variate stochastic process with time index d	$\mathbb{R}^N$
d	Daily time index, $d \in D, d = \{1, \ldots, \bar{D}\}$	
$\{\mathbf{X}_i^{(M)}\}_{i\in I}$	M-variate stochastic process with time index i	$\mathbb{R}^M$
i	Hourly time index, $i \in D, i = \{1, \ldots, \bar{I}\}$	
J	Number of times that index i is related to the same value of index d	
$\{\mathbf{Z}_d\}_{d\in D}$	$(N+M)$-variate stochastic process with index d	$\mathbb{R}^{N+M}$
k	Number of intervals to partition the support $\mathbb{R}$ of each component of $\{\mathbf{Z}_d\}_{d\in D}$	$\mathbb{N}$
$(N+M)^k$	Overall number of states based on $N+M$ partitions of $\mathbb{R}$ into k intervals	
$\mathbf{s}^\nu$	Set of $(N+M)^k$ states $= (s^1, \ldots, s^\nu, \ldots, s^{(N+M)^k})$	
$\{\mathbf{S}_d\}_{d\in D}$	Discretized stochastic $= (s_d^1, \ldots, s_d^N, s_d^{N+1}, \ldots, s_d^{N+M})$	
$p(s^2\|s^1)$	Conditional probability which counts how many times the couple (s^1, s^2) was observed in two *consecutive* realizations of $\{\mathbf{S}_d\}_{d\in D}$, with s^2 following s^1	
g	Gas prices label	
ℓ	Load label	
e	Electricity prices label	
r	Solar radiation label	
H	Set of hours of a day $= \{$*00:00:00*, …, *23:00:00*$\} \equiv \{0, \ldots, 23\}$	

Continued on next page

Table A.2: Symbols Used in Chapter 3 (Continued)

Symbol	Description	Unit
W	Set of weekdays $= \{Monday, \ldots, Sunday\} \equiv \{0, \ldots, 6\}$	
C	Set of calendar days $= \{01\text{-}01, \ldots, 12\text{-}31\} \equiv \{0, \ldots, 364\}$	
M	Set of months of a year $= \{January, \ldots, December\} \equiv \{0, \ldots, 11\}$	
t	Set of historical years $= \{2010, \ldots, 2022\} \equiv \{0, \ldots, 12\}$	
H_{-1}	Set of hours of a day without the first one to deal with collinearity $= \{01{:}00{:}00, \ldots, 23{:}00{:}00\} \equiv \{1, \ldots, 23\}$	
W_{-1}	Set of weekdays without the first one to deal with collinearity $= \{Tuesday, \ldots, Sunday\} \equiv \{1, \ldots, 6\}$	
C_{-1}	Set of calendar days without the first one to deal with collinearity $= \{01\text{-}02, \ldots, 12\text{-}31\} \equiv \{1, \ldots, 364\}$	
M_{-1}	Set of months of a year without the first one to deal with collinearity $= \{February, \ldots, December\} \equiv \{1, \ldots, 11\}$	
$d(i) = d$	Condition to collect the hourly residuals such that the day of hour i is equal to d	
χ	Dummy explanatory variable	
$hr(i) = j$	Functions to transform the index i into elements of H	
$wd(i) = j$	Functions to transform the index i into elements of W	
$wd(d) = j$	Functions to transform the index d into elements of W	
$m(i) = j$	Functions to transform the index i into elements of M	
$m(d) = j$	Functions to transform the index d into elements of M	
$cd(d) = j$	Functions to transform the index d into elements of C	
u_d^g	Original daily historical gas prices	€/MWh
u_i^ℓ	Original hourly historical load (Italian hourly load)	MWh
$\bar{u}^\ell$	Italian hourly average load	kWh
$\bar{o}^\ell$	Hourly average load per household	kWh
f^ℓ	Ratio that links $\bar{o}^\ell$ with the $\bar{u}^\ell$	
u_i^e	Original hourly historical electricity prices	€/MWh
u_i^r	Original hourly historical solar radiation	Wh/m^2
w_d^g	Adjusted daily historical gas prices	€/MWh

Continued on next page

Table A.2: Symbols Used in Chapter 3 (Continued)

Symbol	Description	Unit
w_i^ℓ	Adjusted hourly historical load	MWh
w_i^e	Adjusted hourly historical electricity prices	€/MWh
w_i^r	Adjusted hourly historical solar radiation	Wh/m^2
η_d^g	Daily gas prices residuals after detrendization	
α_0^g	Intercept of detrendization for gas prices	
α_1^g	Slope of detrendization for gas prices	
η_i^ℓ	Hourly load residuals after detrendization	
α_0^ℓ	Intercept of detrendization for load	
α_1^ℓ	Slope of detrendization for load	
η_i^e	Hourly electricity prices residuals after detrendization	
α_0^e	Intercept of detrendization for electricity prices	
α_1^e	Slope of detrendization for electricity prices	
ε_d^g	Daily gas prices residuals after detrendization and deseasonalization	
α_2^g	Intercept of deseasonalization for gas prices	
γ_j^g	Regression coefficient that captures the weekly seasonality with $j \in W_{-1}$	
δ_j^g	Regression coefficient that captures the yearly seasonality on a monthly basis with $j \in M_{-1}$	
ε_i^ℓ	Hourly load residuals after detrendization and deseasonalization	
α_2^ℓ	Intercept of deseasonalization for load	
β_j^ℓ	Regression coefficient that captures the daily seasonality with $j \in H_{-1}$	
γ_j^ℓ	Regression coefficient that captures the weekly seasonality with $j \in W_{-1}$	
δ_j^ℓ	Regression coefficient that captures the yearly seasonality on a monthly basis with $j \in M_{-1}$	
ε_d^ℓ	Daily load residuals after detrendization and deseasonalization by summing hourly values for each day $d \in D$	
ε_i^e	Electricity prices residuals after detrendization and deseasonalization	
α_2^e	Intercept of deseasonalization for electricity prices	
β_j^e	Regression coefficient that captures the daily seasonality with $j \in H_{-1}$	

Continued on next page

Table A.2: Symbols Used in Chapter 3 (Continued)

Symbol	Description	Unit
γ_j^e	Regression coefficient that captures the weekly seasonality with $j \in W_{-1}$	
δ_j^e	Regression coefficient that captures the yearly seasonality on a monthly basis with $j \in M_{-1}$	
ε_d^e	Daily electricity price residuals after detrendization and deseasonalization by averaging hourly values for each day $d \in D$	
ε_d^r	Daily solar radiation residuals after deseasonalization	
α_2^r	Intercept of deseasonalization for solar radiation	
ϕ_j^r	Regression coefficient that captures the yearly seasonality based on calendar days with $j \in C_{-1}$	
$\tilde{\mathbf{s}}_{\bar{c}}$	Series of bootstrapped states at time $\bar{c} \in C$, $= \left(\tilde{s}_{\bar{c}}^g, \tilde{s}_{\bar{c}}^\ell, \tilde{s}_{\bar{c}}^e, \tilde{s}_{\bar{c}}^r\right)$	
$\boldsymbol{\varepsilon}_d$	Series of quadruplets of residuals of gas prices, load, electricity prices and solar radiation $= (\varepsilon_d^g, \varepsilon_d^\ell, \varepsilon_d^e, \varepsilon_d^r)$	
$\tilde{\boldsymbol{\varepsilon}}_c$	Bootstrapped series of residuals before treatments $= (\tilde{\varepsilon}_c^g, \tilde{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \tilde{\varepsilon}_c^r)$	
$\dot{\boldsymbol{\varepsilon}}_c$	Bootstrapped series of residuals after treatments $= (\tilde{\varepsilon}_c^g, \breve{\varepsilon}_c^\ell, \tilde{\varepsilon}_c^e, \breve{\varepsilon}_c^r)$	
$\tilde{\varepsilon}_c^g$	Daily gas prices simulated residuals before reseasonalization and retrendization	
$\breve{\varepsilon}_c^\ell$	Daily load simulated residuals before reseasonalization and retrendization	
$\breve{\varepsilon}_{h,c}^\ell$	Hourly load simulated residuals before reseasonalization and retrendization associated to $\breve{\varepsilon}_c^\ell$	
$\tilde{\varepsilon}^e$	Daily electricity prices simulated residuals before reseasonalization and retrendization	
$\breve{\varepsilon}^r$	Daily solar radiation simulated residuals before reseasonalization and retrendization	
q	Function to count the hours to be simulated	
$\tilde{\eta}_c^g$	Daily gas prices simulated residuals after reseasonalization	
$\tilde{\eta}_{h,c}^\ell$	Hourly load simulated residuals after reseasonalization	

Continued on next page

Table A.2: Symbols Used in Chapter 3 (Continued)

Symbol	Description	Unit
$\tilde{\eta}^{e}_{h,c}$	Hourly electricity prices simulated residuals after reseasonalization	
$\tilde{\eta}^{r}_{c}$	Daily solar radiation simulated residuals after reseasonalization	
$\tilde{w}^{g}_{c}$	Daily gas prices simulated series after reseasonalization and retrendization	
$\tilde{w}^{\ell}_{h,c}$	Hourly load simulated series after reseasonalization and retrendization	
$\tilde{w}^{e}_{h,c}$	Hourly electricity prices simulated series after reseasonalization and retrendization	
$\bar{w}^{r}_{h,c}$	Hourly average solar radiation historically observed for hour h in calendar day c	
$\tilde{w}^{r}_{h,c}$	Hourly solar radiation simulated series after reseasonalization	
$a^{(\dagger)}_{(\cdot)}$	Univariate autoregression coefficients, with $(\dagger) = 0, \ldots, 5$ and $(\cdot) = \ell, e, r$	
$b^{(\dagger)}_{(\cdot)}$	VAR autoregression coefficients, with $(\dagger) = 0, \ldots, 28$ and $(\cdot) = g, \ell, e, r$	
p	Expected rate in Kupiec test	%
$\hat{p}$	Observed rate in Kupiec test	%
n	Number of observations in Kupiec test	$\mathbb{N}$
x	Number of failures in Kupiec test	$\mathbb{N}$
α	Threshold to determine the critical value of the LR distributed according to a χ^2 distribution	$\mathbb{R}$
LR	Likelihood ratio in Kupiec test	$\mathbb{R}$

MODULO DI EMBARGO DELLA TESI
(da compilare solo se si richiede un periodo di segretazione della tesi)

Il/La sottoscritto/a Ruffini Alessandra Nato/a il 06/12/1995
a (indicare anche l'eventuale paese estero) Brescia
provincia di (ovvero sigla del paese estero) Brescia
Dottorato di Ricerca in Analytics for Economics and Management
(Modelli e Metodi per l'Economia e il Management)

DICHIARA

- che il contenuto della tesi **non può essere immediatamente consultabile per il seguente motivo**

motivi di priorità nella ricerca

La motivazione deve essere dettagliata e controfirmata obbligatoriamente dal Primo Supervisore di tesi
(Brevetto, segreto industriale, motivi di priorità nella ricerca, motivi editoriali, altro)

- che il testo completo della tesi potrà essere reso consultabile dopo:

□ 6 mesi dalla data di conseguimento titolo
☑ 12 mesi dalla data di conseguimento titolo

- che sarà comunque consultabile immediatamente l'abstract della tesi, che viene caricato in Esse3, profilo studente.

Luogo e Data
Brescia, 10/06/2024

Firma del Dichiarante

Controfirma del Primo Supervisore di tesi
per la motivazione di embargo e il periodo