ESSAYS ON THE ECONOMICS AND REGULATORY DESIGN OF POWER SYSTEMS

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2015

vorgelegt

von

MSc ETH Simeon Hagspiel

aus

March

Referent	Prof. Dr. Felix Höffler
Korreferent	Prof. Ulrich W. Thonemann, Ph.D.
Tag der Promotion	01.07.2016

Preface

Energy has been the focus of my interest since the first days of my engineering studies at ETH Zurich starting in 2005. Nevertheless, having finished my Master's thesis in 2011, I was somehow unsatisfied with the level of specific knowledge I had reached. So it was quite clear to me that I needed to continue learning as a doctoral student. However, it was not clear to me at all how much more I really had to learn. That I resolved to complement my engineering perspective with a doctoral thesis in economics did not really contribute to simplify the task. Even though I was able to draw on what I had learned so far, I had to start asking questions differently, and to broaden the foundation of my work based on economic principles. Thus, my decision inevitably brought along additional workload and frustration, but also inspiration and satisfaction. In fact, it has probably been exactly that shift of perspective that had a significant, sustainably positive impact on my knowledge about energy and power systems in particular. Moreover, it probably affected the development of my soft skills in a very positive way, too. Overall, my years as a doctoral student at the University of Cologne were very enriching in several respects, and an experience I would not want to miss.

During this adventure, several people and institutions have supported me in different ways, for which I am deeply grateful. First and foremost, I want to thank Felix Höffler for academic supervision and support. His clear line and typical manner of giving dry yet constructive comments forced and motivated me to reflect the existing, and tackle the continuation of my work. His high standards and degree of rationality were certainly reasons why I have learned so much rigorous economic thinking. I am also very grateful to Ulrich Thonemann for academic advice and comments as well as for being the second assessor of my thesis, and to Van Anh Vuong for chairing the defense of my dissertation.

An institutionally and financially stable framework has been provided by the Institute of Energy Economics (EWI) and the University of Cologne, as well as by the German Research Foundation (DFG) with the research grant HO5108/2-1 and the German Federal Ministry for Economic Affairs and Energy (BMWi) and the German Federal Ministry of Education and Research (BMBF) with their Energy Storage Initiative (grant 03ESP239).

Realizing my thesis would have been a much stonier road without the inspiring, persistent and enjoyable teamwork with my coauthors, Joachim Bertsch, Christina Elberg and Lisa Just. Much of the quality and fun is owed to these three people. Special thanks also go to my student assistants Clara Dewes and Michael Franke for their scientific support. Furthermore, I was very lucky to have an exceptionally well-functioning, dynamic and pleasurable working environment at EWI (including managing directors, administration, IT, library, student assistants, research associates, as well as the nice cleaner who reminded me regularly that I should eventually go home in the evening).

Lastly, I am infinitely happy to have my two ladies Lisa and Chiara, as well as my family and friends.

Simeon Hagspiel   Cologne, December 2015

Contents

1 Introduction

Power systems regulation is never at rest. [...] [T]here are quieter periods and more active ones. Present times appear to demand a particularly active regulatory response.

Ignacio J. Pérez-Arriaga (2013)

Power systems have been sparking public attention ever since the first central power plant was commissioned by Thomas Edison on Pearl Street, New York City, in September 1882. They have played an important role in the growth of economies worldwide, and will continue to do so in the future. As an inevitable consequence of public attention and societal importance, power systems have also been subject to intense scrutiny and debates about their best possible organization. After a period of private initiatives and competition at the end of the 19$^{\text{th}}$ and the beginning of the 20$^{\text{th}}$ century, the power sector was for a long time organized by means of vertically integrated monopolies. The entire supply chain – including production, transmission, distribution, and retailing – was hence comprised within a few electric utilities that were either state-owned or private and subject to heavy regulation. However, this paradigm has been overthrown by the deregulation wave beginning in the 1980ies, aiming at a reorganization of power systems to create far-reaching economic benefits. As a consequence, nowadays the production and retail sides of the supply chain typically involve numerous competitive firms. It is a recognized fact that if properly implemented, deregulation was indeed able to involve substantial improvements in the performance of power sectors (Joskow (2008b)).

Nevertheless, regulation remains at the core of power systems. Two of the most important reasons for regulatory interventions to be present in today's power systems are negative environmental externalities from power generation, and the fact that the power grid is a natural monopoly. The former reason has once more gained momentum during the very recent UN Climate Change Conference in Paris. Indeed, to implement the strategies the 195 parties have agreed upon, the supply side of the power sector – with 25% currently the largest single source of global greenhouse gas emissions – will be key for future developments (IPCC (2014)). It is and will be a major challenge to design a regulatory framework for the power sector that incentivizes and manages its transformation towards low-carbon power production (e.g., by means of renewable energies) in an effective and economically efficient way.

The other important reason for regulatory intervention stems from the natural monopoly characteristics of the power grid. Recent advances in the theory of regulation make it imperative to rethink and redesign regulatory approaches for the power grid in order to reap efficiency gains. Especially in the light of the above mentioned fundamental changes in power systems, taking advantage of better regulation gains even more importance. In fact, greenhouse gas emissions in the power sector and grid regulation are closely related when it comes to the expansion of variable renewable electricity sources (such as wind and solar), which typically requires expansion of the electricity network.

Generally speaking, getting the economics and regulation of power system right is important since weak designs may entail significant losses of social welfare. In fact, due to the large turnover of the power sector (e.g., around 455 bn. € in Europe in 2013, representing more than 3% of European GDP[1]), even small *relative* inefficiencies cause a large *absolute* excess burden.

Against this background, it is the goal of this thesis to investigate some aspects of the economics and regulatory design of power systems. The focus lies on the above mentioned fields where regulatory intervention is particularly relevant and economically justified: The integration of variable renewable energies into power systems on the one hand, and electricity network expansion and operation on the other. Overlapping areas are also taken into account. As I will show in this thesis, these fields are promising candidates to yield substantial improvements when being reformed. To this end, novel approaches shall be suggested to identify and tackle economic and regulatory deficits and to improve the related outcomes. In the context of renewable expansion and grid regulation, my thesis contributes to the academic debate with the following four papers that are contained in Chapters 2-5. Three of those chapters are joint work with co-authors whose contribution I shall once more gratefully acknowledge here. If elaborated as joint work, all authors contributed equally to all parts of the corresponding paper.

[1]The annual turnover of the European electricity sector can be estimated with a simple calculation based on 3101 TWh of net electricity production and an average price of electricity for end-consumers of 0.147 € /kWh. European GDP in 2013 was 13520 bn. € (all numbers taken from Eurostat (2015)).

Chapter 2: Spatial dependencies of wind power and interrelations with spot price dynamics (with Christina Elberg). Published in the European Journal of Operational Research, Vol. 241 (1), pp. 260-272, 2015.

Chapter 3: Supply chain reliability and the role of individual suppliers. Forthcoming as EWI Working Paper, in preparation for submission to Management Science.

Chapter 4: Congestion management in power systems – Long-term modeling framework and large-scale application (with Joachim Bertsch and Lisa Just). Published as EWI Working Paper 2015/03, Revised and resubmitted to the Journal of Regulatory Economics.

Chapter 5: Regulation of non-marketed outputs and substitutable inputs (with Joachim Bertsch). Published as EWI Working Paper 2015/06, in preparation for submission to the International Journal of Industrial Organization.

In the following Section 1.1, I will outline the contents of these four chapters and put them into context as well as into relation with each other. In Section 1.2, I discuss the methodologies that are applied in the different chapters. Section 1.3 concludes the introduction.

1.1 Outline of the thesis

In my thesis, I investigate several aspects of the economics and regulatory design of power systems. I focus on the production and transmission sectors (even though the contents of Chapter 5 may also be applied to the distribution sector). Along this part of the supply chain, I analyze different economic activities and how they should be organized in order to induce short- and long-term efficiency. The specific challenges I address in the first two papers stem from the time-varying and interdependent temporal and spatial distribution of production (especially, from variable renewable energies) and demand, while the latter two papers focus on the degree and exchange of information between different players in the supply chain.

Chapter 2 deals with the integration of variable renewable energies, especially wind energy, into power systems. If wind and solar were to (only) receive the wholesale electricity price (and no subsidies via, e.g., fixed feed-in-tariffs), they would face the challenge that their generation is highly correlated: on a windy day in northern

Germany, all wind generators along the coast are able to produce. For that reason, prices earned by renewables will typically be below the average price level. Hence, it may make sense to install renewable capacities where correlation with other producers is favorable, even if the place is less windy.

In this paper, we assess the market value of variable generation assets at different locations using a stochastic simulation model that covers the full spatial dependence structures of generation by using copulas, incorporated into a supply and demand based model for the electricity spot price. This model is calibrated with German data. We find that the specific location of a wind turbine – i.e., its spatial dependence with respect to the aggregated wind power in the system – is of high relevance for its market value. It is reduced by up to 8 € /MWh (i.e., 15%) compared to average spot price levels, and varies by up to 6 € /MWh for the different locations that were analyzed. Many of the locations show an upper tail dependence that adversely impacts the market value. Therefore, a model that assumes a linear dependence structure would systematically overestimate the market value of wind power in many cases. This effect becomes more important for increasing levels of wind power penetration as the price effect of wind power becomes more pronounced.

Regulatory practice so far often ignores the complex dynamics and interactions of generation and electricity prices by offering, e.g., spatially and temporally fixed feed-in-tariffs. In contrast, to reveal the actual value of electricity at each specific time and place – needed to trigger incentives for efficient investments – the dynamics and interactions that are depicted in this section should be represented in the regulatory design of renewable energy integration.

Chapter 3 picks up a related issue arising from the stochastic nature of wind and solar generation. It analyzes the contribution of individual stochastic suppliers to the supply reliability of the overall system, as well as an appropriate payoff scheme. How valuable an individual stochastic supplier really is for a given system is difficult to determine, since it not only depends on the stochastic nature of the individual supplier itself, but also on all other stochastic suppliers that are present. For regulatory purposes, the contribution of intermittent renewables to system reliability is important because it should be the basis for any capacity payment to these generators. In particular, it should influence the payments individual suppliers could receive in any sort of capacity mechanism that have been intensively discussed and implemented in many electricity markets.

To this end, I first investigate the statistical properties of the supply chain, including stochastic and interdependent supply and demand. Based on these finding, I show that an efficient organization of the suppliers is difficult to achieve in a competitive environment (e.g., in the context of capacity mechanisms). To overcome this problem, I propose a payoff scheme based on marginal contributions and the Shapley value which may, for instance, be applied in a centralized procedure. The proposed concept exhibits desirable properties, including static efficiency as well as efficient investment incentives. In order to demonstrate the relevance and applicability of the concepts developed, I investigate an empirical example based on wind power in Germany, thereby confirming my analytical findings. In practice, my approach could improve the design of capacity or renewable support mechanisms. More generally, the approach could be applied to organize supply chains and their reliability more efficiently.

In **Chapter 4**, I turn my attention towards the impact of electricity generation, especially from renewable energies, on the electricity grid. Due to the fact that favorable sites for renewable electricity generation are typically far away from load centers, new grid infrastructures are often needed. However, as generation and grid services are unbundled in today's liberalized power systems, it may be difficult to obtain a system outcome where planning and operational activities in the generation and grid sector are well coordinated. In this context, an appropriate regulatory design can help to improve or even resolve this coordination problem.

Key ingredient to organize the interaction between generation and grid is the way how congestion in the grid is managed. In principle, several congestion management designs are available, differing with respect to the definition of market areas, the regulation and organization of grid operators, the way of managing congestion besides grid expansion, and the type of cross-border capacity allocation. In order to investigate and compare the performance of different designs, we develop a generalized and flexible economic modeling framework based on a decomposed inter-temporal equilibrium model including generation, transmission, as well as their inter-linkages. The model covers short-run operation and long-run investments and hence, allows to assess short and long-term efficiency.

Based on our modeling framework, we are able to identify and isolate implicit frictions and sources of inefficiencies in the different regulatory designs, and to provide a comparative analysis including a benchmark against a first-best welfare-optimal result. Moreover, we provide quantitative results by calibrating and numerically solving our model for a detailed representation of the Central Western European (CWE) region, consisting of 70 nodes and 174 power lines. Analyzing six particularly relevant congestion management designs until 2030, we show that compared to the first-best benchmark, inefficiencies of up to 4.6% arise. Inefficiencies are mainly driven by the national organization of markets and responsibilities for the grid infrastructures, which could be overcome by a coordinated European approach.

Chapter 5 takes a closer look at the specific activities in the grid sector itself, and investigates a problem that was neglected in the previous section: due to the fact that the electricity grid is a natural monopoly, the firm being in charge has its own (profit-maximizing) agenda and hence, needs to be regulated in order to align its activities with social preferences. This can be a challenging task if the firm has exclusive knowledge about the economic and technical characteristics of its activities which it is not willing to share voluntarily. For instance, in the case of electricity grids, the responsible firm may have multiple options to cope with an increasing deployment of renewable energy sources, such as grid expansion or improved grid operation. At the same time, electricity systems are highly complex, and for the regulator it is often hard to observe and judge whether the mix and overall level of measures taken by the firm are adequate, let alone optimal.

This setting appears to be unresolved by the existing literature on regulation. We hence derive the theoretically optimal regulation strategy based on a menu of contracts that is able to make the firm reveal its exclusive knowledge. As an additional contribution, we then contrast our theoretical findings with other regulatory approaches that are practically applied by regulatory authorities in many countries, even though they are seemingly simplistic and outdated from a theoretical perspective. With this comparative analysis, we provide regulators with useful information about ways to improve their strategies. At the same time, we also show for the exemplary case of Germany that the relatively simple (cost-based) regulatory approach may – under certain conditions – in fact be close to the theoretically optimal strategy.

1.2 Discussion of methodological approaches

Methodological approaches were chosen and developed to suitably address the specific research question of each chapter. As discussed in Section 1.1, the specific challenges investigated in Chapters 2 and 3 stem from the interdependent temporal and spatial distribution of supply (especially from variable renewable energies) and demand. Correspondingly, we apply statistical analysis and stochastic simulation models that are able to cover the stochastic nature of the underlying problem. Interdependencies between the random variables are a main focus in both papers. However, they are depicted in greater detail in Chapter 2, where copulas, i.e., *full* multivariate dependence structures, are explicitly analyzed and modeled. In contrast, Chapter 3 represents dependence structures based on covariance, i.e., a *linear* measure of dependence. Hence, Chapter 3 could be extended by covering general, non-linear dependence structures, possibly allowing for interesting additional insights. However, it should be noticed that the analytical tractability – which is indeed one of the main advantages and contributions of Chapter 3 – is probably jeopardized, or at least very complex.

Furthermore, it could be argued that the first two chapters appear to be somewhat "incomplete". Indeed, they both conduct analyses of fundamental drivers on the supply side, especially regarding the availability of renewable resources. Both papers' goal is to determine a system value for the product at hand (value of production in Chapter 2, and value of reliability in Chapter 3). Furthermore, both papers propose payoff mechanisms that are – in contrast to existing designs – in line with these system values, while inducing desirable properties, such as static efficiency and efficient investment incentives. However, even though many implications of these payoff schemes are thoroughly discussed, further analytical techniques could be applied, such as optimization routines or game-theoretical equilibrium analysis. This would be a necessary and interesting next step in order to determine, e.g., the most promising business cases or the economic equilibrium under the suggested payoff schemes as well as possible deviations from a welfare-optimal benchmark.

The latter two papers (Chapters 4 and 5) both focus on the degree of information available to different firms in the system. Specifically, Chapter 4 focuses on information deficits stemming from the spatially aggregated uniform prices in zonal markets. In this setting, asymmetric information is exogenously introduced at the interface between the production and transmission sector by means of different congestion management designs. Importantly, this information deficit cannot be resolved by the parties involved, such that the outcome is necessarily inefficient. The reason lies in the inherent incompleteness of aggregated uniform prices that do not represent real grid scarcities. Representing such an information deficit in a fundamental (partial) equilibrium model is not easy. In fact, classic equilibrium models build upon the assumption of perfect information for all market participants. Hence, in order to introduce this regulatory deficiency, we decompose our problem into the production and transmission sector and limit the amount of information that can be exchanged. This is a novel methodological approach that was derived based on economic principles and proved to be applicable to large-scale problems while producing consistent and robust results. For instance, we are able to confirm the economic inferiority of zonal markets in comparison to nodal pricing that necessarily follows from the introduced information deficit. Nevertheless, it is fair to say that the methodology could benefit from further research. So far, neither the existence and uniqueness of a global optimum of the problem, nor the convergence of the solution algorithm have been analytically proven. Moreover, we build on the assumption of a perfectly regulated (i.e., cost-minimizing) firm responsible for the operation and expansion of the transmission sector. As such, we only consider information deficits induced by a flawed market design and disregard *strategic* withholding of information. While this isolation is indeed intended for this paper, its integration (especially on the transmission side) would represent an interesting extension. However, such an analysis would render the numerical solution even more complex, and would probably come at the cost of a higher level of abstraction to ensure numerical feasibility.

In a certain sense, the methodology of Chapter 5 is on the opposite side of the one applied in Chapter 4. It is a purely theoretical and highly stylized principle-agent model. Yet, it essentially considers a similar problem of information deficits between the involved parties (here, e.g., a monopolistic transmission firm and the regulator). However, in contrast to Chapter 4, the information deficit – even though also introduced exogenously – can here be resolved by means of a suitable contract framework. This endogeneity of information revelation is the reason why the framework is so complex (and interesting) to solve, and why the analytical analysis needs such a high level of abstraction.

Lastly, I shall mention that my entire thesis largely disregards the complexities of the retail sector, and instead assumes an inelastic demand in all chapters. There are essentially two reasons why this assumption is made: First, demand for electricity is indeed relatively inelastic, especially in the short-term. This is mainly due to large shares of consumers not being exposed to short-term price variations, rather paying a uniform price that only changes, e.g., once a year – even though consumers' utility from consuming electricity is in fact often highly time-dependent (for instance, electric lighting is an essential service that people are typically not willing to postpone). Second, assuming an inelastic demand often facilitates the analytical and numerical analysis and solution. For instance, it may under certain conditions allow to formulate a non-linear welfare-maximization as a linear cost-minimization that can be solved much more easily (as done in Chapter 4, for instance). Despite these reasons, the inelastic demand assumption can be seen as a fairly strong assumption. Hence, even though the analysis will quickly become complex, future research could relax this assumption and introduce elastic demand functions.

1.3 Concluding remarks

Nearly all energy-related (research) questions can be posed and answered from different perspectives, using different assumptions, methodologies and ways of interpretation. Unsurprisingly, energy-related research is hence conducted by several academic disciplines, among which the technical and economic sciences are probably the ones that are involved the most. Moreover, interdisciplinary interfaces are often included or even crossed in order to embrace the full scope of a problem.

My thesis is primarily meant as a contribution in the field of energy economics, while in addition, some of the papers reach out to a generalized economic problem and broader area of interest and application.[2] It applies economic ideas and concepts, and provides novel approaches and insights regarding the economics and regulatory design of power systems. Specifically, it focuses on the *efficiency* of design alternatives, which serves as an objective analysis framework and basis of valuation. While doing so, I largely take into account the technological features of the underlying system, e.g., aspects of meteorology, technical functionalities of power plants, or Kirchhoff's laws in power grids. In contrast, however, I focus less on aspects of *equity*, i.e., the question how resources could be distributed throughout society in a way that we consider to be fair. At the same time, I believe that one of the main future challenges in the energy sector will be to complement economically efficient designs with notions of equity. To this end, economic analysis could be joined by political and sociological approaches. As an example, let me mention the European discussion about nodal pricing. Even though many researchers have clearly argued in favor of a nodal pricing regime in terms of efficiency, it seems that (political) decision-makers have been avoiding serious consideration of this issue, perhaps due to a general aversion towards changes, but certainly also because an implementation would entail significant redistributional effects. It will be necessary yet challenging to overcome such barriers, e.g., by means of second-best approaches that still provide a high level of economic efficiency while at the same time being politically feasible.[3] To this end, distributive effects will need to be taken into perspective to make efficient solutions work.

[2] This mainly applies to the contents of Chapter 3 and 5. Both papers take power systems as an (important) exemplary area of application, but refer to more general issues (i.e., supply chain reliability in the former, and monopoly regulation in the latter paper).

[3] In the mentioned example, it could for instance be interesting to only expose the supply side to nodal prices, while retail prices remain uniform throughout the country.

2 Spatial dependencies of wind power and interrelations with spot price dynamics

Wind power has seen strong growth over the last decade and increasingly affects electricity spot prices. In particular, prices are more volatile due to the stochastic nature of wind, such that more generation of wind energy yields lower prices. Therefore, it is important to assess the value of wind power at different locations not only for an investor but for the electricity system as a whole. In this paper, we develop a stochastic simulation model that captures the full spatial dependence structure of wind power by using copulas, incorporated into a supply and demand based model for the electricity spot price. This model is calibrated with German data. We find that the specific location of a turbine – i.e., its spatial dependence with respect to the aggregated wind power in the system – is of high relevance for its value. Many of the locations analyzed show an upper tail dependence that adversely impacts the market value. Therefore, a model that assumes a linear dependence structure would systematically overestimate the market value of wind power in many cases. This effect becomes more important for increasing levels of wind power penetration and may render the large-scale integration into markets more difficult.

2.1 Introduction

The amount of electricity generated by wind power plants has increased significantly during recent years. Due to the fact that wind power is stochastic, its introduction into power systems caused changes in electricity spot price dynamics: Prices have become more volatile and exhibit a correlated behavior with wind power fed into the system. In times of high wind, spot prices are observed to be generally lower than in times with low production of wind power plants. Empirical evidence of this effect has been demonstrated for different markets characterized by high wind power penetration, e.g., by Jónsson et al. (2010) for Denmark, Gelabert et al. (2011) for Spain, Woo et al. (2011) for Texas or Cutler et al. (2011) for the Australian market. Due to the cost-free availability of wind energy, wind power plants are characterized by marginal costs of generation that are close to zero and therefore lower than those for other types of power plants such as coal or gas. Hence – if the wind blows – wind power replaces other types of generation and thus leads to lower spot market prices in such hours. As a consequence, power plants are faced with increasingly difficult conditions and an additional source of price risk when participating in the market. Until now, fluctuating renewable energy technologies (including wind power itself) have often been exempted from this price risk by support mechanisms (e.g., by fixed feed-in-tariff systems) in order to incentivize investments. However, their price risk draws more and more attention as they make up an increasing share of the generation mix and may at some point be fully integrated in the liberalized power market. Therefore, for an individual investor as well as for a social planner it becomes increasingly important to understand the value of wind generation and how it depends on the location of the wind turbine.

The purpose of this paper is to derive revenue distributions and the market value of wind power, i.e., the weighted average spot price wind power is able to achieve when selling its electricity on the spot market, at specific locations. It is clear that the value of a wind turbine at a specific location depends on whether it tends to produce when many other wind turbines at other locations can also produce, or whether it is one out of few producers at a given time. To capture the full stochastic dependence structure of wind power, we use copulas and incorporate the stochastic wind generation in a supply and demand based model for electricity prices. More precisely, we take the following two steps. At first, we develop a stochastic simulation model for electricity spot prices that incorporates the market's aggregated wind power as one of the determinants. We use the residual demand, given by the difference of total demand and aggregated wind power, to establish the relationship between wind power and spot prices. Secondly, we link the market's aggregated wind power to the wind power of single turbines in order to quantify their market value and the revenues depending on their specific location. We use copulas to model this interrelation. The model is calibrated with German data, since Germany already has a high share of wind power.

We find that taking into account the entire spatial dependence structure is indeed necessary, and that considering only correlations between a specific turbine and the aggregate wind power would be misleading. Even if the correlation of a specific turbine is lower compared to another, the resulting market value may be lower due to a non-linear, asymmetric dependence structure. In fact, we find a pronounced upper tail dependence that adversely impacts the market value for many of the locations analyzed. Therefore, a model solely based on linear dependence measures would systematically overestimate the market value of wind power in many cases. Moreover, it is shown that this effect becomes increasingly important for higher levels of wind power penetration.

Our paper contributes to three lines of literature. First, we complement the literature on supply and demand based models. Within this class of models, Bessembinder and Lemmon (2002) were among the first to study the importance of demand and production costs for electricity prices. The model developed by Burger et al. (2006) follows a similar conceptual approach by including a non-linear functional dependence of the electricity spot price on a stochastic demand process as well as a long-term non-stationarity. Howison and Coulon (2009) further extend the number of state variables explaining the electricity spot price by including fuel prices. With our paper, we contribute to this line of literature by including stochastic production quantities of wind power that may impact the supply side and hence electricity prices.

Secondly, we extend the literature employing copulas, especially in the context of wind power applications. Copulas have first been identified by Papaefthymiou (2006) to be a suitable tool for modeling multivariate dependencies of wind speeds. Subsequently, copulas have been employed in different studies to model spatial and temporal dependencies of wind speeds or wind power. Spatial dependencies have been modeled with the help of copulas by Haghi et al. (2010) for PV and wind power as well as system load in an Iranian case study, by Grothe and Schnieders (2011) for wind speeds in an optimization problem minimizing aggregated wind power fluctuations in Germany, by Hagspiel et al. (2012) for wind speeds in a European probabilistic load flow analysis, and by Louie (2014) for power generation from a multitude of pairs of wind turbines in order to identify the best-suited bivariate copula models. In contrast, copulas are used for temporal dependence structures in Pinson and Girard (2012) to model the multivariate stochastic process of short-term wind power trajectories (based on a methodology developed in Pinson et al. (2009)) and in Zhou et al. (2013) to investigate wind power forecasting based on probabilistic kernel densities. We contribute to this second line of literature by applying conditional copulas to model the dependence structure between specific turbines and the aggregated wind power. This approach allows us to specify and investigate interrelations between the physical characteristics of a wind turbine and electricity spot prices, and hence to value wind power assets more appropriately.

In fact, the valuation of power generation assets is the third line of literature our paper complements. So far, research on the valuation of power generation assets has mainly focused on conventional power (e.g., Thompson et al. (2004), Porchet et al. (2009) or Falbo et al. (2010)) and the optimization of hydro power schedules (e.g., García-González et al. (2007) or Densing (2013)). The relatively few papers that deal with the valuation of wind power is primarily based on historical data of wind power and day-ahead market prices (e.g., Green and Vasilakos (2012)). In a recent study presented by Girard et al. (2013), wind power predictability is assessed as a decision factor during the planning phase of a wind power project, showing that the financial loss due to imbalance costs induced by imperfect predictions only represents a low share of revenue in the day-ahead market. Even though they find that the aggregation of wind farms over large distances has an impact on the market value, they do not further elaborate on spatial dependencies. Our paper concentrates on this particular issue and shows that spatial dependencies are indeed crucial for the market value of wind power projects, especially for increasing penetration levels.

The remainder of this article is organized as follows: Section 2.2 provides a short introduction to copula modeling with a particular focus on conditional copula sampling which we apply in our model. The model itself is presented in Section 2.3. Section 2.4 reports the results of the methodology applied to the case of wind power in Germany, namely the revenues and the market value of specific wind turbines. Section 2.5 concludes.

2.2 Stochastic dependence modeling using copulas

In this section, we briefly discuss the modeling of stochastic dependencies with the help of copulas. A more detailed introduction is provided e.g., in Joe (1997), Nelsen (2006) or Alexander (2008). For a comprehensive literature review of the current status and applications of copula models, the interested reader is referred to Genest et al. (2009), Durante and Sempi (2010) and Patton (2012).

2.2.1 Copulas and copula models

A copula is a cumulative distribution function with uniformly distributed marginals on $[0,1]$. Sklar's theorem is the main theorem for most applications of copulas, stating that any joint distribution of some random variables is determined by their marginal distributions and the copula (Sklar (1959)). The bivariate form of Sklar's theorem is as follows: For the cumulative distribution function $F : \mathbb{R}^2 \to [0,1]$ of any random variables X, W, with marginal distribution functions F_X, F_W, there exists a copula $C : [0,1]^2 \rightarrow [0,1]$ such that

$$F(x,w) = C\left(F_X(x), F_W(w)\right). \tag{2.1}$$

Sklar's theorem also holds for the multivariate case of $n > 2$ dimensions. The copula function is unique if the marginals are continuous. Conversely, if C is a copula and F_X and F_W are continuous distribution functions of the random variables X, W, then (2.1) defines the bivariate joint distribution function. From Sklar's theorem, it follows that copulas can be applied with any marginal distributions. Particularly, marginal distributions may differ for each of the random variables considered.

In our application we are interested in the dependence structure of the market's aggregated wind power W and a single turbine wind power X. The copula captures the complete dependence structure of X and W. The selection of an appropriate copula model can be made independent from the choice of the marginal distribution functions. Taking advantage of this, the joint distribution of W and X is determined in a two stage process: First, the marginal distribution functions F_W and F_X are determined, followed by the selection of the most appropriate copula model.

Copula functions are mostly determined in a parametric way. There are different types of parametric copula models that can be used to capture the pairwise dependence. In many applications – such as ours – it is particularly important to differentiate between symmetric or asymmetric, tail or no tail, and upper or lower tail dependence structures. Therefore, one can test several parametric copula models that are able to capture these characteristics: The Gaussian copula is symmetric and has zero or weak tail dependence (unless the correlation is 1). In contrast, the symmetric Student-t copula has a relatively strong symmetric tail dependence. Whereas the Frank copula is another symmetric copula with particularly low tail dependence, Clayton and Gumbel copulas incorporate an asymmetric tail dependence. Lower tail dependence is captured by the Clayton copula, while the Gumbel copula incorporates an upper tail dependence.[1] These copulas are listed in Table 2.1.

The marginals u and v can be interpreted as $F_X(x)$ and $F_W(w)$, respectively. Φ_Σ denotes the multivariate normal distribution function with covariance matrix Σ and $t_{\Sigma,\nu}$ the multivariate Student-t distribution with ν degrees of freedom and covariance matrix Σ. The copula parameters can be estimated based on observed data by

[1]Gaussian and Student-t copulas belong to the group of Elliptical copulas, whereas Frank, Gumbel and Clayton copulas belong to the group of Archimedian copulas. For a more extensive discussion of different copula families, see, e.g., Nelsen (2006)

Copula family	**Copula function** $C(u,v)$
Gaussian	$\Phi_\Sigma\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)$
Student-t	$t_{\Sigma,\nu}\left(t_\nu^{-1}(u),t_\nu^{-1}(v)\right)$
Clayton	$\left(\max\left\{u^{-\theta}+v^{-\theta}-1,0\right\}\right)^{-1/\theta}$
Frank	$\frac{-1}{\theta}\ln\left(1+\frac{\left(e^{-u\theta}-1\right)\left(e^{-v\theta}-1\right)}{e^{-\theta}-1}\right)$
Gumbel	$e^{-\left((-\ln(u))^\theta+(-\ln(v))^\theta\right)^{1/\theta}}$

Table 2.1: Copula models

optimizing the log-likelihood function:

$$\hat{\theta} = \max_{\theta} \sum_t \ln c\left(F_X\left(x_t\right), F_W\left(w_t\right); \theta\right) \tag{2.2}$$

where θ denotes the parameter vector and c the copula density. The selection of the most appropriate copula model can then be determined based on the Akaike Information Criteria (AIC).

2.2.2 Conditional copula and simulation procedure

Like any ordinary joint distribution function, copulas have conditional distribution functions. The conditional copula can be calculated by taking first derivatives with respect to each variable, i.e., for $u = F_X(x)$ and $v = F_W(w)$ we have

$$C(u|v) = \frac{\partial C(u,v)}{\partial v} \quad \text{and} \quad C(v|u) = \frac{\partial C(u,v)}{\partial u}. \tag{2.3}$$

For the application presented in this paper, there is one inherent advantage of using conditional copulas rather than sampling directly from the bivariate copula distribution: Samples can be conditioned on time series that may serve as inputs to the simulation procedure. The time series characteristics can thus be preserved during the simulation process. We use time series of the market's aggregated wind power as an input variable for the spot price model.

We consider the stochastic processes $(X_t)_{t\in\mathbb{N}}$ and $(W_t)_{t\in\mathbb{N}}$. $F_{X_t}(X_t), F_{W_t}(W_t)$ are uniformly distributed random variables on $[0,1]$. For random variables $U_t, V_t \sim U(0,1)$, $F_{X_t}^{-1}(U_t)$ and $F_{W_t}^{-1}(V_t)$ thus follow the distributions of X_t and W_t, respectively. It is important to notice that by applying the inverse distribution functions, the dependence structure is not influenced, i.e., U_t and V_t as well as $F_{X_t}(X_t)$ and $F_{W_t}(W_t)$ have the same copula C.

The conditional sampling procedure can be summarized as follows:

1. Apply the marginal distribution function F_{W_t} to the time series of the market's aggregated wind power $(w_1, w_2, w_3, ...)$ in order to get $\left(v_1^*, v_2^*, v_3^*, ...\right)$.
2. Simulate $(u_1, u_2, u_3, ...)$ from independent uniformly distributed random variables.
3. For each observation $F_{W_t}(w_t) = v_t^*$, apply the inverse conditional copula $C^{-1}_{F_{W_t}(W_t),F_{X_t}(X_t)}\left(\cdot|v_t^*\right)$ to translate u_t into u_t^* by:

$$u_t^* = C^{-1}_{F_{W_t}(W_t),F_{X_t}(X_t)}\left(u_t|v_t^*\right) \tag{2.4}$$

4. Apply the inverse marginal distribution functions to $\left(u_1^*, u_2^*, u_3^*, ...\right)$ in order to obtain the corresponding simulations of the random variable X_t: $\left(F^{-1}_{X_1}\left(u_1^*\right), F^{-1}_{X_2}\left(u_2^*\right), F^{-1}_{X_3}\left(u_3^*\right), ...\right)$.

2.3 The model

We develop a stochastic simulation model for the single turbine wind power and electricity spot prices, including a precise representation of their interrelations. The interrelation is established by the aggregated wind power that is related to both the electricity spot prices as well as the single turbine wind power. Hence, we set up a model that represents these two relationships: First, a supply and demand based model that takes, among others, the aggregated wind power as an input. Second, a stochastic dependence model that links the single turbine wind power to the aggregated wind power. These two parts of the model can be summarized by the following two equations:

$$S_t = h_t\left(D_t - W_t\right) + Z_t \tag{2.5}$$

$$X_t = F^{-1}_{X_t}\left(C^{-1}_{F_{X_t}(X_t),F_{W_t}(W_t)}\left(U_t|F_{W_t}\left(W_t\right)\right)\right) \tag{2.6}$$

where S_t is the hourly stochastic spot price and X_t the hourly single turbine wind power, for $t \in \mathbb{N}$. The spot price S_t is determined by two components: First, the function h_t describes the dependence of the spot price on the residual demand that is determined by the difference of the electricity demand level D_t and the stochastic aggregated wind power W_t. Second, a short term stochastic component adds to the spot price that is denoted by Z_t. As operators of wind power plants are able to curtail their power output in case of negative spot prices, their price is non-negative, i.e., $S_t^W = \max\{0, S_t\}$.

The second part of the model links the hourly single turbine wind power X_t to the aggregated wind power W_t. F_{X_t} and F_{W_t} denote the corresponding marginal distribution functions. The joint distribution function of these two random variables is determined by the corresponding copula, i.e., $F_{X_t,W_t}(x_t, w_t) = C\left(F_{X_t}(x_t), F_{W_t}(w_t)\right)$. Due to the copula's ability to separate marginal distribution functions and the dependence structure, the joint distribution function can be modeled in a two-step process: First, the marginal distribution functions $F_{X,t}$ and $F_{W,t}$ are determined. Second, the appropriate copula $C_{F_{X_t}(X_t),F_{W_t}(W_t)}$ is selected and estimated. We deploy the conditional copula in order to keep the time series properties of the stochastic process $(W_t)_{t\in\mathbb{N}}$. For the simulation procedure, independent $[0,1]$-uniformly random variables U_t are needed. Note that the marginal distribution functions are the same within a month m, i.e., $F_{X_i} = F_{X_j}$ if $i, j \in m$. The same holds for F_{W_t}, h_t and $C_{F_{X_t}(X_t),F_{W_t}(W_t)}$.

Based on Equations (2.5) and (2.6), the amount of hourly wind power produced by a single turbine and the spot prices can be simulated. We sample from these model equations using a Monte Carlo simulation (n=10000) in order to investigate the market value and revenue distributions as well as the relevance of the dependence structure with the aggregated wind power for single turbines at different locations.

While the revenue is simply the sum of the products of electricity generation and prices, the market value of a wind turbine is the average spot price weighted with the electricity generation of the respective wind turbine:

$$MV = \frac{\sum_t X_t S_t}{\sum_t X_t}. \tag{2.7}$$

In the following subsections, we explain the input parameters and the different parts of the model in more detail.

2.3.1 The data

Different data sets are deployed in order to calibrate and estimate the different parts of the model. In the following, we explain the content and origin of these sets, as well as the way in which the data are preprocessed.

Expected and realized generation by the German aggregated wind power: For the supply and demand based model (represented by h_t in Equation (5)), data is needed on the effectively delivered day-ahead prognosis of the German aggregated wind power in 2011. Note that the day-ahead prognosis – and not the actual aggregated wind power – is used, since this is the relevant information for the day-ahead market (Jónsson et al. (2010))). In contrast, for the estimation of the appropriate copula (C) the realized generation of 2011 is used in order to determine its dependence structure with the realized single turbine wind power at different locations. Expected as well as realized generation data are provided by the transmission system operators and published on the EEX Transparency Platform (EEX Transparency Platform (2012)).

Wind speeds of single stations: Hourly mean wind speeds for various stations in Germany are provided via the national climate monitoring of the German Weather Service for the years 1990-2011 (DWD (2014)). The measurement data for 19 locations are used in this project to determine the corresponding power output series of wind turbines.[2] Wind speeds are scaled to the hub height of currently installed wind turbines (100 meters) assuming a power law: $v_{h_1} = v_{h_0}(h_1/h_0)^{\alpha}$, where h_0 is the measurement height, h_1 the height of interest and α the shear exponent. According to Firtin et al. (2011), α is assumed to be 0.14.

Wind power capacities: The development of currently installed wind power capacities per federal state between 1995 and 2011 is available from the German Wind Energy Association (German Wind Energy Association (BWE) (2012)). In 2011, installed wind power capacities in Germany amounted to 27.1 GW.

Electricity demand levels: Hourly electricity demand levels for the German market in 2011 — used as one of the explaining variables for spot prices and denoted by D_t in Equation (2.5) — are provided by ENTSO-E (2012).

Spot prices: EPEX day-ahead prices from 2011 are deployed for the calibration of the spot price model (Equation (2.5)). The EPEX day-ahead market is organized by an auctioning process that matches supply and demand curves once a day, thus determining prices at which electricity is exchanged in each respective hour.

[2] Missing data are interpolated based on the previous and next available value if the missing gap is not exceeding 12 hours. If the gap is longer, the values are replaced by data of the same station and same hours of the previous year.

2.3.2 Derivation of synthetic aggregated wind power

As an important input for the model, curves are needed that describe the wind power that the currently installed wind power capacities would have produced during the last decades (i.e., the long-term stochastic behavior of aggregated wind power in the power system). In the model, the curve is needed for the estimation of the marginal distribution $F_{W,t}$ of the aggregated wind power W_t. It is important to notice that this curve has to be derived synthetically, as wind power capacities changed significantly during the last years.

Based on wind speeds and wind power capacities, the synthetic German aggregated wind power is generated as follows: By applying a power curve capturing the characteristics of the transformation process from wind energy to electrical power, wind turbine power generation profiles can be derived. In this study, the power curve is assumed to be one of a GE 2.5 MW turbine (General Electric (2010)). Alternatively, one could use an average taken from multiple turbines. The transformation is based on a look-up table derived from the power curve and linear interpolation. Furthermore, electrical output is determined as a ratio of installed wind power capacity (i.e., scaled to $[0, 1]$). Multiplying this ratio with the wind power capacity installed in the corresponding federal state yields the wind power. The above steps are repeated for 16 locations (one for each federal state) and all available years (1990–2011), resulting in a time series for what would have been produced during the last 22 years with current wind power capacities. In order to check the plausibility of this approach, historical wind energy time series and volumes can be compared to the model estimates. The comparison for the 2011 time series yields high conformity with an R^2 of 0.84. Another check of consistency is done by calculating the accumulated aggregated wind power production volumes for the past 10 years from the synthetically generated curves, and comparing them to the overall wind power production as reported in Eurostat (2012). We find the deviations to be less than 12%.

2.3.3 Supply and demand based model for the electricity spot price

We develop a supply and demand based model to derive electricity spot prices dependent on the level of wind power. A similar approach has been applied in Burger et al. (2006). The main difference between our and their approach is that we use the residual demand instead of total demand. We are therefore able to integrate the effect of wind power on spot prices.

We describe the non-linear relation between residual demand and spot prices (i.e., h_t in Equation (2.5)) by an empirical function estimated from historical hourly spot prices, demand and wind power data. To derive a functional form for h_t we use spline fits which are suitable to capture the non-linearities in the demand-price dependence. The parameters of h_t are estimated from historical data for the reference year 2011 on a monthly basis in order to capture seasonal differences and variations on the supply side that occur, e.g., because of planned outages or variations in fuel costs.

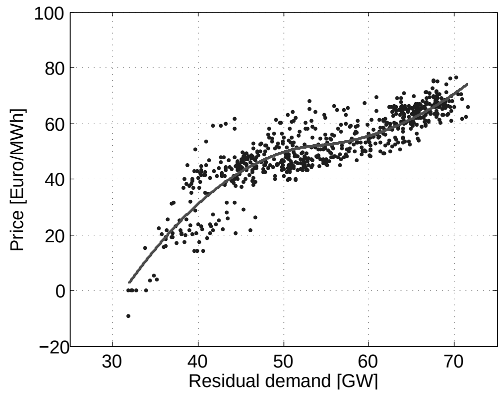


Figure 2.1: Demand-price dependence in February 2011 and spline fit

Note that if demand were totally price-inelastic, the function h_t would approximate the supply curve that represents all available sources of electricity generation ranked in ascending order of their marginal generation costs (excluding wind energy) that is often referred to as the merit order. Even though the electricity is generally very inelastic in the short term, there may be some price-response of demand. Hence, our function h_t should not be seen as an unbiased estimator of the merit order.

The data and the corresponding spline fit are shown in Figure 2.1 for the month of February 2011. All other months of 2011 are presented in Figure 2.10 in the Appendix. As can be observed, the dependence between residual demand levels and prices is characterized by steep ends and a comparatively flat part in between (i.e., for the residual demand ranging between 40 and 70 GW). The steeper part in the lower tail is generally more pronounced than the price increase for higher residual demand levels. Rather moderate price increases in the upper tail may be interpreted by prevailing excess capacity in the German power market, leading to very few instances at which scarcity prices occur.

Besides the functional dependence on (residual) demand, additional stochastic factors influence spot market prices such as speculation, unplanned power plant outages or scarcity prices or demand side management. These effects are lumped together and captured by the residual price process Z_t in Equation (2.5). In the following, we aim at finding a model for Z_t that is capable of capturing the characteristics observed in the data. We can derive the observed residual price component based on h_t, the observations of residual demand and spot prices from $z_t = s_t - h_t$, and use the result for the calibration of the residual price process $(Z_t)_{t\in\mathbb{N}}$. The time series z_t is visually observed to be stationary within the considered time frame, which is confirmed by an augmented Dickey-Fuller test that indicates that the null hypothesis of a unit root can be rejected at the 95% level.

The empirical auto-correlation function of z_t decays slowly, however, with an apparent dependence at a lag of 24 hours. We therefore choose to model Z_t as a seasonal ARIMA (SARIMA) model with a 24 hour seasonality. In order to do so, the ARIMA model needs to be extended to include non-zero coefficients at lag s, where s is the identified seasonality period. SARIMA models can be specified in a multiplicative form, resulting in a more parsimonious model than simply extending ARIMA to s lags.

As the Engle's ARCH test indicates that there is conditional heteroscedasticity in the data, we extend the SARIMA by a GARCH component. GARCH-type models are able to capture conditional heteroscedasticity by splitting the error term ϵ_t into a stochastic component η_t and a time-dependent standard deviation σ_t. The latter can then be expressed dependent on lagged elements of ϵ and σ_t (Engle (1982), Bollerslev (1986)).

Various specifications of SARIMA-GARCH models are estimated and evaluated. Based on the AIC, a GARCH(1,1)-SARIMA$(2,0,2)\times(1,0,1)_{24}$ model is found to perform best. The inclusion of additional parameters hardly improves the fit. Note that no constant needs to be added to the model of Z_t due to the fact that the process has been already centered by applying a spline fit.

Comparing the residual's distribution to the normal distribution yields unsatisfactory results (Figure 2.2, left hand side). Thus, alternatively, the error term can be specified as a t-distribution which leads to an improved match of the distributional shapes (Figure 2.2, right hand side). Instead of $\eta_t \sim \mathcal{N}(\mu, \sigma^2)$ we therefore use $\eta_t \sim t(\nu)$, with ν being the t-distribution's degrees of freedom that are estimated from the data.

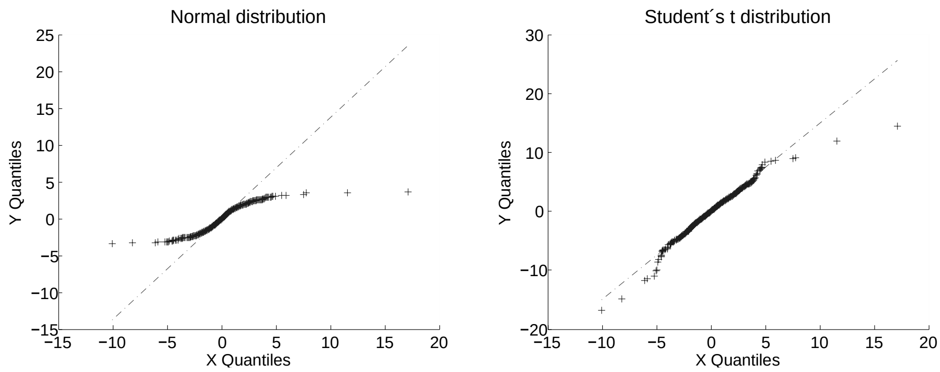


Figure 2.2: QQ-plots of the 2011 residuals compared against a normal distribution and a Student-t distribution

Written explicitly, the model for Z_t now takes the following form:

$$\begin{aligned} Z_t =& \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \Phi_1 Z_{t-24} + \Phi_1(\phi_1 Z_{t-25} - \phi_2 Z_{t-26}) \\ & + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \Theta_1 \epsilon_{t-24} + \Theta_1(\theta_1 \epsilon_{t-25} - \theta_2 \epsilon_{t-26}) \end{aligned} \tag{2.8}$$

$$\epsilon_t = \sigma_t \eta_t \tag{2.9}$$

$$\sigma_t^2 = \alpha + \beta_1 \epsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \tag{2.10}$$

$$\eta_t \sim t(\nu) \tag{2.11}$$

The parameters for the above model are estimated from the time series z_t by optimizing the log-likelihood function. The estimates are presented in Table 2.2.

Parameter	ϕ_1	ϕ_2	Φ_1	θ_1	θ_2	Θ_1	α	β_1	γ_1	ν
Estimate	0.37	0.36	0.97	0.57	0.07	-0.85	2.96	0.30	0.47	3.61
Std. Error	0.15	0.12	0.00	0.15	0.02	0.01	0.03	0.03	0.27	0.16

Table 2.2: Parameter estimates for the residual price process model

2.3.4 Estimation and selection of copula models

In this section, we select and estimate models for the joint distribution of a single turbine wind power and the German aggregated wind power for 19 wind power stations in Germany[3]. We apply the two-stage process introduced in Section 2.2: First,

[3]We determine the models for the joint distribution functions between the German aggregated wind power and the following stations: Aachen, Angermünde, Augsburg, Bremen, Dresden, Emden, Erfurt-Weimar, Idar-Oberstein, Kahler Asten, Kleiner Feldberg, Konstanz, Leipzig-Halle, Magdeburg, Münster-Osnabrück, Oldenburg, Potsdam, Rostock, Saarbrücken and Schleswig.

the marginal distributions are determined, followed by the selection and estimation of the copula model that best describes the dependence structure.

In order to determine the marginal distributions, we consider the hourly synthetic wind power data for the years 1990–2011 for the different stations as well as for the German aggregated wind power. The yearly data is split into monthly intervals in order to capture seasonal differences. We thus obtain 22x12 subsamples from which we get 22x12 empirical distribution functions. With 22 years, the data covers a wide range of weather uncertainties that largely determine the quantity risk of wind power. Furthermore, the extensive database allows us to use the empirical distribution functions as marginal distribution functions (F_{W_t}, F_{X_t}) of the two variables of interest, namely the single turbine wind power and the aggregated wind power.[4]

The copula model $C_{F_{X_t}(X_t),F_{W_t}(W_t)}$ is estimated from the data of the realized German aggregated wind power in 2011 and the corresponding hourly single turbine wind power. We hereby avoid a possible source of imprecision in the dependence structure by relying on observed rather than synthetically generated data. Moreover, with subsamples consisting of approximately 700 observations, the database is sufficiently large for a reliable estimation of the copula parameters. Just as the empirical distribution functions, the copula models are selected and estimated on a monthly basis.

To find the most appropriate copula model, various types are fitted to the data based on the procedure introduced in Section 2.2.1.[5] Table 2.8 in 2.6 report the copulas that provide the best fit to the data in terms of AIC for all stations that are considered in this paper. Note that the data we use may comprise ties, especially for the upper and lower bounds of the single turbine wind power when zero and nominal power output is observed multiple times. As ties can affect the copula estimation quality, we apply the approach described in Kojadinovic and Yan (2010) and construct pseudo-observations by randomly breaking the ties.

In the following, we will first concentrate on particular stations (namely *Bremen*, *Kleiner Feldberg* and *Augsburg*) in order to point out the most important aspects with respect to the dependence structure and the effect on the results. *Bremen* is located in northern Germany where most of the current wind capacity is installed due to generally high average wind speeds. *Kleiner Feldberg* is a mountain in central Germany,

[4]Note that instead of using the empirical distribution function, a parametric distribution, e.g., a beta distribution (which is supported on a bounded interval), could be assumed or estimated. This would become particularly attractive when there is a lack of data.

[5]The following copula models are tested: Gaussian copulas, Frank copulas, Clayton copulas, Gumbel copulas and Student-t copulas for ν=1,2,3,4,5,10,20,30,40,50.

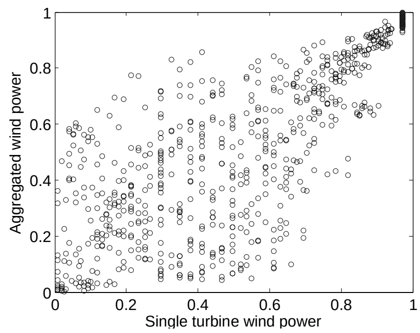
also characterized by comparatively favorable wind speeds but less surrounded by other wind turbines. Finally, we analyze *Augsburg*, which is located in southern Germany and far away from most wind power capacities. *Augsburg* has the fewest full load hours among the three stations considered. Table 2.3 lists the copulas providing the best fit to the data (in terms of AIC) for these three locations in every month. The table reporting the AIC values for all months and all copulas fitted to the data of the three stations considered is provided in Appendix A. For *Bremen* and *Augsburg*, the copula that provides the best fit in almost every month is the Gumbel copula. For these locations, there is a distinctive asymmetric upper tail dependence in the dependence structure of the single turbine wind power and the aggregated wind power. In contrast, there is hardly any tail dependence for the turbine located at *Kleiner Feldberg*. Here, most of the copulas that best fit the data are symmetric (Gaussian, Student-t and Frank copula).

Month	**Augsburg**	**Bremen**	**Kleiner Feldberg**
January	Gaussian	T40	Gaussian
February	Gumbel	Gumbel	Gaussian
March	Gumbel	Gumbel	Frank
April	Gumbel	Gumbel	Frank
May	Gumbel	Gumbel	Frank
June	Gumbel	Gumbel	Clayton
July	Frank	Gumbel	Frank
August	Gumbel	Gumbel	Frank
September	Gumbel	Gumbel	T10
October	Gumbel	Gumbel	Gaussian
November	Gumbel	Gumbel	Frank
December	Frank	T10	Gaussian

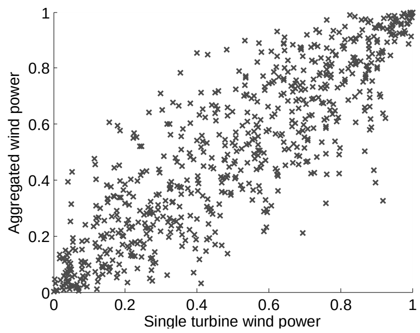
Table 2.3: Copula selection for the three stations of interest

Once the marginal distributions and copulas are estimated, the conditional copula model can be used to simulate the single turbine wind power conditional on the German aggregated wind power, based on the the sampling procedure that was introduced in Section 2.2.2. We loop through the 22 years and the 12 months of data and draw $n = 10000$ samples of the single turbine wind power for each point of the aggregated wind power curve, while applying the corresponding single turbine marginal distribution out of the 22x12 available.

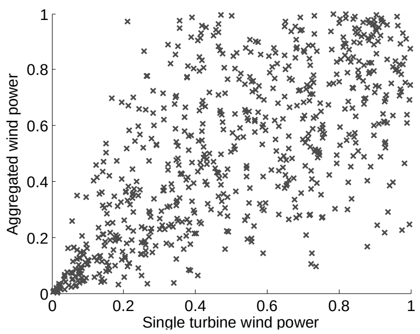
As an example, Figure 2.3 shows the dependence structure of the original data as well as simulations from three different types of copula models for a wind turbine in *Bremen*. Visually, the Gumbel copula provides the best fit to the data, which is confirmed by the comparison of the AIC. It can be observed that there is a distinctive upper tail dependence between the single turbine wind power and the German aggregated wind power. It should be noted that this type of dependence is generally undesirable for wind turbines selling their power on the spot market, as there is a high probability that spot prices are low in case of high power generation.



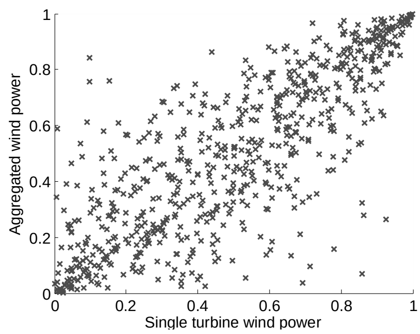
(a) Original data



(b) Gaussian copula



(c) Clayton copula



(d) Gumbel copula

Figure 2.3: Dependence structure of the original data and simulations from three copula models

Figure 2.4 shows the original data together with simulations from the Gumbel copula for the single turbine wind power located in *Bremen* and the aggregated wind power, transformed back to their marginal distributions. The turbine is assumed to be a single GE 2.5 MW turbine. As can be seen, simulations match the original data very well.

2.4 Results

This section presents the results of our simulation with respect to revenue distributions and market values at different locations. In particular, we demonstrate the relevance of the dependence structure for the market value in today's context as well as under increasing wind power penetration levels.

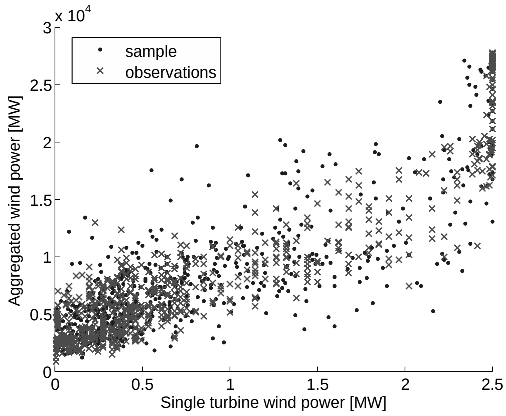


Figure 2.4: Observations and sample of the single turbine wind power and the aggregated wind power

Figure 2.5 presents the yearly revenue distribution for a wind turbine located in *Bremen* together with the 5% value at risk. The expected revenue amounts to 82000 Euro/MW/a, with a standard deviation of 3800 Euro/MW/a and a slightly negative skew. The 5% value at risk is found to be 75000 Euro/MW/a. Note that the distribution of absolute revenue is determined by both the number of full load hours that can be achieved at the specific site of interest and the corresponding market value. However, the scope of this paper lies on the dependence structures of different sites and their impact on the market value, which is thus the main focus in the following analysis.

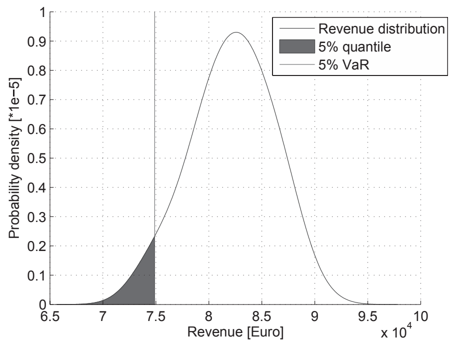


Figure 2.5: Yearly revenue distribution of the *Bremen* station and the 5% value-at-risk

2.4.1 Revenues and market value of different wind turbines

To quantify the effect arising from the dependence structures, distribution functions of the market value are determined and compared for the three stations *Augsburg*, *Bremen* and *Kleiner Feldberg*. Table 2.4 lists the main results for these three stations for the month of February. The expected average spot price of the simulations is 48.52 Euro/MWh. In contrast, the expected market value of the wind turbines is much lower for all turbines due to the dependence between the single turbine wind power and the aggregated wind power, which in turn has a price damping effect. From only the correlation coefficient ρ, one would have anticipated the expected market value of a turbine in *Augsburg* ($\rho = 0.37$) to be much higher than the expected market value of a turbine in *Kleiner Feldberg* ($\rho = 0.51$) which in turn should have a higher market value than a turbine in *Bremen* ($\rho = 0.75$). However, this is not the case: Although the correlation coefficient for a turbine in *Kleiner Feldberg* is much higher than that of a turbine in *Augsburg*, the expected market value is also higher. The reason lies in the dependence structure. As shown in Section 2.3.4, the dependence structure for *Augsburg* in February is best described by a Gumbel copula, thus incorporating an upper tail dependence between the single turbine wind power and the aggregated wind power. In contrast, the dependence structure between the single turbine wind power in *Kleiner Feldberg* and the aggregated wind power is modeled most accurately by a symmetric Gaussian copula. Therefore, *Kleiner Feldberg* benefits from an advantageous dependence structure when selling its wind power at the spot market.

	Augsburg	**Bremen**	**Kleiner Feldberg**
Expected average spot price [Euro/MWh]	48.52	48.52	48.52
Correlation coefficient	0.37	0.75	0.51
Selected copula model	Gumbel	Gumbel	Gaussian
Expected market value [Euro/MWh]	43.10	41.31	44.33
Standard deviation [Euro/MWh]	5.98	6.63	5.63

Table 2.4: Main results for the month of February

The distributions of the yearly market value for the three stations considered are shown in Figure 2.6. Following the same logic as discussed for the specific month of February, the yearly market value of a turbine in *Kleiner Feldberg* is higher than the market value for *Augsburg*. As can be seen in Table 2.3, the dependence structure for *Augsburg* is modeled with a copula incorporating an upper tail dependence in almost every month, whereas the one for *Kleiner Feldberg* is mostly symmetric. Consequently, for the three distributions that are shown in Figure 2.6, the dependence structure reduces the expected yearly market value of the turbines by 3.54, 4.97 and 2.63 Euro/MWh, respectively, compared to the expected average spot price level (49.80 Euro/MWh).

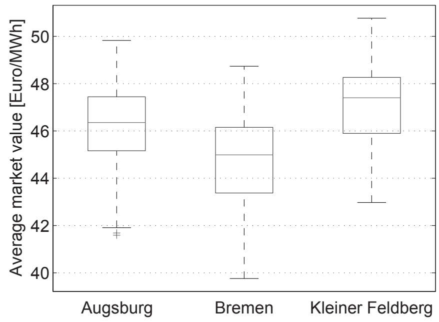


Figure 2.6: Yearly market value of the three turbines

2.4.2 Market value variations in Germany

Germany is characterized by a surface area of 357,021 km^2 and a maximum horizontal width and vertical length of 642 km and 833 km, respectively. Furthermore, there are several diverse geographical regions, suggesting that meteorological conditions may vary substantially when analyzing different locations throughout the country.

With the model developed, we analyze the market value for 19 different stations in Germany, as depicted in Figure 2.7. As the analyzed stations differ with respect to their exact location (and thus with respect to their dependence structure related to the aggregated German wind power), we expect market values to differ as well. Specifically, we expect the market value to be lowest for the stations that are closest to the majority of installed wind power. Indicated by different colors, Figure 2.7 shows the expected market value of the stations that were considered.

Results indicate that the expected market value ranges from 42 to 48 Euro/MWh for the analyzed stations, compared to an expected average spot price level of 49.80 Euro/MWh. Hence, the market value lies between 6 and 15% lower than the average spot price. As expected, lowest values are found for the stations that are closest to the majority of currently installed wind power, i.e., mainly in the area of *Magdeburg* and *Münster-Osnabrück*. For stations in this area, the dependence structure shows a pronounced asymmetric upper tail dependence. It is observed that expected market values are similar for all stations located in the so called 'North German Plain', which is a geographical region in Northern Germany characterized by constant lowlands and hardly any hills. Note that *Aachen* is at the far end of the North German Plain

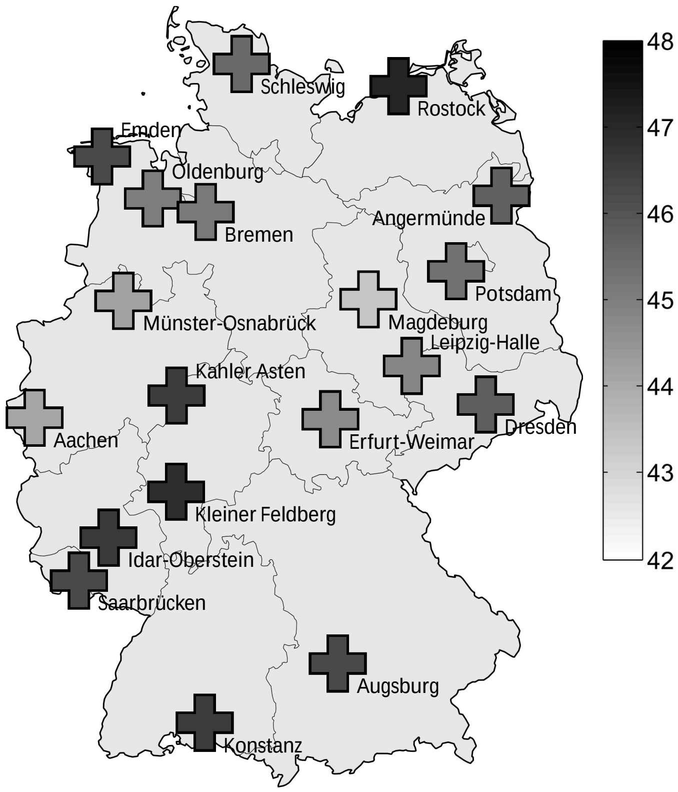


Figure 2.7: Expected market value for 19 stations in Germany (in Euro/MWh)

and, as such, equally characterized by comparatively low expected market values of 43.47 Euro/MWh. In contrast, *Kahler Asten* is located in Germany's Central Uplands, where meteorological conditions are different (e.g., due to pronounced thermals), which is reflected by higher values. Other stations in or south of the Central Uplands show higher expected market values as there are very few installed wind power capacities.

Kahler Asten and *Kleiner Feldberg* are special cases, as they are characterized by advantageous, symmetric dependence structures, resulting in expected market values that are the highest compared to the other stations considered. Similarly, *Emden* and *Rostock* – both located at the seashore – show higher values, compared to other stations in the North German Plain, due to comparatively advantageous dependence structures.

2.4.3 The impact of changing wind power penetration levels

In the previous section, model parameters were set and estimated to reflect the current environment with respect to the physical generation mix and the market conditions. In this section, some of the model parameters are modified to analyze their impact on the outcome. As has been clarified, the effect of wind power on spot market prices largely depends on the quantities of wind power being integrated in the market. With the help of the model presented in this paper, the aforementioned effect is quantified for the case of changing wind power penetration levels in Germany. First, we scale up the wind power penetration up to two times the capacity that is currently installed. Note that this is roughly in line with targets envisaged by the German government, which wants to further extend wind power to 45.8 GW in 2020 (installed capacity was 27.1 GW in 2011). Second, we compare the impact of today's wind power penetration to a situation with no wind power installed. For the analysis, installed wind power capacities are scaled-up stepwise and simulation runs are repeated for each of these steps. The underlying assumptions of this approach are as follows:

- The proportionate geographic distribution of wind power capacities within Germany remains the same. Note that due to the linear up-scaling, the dependence structure is preserved. Alternatively, region-specific changes in installed capacities could be implemented, e.g., for testing the effect of an increased wind power extension in some specific area.
- The functional dependence between residual demand levels and spot prices is again estimated from 2011 data, as explained in Section 2.3.3. This is certainly a strong assumption, as the conventional power sector will dynamically develop with increasing wind power penetration. However, it should be kept in mind that current wind power capacities are being rapidly expanded, whereas the conventional power sector seems to be behind in terms of capacity adjustments. Also note that the functional dependence could also be altered (e.g., by shifting or assuming a different shape). However, this was not implemented in order to focus on the specific impact of the wind power penetration levels.
- The parameter estimates for the residual price process remain the same. Here again, the model could be adjusted in order to represent expectations regarding future price movements.

The resulting distributions of the yearly market value of the *Bremen* station under increasing wind power penetration ranging from 100-200% are shown in Figure 2.8.

As can be observed, the market value distribution is highly affected both in average level and variance. While the expected market value is at 44.83 Euro/MWh at 100% scaling, it decreases to 30.13 Euro/MWh at a scaling of 200%. At the same time, its standard deviation increases from 1.94 to 3.40 Euro/MWh, respectively.

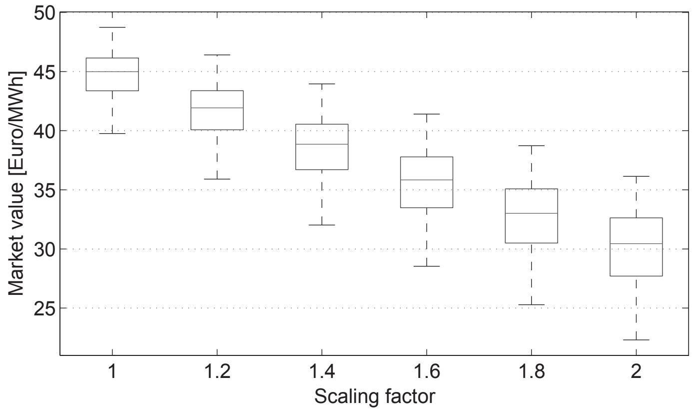


Figure 2.8: Yearly market value of the *Bremen* station under increasing wind power penetration

To achieve further insights regarding the effect of the wind power penetration level, we repeat the simulation for all three stations considered in Section 2.4.1 and a wind penetration level ranging from 0-200%. The relative change in expected values of the resulting market value distributions are presented in Figure 2.9. For completeness, the expected average spot price level is also included. Compared to an expected average spot price of 56.70 Euro/MWh at 0% scaling, the level is reduced by 12% to 49.80 Euro/MWh for today's penetration level. Hence, provided that the rest of the system remains the same, the spot price level would be 7 Euro/MWh higher with no wind power penetration. In this case, resulting market values are above average spot price levels (due to higher wind power infeeds during wintertime when overall demand as well as prices tend to be also higher) and almost equal for any single wind turbine as spot prices are only marginally affected by wind power. Just as average spot price levels, expected market values decrease as the penetration level increases, however, at very different slopes. Whereas the average spot price itself is affected the least, the expected market value decreases corresponding to their dependence structure. They drop below average spot price levels at penetration levels as low as around 30% of today's capacities. A scaling of 100% corresponds to the current situation described in detail in Section 2.4.1. As can be observed, the difference between the average spot price and the market value further increases as the scaling factor approaches 200%, reaching levels of 8.34, 11.63, and 6.20

Euro/MWh for *Augsburg*, *Bremen* and *Kleiner Feldberg*, respectively.

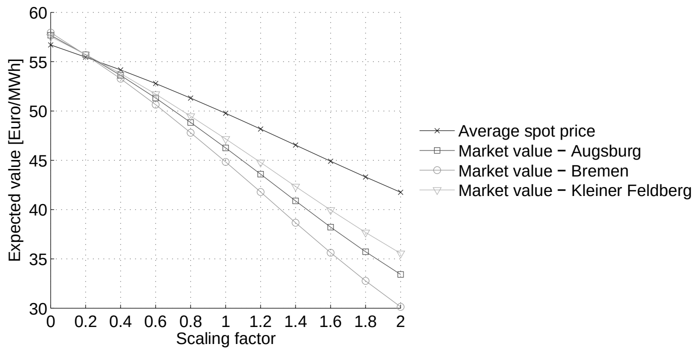


Figure 2.9: Relative change in expected values of the spot price and the market value under changing wind power penetration levels

2.5 Conclusions

The purpose of this paper has been to derive the value of wind power at different locations. In particular, the impact of the dependence structure of wind power on its value has been analyzed. This analysis becomes increasingly important as shares of wind power in electricity markets rise. We therefore developed a model for the simulation of single turbine wind power and electricity spot prices, including a precise representation of their interrelations. Copula theory has been applied to model single turbine wind power and aggregated wind power, thus allowing to decouple their dependence structure from their marginal distributions. The formation of prices has been formulated as a function of the aggregated wind power in a supply and demand based model. As such, the model extends formerly known modeling approaches through the ability to simulate and quantify the price effect of wind power, and hence to determine market values.

We find that the market value highly depends on the specific location and the corresponding dependence structure between the wind power of a single turbine at this location and the aggregated wind power. Whereas most locations are found to be characterized by rather adverse asymmetric dependence structures, some of the locations analyzed are identified as being related to the aggregated wind power such that their realizable selling prices are comparatively high. For the nineteen locations in Germany that we have analyzed in detail, we have shown that the expected market value is reduced by up to 8 Euro/MWh (i.e., 15%) compared to average spot price levels and varies by up to 6 Euro/MWh for the different locations.

Moreover, our results indicate that, in case of increasing wind power capacities, the adverse upper tail dependence structure of many locations has a negative impact on the market value, which makes market integration of wind power even more difficult. Nevertheless, integrating wind power into the market would allow market prices to reveal their key function by indicating the actual value of electricity and thus triggering investments in wind power projects characterized by high realizable spot prices. These projects would deploy balancing potentials much better and reduce the volatility in the electricity spot market as well as in the physical system.

Although a powerful tool to analyze the market value of wind power in a predefined setting, the model reveals its limitations in not being able to determine the dynamic reaction of the power system development in response to changing levels of wind power penetration. Further research could be done by extending the model in order to use it as a forecasting and derivative pricing tool. Moreover, the impact of spatial dependencies on short-term balancing capabilities of wind power could be analyzed (e.g., by extending the work presented by Girard et al. (2013)). This could become particularly relevant as balancing needs in a system depend on the aggregated wind power, while the balancing potential of a single turbine depends on its current output. By combining financial rewards from different markets, future analyses could address a comprehensive valuation and optimization of wind power projects. Besides the relevance for other markets, the methodology may also be applied to other forms of fluctuating renewable energies, such as solar power.

Acknowledgments

We would like to thank Oliver Grothe, Felix Höffler, Christian Growitsch and an anonymous referee for helpful comments and suggestions. Also, we acknowledge valuable discussions at the EURO-INFORMS 2013 as well as at the European IAEE Conference 2013.

2.6 Appendix

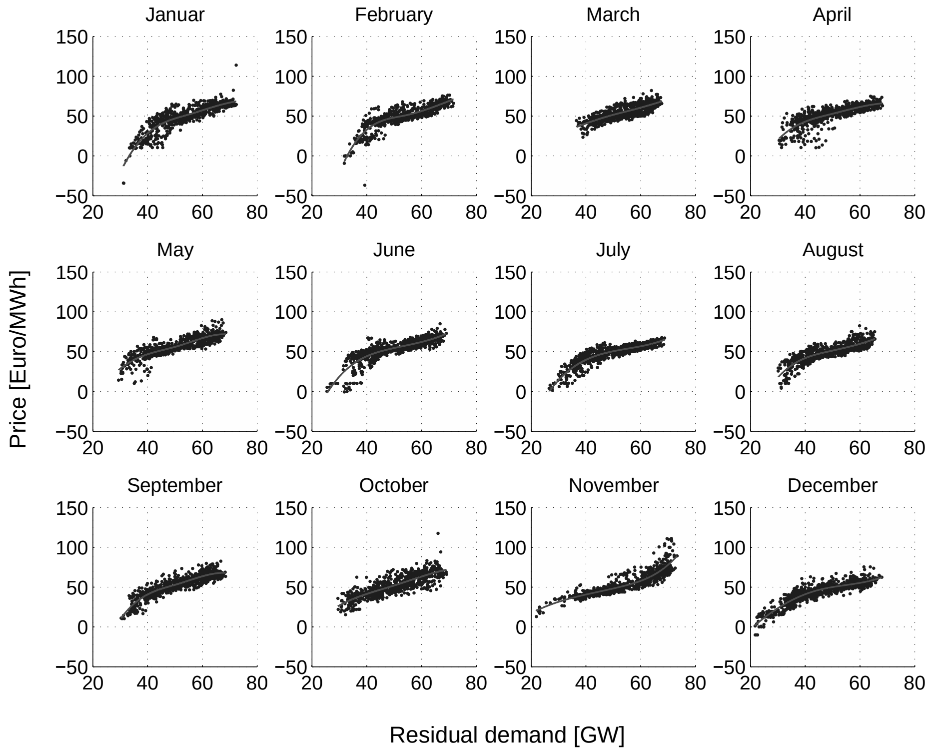


Figure 2.10: Demand-price dependence and spline fits for all months of 2011

Copula	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Clayton	-13.7	-49.4	-15.9	-20.5	-55.5	-103.7	-36.0	-10.8	-6.2	-73.0	-20.9	-112.0
Frank	-34.7	-99.9	-54.8	-107.4	-103.4	-168.5	-67.1	-80.8	-55.8	-145.9	-46.3	-198.2
Gumbel	-19.4	-145.4	-65.1	-172.3	-135.9	-194.0	-43.4	-133.4	-111.4	-163.3	-83.5	-146.6
Gaussian	-36.6	-109.1	-64.2	-117.5	-120.0	-182.5	-59.2	-95.0	-66.7	-156.5	-55.0	-174.0
T1	482.0	48.2	432.2	154.4	191.2	101.1	402.3	245.2	221.6	197.6	167.4	306.7
T2	158.2	-102.8	109.4	-54.8	-40.9	-112.3	92.9	1.9	3.2	-43.8	-26.2	-13.7
T3	77.8	-124.9	32.9	-95.2	-87.3	-155.7	21.0	-49.1	-39.2	-98.4	-58.0	-89.9
T4	43.5	-129.5	1.6	-108.7	-103.6	-170.8	-7.6	-67.8	-53.6	-119.9	-66.4	-120.5
T5	24.8	-129.9	-14.8	-114.5	-111.0	-177.6	-22.2	-76.8	-59.9	-130.9	-68.8	-136.2
T10	-8.4	-124.8	-42.7	-120.4	-120.0	-185.2	-45.3	-89.5	-67.2	-147.8	-67.2	-160.7
T20	-23.1	-118.5	-54.3	-120.3	-121.4	-185.4	-53.5	-93.2	-68.1	-153.5	-62.8	-169.0
T30	-27.7	-115.8	-57.8	-119.7	-121.3	-184.8	-55.7	-94.0	-67.9	-154.8	-60.6	-171.1
T40	-30.0	-114.3	-59.4	-119.3	-121.1	-184.4	-56.7	-94.3	-67.7	-155.4	-59.4	-172.0
T50	-31.3	-113.3	-60.4	-119.0	-120.9	-184.1	-57.3	-94.5	-67.5	-155.7	-58.6	-172.5

Table 2.5: Copula model selection based on AIC for the Station *Augsburg*

Copula	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Clayton	-370.7	-309.2	-263.6	-273.2	-309.4	-193.8	-292.3	-247.0	-224.9	-357.1	-390.3	-652.9
Frank	-511.6	-538.0	-502.5	-641.1	-530.0	-329.1	-685.1	-498.5	-596.2	-664.1	-640.0	-787.9
Gumbel	-539.1	-700.2	-610.7	-746.7	-604.6	-420.5	-692.0	-667.4	-702.0	-727.5	-788.9	-826.3
Gaussian	-554.9	-594.2	-533.6	-614.4	-560.9	-377.4	-637.5	-547.3	-571.0	-668.8	-698.1	-867.0
T1	-333.6	-480.8	-322.3	-402.2	-305.6	-143.3	-283.6	-368.5	-303.5	-435.9	-563.0	-708.9
T2	-477.1	-576.0	-469.5	-565.0	-480.3	-303.6	-500.5	-497.6	-488.1	-591.4	-666.7	-815.8
T3	-514.1	-594.2	-505.8	-602.7	-521.8	-339.2	-559.5	-526.3	-532.2	-631.2	-689.8	-846.6
T4	-530.1	-600.0	-520.4	-616.6	-538.3	-353.3	-585.6	-537.3	-549.6	-647.8	-698.2	-860.1
T5	-538.6	-602.2	-527.6	-622.7	-546.6	-360.5	-599.8	-542.5	-558.2	-656.3	-701.9	-867.1
T10	-551.9	-602.4	-536.8	-627.7	-558.8	-371.7	-623.6	-549.3	-570.0	-668.5	-705.1	-875.8
T20	-555.4	-599.7	-537.5	-624.5	-561.6	-375.4	-632.3	-549.9	-572.3	-670.9	-703.5	-875.1
T30	-555.8	-598.2	-536.8	-622.1	-561.9	-376.3	-634.5	-549.5	-572.3	-670.8	-702.2	-873.6
T40	-555.8	-597.4	-536.2	-620.5	-561.9	-376.7	-635.4	-549.2	-572.2	-670.6	-701.4	-872.4
T50	-555.8	-596.8	-535.8	-619.5	-561.8	-376.8	-636.0	-548.9	-572.0	-670.3	-700.9	-871.6

Table 2.6: Copula model selection based on AIC for the Station *Bremen*

Copula	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Clayton	-76.5	-142.7	-158.9	-142.0	-252.9	-235.2	-202.5	-79.8	-94.8	-141.6	-77.6	-230.5
Frank	-88.3	-189.8	-194.6	-231.9	-308.9	-187.9	-211.1	-110.1	-96.9	-166.3	-80.7	-269.4
Gumbel	-58.2	-164.3	-115.3	-187.4	-197.9	-89.1	-101.3	-60.1	-114.1	-142.5	-32.4	-285.1
Gaussian	-88.7	-199.8	-192.3	-227.0	-288.8	-188.8	-181.4	-100.4	-115.7	-192.4	-70.7	-318.5
T1	288.9	136.8	321.1	173.3	190.2	238.2	286.6	405.0	123.2	275.6	327.8	-1.8
T2	24.8	-79.2	6.0	-80.3	-106.3	-36.4	-20.4	78.8	-65.7	-15.5	56.7	-209.5
T3	-32.8	-131.8	-75.0	-143.6	-181.9	-103.6	-94.8	-0.5	-101.1	-88.0	-5.1	-257.5
T4	-54.6	-153.9	-110.2	-170.5	-214.5	-131.7	-125.2	-33.2	-112.6	-119.2	-29.4	-277.2
T5	-65.4	-165.8	-129.5	-184.9	-232.2	-146.5	-140.9	-50.5	-117.3	-136.3	-41.6	-287.6
T10	-81.2	-185.9	-163.9	-209.4	-263.5	-171.5	-166.2	-79.5	-121.2	-166.8	-60.5	-305.3
T20	-86.1	-193.8	-178.9	-219.2	-276.9	-181.3	-175.3	-91.0	-120.0	-180.2	-66.8	-312.6
T30	-87.3	-196.1	-183.6	-222.1	-281.1	-184.1	-177.7	-94.4	-118.9	-184.4	-68.4	-314.7
T40	-87.7	-197.1	-185.8	-223.4	-283.1	-185.4	-178.8	-96.0	-118.3	-186.5	-69.1	-315.7
T50	-88.0	-197.7	-187.2	-224.2	-284.3	-186.1	-179.4	-96.9	-117.8	-187.7	-69.5	-316.3

Table 2.7: Copula model selection based on AIC for the Station *Kleiner Feldberg*

Station	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Aachen	Frank	Gumbel	Gumbel	Frank	Gaussian	Gumbel	Frank	Frank	Gumbel	Gumbel	Gumbel	T30
	0.63	0.54	0.55	0.53	0.69	0.44	0.58	0.57	0.69	0.61	0.51	0.65
Angermünde	Gumbel	Gumbel	Gaussian	Gumbel	T10	Frank	Gumbel	Gumbel	Gumbel	Gaussian	Gumbel	Gaussian
	0.58	0.61	0.56	0.57	0.49	0.52	0.64	0.59	0.58	0.67	0.53	0.66
Augsburg	Gaussian	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Frank
	0.23	0.37	0.28	0.39	0.37	0.46	0.30	0.35	0.30	0.44	0.26	0.50
Bremen	T40	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	T10
	0.71	0.75	0.70	0.79	0.73	0.61	0.80	0.71	0.78	0.78	0.78	0.81
Dresden	Frank	Gumbel	Frank	Gumbel	T10	Gumbel	Frank	Gumbel	Gumbel	Gaussian	Frank	Clayton
	0.68	0.64	0.59	0.63	0.50	0.58	0.72	0.67	0.61	0.61	0.55	0.60
Emden	Gaussian	Gumbel	Gaussian	Gumbel	Gaussian	Gumbel	Frank	Gumbel	Gumbel	Gaussian	Gumbel	Gaussian
	0.62	0.50	0.68	0.67	0.70	0.49	0.76	0.61	0.73	0.72	0.71	0.68
Erfurt-Weimar	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	Gumbel	T50
	0.74	0.67	0.61	0.70	0.64	0.61	0.78	0.70	0.81	0.74	0.55	0.75
Idar-Oberstein	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Frank
	0.42	0.55	0.49	0.37	0.46	0.35	0.39	0.47	0.55	0.57	0.40	0.68
Kahler Asten	Frank	Gaussian	Gumbel	Frank	Frank	T50	Frank	Frank	Frank	T20	Gumbel	Frank
	0.72	0.69	0.65	0.82	0.76	0.71	0.80	0.74	0.76	0.80	0.72	0.74
Kleiner Feldberg	Gaussian	Gaussian	Frank	Frank	Frank	Clayton	Frank	Frank	T10	Gaussian	Frank	Gaussian
	0.34	0.51	0.50	0.53	0.61	0.49	0.52	0.39	0.35	0.46	0.33	0.56
Konstanz	Frank	Gumbel	Gaussian	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	T30	Frank
	-0.06	0.22	0.19	0.21	0.24	0.39	0.25	0.27	0.41	0.35	0.14	0.37
Leipzig-Halle	Frank	T40	Gumbel	Gumbel	T40	Gumbel	Frank	Gumbel	Gumbel	Gaussian	Gumbel	T5
	0.80	0.78	0.71	0.75	0.70	0.69	0.83	0.77	0.79	0.80	0.70	0.70
Magdeburg	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gaussian
	0.72	0.72	0.63	0.66	0.57	0.65	0.76	0.71	0.67	0.70	0.65	0.76
Münster-Osnabrück	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	Gumbel	T40
	0.67	0.63	0.56	0.66	0.62	0.53	0.73	0.59	0.74	0.74	0.66	0.80
Oldenburg	Gaussian	Gumbel	Gumbel	Gumbel	Gumbel	Gaussian	Gumbel	Gumbel	Gumbel	Gumbel	Gumbel	Gaussian
	0.63	0.56	0.64	0.69	0.65	0.48	0.75	0.56	0.69	0.69	0.75	0.71
Potsdam	Frank	Gumbel	Gaussian	Gumbel	T30	T20	Frank	Frank	Frank	Frank	Gaussian	Gaussian
	0.83	0.76	0.78	0.82	0.72	0.76	0.85	0.81	0.82	0.84	0.78	0.69
Rostock	Gumbel	Gumbel	Clayton	Gumbel	Gaussian	Gaussian	Gaussian	T40	Frank	Frank	Gumbel	Clayton
	0.30	0.52	0.36	0.61	0.56	0.39	0.55	0.43	0.57	0.66	0.70	0.47
Saarbrücken	Gumbel	Gumbel	Gaussian	Gumbel	Gumbel	Gumbel	Frank	Gumbel	Gumbel	Gumbel	Gumbel	Frank
	0.27	0.42	0.48	0.06	0.47	0.39	0.34	0.35	0.47	0.56	0.28	0.64
Schleswig	Frank	Gumbel	Gumbel	Gumbel	Gaussian	Frank	Frank	Gumbel	Gumbel	Gaussian	Gumbel	Gumbel
	0.66	0.60	0.58	0.73	0.63	0.53	0.66	0.58	0.72	0.72	0.71	0.54

Table 2.8: Selected copula models and rank correlation coefficients

3 Supply chain reliability and the role of individual suppliers

We study a one-period supply chain problem consisting of numerous suppliers delivering a homogenous good. Individual supply is uncertain and may exhibit dependencies with other suppliers as well as with the stochastic demand. Assuming that reliability of supply represents an economic value for the customer that shall be paid accordingly, we first derive an analytical solution for the contribution of an individual supplier to supply chain reliability. Second, applying concepts from cooperative game-theory, we propose a payoff scheme based on marginal contributions that explicitly accounts for the statistical properties of the problem. A number of desirable properties is thus achieved, including static efficiency as well as efficient investment incentives. Lastly, in order to demonstrate the relevance and applicability of the concepts developed, we consider the example of payoffs for reliability in power systems that are increasingly penetrated by interdependent variable renewable energies. We investigate empirical data on wind power in Germany, thereby confirming our analytical findings. In practice, our approach could be applied to design and organize supply chains and their reliability more efficiently. For instance, in the field of power systems, the approach could improve designs of capacity or renewable support mechanisms.

3.1 Introduction

While most supply chains face the requirement to provide high levels of reliability, the consequences of supply shortages in power systems can be particularly dramatic. For instance, in the early 2000s insufficient supply capacities caused a series of blackouts in the Californian power system affecting several hundred thousand customers. The State of California was forced to initiate short-term countermeasures to alleviate the crisis, amounting to an estimated 40 bn.$ in additional energy costs from 2001 to 2003 (Weare (2003)). The economy was estimated to slow down by 0.7-1.5%, entailing an increase in unemployment by 1.1% (Cambridge Energy Research Associates (2001)). But also other supply chains, such as manufacturing or food production industries, may suffer severe economic losses from supply shortages.

To manage and foster supply chain reliability, sourcing from multiple suppliers is an effective and often applied practice.[1] While the principle idea of risk diversification is rather simple, organizing and coordinating multiple individual suppliers to ensure envisaged levels of supply chain reliability at reasonable costs is often challenging. An economic approach to tackle this challenge consists in defining reliability as a good that can be provided by individual suppliers. The provision is then paid according to the individual supplier's contribution to supply chain reliability, with a price determined in a competitive auction, for instance. In principle, such an approach ensures and incentivizes an efficient short- and long-term management of supply chain reliability.

An example for such explicit reliability-related payments can be found in so-called capacity mechanisms, nowadays established in many power markets.[2] Key ingredient to all such mechanisms is the determination of an individual supplier's contribution to supply chain reliability, commonly known as prequalification, which is used as a basis for subsequent payments. Despite the apparent relevance of this measure, it appears that existing approaches for its determination lack generality and consistency.[3] The reason may lie in the difficulty to assess an individual supplier's contribution, especially when accounting for the full complexity of the problem which is inherently stochastic. Indeed, a contribution not only depends on the stochastic nature of the individual supplier itself, but also on all other stochastic suppliers that are present.

It is hence the goal of this paper to comprehensively investigate supply chain reliability and the role of individual suppliers therein. For this purpose, we consider a one-period supply chain problem consisting of numerous suppliers delivering a homogenous good.[4] Individual supply is uncertain and may exhibit dependencies with

[1]For a general overview of mitigation strategies, see, e.g., Tang (2006), Tomlin (2006) or Snyder et al. (2014).

[2]In fact, the Californian electricity crisis has triggered much of the worldwide debate and development around capacity mechanisms (for a good overview, see, e.g., Joskow (2008a) or Cramton et al. (2013)). Recently, the discussion about the need and design of capacity mechanisms has been regaining momentum due to the large-scale deployment of variable renewable energies (such as wind or solar power), whose impact on power system reliability is considered a *crucial issue of common interest* (Council of European Energy Regulators (2014)).

[3]For instance, there seems to be no best practice on how to deal with the reliability of intermittent and interdependent renewable resources in power systems, reflected in a variety of different existing approaches. Moreover, all of them incorporate inefficient design features, as we will show in the course of this paper.

[4]The need to source a homogenous good from multiple suppliers may stem, e.g., from capacity constraints or the requirement for risk diversification (e.g., Minner (2003) or Tang (2006)).

other suppliers as well as with the stochastic demand. Due to the stochastic nature of the problem, demand can be served at an intended reliability level, possibly including supply disruptions.[5] Assuming that reliability of supply represents an economic value for the customer and shall be paid accordingly, we are particularly interested in answering the following two questions: First, what is the contribution of individual suppliers to supply chain reliability? And second, how should individual suppliers be paid for what they contribute? In addition, we shall demonstrate practical relevance and applicability of our analytical findings in a novel and pertinent empirical case study based on wind energy in the German power system.

By solving these problems, our paper contributes a general, consistent and applicative framework to address the role of individual suppliers for supply chain reliability. In practice, it could be employed to design and manage supply chain reliability more efficiently. It may proof to be a useful tool for allocation problems in monopsonistic markets or public procurements, for designing bonus payments, for accounting issues in integrated supply chains, or for the assessment of existing or new contract conditions. In the field of power systems, the approach could improve the design of capacity mechanisms which so far appears to be incomplete.

The rest of the paper is structured as follows: Section 3.2 reviews the related literature. Section 3.3 introduces and solves for the contribution of individual suppliers to supply chain reliability. In Section 3.4, we characterize our problem as a cooperative game to develop and investigate a suitable payoff scheme. The empirical case study is presented in Section 3.5. Section 3.6 concludes.

3.2 Related literature

Our paper is closely related to the literature dealing with supply chain reliability and the problem of strategic sourcing of a homogenous good from multiple suppliers, facing either supply disruptions (i.e., a binomial distribution of uncertainty) or capacity uncertainty (i.e., an uncertain upper bound on the actual quantity supplied).[6]

[5]Note that instead of *reliability*, one could also consider the generalized case of supply *quality*. However, for the sake of clarity, we will stick to the term reliability throughout the paper. Nevertheless, the concepts and results derived could also be applied to other dimensions of supply quality, such as time-to-respond, etc.

[6]Note that the former is an extreme case of the latter. Also note that both are different from *yield uncertainty* which incorporates a dependency of the uncertainty on the order quantity. As we consider yield uncertainty as a different problem class, we do not review the related literature here. For a broader review of the literature dealing with supply chain risks, the reader is referred to Tang (2006) and Snyder et al. (2014).

Supply disruptions in a supply chain consisting of multiple suppliers have been studied by Parlar and Perry (1996), Gürler and Parlar (1997), Li et al. (2004) and Tomlin (2006), however, without considering stochastic dependencies among the suppliers. As a natural extension, later papers allow for dependencies between binomial supply disruptions (i.e., Babich et al. (2007), Wagner et al. (2009) and Li et al. (2010)).

Continuous distributions of the suppliers' uncertainties – as used in our paper – were presented by Dada et al. (2007) and Masih-Tehrani et al. (2011). The former paper proves that when selecting among a set of possible suppliers that are different in reliability and costs, cost generally takes precedence over reliability. While this result is derived under the assumption of supply distributions being independent, Masih-Tehrani et al. (2011) captures capacity uncertainty including multivariate dependencies, finding that the buyer's best strategy is risk diversification by choosing suppliers with independent distributions in order to avoid simultaneous supply disruptions.

In contrast to the above literature on supply chain reliability, our paper deviates in several important aspects. First and foremost, instead of analyzing the *costs* of supply chain reliability under *exogenous prices*,[7] we take a different perspective on the problem by *endogenously* determining the individual supplier's *value* (i.e., contribution and corresponding payoff) for supply chain reliability.[8] Note that this is a fundamentally different view on the problem that becomes relevant, e.g., for managing supply chains in public procurements or bonus payments where prices are not fixed ex-ante.

In addition, and in contrast to most papers mentioned, we consider arbitrary and interdependent distributions of (un)availability and demand. Moreover, we study implications for the supply chain organization as well as investment incentives that are, to the best of our knowledge, hardly considered in the existing literature.

Our paper is also related to the more specific field of supply chain reliability in power systems, which has been investigated either from a technical or economic perspective. From a technical perspective, the goal has been to develop methodologies to assess the technical ability to provide reliability of supply (e.g., Garver (1966), Billinton (1970) or Billinton and Allan (1996)). The role of individual units

[7]In the reviewed literature, suppliers' prices are either given as a parameter, or result from some supplier interaction (such as a Cournot game) without the buyer being able to have an influence.

[8]Technically, the difference stems from the fact that we determine the individual supplier's contribution based on an endogenous demand adjustment, while the literature's objective is to serve an exogenous (though, often stochastic) demand level.

for supply reliability, often referred to as capacity credit or capacity value, has also been discussed (for recent surveys, see, e.g., Amelin (2009) or Keane et al. (2011)). However, it appears that these analyses remain very technical and have never been transferred to a broader and generalized supply chain context. Moreover, they almost exclusively deal with the problem numerically rather than in a consistent analytical framework. In contrast, our paper contributes a comprehensive and consistent analytical framework along with generalized implications for managing supply chain reliability.

Economically, power system reliability has been studied with respect to the (in)ability and potential failures of power markets to provide reliability as a market outcome (e.g., Joskow (2008a) or Cramton et al. (2013)). However, even though various designs of capacity-related payoffs have been proposed, the role and implications of stochastic and interdependent suppliers have so far been disregarded. Our paper fills this gap by suggesting suitable approaches to incorporate those suppliers into reliability-related mechanisms in order to ensure economically efficient outcomes.

3.3 Supply chain reliability and the contribution of individual suppliers

3.3.1 Supply chain reliability

We consider a one-period supply chain $\mathscr{S}$ consisting of numerous suppliers delivering a homogenous good. Suppliers are characterized by their joint stochastic availability of supply capacity C.[9] Demand D is also assumed stochastic. Due to considering only one period without additional backup (such as, e.g., inventory storage, emergency service, etc.), supply shortages occur whenever C is unable to cover D. Consequently, we define supply chain reliability in probabilistic terms as follows:[10]

[9] Here and in the following – unless indicated differently – capital letters are used for random variables.

[10] Note that we implicitly assume an inelastic demand with no reaction as capacity becomes scarce. If the good is marketed, this implies that market clearing cannot be guaranteed, e.g., because of the lack of real time pricing. Consequently, there is a risk of situations with all available capacities producing, but still being unable to fully serve demand – irrespective of the price level.

Definition 1. ***Supply chain reliability:*** *Probability of a supply chain $\mathscr{S}$, characterized by the stochastic overall availability of supply capacity C, to be able to cover a stochastic demand D, i.e.:*

$$R^{\mathscr{S}} = \Pr(D \leq C) = \Pr(D - C \leq 0) = \Pr(X \leq 0) = F_X(0). \tag{3.1}$$

In Equation (3.1), we have used F_X for the cumulative distribution function (cdf) of the capacity shortage $X = D - C$. Note that stochastic dependencies between the random variables are so far unspecified. For instance, the overall availability of supply capacity C may result from multiple interdependent supply capacities of numerous individual suppliers. Furthermore, C may exhibit dependencies with the stochastic demand D.

It is worthwhile to mention structural correspondence of our Definition 1 with another well-known risk measure, i.e., Value-at-Risk (VaR), defined as $\Pr(X \leq VaR) = R^{\mathscr{S}}$ (e.g., Jorion (2007)). The measures coincide for VaR being normalized to zero, i.e., when $R^{\mathscr{S}}$ corresponds to the probability that X is lower than zero.

3.3.2 The contribution of individual suppliers

We now investigate the contribution of an individual supplier with random production capacity $Y < C$ to the reliability of supply chain $\mathscr{S}$. To this end, we remove it from the existing system $\mathscr{S}$, and subsequently determine the reliability of supply of the new diminished supply chain $\mathscr{T}$ as:

$$R^{\mathscr{T}} = \Pr(D \leq C - Y) \tag{3.2}$$

Note that as Y is positive, $R^{\mathscr{T}} \leq R^{\mathscr{S}}$ always holds. With the goal to capture the contribution of Y to supply chain reliability, we follow the concept of incremental VaR. I.e., we capture the change in risk exposure induced by the adjustment of X by Y while requiring the corresponding probability to remain at the original level (e.g., Tasche and Tibiletti (2003) or citeJorion2007). In other words, we measure the demand reduction necessary to bring $R^{\mathscr{T}}$ back to the original reliability level $R^{\mathscr{S}}$.[11,12] Analytically, we define the contribution of an individual supplier as follows:

[11] Noticeably, this approach has for long been used in the context of power systems under the name of effective load carrying capability, which was originally developed by Garver (1966).

[12] As an alternative measure, one could also consider the change in reliability induced by removing supply capacity Y, i.e., $\Delta R = R^{\mathscr{S}} - R^{\mathscr{T}}$, however, without altering the principle results derived hereafter.

Definition 2. ***Contribution of an individual supplier to supply chain reliability:*** *Level of demand v by which D needs to be reduced in order to maintain the original level of reliability, i.e.:*

$$\Pr(D - v \leq C - Y) = R^{\mathscr{S}} \tag{3.3}$$

Due to its complexity, Equation (3.3) has commonly been solved for v numerically by means of iteration (see, e.g., Wang (2002) or Keane et al. (2011)). Advantages of the numerical solution include its straightforward implementability as well as the implicit coverage of statistical dependencies when using concurrent observations of demand and supply. However, numerical solutions – even if conducted for a wide range of parameter constellations and application examples – do not allow for generalizations of the results obtained. Moreover, the numerical solution often entails a high computational burden.

In contrast, analytical solutions allow to readily calculate the desired results and provide further general insights. Nevertheless, only very few authors have engaged with the analytical analysis of Equation (3.3). In the literature related to portfolio risks, the incremental VaR is conveniently solved analytically up to a first order approximation (e.g., Tasche and Tibiletti (2003) or citeJorion2007). In contrast, Dragoon and Dvortsov (2006) propose the z-method, considering higher order terms and the special case of a normally distributed capacity shortage over demand and independence with the individual supplier (i.e., $X \sim \mathscr{N}$ and $X \perp Y$). Extending this approach, Zachary and Dent (2011) present a closed-form solution with an arbitrary distribution of X, but for independent X and Y. Even though the latter paper discusses the natural extension to the case of dependent distributions, the formal proof is not included. Hence, in the following proposition we present the generalized solution of Equation (3.3) for v, with arbitrary dependence between the distributions of X and Y, and $\sigma_{(\cdot)}$ being their standard deviation.

Proposition 1. *For $\sigma_Y \ll \sigma_X$, the contribution v of an individual capacity Y to supply chain reliability is approximated by*

$$v = \mu_{Y|X\approx 0} - \frac{\sigma^2_{Y|X\approx 0}}{2} \frac{f'_X(0)}{f_X(0)}. \tag{3.4}$$

Proof. From Equations (3.3) and (3.1) it follows that the equation to be solved for v is

$$\Pr(D - v \leq C - Y) = \Pr(X + Y \leq v) = \Pr(X \leq 0) = F_X(0). \tag{3.5}$$

The cumulative distribution function $\Pr(X + Y \leq v)$ can be expressed as integrals over the joint probability density function $f_{X,Y}$, with X, Y being arbitrary dependent distributions. Reformulating using conditional distributions and integrating over x, we obtain

$$\Pr(X+Y \leq v) = \int_{-\infty}^{\infty} \int_{-\infty}^{x=v-y} f_{X,Y}(x,y) dx dy \tag{3.6}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x=v-y} f_{Y|X}(y|x) f_X(x) dx dy \tag{3.7}$$

$$= \int_{-\infty}^{\infty} f_{Y|X \leq v-y}(y|X \leq v-y) F_X(v-y) dy = F_X(0). \tag{3.8}$$

Instead of continuing with the explicit form of the cumulative distribution function F_X, we approximate it via Taylor expansion around the critical point $v-y=0$ up to the second order polynomial degree, i.e.,

$$F_X(v-y) \approx F_X(0) + f_X(0)(v-y) + \frac{f_X'(0)}{2}(v-y)^2. \tag{3.9}$$

Noticeably, in the above equation we have induced and accepted an approximation error of $o((v-y)^2)$, which occurs if derivatives of order two or higher are non-zero.

Next, note that if $\sigma_Y \ll \sigma_X$, it follows that $\Pr(X+Y \leq v) \approx \Pr(X+\mu_Y \leq v)$ and hence, that $v \approx \mu_Y$ in Equation (3.5). We now insert (3.9) in (3.8), and reformulate using the concept of conditional expectations:

$$\int_{-\infty}^{\infty} f_{Y|X \leq v-y}(y|X \leq v-y) \left[F_X(0) + f_X(0)(v-y) + \frac{f_X'(0)}{2}(v-y)^2 \right] dy \tag{3.10}$$

$$= F_X(0) + f_X(0)\mathbb{E}[(v-Y)|X \approx 0] + \frac{f_X'(0)}{2}\mathbb{E}\left[(v-Y)^2|X \approx 0\right] \approx F_X(0), \tag{3.11}$$

which can be simplified using standard deviation σ as well as expected values μ to

$$f_X(0)(v-\mu_{Y|X\approx 0}) + \frac{f_X'(0)}{2}\left((v-\mu_{Y|X\approx 0})^2 + \sigma^2_{Y|X\approx 0}\right) \approx 0. \tag{3.12}$$

Equation (3.12) represents a quadratic equation in $(v-\mu_{Y|X\approx 0})$ that can readily be solved for v based on the assumption of small σ_Y (so that the error of order $\mathcal{O}(\sigma^4_{Y|X\approx 0})$ is small), such that Equation (3.4) follows. □

From Equation (3.4), we observe that two terms including two different statistical features of the individual supplier are decisive for its contribution to supply chain reliability: its average availability μ on the one hand, and its standard deviation σ on the other, both at times of scarce capacity, i.e. for $X \approx 0$. Specifically, higher average availability at times of critical capacity directly contributes to reliability. In contrast, the effect of the standard deviation getting larger may be either positive or negative for reliability, depending on the sign of $f_X'(0)$ (i.e., the convexity of the cdf).[13,14] It should typically hold true that high levels of reliability are required from supply chains, such that the critical point $X = 0$ is located at the left hand side of the distribution where the probability for insufficient capacity is still low and increasing in x, which yields $f'(x) > 0$. Then, the individual supplier's contribution to reliability decreases in its standard deviation. Somewhat counterintuitive, however, the individual supplier's contribution to reliability may also *benefit* from a high standard deviation. This is the case for $f'(x) < 0$, resulting in a situation where positive deviations weight more than negative ones. This could hold true and represent an interesting feature, e.g., for supply chains that are still at an early stage of development while already facing high levels of demand, or for well-established systems that underwent a sudden and substantial increase in demand.

Two further points are worth mentioning. First, the contribution of an individual supplier to supply chain reliability may – instead of absolute numbers – be reported relative to its maximum available capacity $\overline{y} = \max Y$, i.e., as a fraction $\tilde{v} = v/\overline{y}$, with $0 \leq \tilde{v} \leq 1$. Second, Equation (3.4) can readily be extended to multiple (n) suppliers contributing jointly to supply chain reliability. In this case, define Y as the sum of the individual unit's generation, i.e., $Y = \sum_{i=1}^{n} Y_i$. Denoting with N the set of all n suppliers in the system contributing jointly to generation adequacy, we will write $v(N)$ for their joint contribution. To determine the joint contribution of a few units only, we will use the set $S \subseteq N$, and denote the corresponding contribution by $v(S)$. The joint contribution of multiple units (or, in other words, a *coalition* of units) will become important for large parts of the subsequent analysis.

[13]Note that the density function $f(x)$ is always positive by definition.

[14]The effect of the standard deviation essentially stems from the difference in impact from positive and negative deviations from the average availability. If positive and negative deviations had an equal impact and would outweigh each other, the *expected* overall contribution would not change. However, the impact of the standard deviation on the reliability contribution increases with the absolute level of $f'(x) = F''(x)$, i.e., with the level of convexity of the cumulative distribution function $F(x)$. It also increases with $f(x) = F'(x)$ (i.e., the slope of the cdf) decreasing, as the difference in impact from positive and negative deviations then becomes more important.

3.3.3 Statistical properties of the contribution

In the following, three corollaries shall be discussed that describe essential properties of the individual supplier's contribution v. For the formal proofs, we build on the explicit formulation for the supplier's contribution to supply chain reliability as presented in Proposition 1. Even though derived statistically, we will also discuss first economic implications the identified properties may have.

Corollary 1. *The contribution of an individual supplier is generally subject to changing returns to scale.*

Proof. Inserting a scaled random production capacity aY in Equation (3.4), and using the fact that both, mean and standard deviation scale directly with the scaling factor of the random variable, it follows that

$$v(aS) = \mu_{aY|X\approx 0} - \frac{\sigma^2_{aY|X\approx 0}}{2} \frac{f'_X(0)}{f_X(0)} \tag{3.13}$$

$$= a\mu_{Y|X\approx 0} - \frac{a^2\sigma^2_{Y|X\approx 0}}{2} \frac{f'_X(0)}{f_X(0)}. \tag{3.14}$$

Whereas the first term on the right hand side increases linearly with a, the second term decreases with higher order a^2 as long as $\sigma_{Y|X\approx 0} > 0$ and $f'(x) > 0$. Under these conditions, it holds that $v(aS) < av(S)$, i.e. decreasing returns to scale result when considering an increasing amount of capacity with equal availability Y. In contrast, $\sigma_{Y|X\approx 0} > 0$ and $f'(x) < 0$ yields increasing returns to scale. □

Economically, Corollary 1 may be of particular relevance for a supply chain if it shall increasingly rely on one particular supplier. For illustration, imagine a power system that aims at replacing an increasing number of fossil power plants with wind power capacities while keeping its original reliability level.[15] Corollary 1 implies that with each additional unit of wind power installed, a decreasing amount of fossil power can be safely removed from the system.

Corollary 2. *Gains of diversification may apply for the contribution of individual suppliers.*

Proof. We assess the contribution of some capacity Y_1 to supply chain reliability. In order to analyze possible gains of diversification – without loss of generality – we

[15] Suppose a reliable power system with $f'_X(0) > 0$, and variable wind power resources with $\sigma_{Y|X\approx 0} > 0$.

assume that some share $\alpha \in [0,1]$ of capacity Y_1 can be sourced from an alternative supplier S_2 instead of S_1. Depending on the choice of α, the joint production capacity becomes $Y_{12} = (1-\alpha)Y_1 + \alpha Y_2$. The joint contribution of this portfolio to supply chain reliability writes as

$$\begin{aligned} v((1-\alpha)S_1 \cup \alpha S_2) = & \mu_{Y_{12}|X} - \frac{\sigma^2_{Y_{12}|X}}{2} \frac{f'_X(0)}{f_X(0)} && (3.15) \\ = & (1-\alpha)\mu_{Y_1|X\approx 0} + \alpha \mu_{Y_2|X\approx 0} \\ & - \left((1-\alpha)^2 \sigma^2_{Y_1|X\approx 0} + \alpha^2 \sigma^2_{Y_2|X\approx 0} + 2\alpha(1-\alpha)\sigma_{Y_1,Y_2|X\approx 0}\right) \frac{f'_X(0)}{2 f_X(0)}. && (3.16) \end{aligned}$$

To identify gains of diversification, we need to check the derivative of v with respect to α in the region of $\alpha = 0$, i.e.,

$$\begin{aligned} \frac{\partial}{\partial \alpha} v((1-\alpha)S_1 \cup \alpha S_2)\Big|_{\alpha=0} = & - \mu_{Y_1|X\approx 0} + \mu_{Y_2|X\approx 0} + (\,(1-\alpha)\sigma^2_{Y_1|X\approx 0} \\ & - \alpha \sigma^2_{Y_2|X\approx 0} - (1-2\alpha)\sigma_{Y_1,Y_2|X\approx 0}\,) \frac{f'_X(0)}{f_X(0)}\Big|_{\alpha=0} && (3.17) \\ = & - \mu_{Y_1|X\approx 0} + \mu_{Y_2|X\approx 0} + \left(\sigma^2_{Y_1|X\approx 0} - \sigma_{Y_1,Y_2|X\approx 0}\right) \frac{f'_X(0)}{f_X(0)}. && (3.18) \end{aligned}$$

From the derivative evaluated at $\alpha = 0$, we observe that it is positive (and hence, $v((1-\alpha)S_1 \cup \alpha S_2)$ increasing) as long as the average availability of supplier S_2 is similar to that of supplier S_1, and their covariance is either *not* particularly strong in case of $f'_X(0) > 0$ or particularly strong otherwise. Under those conditions, the joint contribution is subject to gains of diversification. □

The meaning of gains from diversification, as identified in Corollary 2, is quite intuitive, and essentially follows from balancing effects that may result from statistical aggregation. Hence, a diversified portfolio with the good sourced from multiple suppliers is often better able to reliably supply a certain load level. This holds true as long as the expected yield is similar and they are not highly correlated (for the typically more relevant case of $f_X'(0) > 0$).[16]

Corollary 3. *The contributions of individual suppliers are generally non-additive.*

Proof. Similar to Corollary 2, without loss of generality, let us consider a supply chain comprising two suppliers S_1 and S_2 with (possibly dependent) production capacities Y_1 and Y_2, respectively. The joint contribution of S_1 and S_2 to supply chain reliability writes as

$$v(S_1 \cup S_2) = \mu_{Y_{12}|X} - \frac{\sigma^2_{Y_{12}|X}}{2} \frac{f_X'(0)}{f_X(0)} \tag{3.19}$$

$$= \mu_{Y_1|X\approx 0} + \mu_{Y_2|X\approx 0} - \left(\sigma^2_{Y_1|X\approx 0} + \sigma^2_{Y_2|X\approx 0} + 2\sigma_{Y_1,Y_2|X\approx 0}\right) \frac{f_X'(0)}{2f_X(0)}. \tag{3.20}$$

In contrast, if these suppliers were to be assessed independently, we can simply apply Equation (3.4) to each of them to get

$$v(S_1) = \mu_{Y_1|X\approx 0} - \frac{\sigma^2_{Y_1|X\approx 0}}{2} \frac{f_X'(0)}{f_X(0)} \tag{3.21}$$

$$v(S_2) = \mu_{Y_2|X\approx 0} - \frac{\sigma^2_{Y_2|X\approx 0}}{2} \frac{f_X'(0)}{f_X(0)}. \tag{3.22}$$

From the above equations, we see that as long as $\sigma_{Y_1,Y_2|X\approx 0} \neq 0$, it follows that $v(S_1 \cup S_2) \neq v(S_1) + v(S_2)$. Hence, any type of (positive or negative) dependence changes the joint contribution compared to the sum of independent contributions. Specifically, the sum of the independent assessments is then inconsistent with the joint contribution. Generalizing to all n suppliers in the system, $\sum_{i=1}^{n} v(i) \neq v(N)$ as long as $\sigma_{Y_j,Y_k|X\approx 0} \neq 0$ for any pair j, k. □

From Corollary 3, it follows that a comprehensive and coordinated approach is

[16]Note that in finance, the concept of risk spreading is well-known from Markowitz' portfolio theory that shows similar characteristics when valuing properties of multi-asset portfolios (assets are here the equivalent to our suppliers). However, while the interest in a Markowitz portfolio lies in finding an efficient *tradeoff* between risk and expected returns, we are here interested in reliability contributions which are, as found in Equation (3.4), dependent on *both* variance *and* expectation (together with additional characteristics captured by f_X which are not occurring in Markowitz). Also note that one of the main criticisms with respect to Markowitz' portfolio theory, namely the linear dependence assumption between the joint distributions, also applies to our case here. In fact, non-linear dependencies can indeed be relevant for supply chain risks, as shown, e.g., by Wagner et al. (2009), Masih-Tehrani et al. (2011) or Elberg and Hagspiel (2015). While this would be an interesting extension of our analysis, it would go beyond the scope of this paper and is left for future research.

imperative to assess supply chain reliability in order to reach a consistent representation of all dependent suppliers. In other words, the *joint* contribution of all suppliers is essentially the only relevant one for supply chain reliability. However, with the goal to provide explicit reliability-related payments in order to manage and incentivize supply chain reliability, allocations need to be made to *individual* suppliers according to their *individual value*. Thus, Corollary 3 represents a particular challenge that will be tackled by a suitable payoff scheme in the next Section.

3.4 Payoff scheme

To approach the problem of designing a suitable payoff scheme, we apply concepts from cooperative game-theory. Specifically, we characterize the (joint) value of one or multiple suppliers for supply chain reliability as the output of a coalitional game (N, v) with transferable utility, where N is a finite set of units in the system, and v a characteristic function. $v(\cdot)$ measures the (joint) contribution of a nonempty coalition of suppliers $S \subseteq N$ as defined implicitly in Equation (3.3), or explicitly in Equation (3.4). Note that $v(S) \in \mathbb{R}^s$, i.e., for every coalition S, a corresponding contribution can be determined. A solution concept for this coalitional game is a payoff vector $\Phi \in \mathbb{R}^N$ allocating the joint value $v(N)$ to the coalition members.

The subsequent analysis consists of three steps: First, we assess the properties of the coalitional game (N, v) and implications regarding the organizational design of the supply chain. Second, we develop our solution concept, i.e., how payoffs are allocated to individual suppliers. Lastly, we investigate investment incentives resulting from our allocation rule.

3.4.1 Supply chain organization

Here and in the following, let us suppose it is the goal for the supply chain to reach or sustain some envisaged level of reliability $R^{\mathscr{S}}$. In an appealing approach to organize the supply chain, the buyer would determine his/her envisaged level of reliability, and invite tenders among the suppliers to reach it. In other words, (s)he would implement a competitive procurement of reliability contributions. For instance, in the context of a capacity procurement auction in power systems, the state, regulator, or some other independent authority could fix the envisaged reliability of power supply, and implement a platform that provides suppliers with the possibility to bid their contribution – either individually or jointly by forming coalitions.[17] Hence, regulatory intervention by the authority would be limited, and the need to administratively determine individual allocations be avoided. Individual bids would be relatively easy to verify, e.g., by means of random checks and sufficiently high penalties, hence ensuring truthful individual bidding.[18] However, this approach could result in unsatisfactory outcomes, as stated and proven in the following Proposition.

Proposition 2. *The equilibrium in a competitive reliability procurement is inefficient if suppliers are positively related and the buyer only verifies individual bids.*

Proof. Consider a supply chain organization that lets suppliers bid their individual contribution in a competitive procurement of reliability that features a multiunit auction. The auction is cleared by accepting all bids that are necessary to achieve the requested reliability level. To assess the quality of the auction outcome, we need to identify the suppliers' optimal bidding strategy.

To this end, without loss of generality, consider the case of two suppliers S_1 and S_2 for which the joint contribution $v(S_1 \cup S_2)$ has been derived in Equation (3.20) – in contrast to their individual contributions $v(S_1)$ and $v(S_2)$ (Equations (3.21) and (3.22)). As we see, a positive dependency (i.e., covariance $\sigma_{Y_1,Y_2|X\approx 0} > 0$) implies sub-additivity of the contributions (i.e., $v(S_1 \cup S_2) < v(S_1) + v(S_2)$) and hence, disincentives to cooperate. Technically, sub-additivity implies an empty core of our game (N, v), i.e., in every situation there is an alternative coalition able to improve the payoffs of all its members. As a consequence, the optimal bidding strategy consists of individual or coalitions of suppliers with positive dependencies bidding by themselves.

While the buyer is able to verify these individual bids (and hence, to avoid overstatements by imposing sufficiently high penalties), their sum does not correspond to the true joint contribution. In fact, the buyer accepts bids whose sum is *larger* than their actual joint contribution. Consequently, even though a competitive equilibrium exists, an insufficient amount of supply bids is contracted, and the reliability target undershot. Moreover, contracted suppliers would be overpaid for what they essentially deliver to the buyer. Hence, the auction outcome under positively related suppliers is unavoidably inefficient. □

[17] For instance, the PJM capacity market has implemented such a design. It offers the possibility to make coupled offers (PJM (2015)).

[18] We abstract here from additional problems such as limited liabilities, etc.

It is worthwhile to make a couple of remarks regarding this important proposition. First, note that the inefficiency of the procurement essentially stems from negative externalities emerging from positive dependencies that are not internalized in the competitive equilibrium. Also note that if all suppliers were independent, externalities would be non-existent, and they would be strictly positive with negative relations. In the former case, suppliers would still bid individually (thus avoiding coordination effort), while in the latter case they would bid jointly into the auction (thus taking advantage of the coalition surplus). Both cases would be consistent with the true joint contribution and hence, yield an efficient equilibrium.

To overcome the inefficiency revealed in Proposition 2, two strategies could be followed: Either, the buyer could require consistency of all (accepted) bids with their true joint contribution. Suppliers would then be forced to form and bid as a coalition that is truly able to deliver the required level of reliability. This would in principle entail an efficient auction outcome. However, in order to ensure the optimal coalition really emerges, all suppliers would need to know and process the properties of all other suppliers, including their interdependencies. In practice, ensuring such a level of complex information and information processing among the suppliers would probably be hard to achieve. Moreover, related transaction costs (that would subsequently be internalized in the bids and paid by the buyer) would probably be very high. Hence, as an alternative, the buyer could centralize the process of determining the optimal coalition in the auction clearing, thus only requiring the physical properties of the suppliers along with their prices. The buyer would then determine the joint contribution of the suppliers for reliability *himself* instead of inviting tenders for them.

Either way, it is the joint contribution that essentially counts for the buyer and that suppliers should be paid for. However, in order to design at a corresponding payoff scheme, the joint contribution (or value) needs to be split and allocated to the suppliers. A suitable allocation rule that disentangles the value of individual suppliers under complex interrelations will be discussed hereafter.

3.4.2 Allocation rule

There is an infinite number of allocation rules that could be applied to split the joint contribution and allocate it to individual suppliers. To name a few examples, the joint contribution could simply be split into equal parts, or be allocated according to average availabilities of the suppliers. However, instead of choosing an arbitrary allocation rule, it is important to note that the specific choice may have an impact that feeds back to the supply chain performance. For instance, consider a rule that is perceived particularly unfair by some suppliers. This could, as a consequence, reduce the exerted effort or even conscious insurrection (e.g. in form of a strike) of those suppliers, and hence, trigger a suboptimal supply chain performance. Equally important are long-term investment incentives that are implicitly provided by the choice of an allocation rule. Hence, we aim at identifying an allocation rule that incorporates a number of desirable properties to suitably address the aforementioned aspects.

Proposition 3. *A normatively fair allocation of reliability-related payoffs to interdependent individual suppliers can be obtained by the Shapley value:*

$$\Phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!}(v(S \cup \{i\}) - v(S)) \tag{3.23}$$

Proof. Taking as given the physical and strategic realities captured by the characteristic function, the Shapley value is an attempt to distribute the joint contribution of a coalition in a reasonable, or fair way. The motivation of the Shapley value is normative, and the criterion of fairness applied is egalitarianism (Mas-Colell et al. (1995)). It pays off the average marginal contribution of a supplier i to the set of predecessors, with the average taken over all orderings. It aims at incorporating the following desirable properties:

- **Static efficiency:** $\sum_i \Phi_i(v) = v(N)$, i.e., the joint value of the grand coalition is distributed and no utility is wasted. In our application, this property ensures that the sum of the individual payoffs according to the Shapley value corresponds to the joint contribution. Hence, the Shapley value ensures static efficiency of the allocation rule.
- **Symmetry:** If two suppliers i and j have an equivalent position in the game (i.e., if $v(S \cup \{i\}) = v(S \cup \{j\})$), or, in other words, if their individual contribution for reliability is the same, then their Shapley value is identical.
- **Linearity:** $\Phi_i(v_1 + v_2) = \Phi_i(v_1) + \Phi_i(v_2)$ for two characteristic functions v_1 and v_2, and $\Phi_i(av) = a\Phi_i(v)$ for any real number a. Hence, individual payoffs are equally affected if the coalition payoff (e.g., the price in an auction for reliability) changes.
- **Dummy:** If a supplier i offers no contribution to reliability (i.e., if $v(S \cup \{i\}) = v(S)$ for all coalitions S), then the Shapley value of this supplier is zero.

All the above properties are meaningful and relevant for allocations to individual suppliers according to their reliability contribution. Shapley (1953) showed that the payoff allocation according to the Shapley value given in Equation (3.23) is the unique value satisfying the efficiency, symmetry, linearity and dummy axioms. □

3.4.3 Investment incentives

We investigate the investment incentives induced by the above allocation rule (Equation (3.23)) based on the concept of biform games, studying the impact of surplus division on non-cooperative investment decisions (Brandenburger and Stuart (2007)). Investments take place in a stage prior to surplus allocation, such that the suppliers' profits are maximized. To assess the incentives of our allocation rule with respect to investment behavior, we focus on the relative weight placed on different investment options (i.e., suppliers). I.e., we analyze how investments would be *distributed*, instead of the *absolute* level of investment.[19] We compare the distribution of investments under two different premises. First, let us state the first-best benchmark. From an overall supply chain perspective, the aim is to maximize a weighted trade-off between high reliability (i.e., high v) and high production volumes (i.e., high μ), with parameter $\beta \in [0,1]$ reflecting the relative weight given to high generation volumes and reliability, respectively. Without loss of generality, let us suppose the case of two suppliers, and the possibility to distribute investments among this set of suppliers by adjusting the choice variable $\alpha \in [0,1]$. Then the first-best benchmark is obtained by maximizing the following objective function:

$$\max_{\alpha} \ \beta[v((1-\alpha)S_1 \cup \alpha S_2)] + (1-\beta)[(1-\alpha)\mu_{Y_1} + \alpha\mu_{Y_2}]. \tag{3.24}$$

Second, we consider the case of a price-taking independent investor who wants to extend an amount $I \in [0, \inf)$ of supply capacity, and needs to decide on the share $\gamma \in [0,1]$ (respectively $1-\gamma$) to be invested in supplier 1 (2). Under the assumption that reliability-related payoffs according to the Shapley value are being complemented by revenues from a production-related procurement or market, the

[19]Reaching a certain *absolute* investment level would simply require the allocation of an investment volume that is sufficiently high (especially, sufficiently high to cover the costs of the investment). The linearity property of the Shapley value would ensure that increasing the investment volume would affect individual suppliers equally, and hence, that the *distribution* of investment is not altered.

investor's profit function becomes

$$\Pi = p_r((1-\alpha)\Phi_1 + \alpha\Phi_2) + p_p((1-\alpha)\mu_{Y_1} + \alpha\mu_{Y_2}) - Ic, \tag{3.25}$$

with p_r the price paid per reliable capacity (e.g., determined through a capacity auction, as previously discussed), p_p the price per production volume, and c the capacity cost. We assume that p_r, p_p and c are constant and independent of the supplier.

Proposition 4. *The optimal distribution of independent investments coincides with the first-best benchmark if the ratio of prices for reliability and production is properly set.*

Proof. **First-best benchmark:** From a social perspective, we shall derive the system-optimal diversification strategy $\alpha^* \in [0,1]$ for two suppliers S_1 and S_2 with production capacity $(1-\alpha)Y_1$ and αY_2, respectively. Hence, we differentiate (3.24) with respect to α, which – after a few calculations – yields:

$$\alpha^* = \frac{\left[(\mu_{Y_2} - \mu_{Y_1})\frac{1-\beta}{\beta} + (\mu_{Y_2|X\approx 0} - \mu_{Y_1|X\approx 0})\right]\frac{f_X(0)}{f_X'(0)} + \sigma^2_{Y_1|X\approx 0} - \sigma_{Y_1,Y_2|X\approx 0}}{\sigma^2_{Y_1|X\approx 0} + \sigma^2_{Y_2|X\approx 0} - 2\sigma_{Y_1,Y_2|X\approx 0}}. \tag{3.26}$$

Independent investor: Now consider the case of a price-taking independent investor characterized by the profit function (3.25). For the case of two suppliers, the Shapley values Φ_1, Φ_2 can easily be calculated as

$$\Phi_1 = \frac{1}{2}v(S_1) + \frac{1}{2}(v(S_1 \cup S_2) - v(S_2)) \tag{3.27}$$

$$\Phi_2 = \frac{1}{2}v(S_2) + \frac{1}{2}(v(S_1 \cup S_2) - v(S_1)). \tag{3.28}$$

Inserted in Equation (3.25), the optimal weight for the distribution of the investment can be derived from the first-order condition $\frac{\partial \Pi}{\partial \gamma} = 0$, which – after a few calculations – yields γ^*. We find that $\gamma^* = \alpha^*$, i.e., equality of the investor's optimal choice and the first-best benchmark, however, only as long as the ratio of prices ($\frac{p_r}{p_p}$) equals the ratio of supply chain preferences for reliability and production ($\frac{\beta}{1-\beta}$).

Note that we have so far assumed a greenfield investment with no supply capacity being present. However, it is straightforward to relax this assumption, and instead allow for some arbitrarily distributed capacity already being installed. These capacities will, due to their stochastic interdependence, impact the investor's profit (and hence, investment decisions) via the amount and shares in payoffs allocated to the investor's capacities. However, the impact on the independent investor's decision induced by the Shapley value is such that the additional capacity is ensured to be installed where it is supply chain optimal. The reason is that an investment in line with supply chain optimality yields highest overall contributions to reliability, which in turn is distributed among the coalition members. As the Shapley value strictly increases with the overall surplus (linearity axiom), it follows that an independent investor aligns its strategy with the supply chain optimal result. Hence, the Shapley value is able to ensure that additional investments bring the system towards the first-best optimal distribution of capacities.

Lastly, generalization to $n > 2$ suppliers is straightforward using $\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{Y_i,Y_j} = \sum_{i=1}^n \sigma_{Y_i}^2 + 2\sum\sum_{1\leq i<j\leq n} \sigma_{Y_i,Y_j}$. □

Let us briefly discuss the plausibility of the results derived in Proposition 4 and its proof. Consider the extreme case of $\beta = 0$, i.e., when full preference is given to production volumes. Then, the first-best result becomes

$$\tilde{\alpha}^* \mid_{\beta\to 0} = \begin{cases} 0 & : \mu_{Y_2} < \mu_{Y_1} \\ 1 & : \mu_{Y_2} > \mu_{Y_1} \\ \text{indifferent} & : \mu_{Y_2} = \mu_{Y_1}, \end{cases} \tag{3.29}$$

i.e., – as expected – a simple preference for the location with higher average availability if there is any, and indifference otherwise.

For the case of some preference being given to reliability, i.e., for $\beta \neq 0$, the result reveals that the optimal share depends upon the difference between the average availabilities of the two suppliers, their variances, as well as their covariance. Specifically, the following characteristics require a stronger redistribution from supplier S_1 to S_2: high average availability of S_2; low average availability of S_1; large variability of S_1; and low covariance between S_1 and S_2.

Some further remarks concerning Proposition 4 are worthwhile: First, note that it always needs a certain price level to incentivize investments. For reliability-payoffs only, it would be individually rational to invest as long as $\frac{p_r}{c} > \frac{I}{v(\gamma S_1 \cup (1-\gamma)S_2)}$, whereas for complementary payoffs, the individual rationality constraint becomes $\frac{p_r}{c}(\Phi_1 + \Phi_2) + \frac{p_p}{c}(\mu_{Y_1} + \mu_{Y_2}) > I$. Second, we want to point the readers interest to possible extensions of our analysis of investment incentive effects: Besides the case of one investor, it would also possible to consider one or multiple players per supplier optimizing their payoffs by choosing an investment level while strategically considering the action performed by the other investors. For the case of one player per supplier, we have found an underinvestment of individual players compared to the supply chain optimal outcome, also impacting the distribution of capacities towards an inefficient balance.[20] As expected, this inefficiency increases with the level of interdependencies among the locations, and vanishes for independent resources. This result might indicate possible problems occurring from the exercise of market power within the Shapley value approach.

Lastly, we remark that from a supply chain organization perspective, properly setting the ratio of prices for reliability and production may indeed be challenging. An appropriate weighting of prices in the case of complementary reliability- and production-related payoffs could be ensured, e.g., through efficient prices stemming from appropriately designed markets.

3.5 Empirical case study: Wind power in Germany

This section presents an empirical case study for wind power embedded in the German power system. Based on real-world data, it is meant to demonstrate applicability of the concepts developed in the previous sections as well as their practical relevance, and to empirically confirm our analytical results. The case of wind power in Germany appears to be novel and pertinent. With 34.02 out of 188.68 GW production capacity being installed, wind power plays a major role in Germany's generation portfolio, while it is a declared goal to integrate the technology into general market structures. Moreover, power system reliability as well as possible regulatory interventions, such as capacity mechanisms, have been heavily discussed for several years.

3.5.1 Estimation procedure

Recall Equation (3.3) that needs to be solved for v in order to obtain the contribution of individual suppliers. We assume random variable C, i.e., the availability of the power supply fleet apart from wind power, to be independent from wind power Y and demand D, and that its distribution can be determined through convolution of the suppliers' outage probabilities (see below). In contrast, the joint distribution of Y and D is estimated from simultaneous historical observations. As we only have one observation per instant in time t, we need to extend Equation (3.3) to multiple

[20]The result shows strong similarities with the withholding situation in a Cournot-Duopoly.

time periods, such that random variables Y, D may be replaced by corresponding observations y_t, d_t:[21]

$$\sum_{t=1}^{T} \Pr(d_t + v \leq C + y_t) \equiv \sum_{t=1}^{T} \Pr(d_t \leq C) \tag{3.30}$$

Note that summing up the probabilities over time yields the expected value during the considered number of hours T. This measure is often applied to formulate or benchmark reliability levels. For instance, a 1-day-in-10-years criterion has often been used as a benchmark or target value, both in the academic literature (e.g., see Keane et al. (2011)) and in practice (e.g., by the Midcontinent ISO or the ISO New England).[22]

3.5.2 Data

The necessary data can be classified into three main areas: First, detailed information is needed about installed capacities and availability factors of dispatchable power suppliers apart from wind (C in Equation (3.30)). Second, the analysis requires high-resolution data on wind power capacities as well as their infeed profiles (y_t). Third, load levels with the same temporal resolution and regional coverage are needed to perform the calculations (d_t). Descriptions of the data along with some preparatory calculations can be found in the 3.7. Importantly, we find clearly positive dependencies among all wind power profiles (see 3.7 for a detailed analysis).

Regarding the level of detail in our analysis, we would ideally opt for a representation of each individual supplier in the system. However, for large systems with many interdependent suppliers, this would quickly involve impractical data requirements and calculation efforts. This is mainly due to the fact that each of the $2^n - 1$ possible coalitions needs to be calculated. Hence, some administrative division may provide a satisfactory level of disaggregation while still being manageable.[23] In our

[21] The validity and consistency of the result obtained from this reformulation may be justified by the central limit theorem (Zachary and Dent (2011)).

[22] Alternative economic approaches would try to estimate efficient reliability levels from the value of lost load (VOLL) and costs of maintaining a certain level of supply chain reliability (Telson (1975)). A thorough discussion would clearly be beyond the scope of this paper, such that the interested reader is referred to Stoft (2002) for the necessary calculations, and attempts to quantify the VOLL, e.g., by Anderson and Taylor (1986) for Sweden or by Growitsch et al. (2014) for Germany. Noticeably, due to the fact that data requirements and estimation procedures are far from being straightforward, it is not surprising that rules of thumb and common practice, such as the 1-day-in-10-years, are often applied – both by academics as well as practitioners.

[23] The same (relative) Shapley value would then apply to all units within that area.

empirical example, we will hence aggregate wind power on the federal state level.[24]

3.5.3 Results

Reliability of Germany's dispatchable power fleet

To determine the cumulative distribution function F_C describing the availability distribution of Germany's power supply fleet apart from wind power, we assume that each supplier is either fully available or not,[25] and that individual failure probabilities are independent. Based on these assumptions, F_C can be derived via convolution of the individual suppliers' failure probabilities. We implemented the algorithm developed in Hasche et al. (2011) which proved to be fast enough to calculate the cdf for our supply chain (consisting of nearly 900 dispatchable power supply units) in less than a minute on a standard laptop. The resulting (complementary) cumulative distribution function is shown in Figure 3.1, together with a histogram of load levels. Load levels are for the most part far left of the critical range of around 90 GW where the complementary cdf begins to drop sharply.

From Equation (3.30) and our data, we obtain a basically perfect reliability of supply of $R^{\mathscr{S}} = 1 - 1.15e^{-12}$ hours/year.[26] Hence, installed capacities appear to be largely sufficient to reliably cover today's load profile.

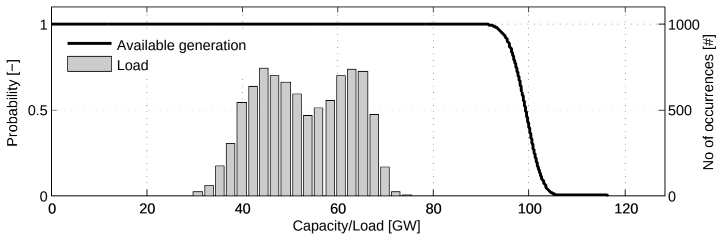


Figure 3.1: Complementary cdf of Germany's power plants without wind power together with a histogram of 2013 load levels

Having previously mentioned the often applied 1-day-in-10-years target, we find that a 30% increase in current (2013) load levels could be sustained in order to

[24]Germany consists of 16 federal states: Baden-Württemberg (BW), Bayern (BY), Berlin (BE), Brandenburg (BB), Bremen (HB), Hamburg (HH), Hessen (HE), Mecklenburg-Vorpommern (MV), Niedersachsen (NI), Nordrhein-Westfalen (NW), Rheinland-Pfalz (RP), Saarland (SL), Sachsen (SN), Sachsen-Anhalt (ST), Schleswig-Holstein (SH), and Thüringen (TH).

[25]I.e., no partial failures are taken into account.

[26]Note that $R^{\mathscr{S}}$ is a risk measure, not implying that a certain number of load shedding events effectively occurs. Hence, the figures presented here and in the following should not be confused with realized statistical numbers, such as the Average System Interruption Duration Index (ASIDI).

reach that threshold, again indicating large amounts of over-capacity in the German system.

The contribution of wind power to reliability of supply

We find the joint contribution of Germany's aggregated wind power to reliability of supply $v(N)$ to be 3376 MW, corresponding to $\tilde{v}(N) = 9.9\%$ relative to the installed capacity of 34.02 GW.

To provide empirical evidence for the decreasing returns to scale found in Corollary 1, we scale today's wind power capacity by factors of 0 to 3, while assuming unchanged characteristics of load and dispatchable power (Figure 3.2, left hand side). As expected, the contribution decreases monotonically along a convex function. To empirically confirm gains of diversification (Corollary 2), we first calculate the contribution of wind capacities being installed in two states (S_1 and S_2) separately, ranging from 100 to 900 MW. We take Sachsen (SN) and Thüringen (TH) as an example.[27] We then calculate the joint contribution of a diversified portfolio of the same aggregated capacity being installed in both the states (capacity split half-half, i.e., $v(\frac{1}{2}S_1 \cup \frac{1}{2}S_2)$). Whereas the joint contribution of the diversified portfolio lies in between the separate states for small capacities, it clearly yields higher contributions for increasing penetration levels. Especially, the rate at which the diversified contribution drops is significantly smaller.

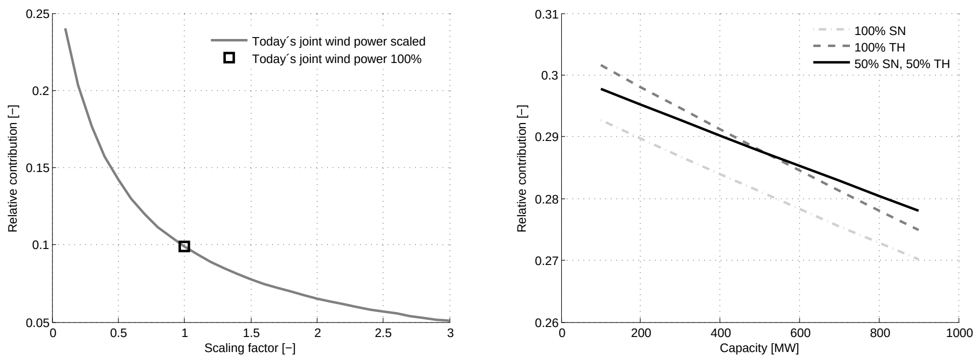


Figure 3.2: Decreasing returns to scale (left); Gains of diversification (right)

The following Table 3.1 presents today's installed wind power capacities per state, along with the individual absolute ($v(S_i)$) and relative contributions ($\tilde{v}(S_i)$). Individual relative contributions vary significantly, ranging from 13.9 to 31.3%. Sum-

[27] These two states have a similar average availability of wind power and a correlation coefficient of 0.59.

ming up the individual contributions would clearly yield a false joint contribution of 19.6%. Compared to the previously determined consistent joint contribution of 9.9%, this is a huge overestimation that would be induced by neglecting the positive interdependencies between the resource availability at different locations. This empirical finding underlines the importance of the non-additivity property, as stated in Corollary 3.

State	Installed Capacity [MW]	Absolute Contribution [MW]	Relative Contribution
BW	624	156	24.9%
BY	1066	158	14.8%
BE	2	1	31.3%
BB	5233	1013	19.4%
HB	114	34	30.0%
HH	55	16	28.1%
HE	961	145	15.1%
MV	2301	588	25.6%
NI	7676	1392	18.1%
NW	3473	483	13.9%
RP	2366	395	16.7%
SL	223	44	19.9%
SN	1055	281	26.6%
ST	4093	955	23.3%
SH	3683	723	19.6%
TH	1097	295	26.9%

Table 3.1: Installed capacity, absolute and relative contributions calculated for each state individually

Supply chain organization

For our case study, empirical evidence shows that our game is largely subadditive, and that the core is empty. The following Figure 3.3 is meant to illustrate the subadditive nature of our problem from two perspectives. The left hand side shows the coalition's joint contribution vs. its size (in terms of coalition members) for all 65535 possible coalitions,[28] clearly indicating an inverse relation. Similarly, the right hand side shows the error that would be induced by neglecting the (positive) interdependencies by plotting each possible coalition's sum of individual contributions vs. its consistent joint contribution.[29] Whereas superadditivity would require all points to be above the perfect additivity line, we find that virtually all points lie well below, with the error substantially increasing for larger individual contributions (essentially due to decreasing returns to scale, i.e., in line with Corollary 1).

[28]For a game with 16 players, the number of possible coalitions is $2^{16} - 1 = 65535$.

[29]The most exterior point, i.e. the grand coalition of all states, has already been discussed in the previous paragraph.

The results of both figures are driven by the positive interdependencies among the 16 states that are illustrated in Figure 3.9. They demonstrate the empirical relevance of Proposition 2, i.e., the inefficiency of a competitive equilibrium in a reliability procurement where only individual bids are verified.

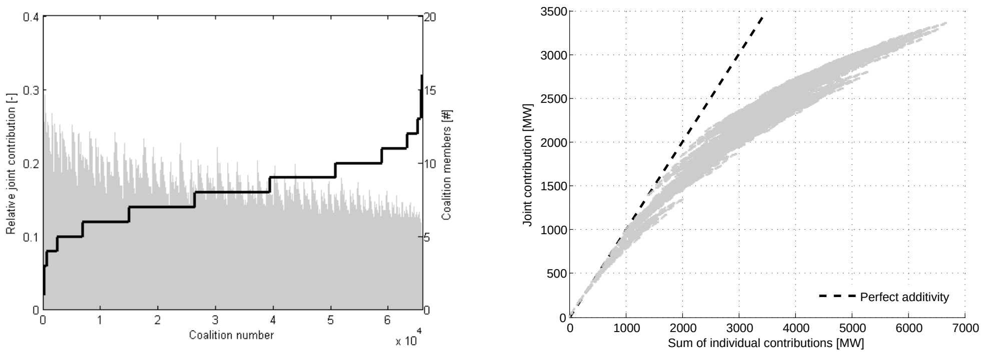


Figure 3.3: Relative joint contribution vs. size of coalition (left); Sum of individual contributions vs. joint contribution (right)

Payoff allocation according to the Shapley value

Turning to the core result, Figure 3.4 presents the relative Shapley value $\tilde{\Phi}_i(v(N))$, i.e. the payoffs relative to installed capacities allocated to each of the 16 states (green squares). To put these values into perspective, they are presented in combination with the suppliers' average availability (μ_i) as well as with their separately calculated individual contribution ($\tilde{v}(S_i)$). We find that the individual contribution as well as the Shapley value tends – as expected – to decrease with average availability. However, the Shapley values also show pronounced deviations from the joint contribution (9.9%), as well as from the ordering of the individual contributions and average availabilities. For instance, Niedersachsen (NI) and Sachsen-Anhalt (ST) both have high average availabilities, but also large installed capacities as well as pronounced positive dependencies with other states – and hence comparatively low Shapley values. Overall, the Shapley values range from 15.1% for Baden-Württemberg (BW) and Sachsen (SN) to 5.7% for Bavaria (BY).

Relating the Shapley value to the possibility of emerging coalitions in a competitive environment (Proposition 2), we find that 50337 out of 65535 possible coalitions would be able to block this allocation, due to the subadditivity found above. Hence, it seems indeed reasonable to transfer the process of determining the joint contribution of the coalition and allocating individual payoffs to an independent central authority.

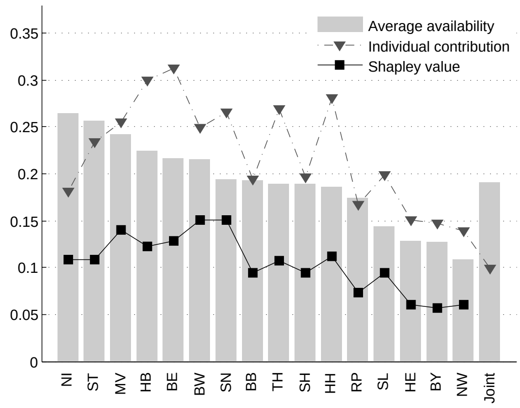


Figure 3.4: Average availabilities, individual contributions and Shapley values (all relative)

In practice, it will be most plausible to pay power supply units according to a weighted production- and reliability-related preference (i.e., if existing markets for energy are or were to be complemented by a capacity mechanism). As discussed in Section 3.4.3 and Proposition 4, it would then be important to find a balance between high production volumes (i.e., high average availability) and high Shapley values. Figure 3.5 presents this tradeoff, showing that states perform strikingly different on both dimensions. Separating the field into quadrants, we find that Baden-Württemberg (BW) and Sachsen (SN) perform particularly well for the case of more preference given to reliability, Niedersachsen (NI) and Sachsen-Anhalt (ST) for the case of more preference on production volumes, and Mecklenburg-Vorpommern (MV) for an equal balance. Nordrhein-Westfalen (NW), Bavaria (BY) and Hessen (HE) perform poorly with respect to both properties. Moreover, it should be noticed that even though a general positive trend can be found, the tradeoff is widely scattered around a straight line. In practice, this property would offer the possibility to effectively incentivize investments that contribute much better to reliability than purely production-based decisions by means of appropriate payoffs reflecting (weighted) preferences for production volumes *and* reliability.

3.6 Conclusions

Supply chain reliability is a timely and relevant subject that has been studied intensively in the academic literature. However, the specific role of individual, possibly stochastic and interdependent suppliers for supply chain reliability has so far been disregarded. To fill this gap, we have thoroughly investigated the issue based on

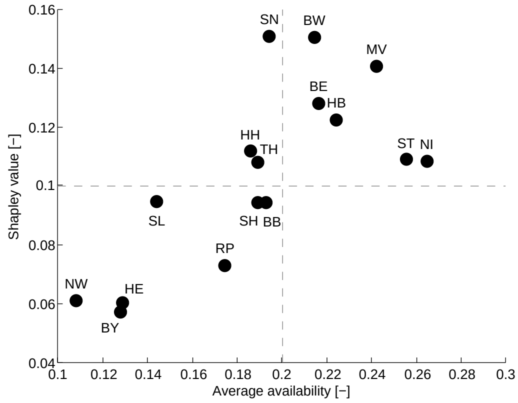


Figure 3.5: Average availability vs. Shapley value

statistical and economic analyses. Especially, we have derived an analytical solution for the contribution of individual suppliers to reliability as well as a corresponding payoff scheme that accounts for the statistical properties of the problem. Practical applicability and relevance has been demonstrated with a numerical example based on wind energy in the German power system, thereby confirming our analytical findings.

With the concepts developed, the reliability of supply chains could in the future be designed and managed more efficiently. For instance, our payoff scheme could be applied to write payoff contracts in monopsonistic markets or public procurements from various suppliers, or to allocate individual payoffs in a bonus system. To name a practical example, the approach could improve the design of capacity mechanisms in power systems, where the state or regulator procures reliability from power supply units. In fact, our analysis reveals that existing designs to incorporate variable renewable energies into capacity-mechanisms contain substantial shortcomings.[30] In contrast, the properties of the approach developed in this paper would ensure a

[30] For instance, the selective Spanish capacity mechanism excludes intermittent resources by allocating zero reliable capacity, hence inducing a situation of static inefficiency as soon as there is any positive contribution – which is likely to be the case as presented in our empirical case study. Moreover, the Spanish design fails to incentivize investments into suppliers that are beneficial for reliability, hence also inducing an additional long-term inefficiency. As a serious design flaw in the PJM capacity auction, they attempt to prequalify suppliers individually – which contradicts the non-additivity we have found in Corollary 3. Moreover, while PJM allows suppliers using intermittent resources (i.e., wind and solar power) to cooperate and bid jointly, there will be disincentives to do so as implied in Proposition 2. These shortcomings will necessarily entail an inefficient outcome. Drawbacks in the UK and French capacity mechanisms can be found in the average payoff allocated to each unit in the same technology-class. While the approach ensures static efficiency by complying with the consistent joint contribution, it lacks the provision of locational investment incentives. In addition, the approach may face opposition from underrated suppliers whose individual contribution to reliability lies well above the average level.

level playing field for all suppliers, and clearly outperform existing, currently applied alternatives.

Our analysis could be extended in several directions: It would be worthwhile to consider a strict optimization applied within our framework with costs attached to individual suppliers, e.g., to determine an optimal subset among a broader set of interdependent suppliers. The investment incentives provided by our Shapley value approach should be studied in more detail, including variations in the number and role of market participants (e.g., market power) as well as a more sophisticated representations of investment decisions. The effect of an endogenous demand-side response on reliability would also be interesting. Our approach could be applied to other supply chains incorporating stochastic suppliers and demand, especially those with positive dependencies. For instance, it would be interesting and relevant to assess the value of an additional call center for reliable customer services, or to design suitable payoffs for individual drivers in a large network providing transport services.

Acknowledgments

Special thanks go to Joachim Bertsch, Clara Dewes, Felix Höffler and Ulrich Thonemann for valuable discussions and support. Funding of the German research society DFG through research grant HO5108/2-1 is gratefully acknowledged.

3.7 Appendix: Data and preparatory calculations

Installed capacities and availability factors

For information about currently installed power supply units, we use the List of Power Plants prepared and provided by the Federal Network Agency (Bundesnetzagentur (2014)). It lists all existing units in Germany with a net nominal electricity capacity of at least 10 MW.[31] Moreover, supply facilities of less than 10 MW are also included on an aggregated basis grouped by energy source. Extracted net nominal capacities by fuel type are depicted in Table 3.2. As for wind power, 34.02 GW were installed by mid 2014, distributed among the 16 federal states as shown in Figure 3.6.

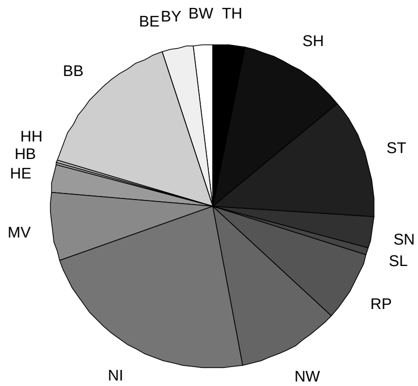


Figure 3.6: Distribution of installed wind power capacities (34.02GW total)

Regarding the suppliers' availabilities, we assume fuel-type specific factors according to historical observations, taken from VGB and Eurelectric (2012), and complemented with dena (2010) for Hydro, Geothermal and Biomass, as reported in Table 3.2. For PV, we assume an availability factor of 0%.[32]

Wind speed data and wind to power conversion

As we want to focus on the supply side uncertainty of wind power in our empirical example, coverage and resolution of the wind data are crucial for obtaining reliable results. Consequently, we use hourly wind speed data of 32 years (1982-2013) to

[31] In addition, it also comprises capacities directly feeding into the German grid from Luxembourg, Switzerland and Austria (which we assume to contribute to Germany's generation adequacy)

[32] This is a conservative estimate, however, consistent with the observation that highest load hours occur in the late evening during winter time.

Fuel type	Availability	Capacity [GW]
Biomass	88.0%	6.38
Coal	83.9%	27.73
Gas	88.3%	25.42
Geothermal	90.0%	0.03
Hydro (pump) storage	90.0%	10.63
Hydro run-of-river	40.0%	3.92
Lignite	85.3%	20.95
Nuclear	83.3%	12.07
Oil	89.2%	4.14
Others (Waste, Landfill gas, etc.)	90.0%	5.32
PV	0.0%	37.45
Wind	*to be calculated*	34.02

Table 3.2: Availability factors and installed capacities per fuel type

cover a broad range of possible wind patterns in Germany, provided by the national climate monitoring of the German Weather Service (DWD (2014)).[33] We select one representative location per federal state according to the agglomeration of wind turbines within each state.[34]

State	DWD-ID	Observatory name
BW	4887	Stötten
BY	5705	Würzburg
BE	3987	Potsdam
BB	164	Angermünde
HB	691	Bremen
HH	1975	Hamburg-Fuhlsbüttel
HE	1420	Frankfurt
MV	4271	Rostock
NI	891	Cuxhaven
NW	2483	Kahler Asten
RP	2385	Idar-Oberstein
SL	4336	Saarbrücken
SN	1048	Dresden
ST	1957	Halle-Kröllwitz
SH	4466	Schleswig
TH	1270	Erfurt-Weimar

Table 3.3: Selected DWD observatories

The conversion of wind speed to electrical power output is described by a turbine-specific power curves. As a representative power curve, we use the Nordex S77 turbine with a hub height of 77 meters and a rated power of 1.5kW (Nordex (2007)).

[33] In case of missing data, empty entries are replaced by interpolations based on the previous and next available value if the empty space is not exceeding 12 hours. If the gap is longer, entries are replaced by data of the same station and same hours of the previous year. As measurements are taken a couple of meters above ground only, wind speeds are scaled to the wind turbines' hub height assuming a power law: $v_{h_1} = v_{h_0}(h_1/h_0)^{\alpha}$, where h_0 is the measurement height, h_1 the height of interest and α the shear exponent. According to Firtin et al. (2011), α is assumed to be 0.14.

[34] The list of selected observatories is presented in the following Table 3.3.

Wind speed to power conversion is implemented via lookup-tables with linear interpolation.

To validate this relatively simple model for generating wind power profiles, we compare statistical profiles and volumes with our synthetic model-generated data. Figure 3.7 compares statistical yearly production per federal state (available from Agentur für erneuerbare Energien (2014) for the years 2001-2012, except for 2007) against our modeling results (based on historically installed wind capacities taken from German Wind Energy Association (BWE) (2012) and wind speeds from the corresponding years). As can be seen, our model overestimates production for the first years, whereas satisfactory conformity is reached for more recent periods. This is probably due to improvements in turbine technologies over the years (remember that we have applied the power curve of a relatively modern turbine).

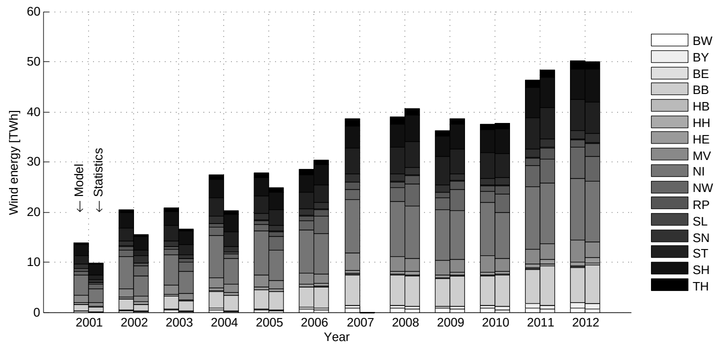


Figure 3.7: Yearly wind power production per federal state (model: left, statistics: right)

Another step of validation has been carried out for the hourly profiles, based on a comparison of the distributions of the model-based data with historical production profiles (available for the years 2010-2013) by means of QQ-plots (Figure 3.8). Distributions are found to be very similar for those years, however, with slightly decreasing conformity for upper quantiles. Conducting a simple regression analysis yields an R^2 of 0.80, 0.81, 0.82 and 0.81 for the years 2010-2013. For completeness, summary statistics of the obtained profiles (based on 2014 wind capacities) are also provided in Table 3.4.

Load profiles

Germany's load levels are reported on an hourly basis by ENTSO-E (2012), representing the hourly average active power absorbed by all installations connected to the transmission or distribution network. Instead of using multiple years of load

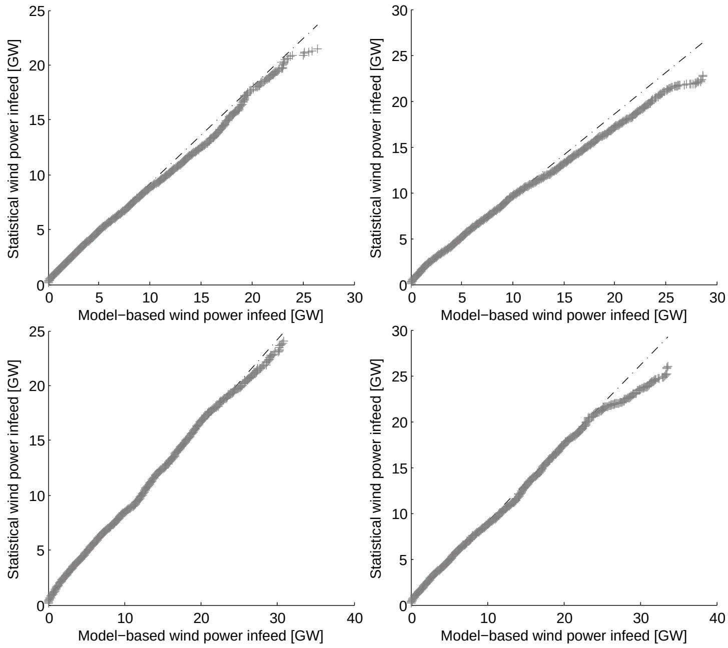


Figure 3.8: QQ-plots of model-based and historical production profiles, for the years 2010-2013

Profile	Min	Max	Mean	Std
Load Germany (2013)	29.55	75.623	52.8916	9.6327
Wind Power BW (1982-2013)	0	0.6241	0.1340	0.1828
Wind Power BY (1982-2013)	0	1.0666	0.1363	0.2513
Wind Power BE (1982-2013)	0	0.002	0.0004	0.0005
Wind Power BB (1982-2013)	0	5.2331	1.01	1.4062
Wind Power HB (1982-2013)	0	0.1142	0.0256	0.0321
Wind Power HH (1982-2013)	0	0.0551	0.0102	0.0138
Wind Power HE (1982-2013)	0	0.961	0.1239	0.2147
Wind Power MV (1982-2013)	0	2.3014	0.5572	0.731
Wind Power NI (1982-2013)	0	7.6757	2.0334	2.3978
Wind Power NW (1982-2013)	0	3.4732	0.3751	0.7232
Wind Power RP (1982-2013)	0	2.3661	0.4126	0.6694
Wind Power SL (1982-2013)	0	0.2233	0.0322	0.051
Wind Power SN (1982-2013)	0	1.0545	0.2048	0.2865
Wind Power ST (1982-2013)	0	4.0925	1.0465	1.3257
Wind Power SH (1982-2013)	0	3.6828	0.6963	0.932
Wind Power TH (1982-2013)	0	1.0965	0.2073	0.3079

Table 3.4: Summary statistics of load and wind power profiles in [GW]

data, we restrict our attention to the most recent year 2013 in order to focus on the supply side uncertainty. We hence refer to reliability under current load levels and profiles. Summary statistics of the load profile are comprised in Table 3.4.

Correlation analysis

As has been noticed several times during the theoretical analysis in Sections 3.3 and 3.4, covariance among the power production of the different locations crucially impacts the properties and results of the problem. In order to get an impression of the dependencies characterizing wind power and load in Germany, Figure 3.9 shows the matrix of linear correlation coefficients ρ.

As can be seen, ρ among the wind power profiles is in a range of $[0.10, 0.77]$, with a mean of 0.48 (excluding diagonal values). Correlations between wind power and load are in a range of $[0.06, 0.17]$, with a mean of 0.13. Hence, all dependencies are clearly positive.[35]

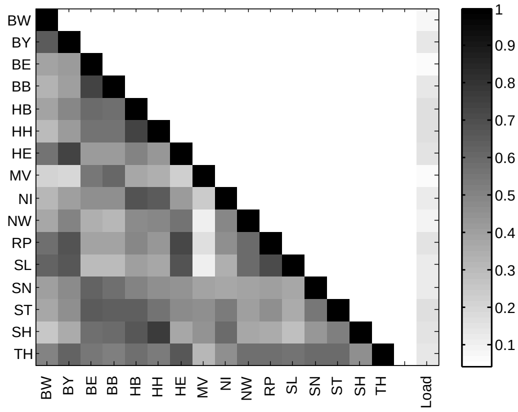


Figure 3.9: Correlation matrix of wind power profiles

[35] As reliability of supply is particularly relevant during tight capacities, we recalculate the same numbers for the data corresponding to the 5% highest load hours, resulting in a range of $[0.15, 0.79]$ and a mean of 0.51 for correlations among wind power, and $[-0.02, 0.03]$ and a mean value of 0.00 for correlations between wind power and load. Hence, wind power dependencies are even slightly more pronounced during high-load-hours, whereas wind power and load are independent.

4 Congestion management in power systems - Long-term modeling framework and large-scale application

In liberalized power systems, generation and transmission services are unbundled, but remain tightly interlinked. Congestion management in the transmission network is of crucial importance for the efficiency of these inter-linkages. Different regulatory designs have been suggested, analyzed and followed, such as uniform zonal pricing with redispatch or nodal pricing. However, the literature has either focused on the short-term efficiency of congestion management or specific issues of timing investments. In contrast, this paper presents a generalized and flexible economic modeling framework based on a decomposed inter-temporal equilibrium model including generation, transmission, as well as their inter-linkages. The model covers short-run operation and long-run investments and hence, allows to analyze short and long-term efficiency of different congestion management designs that vary with respect to the definition of market areas, the regulation and organization of TSOs, the way of managing congestion besides grid expansion, and the type of cross-border capacity allocation. We are able to identify and isolate implicit frictions and sources of inefficiencies in the different regulatory designs, and to provide a comparative analysis including a benchmark against a first-best welfare-optimal result. To demonstrate the applicability of our framework, we calibrate and numerically solve our model for a detailed representation of the Central Western European (CWE) region, consisting of 70 nodes and 174 power lines. Analyzing six different congestion management designs until 2030, we show that compared to the first-best benchmark, i.e., nodal pricing, inefficiencies of up to 4.6% arise. Inefficiencies are mainly driven by the approach of determining cross-border capacities as well as the coordination of transmission system operators' activities.

4.1 Introduction

The liberalization of power systems entails an unbundling of generation and grid services to reap efficiency gains stemming from a separate and different organization. While there is competition between generating firms, transmission grids are considered a natural monopoly and are operated by regulated transmission system operators (TSOs). However, strong inter-linkages remain between these two parts of the power system: From a transmission perspective, TSOs are responsible for non-discriminatory access of generating units to transmission services while maintaining a secure grid operation. They are thus strongly influenced by the level and locality of generation and load. Furthermore, due to Kirchhoff's laws, operation and investment decisions of one TSO may affect electricity flows in the area of another TSO. From a generation firms' perspective, activities are impacted by restrictions on exchange capacities between markets or operational interventions by the TSOs to sustain a reliable network.

An efficient regulatory design of those inter-linkages between generation and grid will positively affect the overall efficiency of the system, for instance by providing locational signals for efficient investments into new generation or transmission assets. To ensure an efficient coordination of short (i.e., operational) and long-term (i.e., investment) activities in the generation and grid sectors, congestion management has been identified to be of utmost importance (e.g., Chao et al. (2000)). Different regulatory designs and options are available to manage congestion, including the definition of price zones as well as various operational and investment measures. Because it is able to deliver undistorted and hence efficient price signals, nodal pricing is a powerful market design to bring along efficiency. This was shown in the seminal work of Schweppe et al. (1988) and Hogan (1992). Nevertheless, many markets deviate and pursue alternative approaches, e.g., due to historical or political reasons. For instance, most European countries deploy national zonal market areas with uniform electricity prices. Implicitly, several challenges are thus imposed upon the system: First, in zonal markets, intra-zonal network congestion remains unconsidered by dispatch decisions. However, if a dispatch induces intra-zonal congestion (which is typically often the case), it might be necessary to reconfigure the dispatch, known as re-dispatch. Alternatively, the dispatch can be impacted by charging grid costs directly to generators in order to avoid congestion in the market clearing process (a so-called generator- or g-component, also known as grid connection charge). Such charges reflect the locational scarcity of the grid, and are thus conceptually similar to nodal prices, depending on the calculation method applied (see Brunekreeft et al. (2005) for a comprehensive discussion). Second, cross-border capacity needs to be managed. Whereas historically, cross-border capacities have often been auctioned explicitly, many market areas are now turning to implicit market coupling based on different allocation routines, such as net-transfer capacities (NTC) or flow-based algorithms (Brunekreeft et al. (2005), Oggioni and Smeers (2012), Oggioni and Smeers (2013)).[1]

The literature has investigated various regulatory designs to manage congestion in power systems from different perspectives. Static short term efficiency of nodal pricing – as shown by Schweppe et al. (1988) – was confirmed, e.g, by van der Weijde and Hobbs (2011) who compare nodal pricing and NTC based market coupling in a stylized modeling environment. Furthermore, several papers have quantified the increase in social welfare through a switch from zonal to nodal pricing for static real world case studies (see for example: Green (2007), Leuthold et al. (2008), Burstedde (2012), Neuhoff et al. (2013)). Similarly, Daxhelet and Smeers (2007) show that generator and load components reflecting their respective impact on congestion have a positive effect on static social welfare (as well as its distribution), while Oggioni and Smeers (2012) investigate different congestion management designs in a six node model and find that a single TSO or multi-lateral arrangements for counter-trading between several TSOs may improve efficiency. Oggioni et al. (2012) and Oggioni and Smeers (2013) show that in a zonal pricing system, the configuration of zones as well as the choice of counter-trading designs have a significant impact on efficiency.

A second line of literature deals with the dynamic long-term effects of congestion management, i.e., the investment perspective. On the one hand, issues of timing (e.g., due to uncertainty or commitment) in settings consisting of multiple players (such as generation and transmission) have been addressed. Höffler and Wambach (2013) find that long-term commitment of a benevolent TSO may lead to inefficient investment decisions due to the locational decisions of investments in generation. In contrast, Sauma and Oren (2006) and Rious et al. (2009) formulate the coordination problem between a generation and a transmission agent as a decomposed problem, and find that a prospective coordinated planning approach as well as transparent price signals entail efficiency gains, though some inefficiencies remain and the first-best is not realized. On the other hand, imperfect simultaneous coordination (e.g., due to strategic behavior or hidden information) has been investigated by Huppmann and Egerer (2014) for the case of multiple TSOs being active in an interconnected system. They find that a frictionless coordinated approach outperforms the system outcome with strategic TSOs maximizing social welfare within their own jurisdiction.

[1] Under implicit market coupling, cross-border capacities and prices are implicitly taken into account during the joint clearing process of coupled markets.

With this paper, we contribute to the above literature with a generalized and flexible economic modeling framework for analyzing the short as well as long-term effects of different congestion management designs in a decomposed inter-temporal equilibrium model including generation, transmission, as well as their inter-linkages. Specifically, with our framework we are able to represent, analyze and compare different TSO organizations, market areas (i.e., nodal or zonal pricing), grid expansion, redispatch or g-components, as well as calculation methods for cross-border capacity allocation (i.e., NTC and flow-based). A major advantage of our analytical and numerical implementation is its flexibility to represent different congestion management designs in one consistent framework. We are hence able to identify and isolate frictions and sources of inefficiencies by comparing these different regulatory designs. Moreover, we are able to benchmark the different designs against a frictionless welfare-optimal result, i.e., the first-best. In order to exclusively focus on the frictions and inefficiencies induced by the congestion management designs, we do not address issues of timing, such as uncertainty or sequential moving. Instead, we assume perfect competition, perfect information, no transaction costs, utility-maximizing agents, continuous functions, inelastic demand and an environment where generation and grid problems are solved simultaneously. As an additional contribution, we calibrate and numerically solve our model for a large-scale problem. Specifically, we investigate a detailed representation of the Central Western European (CWE) region.[2] To tackle the complex nature of the optimization problem, we develop a numerical solution algorithm based on decomposition, while a detailed analysis of the convergence behavior suggests that the results obtained are robust. Thereby, we offer a sound indication on how different congestion management designs perform in practice, and provide empirical evidence that nodal pricing is the efficient benchmark while alternative designs imply inefficiencies of up to 4.6% until 2030.

The paper proceeds as follows: In Section 4.2, we analytically develop our modeling framework. In Section 4.3, a numerical solution method to solve this framework is proposed. In Section 4.3, we apply the methodology to a detailed representation of the CWE region in scenarios up to the year 2030. Section 4.5 concludes and provides an outlook on future research.

[2]The CWE region is one of seven regional initiatives to bring forward European market integration. The countries within this area are Belgium, France, Germany, Luxemburg and the Netherlands.

4.2 Economic framework

In order to develop a consistent analytical modeling framework for different congestion management designs, we start with the well-known model for an integrated optimization problem for planning and operating a power system.[3] By design, this model does not contain any frictions and inefficiencies. Hence, the results obtained are necessarily first-best and may serve as the efficient benchmark for alternative settings. Moreover, it corresponds to the concept of nodal pricing as introduced by Schweppe et al. (1988).[4]

To depict various congestion management designs, we make use of the possibility to separate an integrated optimization problem into multiple levels (or, in other words, subproblems). Even though the model structure is then different, it can be shown that both formulations of the problem yield the same results. However, in the economic interpretation we can take advantage of the separated model structure representing unbundled generation and transmission sectors. On the generation stage, competitive firms decide about investments in and dispatch of power plants, whereas the transmission stage consists of one or multiple TSOs that efficiently expand and operate transmission grid capacities.[5] Lastly, with generation and transmission separated, we are able to introduce six practically relevant congestion management designs through the manipulation of the exchange of information between and among the two levels, and show how they deviate from the first-best.

Even though the modeling framework would allow to study an extensive range of congestion management designs, we restrict our attention to four settings (and two additional variations) that are both, relevant in practical applications and sufficiently different from each other. Specifically, our settings vary in the definition of market areas (nodal or coupled zonal markets), the regulation and organization of TSOs (one single TSO for all zones or several zonal TSOs), the way of managing congestion besides grid expansion (redispatch and g-component) and different alternatives for cross-border capacity allocation (NTC vs. flow-based market coupling). We consider Net Transfer Capacity (NTC) and flow-based market coupling as cross-border capacity allocation algorithms because they have been used extensively in the European context (see, e.g., Glachant (2010)). NTCs are a rather simplified version

[3]Such a model is typically applied to represent the optimization problem of a social planner or an integrated firm optimizing the entire electricity system, including generation and transmission.

[4]One main difference in our model is the assumption of an inelastic demand which was necessary to formulate and solve the model as a linear program. We will elaborate on this issue in Section 4.2.1.

[5]Efficient in this context means that the TSO(s) are perfectly regulated to expand and operate the grid at minimal costs.

of cross-border trade restrictions, widely neglecting the physical properties of the grid as well as its time-varying characteristics. Under flow-based market coupling, cross-border transmission capacities are calculated taking into account the impact of (cross-border) line flows on every line in the system (e.g., Oggioni and Smeers (2013)), hence providing a much better consideration of the physical grid properties which is crucially important in case of meshed networks. As a consequence, more capacity can generally be offered for trading between markets, and a better usage of existing infrastructures is achieved. The analyzed settings are summarized in the following Table 4.1.

	Market area and coupling	TSO scope	TSO measures
I	Nodal markets	One TSO	Grid expansion
II - NTC	Zonal markets, NTC-based coupling	One TSO	Grid expansion, redispatch
II - FB	Zonal markets, Flow-based coupling	One TSO	Grid expansion, redispatch
III - NTC	Zonal markets, NTC-based coupling	Zonal TSOs	Grid expansion, redispatch
III - FB	Zonal markets, Flow-based coupling	Zonal TSOs	Grid expansion, redispatch
IV	Zonal markets	Zonal TSOs	Grid expansion, g-component

Table 4.1: Analyzed congestion management designs

Noticeably, despite the separated generation and transmission levels, agents are in all settings assumed to act rationally and simultaneously while taking into account the activities of the other stage.[6] Furthermore, we assume perfect competition on the generation stage and perfect regulation of the TSOs in the sense that TSO activities are aligned with social objectives. TSOs as well as generators are price taking, with an independent institution (e.g., the power exchange) being responsible for coordinating the activities of the different participating agents and for market clearing.[7] Importantly, while in the first-best design all information is available to all agents, alternative congestion management designs may induce an adverse (e.g., aggregated) availability of information. The solution of the problem is an intertemporal equilibrium which is unique under the assumption of convex functions. We will thoroughly discuss issues of convexity in the context of the numerical implementation in Section 4.3. Noticeably, with the above assumptions, our general modeling approach can be thought of as a way to compare today's and future performances of different congestion management designs based on today's state of the system, today's information horizon, as well as rational expectations about future developments and resulting

[6] I.e., sequential moving and issues of timing are not considered.

[7] By assuming perfect competition and an inelastic demand, we are able to treat the general problem as a cost minimization problem. This assumption is commonly applied for formulation of electricity markets in the literature. An alternative formulation with a welfare maximization approach would be possible, but wouldn't impact the general conclusions.

investment decisions.[8]

For the sake of readability (and in contrast to the large-scale application presented in Section 4.4), we make some simplifications in the theoretical framework: dispatch decisions are realized in several points of time, but invest decisions are undertaken only once. Furthermore, we neglect different types of generation technologies that may be available at a node. This simplification does not change any of the conclusions drawn from the theoretical formulation.[9]

For developing the economic modeling framework in the following subsections, we will deploy parameters, variables and sets as depicted in Table 4.2 in the Appendix 4.6.

4.2.1 Setting I – First-Best: Nodal pricing with one TSO

By design, nodal pricing avoids any inefficiency by covering and exchanging all information present within the problem – leading to a welfare-optimal electricity system. It hence represents the first-best setting in our analysis of different congestion management designs. With the assumption of a social planner or perfect competition and regulation, nodal prices can be derived from locational marginal costs (of generation and capacity) in a market clearing that implicitly considers the physical properties of the electricity network (specifically, loop flows). Abstracting from economies of scale and lumpiness of investment, it can be shown that an efficient and unique equilibrium exists under nodal prices (Caramanis (1982), Joskow and Tirole (2005), Rious et al. (2009)). In line with these findings, we assume constant marginal grid costs as well as continuous generation and transmission expansion.[10] Another assumption in our formulation is an inelastic (yet time-varying) demand. The reason for assuming an inelastic demand is mainly triggered by the excessive computational burden that would be induced by an elastic demand in the numerical solution approach (an inelastic demand allows us to formulate and solve the model as a linear instead of a non-linear program). As a drawback, the assumption of an inelastic demand differs from the formulation in Schweppe et al. (1988) and

[8]In our numerical application, this approach is supplemented with discounted future cash flows. See Section 4.4 for further details.

[9]To include multiple instances in time for investments, the formulation could easily be adapted by adding an index to all parameters, variables and equations related to installed capacities (generation and transmission). In the same vein, an additional index could be inserted to account for different types of generation technologies.

[10]This assumption is certainly more critical for transmission investments which require a certain magnitude to be realized. Generation investment might also be lumpy, but smaller plant sizes are possible.

leads to the artifact that demand can never set the price. However, scarcity rents to cover capacity costs are still possible under perfect information and competition (including entry and exit of generators). For instance, consider the bid of a peak load plant during a single peak load hour when it is dispatched and pivotal. The bid will consist of the variable costs plus the long-term marginal costs of the capacity. If the bid was lower, the peak load plant would leave the market due to an overall loss. If the bid was higher, another peak load plant would enter the market due to the possibility of making a profit. This forces the peak load plant to bid its true variable plus marginal capacity costs. Once accepted, this bid can be interpreted as the resulting market prices under capacity scarcity. Lastly, note that off-peak hours can also have capacity components in prices if there is a diversified mix of generation technologies, characterized by different cost structures.

The following optimization problem *P1* is similar to the formulation of an integrated problem for operating generation and transmission as in Schweppe et al. (1988), except for the major change of demand being inelastic. In this formulation, a social planner or an integrated firm minimizes total system costs of the operation and investment of generation and transmission.

P1 Integrated Problem

$$\min_{\overline{G}_i, G_{i,t}, T_{i,j,t}, \overline{P}_{i,j}} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} \tag{4.1a}$$

$$\text{s.t.} \quad G_{i,t} - \sum_j T_{i,j,t} = d_{i,t} \qquad \forall i,t \qquad | \lambda_{i,t} \tag{4.1b}$$

$$G_{i,t} \leq \overline{G}_i \qquad \forall i,t \tag{4.1c}$$

$$|T_{i,j,t}| = |P_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, d_{k,t})| \leq \overline{P}_{i,j} \qquad \forall i,j,t \qquad | \kappa_{i,j,t} \tag{4.1d}$$

$$T_{i,j,t} = -T_{j,i,t} \qquad \forall i,j,t \tag{4.1e}$$

Indices i, j, k, l represent nodes in the system. Generation $G_{i,t}$, generation capacity $\overline{G}_i$, trade $T_{i,j,t}$ and transmission capacity $\overline{P}_{i,j}$ are optimization variables. Additional capacities can be installed at the costs of δ_i for generation and $\mu_{i,j}$ for transmission. Nodal prices are derived from the dual variables $\lambda_{i,t}$ of the equilibrium constraint which states that the demand level $d_{i,t}$ at node i can be either satisfied by generation at the same node or trade between nodes (Equation (4.1b)). Equations (4.1c) and (4.1d) mirror that generation is restricted by installed generation capacities, and physical flows by installed transmission capacities. Furthermore, trades from node i to node j are necessarily equal to negative trades from node j to node i (Equation

(4.1e)). As the market clearing fully accounts for the transmission network in the nodal pricing regime, trade between adjacent nodes is equal to physical flows on the respective line, i.e., $T_{i,j,t} = P_{i,j,t}$ (Equation (4.1d)).

Load flows on transmission lines are based on Kirchhoff's law, which we represent based on a linearized load flow approach.[11] Thereby, flows are impacted by generation ($G_{k,t}$) and demand ($d_{k,t}$), i.e., power balances of all nodes in the system, as well as by the physical properties of the transmission system, represented by installed transmission capacities $\overline{P}_{k,l}$. Thus, there is a functional dependency of flows and trades on generation, demand, and line capacities throughout the system, i.e., $T_{i,j,t} = T_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, d_{k,t})$.

As has been shown, e.g., by Conejo et al. (2006), an integrated optimization problem can be decomposed into subproblems which are solved simultaneously, while still representing the same overall situation and corresponding optimal solution. In our application, we take advantage of this possibility to represent separated generation and transmission levels in problem *P1'*. The generation stage *P1'a* states the market clearing of supply and demand while respecting generation capacity constraints. As in *P1*, the same nodal prices are obtained by the dual variable $\lambda_{i,t}$ of the equilibrium constraint (4.2b). Instead of including the explicit grid expansion costs in the cost minimization, the objective function of the generation stage now contains transmission costs which assign transmission prices $\kappa_{i,j,t}$ to trade flows between two nodes i and j. These prices are derived from the dual variable of the equilibrium constraint on the transmission stage (Equation (4.2g)). We assume that the TSO is perfectly regulated to minimize costs of grid extensions accounting for the physical feasibility of the market clearing as determined on the generation stage while considering all grid flows and related costs (problem *P1'b*). As trade is a function of $\overline{P}$, which in turn is the decision variable in the transmission problem, the market clearing conditions need to reoccur in the transmission problem.

[11] We will use the PTDF approach shown in Appendix 4.6 in our numerical implementation in Section 4.3, as this enables a linearization of the generally non-linear load flow problem, given a fixed transmission network (cf. Hagspiel et al. (2014)).

P1'a *Generation*

$$\min_{\overline{G}_i, G_{i,t}, T_{i,j,t}} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{i,j,t} \kappa_{i,j,t} T_{i,j,t} \tag{4.2a}$$

$$\text{s.t.} \quad G_{i,t} - \sum_j T_{i,j,t} = d_{i,t} \qquad \forall i,t \qquad | \lambda_{i,t} \tag{4.2b}$$

$$G_{i,t} \leq \overline{G}_i \qquad \forall i,t \tag{4.2c}$$

$$T_{i,j,t} = -T_{j,i,t} \qquad \forall i,j,t \tag{4.2d}$$

P1'b *Transmission*

$$\min_{\overline{P}_{i,j}} \quad Y = \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} \tag{4.2e}$$

$$\text{s.t.} \quad G_{i,t} - \sum_j T_{i,j,t} = d_{i,t} \qquad \forall i,t \tag{4.2f}$$

$$|T_{i,j,t}| = |P_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, d_{k,t})| \leq \overline{P}_{i,j} \qquad \forall i,j,t \qquad | \kappa_{i,j,t} \tag{4.2g}$$

$$T_{i,j,t} = -T_{j,i,t} \qquad \forall i,j,t \tag{4.2h}$$

As can be seen, all terms of *P1* reappear in *P1'*, however, allocated to two separated levels. Mathematically, the equivalence of *P1* and *P1'* is shown in Appendix 4.6, where the first order conditions of both formulations are compared.

4.2.2 Setting II: coupled zonal markets with one TSO and zonal redispatch

In zonal markets, a number of nodes are aggregated to a market with a uniform price. In contrast to nodal pricing, coupled zonal markets only consider aggregated cross-border capacities between market zones during market clearing (instead of all individual grid elements). Thus, the obtained prices for generation do not reflect the true total costs of the entire grid infrastructure. This is due to the fact that zonal prices only reflect those cross-border capacities that limit activities between zonal markets. Cross-border capacities can be allocated in different ways. We consider Net Transfer Capacity (NTC) and the more sophisticated flow-based market coupling as cross-border capacity allocation algorithms (see Oggioni and Smeers (2013)). Under the latter regime, more capacity can generally be offered for trading between markets, and a better usage of existing infrastructures is achieved.

Because intra-zonal congestion is neglected in the zonal market-clearing, it needs to be resolved in a subsequent step by the TSO. Besides the expansion of grid capacities, in *Setting II* we provide the TSO with the opportunity of zonal redispatch. The TSO may instruct generators located behind the bottleneck to increase production (positive redispatch), and another generator before the bottleneck to reduce production (negative redispatch).[12] We assume here a perfectly discriminating redispatch: the TSO pays generators that have to increase their production their variable costs, and in turn receives the avoided variable costs of generators that reduce their supply. As the generator with positive redispatch was not part of the original dispatch, it necessarily has higher variable costs than the generator that reduces supply. Thus, the TSO has to bear additional costs that are caused by the redispatch which amount to the difference between the variable costs of the redispatched entities. Assuming further that the TSO has perfect information about the variable costs of the generating firms, redispatch measures of the TSO have no impact on investment decisions of generating firms as the originally dispatched generation capacity is still able to cover capital costs from the spot market result. Hence, additional costs for the economy are induced by inefficient investment decisions of those generators that are not aligned with the overall system optimum due to missing locational price signals.

In the formulation of problem *P2a* zonal pricing is represented by the zonal market indices n, m, each containing one or several nodes i. Market clearing, depicted by the equilibrium Equation (4.3f), now takes place on zonal instead of nodal markets. The corresponding dual variable $\lambda_{m,t}$ represents zonal prices, which do not include any grid costs except for cross-border capacities. This is indicated by the term $\sum_{m,n,t} \kappa_{m,n,t} T_{m,n,t}$ instead of the nodal formulation (with $\kappa_{i,j,t}$) above. Transmission prices are determined on the transmission stage (Equation (4.3j)). However, contrary to nodal pricing, these prices are calculated based on some regulatory rule (e.g., NTC or FB) and are thus inherently incomplete since they do not represent real grid scarcities.[13] In addition to grid expansion, the TSO may relieve intra-zonal congestion and optimize the situation by means of redispatch measures $R_{i,t}$ at costs of $\gamma_{i,t} R_{i,t}$.

[12]Redispatch is always feasible due to the fact that the TSO can foresee congestion and hence, counteract by expanding line capacities.

[13]Note that the duality of the problem would also allow for an alternative formulation of the cross-border transmission constraint by means of quantity constraints instead of prices. Hence, the cost of transmission in the objective function of the generation stage ($\sum_{m,n,t} \kappa_{m,n,t} T_{m,n,t}$) would disappear and an additional constraint for trading would be implemented ($|T_{m,n,t}| \leq C_{m,n}, \forall m, n, t$). The restriction of trading volumes $C_{m,n,t}$ would be calculated on the transmission stage *P2b* via a constraint $C_{m,n} = h(\overline{P}_{i,j})$ instead of the prices $\kappa_{m,n,t}$. These prices would then be the dual variable of the volume constraint on the generation stage, and necessarily coincide with $\kappa_{m,n,t}$.

P2a Generation

$$\min_{\overline{G}_i, G_{i,t}, T_{m,n,t}} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{m,n,t} \kappa_{m,n,t} T_{m,n,t} \tag{4.3a}$$

$$\text{s.t.} \quad \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_n T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \quad \forall m,t \quad | \, \lambda_{m,t} \tag{4.3b}$$

$$G_{i,t} \leq \overline{G}_i \quad \forall i,t \tag{4.3c}$$

$$T_{m,n,t} = -T_{n,m,t} \quad \forall m,n,t \tag{4.3d}$$

P2b Transmission

$$\min_{\overline{P}_{i,j}, R_{i,t}} \quad Y = \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} + \sum_{i,t} \gamma_{i,t} R_{i,t} \tag{4.3e}$$

$$\text{s.t.} \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_n T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \quad \forall m,t \tag{4.3f}$$

$$|T_{i,j,t}| = |P_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, R_{k,t}, d_{k,t})| \leq \overline{P}_{i,j} \quad \forall i,j,t \quad | \, \kappa_{i,j,t} \tag{4.3g}$$

$$\sum_{i \in \mathbf{I_m}} R_{i,t} = 0 \quad \forall m,t \tag{4.3h}$$

$$0 \leq G_{i,t} + R_{i,t} \leq \overline{G}_i \quad \forall i,t \tag{4.3i}$$

$$\kappa_{m,n,t} = g(\kappa_{i,j,t}) \tag{4.3j}$$

$$T_{m,n,t} = -T_{n,m,t} \quad \forall m,n,t \tag{4.3k}$$

The following two examples illustrate the fundamental differences between *Setting I* and *II*.

Example for 2 nodes and 2 markets: If the electricity system consists of 2 nodes and 2 markets (Figure 4.1, left hand side), *Setting I* and *II* are identical: There is only one element $i \in \mathbf{I_m}$, such that Equation 4.3h fixes variable R_i to zero. Equation 4.3i is then no longer relevant, and the cost term of redispatch in the objective function ($\sum_{i,t} \gamma_{i,t} R_{i,t}$) becomes zero. The only difference remaining between *1'b P2b* is then Equation 4.3j. However, due to $I = M$, it follows that $\kappa_{m,n,t} = \kappa_{i,j,t}$, which, inserted on the generation level, yields equivalence of problems *P1'* and problem *P2* for the chosen example.

Example for 3 nodes and 2 markets: Figure 4.1, right hand side, shows an electricity system consisting of two markets m and n, where m includes one node (1) and n two nodes $(2,3)$ at a point in time t. Function g for calculating the transmission price $\kappa_{m,n,t}$ (Equation (4.3j)) between the markets has to be defined, e.g., by averaging the single line prices $\kappa_{m,n,t} = (\kappa_{1,2,t} + \kappa_{1,3,t})/2$. Still, the TSO cannot supply the locational fully differentiated prices $\kappa_{1,2,t}$, $\kappa_{1,3,t}$ and $\kappa_{2,3,t}$ to the market, and hence, efficient allocation of investments is (partly) achieved *between* the markets, but not *within* the markets. Redispatch does not fully solve this problem, because it is revenue-neutral and does not affect the investment decision.

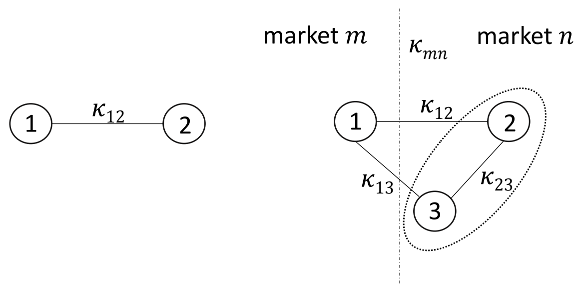


Figure 4.1: Two simple examples. Left: 2 nodes, 2 markets. Right: 3 nodes, 2 markets

Overall, *Settings I* and *II* differ in the way grid costs are reflected on the generation stage. Specifically, *Setting II* lacks locational differentiated prices, thus impeding efficient price signals $\kappa_{i,j,t}$ for the generation stage. Of course, the level of inefficiency depends substantially on the regulatory rule determining the calculation of prices based on a specification of function $g(\kappa_{i,j,t})$. In general, it is clear that the closer the specification of g reflects real-time conditions and the more it enables the full usage of existing grid infrastructures, the more efficiently the general problem will be solved. While we limit our analysis in this section to this general finding, we will discuss two possible specifications often implemented in practice (NTC and flow-based market coupling) in the empirical example in Section 4.4. Given the inefficiency induced by the specification of function g, the question remains whether and how redispatch measures may help to relieve the problem. We find that the resulting inefficiency cannot be fully resolved by redispatch because the latter remains a zonal measure (Equation (4.3h)). Hence, the TSO cannot induce an efficient usage of generation and transmission across zonal borders. Furthermore, investments into generation capacities are not influenced by redispatch and only zonal prices as well as their costs are considered.[14] Hence, the setting lacks locational signals for efficient generation investments within zonal markets.

[14]For obtaining a unique equilibrium we assume that costs differ over all nodes, such that decisions for generation and investments are unambiguously ordered.

4.2.3 Setting III: coupled zonal markets with zonal TSOs and zonal redispatch

In this setting, we consider zonal markets with zonal TSOs being responsible for grid expansion as well as a zonal redispatch. Thus, the problem on the generation stage remains exactly the same as in the previous setting (i.e., $P3a = P2a$). However, the transmission problem changes, such that now multiple zonal TSOs are considered. Each TSO solves its own optimization problem according to the national regulatory regime (in our case corresponding to a cost-minimization within the zones). Formally, problem *P3b*, now consists of multiple separate optimization problems for each zonal TSO, with the objective to minimize costs from zonal grid as well as from zonal redispatch measures. However, cross-border line capacities are also taken into account. As these are by definition located within the jurisdiction of two adjacent market areas, the two corresponding TSOs have to negotiate about the extension of these cross-border capacities. In fact, cross-border capacities built by two different TSOs may be seen as a Leontief production function, due to the fact that the line capacities built on each side are perfect complements. Corresponding costs from inter-zonal grid extensions are assumed to be shared among the TSOs. Due to the fact that situations may arise where an agreement on specific cross-border lines between neighboring TSOs cannot be reached (which would imply that an equilibrium solution cannot be found), we assume the implementation of a regulatory rule that ensures the acceptance of a unique price for each cross-border line by both of the neighboring TSOs. For instance, the regulatory rule may be specified such that both TSOs are obliged to accept the higher price offer, or, equivalently, the lower of the two capacities offered for the specific cross-border line.

As a consequence, grid capacities, especially cross-border capacities, are extended inefficiently as they do not result from an optimization of the entire grid infrastructure. In addition – just as in the previous setting – inefficient investment incentives for generation and grid capacities are caused by the lack of locational differentiated prices. Hence, overall, system outcomes in *Setting III* must be inferior or at most equal to those of *Setting II*.[15]

The mathematical program as well as further technical details of *Setting III* can be found in the Appendix 4.6.

[15]The only mathematical difference of problem *P3b* compared to *P2b* is that the transmission level is partitioned into several optimization problems that are solved separately from each other. Hence, compared to problem *P2b* where the transmission level is solved comprehensively, this represents a more restrictive problem that must be inferior (or at most equal) to the one of *P3b*.

4.2.4 Setting IV: coupled zonal markets with zonal TSOs and generator component

In this last setting, we again consider coupled zonal markets with zonal TSOs. However, instead of having the possibility to perform a zonal redispatch (as in *Setting III*), zonal TSOs may now determine local, time-varying prices for generators, i.e., a g-component, at each node belonging to its zone to cope with intra-zonal congestion. A g-component charges grid costs directly to generators in order to avoid congestion in the market clearing process reflecting the impact of generators on the grid at each node and each instant of time. Thus, grid costs are being transferred to the generating firms which consider them in their investment and dispatch decision. In other words, TSOs are able to provide locationally differentiated prices (and hence, generation and investment incentives) for generators within their zone. Noticeably, we do not consider an international g-component here as this would yield the same results as a nodal pricing regime due to generators considering the full set of information concerning grid costs. However, two frictions that may cause an inefficient outcome of this setting remain. When determining nodal g-components, zonal TSOs only consider grid infrastructures within their zone, and not within the entire system. Furthermore, as in *Setting III*, the desired expansion of cross-border lines, which is here assumed to be solved by some regulatory rule ensuring successful negotiation, may deviate between/across neighboring TSOs.

The mathematical program as well as further technical details of *Setting IV* can be found in the Appendix 4.6.

4.3 Numerical solution approach

Our approach to numerically solve the problem depicted in the previous section builds on the concept of decomposition. In fact, it follows the approach already applied in the context of *Setting I* (Section 4.2.1), where we decomposed the integrated problem into two separate levels that are solved simultaneously and showed that they can – in economic terms – be interpreted as generation and transmission levels. Algorithmically, according to Benders (1962), decomposition techniques can be applied to optimization problems with a decomposable structure that can be advantageously exploited. The idea of decomposition generally consists of splitting the optimization problem into a master and one or several subproblems that are solved iteratively. For the problem we are dealing with, namely the simultaneous optimization of generation and grid infrastructures under different congestion management designs and a varying number of TSOs, decomposing the overall problem entails two major advantages: First, the decomposition allows to easily implement variations of the generation and transmission levels including the underlying congestion management design. Hence, the model can be flexibly adjusted to represent the various settings described in the previous section. Second, the iterative nature of the solution process resulting from the decomposition allows to readily update PTDF matrices every time changes in the grid infrastructure have been made, according to Equation (4.15) and the PTDF calculation procedure presented in Section 4.6. This iterative update of the grid properties, as applied in Hagspiel et al. (2014) and Ozdemir et al. (2015), successively linearizes the non-linear optimization problem to ensure a consistent representation of generally non-linear grid properties, and allows for solving a corresponding linear problem.[16] In turn, linear problems can be solved effectively for global optima using standard techniques, such as the Simplex algorithm (e.g., Murty (1983)).

Even though the PTDF update ties in nicely with the iterative solution of the decomposed problem, it also imposes a particular challenge stemming from the non-linearity in the PTDF calculation (see Appendix 4.6). Specifically, despite the successive linearization and iterative solution, the non-linearity of the transmission expansion problem remains. Hence, neither the existence and uniqueness of a global optimum of the problem, nor the convergence of the solution algorithm can generally be guaranteed (e.g., Bazaraa et al. (2006)). This would change, however, if the problem was convex. Then, there would be a unique equilibrium, corresponding to a global optimum. Furthermore, deploying a Benders-type decomposition, the algorithm would preserve convexity and guarantee that the iterative solution converges towards this global optimum (Benders (1962) and, e.g., Conejo et al. (2006) for a general overview). Unfortunately, to the best of our knowledge, a formal proof of the (non-)convexity of the transmission expansion problem is still missing. Meanwhile, it would also be beyond the scope of this paper to approach this challenging problem. As an alternative, we build on numerical experience that has been gained by two papers that are closely related to ours in terms of the algorithmic approach: The analysis in Hagspiel et al. (2014) is closest to our application as they deploy the same successive PTDF update to co-optimize generation and transmission assets (including operation and investment). They show that the algorithm converges in a large number of configurations, including small analytically tractable test sys-

[16]Accordingly, in our model $PTDF$ is depicted as a parameter that is updated in each iteration instead of a variable.

tems as well as large-scale applications. Furthermore, they do not detect issues of multiple equilibria in their analysis. In a very similar vein, Ozdemir et al. (2015) develop a methodology based on successive linear programming and Gauss-Seidel iteration to jointly optimize transmission and generation capacities. They report that even though they cannot guarantee convergence or global optimality either, their approach shows good performance. In the course of preparing the results presented in this paper, we were able to confirm the above findings in several model runs where we varied starting values over a broad range and did not find evidence neither against convergence nor against uniqueness of our optimum. Hence, even though not guaranteed, empirical evidence indicates that we are facing a numerical problem that we are able to reliably solve with our algorithm while converging towards an optimal solution. In our application, the obtained solution represents an intertemporal equilibrium without uncertainty. Interestingly, in economic terms, the iterative algorithm to solve the decomposed problem can be readily interpreted as a price adjustment by a Walrasian auctioneer, also know as tatonnement procedure (e.g., Boyd et al. (2008)).

With some minor modifications, we can directly follow the (economically intuitive) formalization developed in the previous section and implement separate optimization problems representing the different tasks of generation and grid as well as the various settings (*I-IV*). We follow the Benders decomposition approach described in Conejo et al. (2006), while considering the transmission capacities as complicating variables. We define the generation stage as the master problem, whereas the subproblem covers the transmission stage.[17] The principle idea of the solution algorithm is to solve the simultaneous generation and transmission stage problem iteratively, i.e., in a loop that runs as long as some convergence criterion is reached. In this process, optimized variables and marginal values are exchanged between the separated generation and grid levels reflecting the configuration of congestion management and TSO organization. For the settings described in the previous section, prices, which are iterated and thus adjusted, differ with respect to the information they contain and hence determine to which degree efficiency can be reached. Compared to nodal pricing (*Setting I*), the other settings provide prices or products that describe the underlying problem only incompletely – and hence, entail an inefficient outcome.

The numerical algorithm to solve the nodal pricing model is sketched below. Parameters that save levels of optimal variables for usage in the respective other stage

[17]Noticeably, the model could be inverted such that the master problem represents the grid sector which would, however, not change any of the results obtained.

are indicated by $^{(\cdot)}$. It should be noticed that for the sake of comprehensibility, we still represent a simplified version of a more complete power system model that would need to account for multiple instances in time for investments, multiple generation technologies, etc. However, the extension is straightforward and does not change the principle approach depicted here.

Information passed from the transmission to the generation stage is captured by α, for which a benders cut (lower bound constraint) is added in each iteration u up to the current iteration v (Equation (4.4e)). This benders cut consists of total grid costs $Y^{(u)}$ as well as the marginal costs each unit of trade $T_{i,j,t}$ is causing in the grid per node, denoted by $\kappa^{(u)}_{i,j,t}$. Both pieces of information are provided in the highest possible temporal and spatial resolution. As these components occur in the objective function of the generation stage (via α), the optimization will try to avoid the additional costs it is causing on the transmission stage, e.g., by moving power plant investments to alternative locations. The variable α is needed to correctly account for the impact of the transmission on the generation stage. On the transmission stage, the TSO is coping with the exchange (i.e., trade) of power stemming from the dispatch situation delivered by the master problem, thereby determining the marginal costs the trade is causing on the transmission stage, i.e., κ. Power flows are calculated by linearized load-flow equations represented by PTDF matrices mapping. The TSO then expands the grid such that it supports the emerging line flows at minimal costs.

$v = 1$; convergence=false

While(convergence=false) {

Master problem: generation

$$\min_{\overline{G}_i, G_{i,t}, T_{i,j,t}, \alpha} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \alpha \tag{4.4a}$$

$$\text{s.t.} \quad G_{i,t} - \sum_j T_{i,j,t} = d_{i,t} \quad \forall i, t \tag{4.4b}$$

$$G_{i,t} \le \overline{G}_i \quad \forall i, t \tag{4.4c}$$

$$T_{i,j,t} = -T_{j,i,t} \quad \forall i, j, t \tag{4.4d}$$

$$Y^{(u)} + \sum_{i,j} \kappa^{(u)}_{i,j,t} \cdot (T_{i,j,t} - T^{(u)}_{i,j,t}) \le \alpha \quad \forall u = 1, ..., v-1 | v > 1 \tag{4.4e}$$

$$G^{(v)}_{i,t} = \text{Optimal value of } G_{i,t} \quad \forall i, t \tag{4.4f}$$

Sub-problem: transmission

$$\min_{\overline{P}_{i,j},T_{i,j,t}} \quad Y = \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} \tag{4.4g}$$

$$\text{s.t.} \quad \left|P_{i,j,t}\right| = \left|\sum_k PTDF^{(v)}_{k,i,j} \cdot (G^{(v)}_{k,t} - d_{k,t})\right| \leq \overline{P}_{i,j} \quad \forall i,j,t \quad | \kappa^{(v)}_{i,j,t} \tag{4.4h}$$

$$Y^{(v)} = \text{Optimal value of } Y \tag{4.4i}$$

$$PTDF^{(v)} = \text{PTDF matrix calculated based on } \overline{P}_{i,j} \tag{4.4j}$$

if(convergence criterion < threshold; convergence=true)

$v = v + 1$

};

As regards the representation of settings *II-IV*, only very few modifications are needed compared to the nodal pricing regime (*Setting I*). The numerical algorithmic implementation of the various settings and modifications directly follows the procedure discussed in Section 4.2 and is thus not discussed again in detail here.[18]

4.4 Large-scale application

In this section, we apply the previously developed methodology to a detailed representation of the power sector in the Central Western European (CWE) region up to the year 2030. The application demonstrates the suitability of the modeling framework for large-scale problems and allows to assess and quantify the welfare losses in the considered region caused by different congestion management designs.

Given its historical, current and foreseen future development, the CWE region appears to be a particularly timely and relevant case study for different congestion management designs. In order to increase the market integration of European electricity markets towards an internal energy market, the European Union (EU) has declared the coupling of European electricity markets, which are organized in uniform price zones, an important stepping stone (see, e.g., Glachant (2010)). As for the cross-border capacity allocation, after a phase of NTC (Net Transfer Capacities) based market coupling, the CWE region is currently implementing a flow-based mar-

[18] Nevertheless, for the sake of completeness and reproducibility, we have included one more complete model formulation illustrating the main differences of the other settings in Appendix 4.6.

ket coupling which is expected to increase the efficiency of the utilization of transmission capacities as well as overall social welfare (Capacity Allocating Service Company (2014)). Even though nodal pricing regimes have often been discussed for the European power sector (see, e.g., Ehrenmann and Smeers (2005) or Oggioni and Smeers (2012)), it can be expected that uniform price zones that correspond to national borders will remain. In fact, zonal markets coupled via a flow-based algorithm have been declared the target model for the European power sector (ACER (2014)).

In each zonal market, the respective zonal (i.e., national) TSO is responsible for the transmission network. Thereby, TSOs are organized and regulated on a national level, such that they can be assumed to care mainly about grid operation and expansion planning within their own jurisdiction. Although there are an umbrella organization (ENTSO-E) and coordinated actions, such as the (non-binding) European Ten-Year-Network-Development-Plan (TYNDP), the incentives of the national regulatory regime to intensify cross-border action might fall short of effectiveness. At the same time, Europe is heavily engaged in the large-scale deployment of renewable energies, hence causing fundamental changes in the supply structure. Generation is now often built with respect to the availability of primary renewable resources, i.e., wind and solar irradiation, and not necessarily close to load. This implies that the current grid infrastructure is partly no longer suitable and needs to be substantially redesigned, rendering an efficient congestion management even more important than before.

4.4.1 Model configuration and assumptions

The applied model for the generation stage belongs to the class of partial equilibrium models that aim at determining the cost-optimal electricity supply to customers by means of dispatch and investments decisions based on a large number of technological options for generation. As power systems are typically large and complex, these models are commonly set up as a linear optimization problem which can efficiently be solved. Our model is an extended version of the linear long-term investment and dispatch model for conventional, renewable, storage and transmission technologies as presented in Richter (2011) and applied in, e.g., Jägemann et al. (2013) or Hagspiel et al. (2014). In contrast to previous versions, the CWE region, i.e., Belgium, France, Germany, Luxembourg and Netherlands, is considered with a high spatial (i.e., nodal) resolution. In order to account for exchanges with neighboring countries, additional regions are defined, but at an aggregated level: Southern Europe (Austria, Italy and Switzerland), South-West Europe (Portugal and Spain), North-West Europe (Ireland and UK), Northern Europe (Denmark, Finland, Norway and Sweden), and Eastern Europe (Czech Republic, Hungary, Poland, Slovakia and Slovenia). Figure 4.2 depicts the regional coverage and aggregation as they are represented in the model. In total, the model represents 70 nodes (or markets) and 174 power lines (AC and DC).

The model determines a possible path of how installed capacities will develop and how they are operated in the future assuming that electricity markets will achieve the cost-minimizing mix of different technologies which is obtained under perfect competition and the absence of market failures and distortions. Among a number of techno-economic constraints, e.g., supply coverage or investment decisions, the model also includes a number of politically implied constraints: nuclear power is phased-out where decided so, and then only allowed in countries already using it; a CO_2-Quota is implemented corresponding to currently discussed targets for the European energy sector, i.e., 20% reduction with respect to 1990 levels in 2020, and 40% in 2030 (European Commission (2013, 2014)); nation-specific 2020 targets for renewable energy sources are assumed to be reached until 2020 whereas from 2020 onwards there are no further specific renewable energy targets. At the same time, endogenous investments into renewable energy technologies are always possible.

The utilized model for the transmission stage is based on PTDF matrices which are calculated using a detailed European power flow model developed by Energynautics (see Ackermann et al. (2013) for a detailed model description). The number of nodes (70) corresponds to the nodal markets implemented in the generation market model and represents generation and load centers within Europe at an aggregated level. Those nodes are connected by 174 high voltage alternating current (AC) lines (220 and 380kV) as well as high voltage direct current (HVDC) lines. Even though the model is generally built for AC load flow calculations, it is here used to determine PTDF matrices for different grid expansion levels. Details on how the PTDF matrices are calculated can be found in Appendix 4.6.

As a starting point, the optimization takes the situation of the year 2011, based on a detailed database developed at the Institute of Energy Economics at the University of Cologne which in turn is largely based on the Platts WEPP Database (Platts (2009)). From these starting conditions, the development for the years 2020 and 2030 is optimized.[19] As for the temporal resolution, we represent the operational phase by nine typical days representing weekdays and weekend as well as variations

[19]Technically, we implement the optimization routine up to 2050, but only report results until 2030. This is necessary to avoid problematic results at the end of the optimization timeframe.

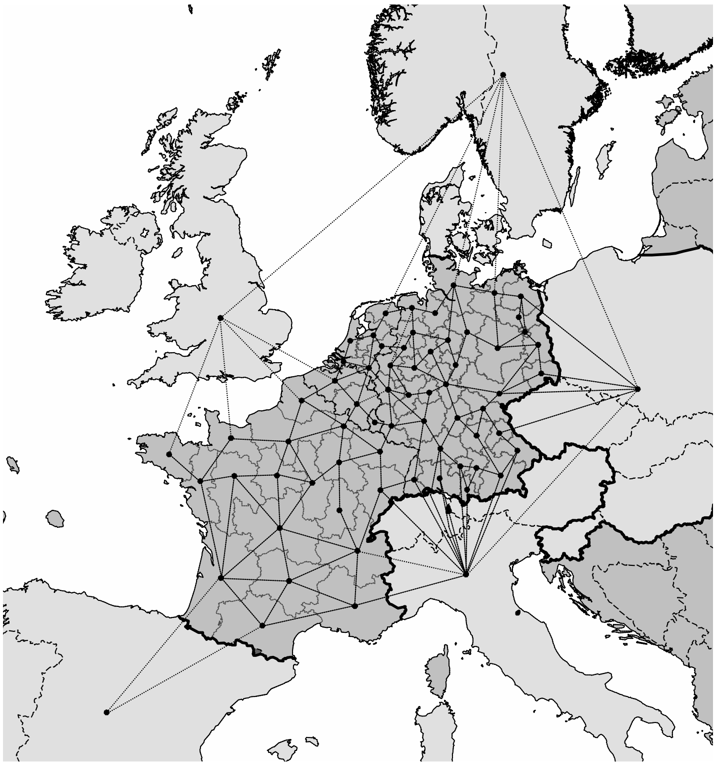


Figure 4.2: Representation of the CWE and neighboring regions in the model

in and interdependencies between demand and power from solar and wind. One of the typical days represents an extreme day during the week with peak demand and low supply from wind and solar. Specific numerical assumptions for the generation and transmission model can be found in the 4.6.

As in *Settings II-IV* zonal markets are being considered, assumptions about the cross-border price function $g(\kappa_{i,j,t})$ are necessary. For the NTC-based coupling of market zones, we define function $g(\kappa_{i,j,t}) = 1.43 \cdot \frac{\kappa_{i,j,t}\overline{P}_{i,j}}{\sum_{i,j}\overline{P}_{i,j}} \ \forall i,j \in \mathbf{I_{m,cb}}$ for each market border. The function consists of the weighted average of cross-border line marginals multiplied by a security margin. The security margin is the inverse of the ratio of NTC capacity to technical line capacity and has been derived heuristically by comparing currently installed cross-border grid capacities with NTC values reported by ENTSO-E for the CWE region. For flow-based market coupling, we set this security margin to one, in order to account for enhanced cross-border capacities provided to the power market.[20] In the case of zonal TSOs, we have made the following two assumptions: Differing interest of TSOs regarding cross-border line extensions are aligned by taking the smaller one of the two expansion levels.[21] The costs of cross-border lines are shared half-half by the two TSOs concerned, i.e., $\sigma_{i,j} = 0.5$.

4.4.2 Results and discussion

As usual in a Benders decomposition, we trace convergence based on the difference between an upper (i.e., the objective value of the integrated problem with solution values of the current iteration) and a lower bound (i.e., the objective of the master problem with the same solution values). We found that all settings undershoot a convergence threshold of 2.5% within 20 to 60 iterations (corresponding to a solution time of 2 to 7 days).[22] For practical reasons, we let all settings solve for one week and – after having double-checked that the convergence threshold of 2.5 % is met – take the last iteration to obtain our final results. The convergence threshold is chosen to keep the solution process computationally treatable, but is also based on empirical observations as well as expected convergence behavior. In fact, a lower convergence criterion increases computational time significantly, while further improvements on the objective value and optimized capacities are hardly observable.

To illustrate the convergent behavior of our problem, Figure 4.3, left hand side, shows the development of the optimality error (relative difference between the upper and lower bound of the optimization), along with the (absolute) rate of change of the lower bound obtained during the iterative solution of the nodal pricing setting. The lower bound is observed to change only slightly, reaching change rates smaller than 0.01% after some 40 iterations. Moreover, as can be derived from the interpolation curves presented in Figure 4.3, left hand side, the relative error decreases at much faster rates with a ratio of approximately 200 for an estimated exponential

[20] Of course, this is just a simple representation of the cross-border capacity allocation. However, a more detailed representation is rather complex and would go beyond the scope of this paper. For more sophisticated models of flow-based capacity allocation, the reader is referred to Kurzidem (2010).

[21] Equation (4.24m) in Appendix 4.6. Note that this assumption may influence the equilibrium solution of the coordination between the TSOs. Due to the fact that the minimum of the line capacities is chosen, the solutions for the TSOs are no longer continuous. Hence, some equilibria might be omitted during the iterative solution of the problem. We accept this shortfall in our numerical approach for the sake of the large-scale application. The general approach, however, remains valid, and a process for determining all equilibria could be implemented in the numerical solution method (e.g., through randomized starting values).

[22] All models were coded in GAMS 24.2.2 and solved with CPLEX 12.6 on a High Performance Computer with two processors (1600 and 2700Mhz) and physical/virtual memory of 98/150GB.

trend and an iteration count of 60. Based on the fact that in a Benders decomposition the lower bound is non-decreasing (i.e., change rates are always positive as demonstrated in Figure 4.3, left hand side), and the empirically observed behavior of the lower bound, it can be concluded that the error further decreases mainly due to changes in the upper bound. Hence, we argue that the lower bound can be taken as a good approximation of the optimal objective value as soon as our convergence criterion is met. To support this argument and to deepen our insights, we closely analyzed optimized levels of the variables, observing that they reach fairly stable levels in the last iterations before reaching the convergence criterion.[23] As an example, the right hand side of Figure 4.3 shows aggregated AC line capacities obtained in the final runs of the nodal pricing setting.

Based on the interpolation curves estimated from the observed changes in the optimality error, a 1% threshold is expected to be reached after around 150 iterations. The estimated increase of the lower bound and hence, the improvement of the optimal solution, will then be around 0.21% higher compared to our obtained value. At around 300 iterations, the optimal solution will deviate by about 0.24% from our obtained value, and further improvements of the optimal solution would be negligible. Considering the extensive computational burden as well as the expected limited improvements, we do not consider a smaller convergence threshold and rather accept some level of uncertainty regarding the different levels of optimality achieved in the different settings.

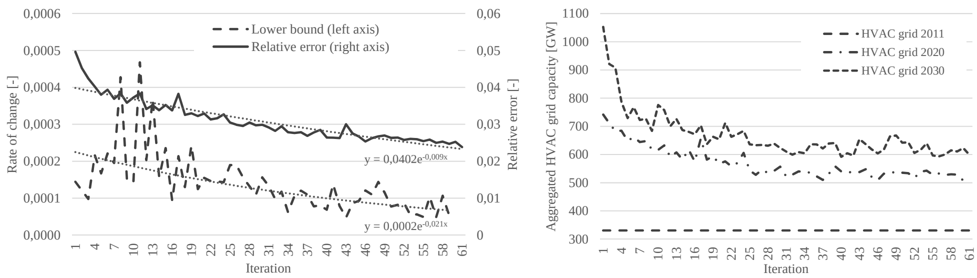


Figure 4.3: Development of lower bound, optimality error and aggregated AC line capacities during the iteration in *Setting I*

Costs are reported as accumulated discounted system costs.[24] In the generation sector, costs occur due to investments, operation and maintenance, production as well as ramping, whereas in the grid sector, investment as well as operation and

[23]Note that this argument is also supported by the analysis of convergence in a very similar setting published in Hagspiel et al. (2014).

[24]The discount rate is assumed to be 10% throughout all calculations.

maintenance costs are considered. Overall costs of electricity supply can be considered as a measure of efficiency and are reported in the following Figure 4.4 for the different settings. Besides the absolute costs, which are subdivided into generation and grid costs, the relative cost increase with respect to the overall costs of the nodal pricing setting is also depicted.

Considering the optimality error in the obtained solution, it should be stressed that the exact differences reported here do not necessarily persist after full convergence. However, based on the above discussion about convergence, the general conclusions and order of magnitude are expected to remain valid.

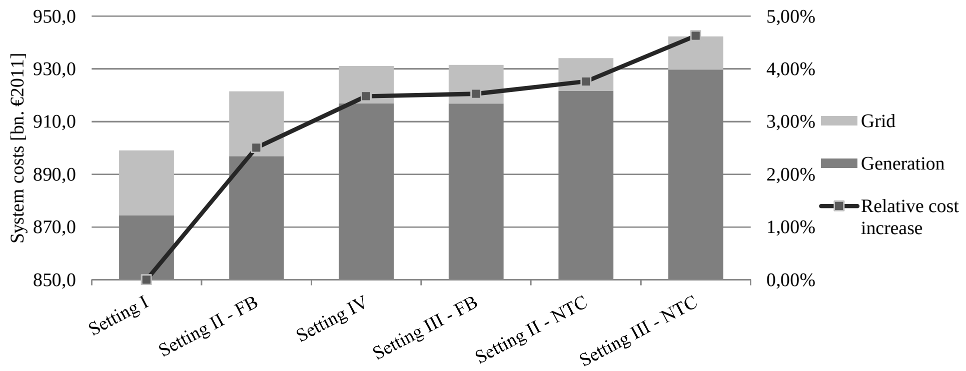


Figure 4.4: Total costs and relative performance of the different settings

As expected, nodal pricing (*Setting I*) is most efficient, with total costs summing up to 899.0 bn. €$_{2011}$ (874.3 bn. for generation and 24.7 bn. for the grid). Overall, costs increase by up to 4.6% relative to *Setting I* for the other settings. Thereby, NTC-based market coupling induces highest inefficiencies of 3.8% and 4.6% for one single TSO or zonal TSOs, respectively, both with the possibility to do redispatch on a national basis (*Setting II-NTC* and *Setting III-NTC*).[25] Hence, offering few amounts of trading capacity to the generation market, as implied by NTC-based market coupling, induces significant inefficiencies. In fact, by increasing trading capacities via flow-based market coupling, system costs can be lowered and inefficiencies amount to 2.5% for the single TSO, respectively 3.5% for zonal TSOs compared to nodal pricing (*Setting II-FB* and *Setting III-FB*). Hence, efficiency gains of 1.1-1.3 % of total system costs can be achieved by switching from NTC to flow-based market coupling. In turn, enhanced trading activities induced by flow-based market coupling entail greater

[25]Since topology control (as, e.g., in Kunz (2013) is not considered, costs of redispatch could possibly be lower. However, since topology control would also be available in the market clearing of the nodal pricing, efficiency gains would persist for all regimes. Hence, the reported differences between the inefficiencies should be similar.

TSO activity, both in the expansion as well as in the redispatch. For this reason, TSO costs are higher for flow-based than for NTC-based market coupling. However, these additional costs are overcompensated by lower costs in the generation sector. The net effect of a switch from NTC to flow-based market coupling is beneficial for the overall system.

Somewhat surprisingly, the national g-component (*Setting IV*) hardly performs better than the same setting with redispatch (*Setting III-FB*). Hence, the optimal allocation of power generation within market zones is hardly influenced by grid restrictions within that zone. In contrast, the optimal allocation induced by nodal prices throughout the CWE region entails substantial gains in efficiency due to reduced system costs. The setting that comes closest to nodal pricing consists of flow-based coupled zonal markets with a single TSOs and induces an inefficiency of 2.5% in comparison to nodal pricing (*Setting II-FB* vs. *Setting I*).

Even though the share of TSO costs on total costs is very small compared to the share of generation costs in all settings (1.3-2.7%)[26], the amount of grid capacities varies greatly between the different settings. Figure 4.5 shows the aggregated high voltage (HV) AC and HVDC line capacities.

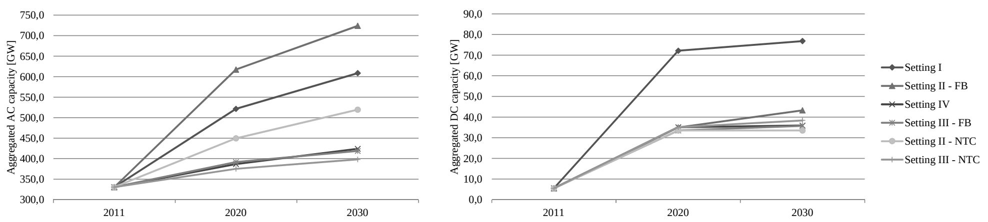


Figure 4.5: Aggregated line capacities AC and DC

Grid capacities are generally lower in the case of zonal TSOs where they only agree on the smaller of the two proposed expansion levels for cross-border lines (*Setting III-FB* and *Setting III-NTC*). In these cases, overall AC grid capacities increase from 331GW in 2011 to 398GW (*Setting III-NTC*) respectively 418GW (*Setting III-FB*) in 2030, corresponding to an increase of 20-28%. In case of a single TSO, cross-border along with overall line expansions are significantly higher compared to zonal TSOs, with 2030 levels reaching 519GW (*Setting II-NTC*) to 724GW (*Setting II-FB*). Especially in *Setting II-FB*, the TSO is obliged to cope with inefficiently allocated generation plants by excessively expanding the grid, while not being able to avoid those measures with suitable price signals. DC line expansions appear to be crucial

[26]The rather minor role of grid costs compared to costs occurring in the generation sector has already been identified, e.g., in Fürsch et al. (2013).

for an efficient system development, especially towards the UK where large wind farms help to reach CO_2-targets and to supply the UK itself as well as the continent with comparatively cheap electricity. Thereby, the high DC expansion level in the nodal pricing regime is remarkable. Whereas in zonal markets prices are averaged across the zone, nodal prices reveal the true value of connecting specific nodes via DC-lines and thus enable efficient investments in those projects. In consequence, in the nodal pricing regime, DC line capacities are about double as high as in the other settings. This helps to reduce overall costs to a minimum (*Setting I*).

Besides the overall level of grid and generation capacities, their regional allocation also differs between the various settings, mainly due to differences in the (local) availability of transmission upgrades. As has been seen, higher grid expansion levels result from a single TSO (*Setting I and Setting II*), enabling a better utilization of renewable energies at favorable sites (i.e., sites where the specific costs of electricity generation are lowest). In Figure 4.6, we exemplarily illustrate this effect based on a cross-border line between France and Germany (line 80 in our model). However, the same effect is observable for other interconnections, e.g., between France and Belgium. Higher grid capacities allow the use of high wind speed locations in Northern France and thus foster more expansion of wind capacities in this area. In case of zonal TSOs (*Setting III and Setting IV*) only low amounts of wind capacity are built in France (e.g., in node FR-06) as these areas cannot be connected with the rest of the system. To still meet the European CO_2-target, PV power plants are built in the southern part of Germany (e.g., in node DE-27). Obviously, these locations are non-optimal with respect to other options as they are not used in the setting with one TSO. Thus, implemented market designs significantly influence the amount and location of renewable energies within the system.[27]

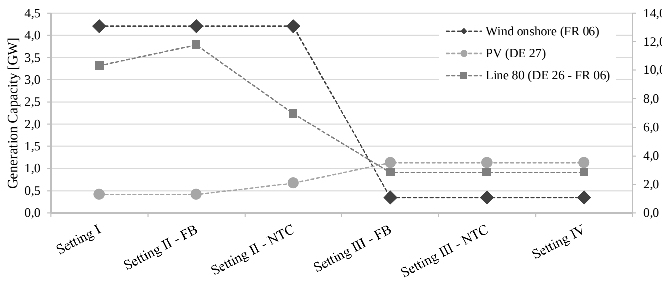


Figure 4.6: Exemplary grid expansion and regional allocation of renewable energies

[27]Conventional capacities are also affected. However, the effect is less pronounced as the differences between the site-specific costs of generation are smaller.

4.5 Conclusions

In the context of liberalized power markets and unbundled generation and transmission services, the purpose of this paper was to develop a modeling framework for different regulatory designs regarding congestion management including both, the operation as well as the investment perspective in the generation and transmission sector. We have presented an analytical formulation that is able to account for different regulatory designs of market areas, a single or zonal TSOs, as well as different forms of measures to relieve congestion, namely grid expansion, redispatch and g-components. We have then proposed an algorithm to numerically solve these problems, based on the concept of decomposition. This technique has shown to entail a number of characteristics that work to our advantage, especially flexible algorithmic implementation as well as consistency of the grid flow representation through PTDF update.

Calibrating our model to the CWE region, we have demonstrated the applicability of our numerical solution algorithm in a large-scale application consisting of 70 nodes and 174 lines along with a detailed bottom-up representation of the generation sector. Compared to nodal pricing as the efficient benchmark, inefficiencies induced by alternative settings reach additional system costs of up to 4.6%. Major deteriorative factors are TSOs activities restricted to zones as well as low trading capacities offered to the market. These findings may serve as a guideline for policymakers when designing international power markets. For instance, our results confirm ongoing efforts to implement flow-based market coupling and to foster a closer cooperation of TSOs in the CWE region. In fact, we find that such a regulatory design could come close to the nodal pricing benchmark, with an efficiency difference of only 2.5%. Reported cost differences might be impacted by numerical imprecision in the solution algorithm, although empirical observations of the convergence behavior suggest that the general effects as well as the order of magnitude persist. Noticeably, the magnitude of these results should be interpreted as the lower bound of efficiency gains, since we focus on frictions in the congestion management only.

More generally, we find that a single TSO (or enhanced coordination between the zonal TSOs) is key for an efficient development of both, grid and generation infrastructures. Whereas the expansion of grid infrastructure is immediately affected, the generation sector indirectly takes advantage of increased grid capacities and hence, can develop more efficiently. Better allocation of generation units with respect to grid costs through high resolution price signals gains importance for larger geographical areas and larger differences between generation costs and expansion potentials (such as wind or solar power). This has been found for the CWE region, and may prove even more important for the whole of Europe. It should be noted, however, that efficiency gains need to be put into the context of transaction costs occurring from the switch to a different congestion management design. In addition, socio-economic factors such as acceptance for grid expansion are not considered in the analysis, but might also play a role considering the large differences of necessary grid quantities.

Limitations of our approach that leave room for extensions and improvement stem from the fact that we assume linear transmission investments, and do not consider strategic behavior of individual agents, imperfectly regulated TSOs, or uncertainty about future developments (e.g., delays in expansion projects). The assumption of an inelastic demand probably reduces the magnitude of the measured inefficiencies, since demand does not react to any price changes and hence only supply-side effects are captured. Algorithmically, the effectiveness of our solution process could be further improved, e.g., through better usage of numerical properties of the problem (such as gradients, etc.). Nevertheless, in its present form, our framework may serve as a valuable tool to assess a number of further relevant questions, such as the tradeoff between different flexibility options (such as grids, storages or renewable curtailment), the impact of different forms of congestion management in other European regions, or the valuation of grid expansion projects.

Acknowledgments

Financial support by the Energy Storage Initiative of the German Federal Ministry for Economic Affairs and Energy and the German Federal Ministry of Education and Research through grant 03ESP239 is gratefully acknowledged.

4.6 Appendix

Notation list

Abbreviation	Dimension	Description
Model sets		
$i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{K}, l \in \mathbf{L}$		Nodes, $\mathbf{I}, \mathbf{J}, \mathbf{K}, \mathbf{L} = [1, 2, ...]$
$m, n \in \mathbf{M}$		Zonal markets, $\mathbf{M} = [1, 2, ...]$
$i \in \mathbf{I_m}, j \in \mathbf{J_m}$		Nodes that belong to zonal market m, $\mathbf{I_m} \subset \mathbf{I}$, $\mathbf{J_m} \subset \mathbf{J}$
$i \in \mathbf{I_{m,cb}}, j \in \mathbf{J_{m,cb}}$		Nodes that belong to zonal market m and are connected to a another zone n by a cross-border line, $\mathbf{I_{m,cb}} \subset \mathbf{I_m}$, $\mathbf{J_{m,cb}} \subset \mathbf{J_m}$
$t \in \mathbf{T}$		Point in time for dispatch decisions (e.g., hours)
Model parameters		
δ_i	EUR/kW	Investment and FOM costs of generation capacity in node i
$\gamma_{i,t}$	EUR/kWh	Variable costs of generation capacity in node i
$\mu_{i,j}$	EUR/kW	Investment costs of line between node i and node j
$d_{i,t}$	kW	Electricity demand in node i
$PTDF_{k,i,j}$	–	Power Transfer Distribution Factor (impact of the power balance in node k on flows on line i, j)
$\sigma_{i,j}$	%	Cost share for an interconnector capacity between node i and node j, $i \in \mathbf{I_{m,cb}}$, $j \in \mathbf{J_{m,cb}}$
Model primal variables		
$\overline{G}_i$	kW	Generation capacity in node i, $\overline{G}_i \geq 0$
$G_{i,t}$	kW	Generation dispatch in node i, $G_{i,t} \geq 0$
$T_{i,j,t}, T_{m,n,t}$	kW	Electricity trade from node i to node j, or market m to market n
X	EUR	Costs of generation
Y	EUR	Costs of TSO
$\overline{P}_{i,j}$	kW	Line capacity between node i and node j, $\overline{P}_{i,j} \geq 0 \;\; \forall\, i, j \neq j, i$
$P_{i,j,t}$	kW	Electricity flow on line between node i and node j
$R_{i,t}$	kW	Redispatch in node i
α	EUR	Helping variable to include transmission costs of the current iteration in the master problem
Model dual variables		
$\kappa_{i,j,t}, \kappa_{m,n,t}$	EUR/kW	price for transmission between nodes (i and j) or zones (m and n)
$\lambda_{i,t}, \lambda_{m,t}$	EUR/kW	nodal or zonal price for electricity

Table 4.2: Model sets, parameters and variables

Derivation of the load flow equations by means of PTDFs

Power Transfer Distribution Factors (PTDFs) are a well-established method to account for load flows in meshed electricity networks by means of linearization. They can be derived from the network equations in an AC power network that write as follows:[28]

$$P_i = U_i \sum_{j \in \mathscr{I}} U_j (g_{i,j} \cos(\varphi_i - \varphi_j) + b_{i,j} \sin(\varphi_i - \varphi_j)) \tag{4.5}$$

$$Q_i = U_i \sum_{j \in \mathscr{I}} U_j (g_{i,j} \sin(\varphi_i - \varphi_j) - b_{i,j} \cos(\varphi_i - \varphi_j)) \tag{4.6}$$

$$P_{i,j} = U_i^2 g_{i,j} - U_i U_j g_{i,j} \cos(\varphi_i - \varphi_j) - U_i U_j b_{i,j} \sin(\varphi_i - \varphi_j) \tag{4.7}$$

$$Q_{i,j} = -U_i^2 (b_{i,j} + b_{i,j}^{sh}) + U_i U_j b_{i,j} \cos(\varphi_i - \varphi_j) - U_i U_j g_{i,j} \sin(\varphi_i - \varphi_j). \tag{4.8}$$

P_i and Q_i represent the net active and reactive power infeed (i.e., nodal power balances), and $P_{i,j}$ and $Q_{i,j}$ the active and reactive power flows between node i and j. Voltage levels U and phase angles φ of the nodes as well as series conductances g and series susceptances b of the transmission lines determine active and reactive power flows in a highly nonlinear way.

In order to linearize the above equations, a number of assumptions are made:

- All voltages are set to 1 p.u.
- Voltage angles are all similar (and hence, $\sin(\varphi_i - \varphi_j) \approx \varphi_i - \varphi_j$).
- Reactive power is neglected (i.e., $Q_i = Q_{i,j} = 0$).
- Losses are neglected and line reactances are much larger than their resistance, such that $x \gg r \approx 0$.

Under these assumptions and using Kirchoff's power law, the network equations can be simplified to

$$P_{i,j} \approx \frac{1}{x_{i,j}} (\varphi_i - \varphi_j) \tag{4.9}$$

$$P_i \approx \sum_{j \in \Omega_i} \frac{1}{x_{i,j}} (\varphi_i - \varphi_j), \tag{4.10}$$

with Ω_i representing the nodes adjacent to i. If there are multiple nodes and branches, this can be written in a more convenient matrix notation as $\tilde{\boldsymbol{P_i}} = \tilde{\boldsymbol{B}} \cdot \tilde{\boldsymbol{\Theta}}$, with $\tilde{\boldsymbol{P_i}}$ being the vector of net active nodal power balances P_i, $\tilde{\boldsymbol{\Theta}}$ the vector of phase

[28]The following is based on Andersson (2011), even though the general approach can be found in most electrical engineering textbooks.

angles, and $\tilde{\boldsymbol{B}}$ the nodal admittance matrix with the following entries:

$$\tilde{B}_{i,j} = -\frac{1}{x_{i,j}} \tag{4.11}$$

$$\tilde{B}_{i,i} = \sum_{j \in \Omega_i} \frac{1}{x_{i,j}}. \tag{4.12}$$

By deleting the row and column belonging to the reference node (thus assuming a zero reference angle at this node), the previously singular matrix $\tilde{\boldsymbol{B}}$ becomes $\boldsymbol{B}$, the vector of phase angles $\boldsymbol{\Theta}$, and the vector of net active nodal power balances $\boldsymbol{P_i}$. We can now solve for $\boldsymbol{\Theta}$ by matrix inversion:

$$\boldsymbol{\Theta} = \boldsymbol{B}^{-1} \cdot \boldsymbol{P_i}. \tag{4.13}$$

Defining $H_{ki} = 1/x_{i,j}$, $H_{kj} = -1/x_{i,j}$ and $H_{km} = 0$ for $m \neq i, j$ (with k running over the branches i, j), Equation (4.9) can be rewritten in matrix form as $\boldsymbol{P_{i,j}} = \boldsymbol{H} \cdot \boldsymbol{\Theta}$. Inserting $\boldsymbol{\Theta}$ from Equation (4.13) finally yields

$$\boldsymbol{P_{i,j}} = \boldsymbol{H} \cdot \boldsymbol{\Theta} = \boldsymbol{H} \cdot \boldsymbol{B}^{-1} \cdot \boldsymbol{P_i} = PTDF \cdot \boldsymbol{P_i} \tag{4.14}$$

The elements of $PTDF$ are the power transfer distribution factors that constitute a linear relationship between nodal power balances and load flows. Note that the size of the $PTDF$ matrix is determined by the size of the system, with the number of matrix lines corresponding to the number of transmission lines, and the number of matrix columns representing the number of nodes. The matrix entry $PTDF_{k,i,j}$ represents the impact of the power balance in node k on power flows on line between node i and j. Also note that $PTDF$ essentially depends (only) on the line impedances $x_{i,j}$ in the system that in turn depend primarily on the respective line capacities $\overline{P}_{i,j}$. Hence, as done, e.g., in Hogan et al. (2010), we apply the law of parallel circuits to adjust line reactances when altering transmission capacities, i.e.,

$$x_{i,j} = \frac{\overline{P}^0_{i,j}}{\overline{P}_{i,j}} x^0_{i,j}, \tag{4.15}$$

where $\{\overline{P}^0_{i,j}, x^0_{i,j}\}$ is a point of reference taken from the original configuration of the transmission network. Overall, this yields a functional dependency of power flows on nodal balances (determined by generation G_k and load d_k in all nodes) as well as line capacities $\overline{P}_{k,l}$ of all lines in the system, i.e., $P_{i,j} = P_{i,j}(\overline{P}_{k,l}, G_k, d_k)$.

Equivalence of Problem P1 and P1'

To show the equivalence of the optimal solution of *P1* and *P1'*, we compare the problems by means of their Karush-Kuhn-Tucker (KKT)conditions. If they are equal, the optimal solution has to be equal, too (e.g., Bazaraa et al. (2006)). For the derivations, note that trade is a function of line capacity, generation and demand, i.e., $T_{i,j,t} = T_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, d_{k,t})$, and that $T_{i,j,t} = -T_{j,i,t}$. The following is the Lagrangian function belonging to Problem *P1*:

$$\begin{aligned} L(\overline{G}_i, G_{i,t}, T_{i,j,t}, \overline{P}_{i,j}, \lambda_{i,t}, \tau_{i,t}, \kappa_{i,j,t}) = & \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} \\ & + \sum_{i,t} (\lambda_{i,t}(G_{i,t} - \sum_j T_{i,j,t} - d_{i,t}) \\ & + \tau_{i,t}(G_{i,t} - \overline{G}_i)) + \sum_{i,j,t} (\kappa_{i,j,t}(|T_{i,j,t}| - \overline{P}_{i,j})) \end{aligned} \tag{4.16}$$

The corresponding KKT conditions are:

$$\frac{\partial L}{\partial \overline{G}_i} = \delta_i - \sum_t \tau_{i,t} \leq 0, \ \ \overline{G}_i \geq 0, \ \ \overline{G}_i(\frac{\partial L}{\partial \overline{G}_i}) = 0 \quad \forall i \tag{4.17a}$$

$$\frac{\partial L}{\partial G_{i,t}} = \gamma_{i,t} + \lambda_{i,t}(1 - \sum_j \frac{\partial T_{i,j,t}}{\partial G_{i,t}}) + \tau_{i,t} + \sum_j \kappa_{i,j,t} \frac{\partial T_{i,j,t}}{\partial G_{i,t}} \leq 0, \tag{4.17b}$$

$$G_{i,t} \geq 0, \ G_{i,t}(\frac{\partial L}{\partial G_{i,t}}) = 0 \ \ \forall i,t \tag{4.17c}$$

$$\frac{\partial L}{\partial \overline{P}_{i,j}} = \mu_{i,j} - \sum_t \lambda_{i,t} \frac{\partial T_{i,j,t}}{\partial \overline{P}_{i,j}} + \sum_t \kappa_{i,j,t}(\frac{\partial T_{i,j,t}}{\partial \overline{P}_{i,j}} - 1) \leq 0, \tag{4.17d}$$

$$\overline{P}_{i,j} \geq 0, \ \ \overline{P}_{i,j}(\frac{\partial L}{\partial \overline{P}_{i,j}}) = 0 \quad \forall i,j \tag{4.17e}$$

$$\frac{\partial L}{\partial \kappa_{i,j,t}} = |T_{i,j,t}| - \overline{P}_{i,j} \leq 0, \quad \kappa_{i,j,t} \geq 0, \quad \kappa_{i,j,t}(\frac{\partial L}{\partial \kappa_{i,j,t}}) = 0 \ \ \forall i,j,t \tag{4.17f}$$

$$\frac{\partial L}{\partial \tau_{i,j}} = G_{i,t} - \overline{G}_i \leq 0, \quad \tau_{i,j} \geq 0, \quad \tau_{i,j}(\frac{\partial L}{\partial \tau_{i,j}}) = 0 \qquad \forall i,j \tag{4.17g}$$

$$\frac{\partial L}{\partial \lambda_{i,t}} = G_{i,t} - \sum_j T_{i,j,t} - d_{i,t} = 0 \qquad \forall i,t \tag{4.17h}$$

$$\frac{\partial L}{\partial T_{i,j,t}} = \kappa_{i,j,t} - \lambda_{i,t} + \lambda_{j,t} = 0 \qquad \forall i,j,t \tag{4.17i}$$

The Langragian functions for *P1'* are:

$$L'^a(\overline{G}_i, G_{i,t}, T_{i,j,t}, \lambda_{i,t}, \tau_{i,t}) = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{i,j,t} \kappa_{i,j,t} T_{i,j,t} + \sum_{i,t} (\lambda_{i,t}(G_{i,t} - \sum_{j,t} T_{i,j,t} - d_{i,t}) + \kappa_{i,j,t}(G_{i,t} - \overline{G}_i)) \tag{4.18}$$

$$L'^b(\overline{P}_{i,j}, \kappa_{i,j,t}) = \sum_{i,j} \mu_{i,j} \overline{P}_{i,j} + \sum_{i,t} (\lambda_{i,t}(G_{i,t} - \sum_{j,t} T_{i,j,t} - d_{i,t})) + \sum_{i,j,t} (\kappa_{i,j,t}(|T_{i,j,t}| - \overline{P}_{i,j})) \tag{4.19}$$

The KKT conditions of *P1'a* are:

$$\frac{\partial L'^a}{\partial \overline{G}_i} = \delta_i - \sum_t \tau_{i,t} \leq 0, \;\; \overline{G}_i \geq 0, \;\; \overline{G}_i(\frac{\partial L}{\partial \overline{G}_i}) = 0 \quad \forall i \tag{4.20a}$$

$$\frac{\partial L'^a}{\partial G_{i,t}} = \gamma_{i,t} + \lambda_{i,t}(1 - \sum_j \frac{\partial T_{i,j,t}}{\partial G_{i,t}}) + \tau_{i,t} + \sum_j \kappa_{i,j,t} \frac{\partial T_{i,j,t}}{\partial G_{i,t}} \leq 0, \tag{4.20b}$$

$$G_{i,t} \geq 0, \; G_{i,t}(\frac{\partial L}{\partial G_{i,t}}) = 0 \;\; \forall i,t \tag{4.20c}$$

$$\frac{\partial L'^a}{\partial \tau_{i,j}} = G_{i,t} - \overline{G}_i \leq 0, \quad \tau_{i,j} \geq 0, \quad \tau_{i,j}(\frac{\partial L'^a}{\partial \tau_{i,j}}) = 0 \qquad \forall i,j \tag{4.20d}$$

$$\frac{\partial L'^a}{\partial \lambda_{i,t}} = G_{i,t} - \sum_j T_{i,j,t} - d_{i,t} \qquad \forall i,t \tag{4.20e}$$

$$\frac{\partial L'^a}{\partial T_{i,j,t}} = \kappa_{i,j,t} - \lambda_{i,t} + \lambda_{j,t} = 0 \qquad \forall i,j,t \tag{4.20f}$$

The KKT conditions of *P1'b* are:

$$\frac{\partial L'^b}{\partial \overline{P}_{i,j}} = \mu_{i,j} - \sum_t \lambda_{i,t} \frac{\partial T_{i,j,t}}{\partial \overline{P}_{i,j}} + \sum_t \kappa_{i,j,t}(\frac{\partial T_{i,j,t}}{\partial \overline{P}_{i,j}} - 1) \leq 0, \tag{4.21a}$$

$$\overline{P}_{i,j} \geq 0, \;\; \overline{P}_{i,j}(\frac{\partial L}{\partial \overline{P}_{i,j}}) = 0 \quad \forall i,j \tag{4.21b}$$

$$\frac{\partial L'^b}{\partial \kappa_{i,j,t}} = |T_{i,j,t}| - \overline{P}_{i,j} \leq 0, \quad \kappa_{i,j,t} \geq 0, \quad \kappa_{i,j,t}(\frac{\partial L}{\partial \kappa_{i,j,t}}) = 0 \;\; \forall i,j,t \tag{4.21c}$$

Comparing the KKT conditions of problem *P1* to the ones of *P1a* and *P1b*, we can conclude that the problems are indeed equivalent.

Model of Setting III: coupled zonal markets with zonal TSOs and zonal redispatch

Mathematically, the model of *Setting III*, representing coupled zonal markets with zonal TSOs and zonal redispatch, is formulated as follows:

P3a Generation

$$\min_{\overline{G}_i, G_{i,t}, T_{m,n,t}} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{m,n,t} \kappa_{m,n,t} T_{m,n,t} \tag{4.22a}$$

$$\text{s.t.} \quad \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_{n,t} T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \qquad \forall m, t \qquad | \lambda_m \tag{4.22b}$$

$$G_{i,t} \leq \overline{G}_i \qquad \forall i, t \tag{4.22c}$$

$$T_{m,n,t} = -T_{n,m,t} \qquad \forall m, n, t \tag{4.22d}$$

P3b Transmission

$$\min_{\overline{P}_{i,j \in \mathbf{I_m}}, R_{i \in \mathbf{I_m},t}} \quad Y_m = \sum_{i,j \in \mathbf{I_m}} \mu_{i,j} \overline{P}_{i,j} + \sum_{i,j \in \mathbf{I_{m,cb}}} \sigma_{i,j} \mu_{i,j} \overline{P}_{i,j} + \sum_{i \in \mathbf{I_m},t} \gamma_{i,t} R_{i,t} \qquad \forall m \tag{4.22e}$$

$$\text{s.t.} \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_{n} T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \qquad \forall m, t \tag{4.22f}$$

$$|T_{i,j,t}| = |P_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, R_{k,t}, d_{k,t})| \leq \overline{P}_{i,j} \qquad \forall t, i, j \in \mathbf{I_m} \qquad | \kappa_{i,j \in \mathbf{I_m}} \tag{4.22g}$$

$$\sum_{i \in \mathbf{I_m},t} R_{i,t} = 0 \tag{4.22h}$$

$$0 \leq G_{i,t} + R_{i,t} \leq \overline{G}_i \qquad \forall t, i \in \mathbf{I_m} \tag{4.22i}$$

$$\kappa_{m,n,t} = g(\kappa_{i,j,t}) \tag{4.22j}$$

$$T_{m,n,t} = -T_{n,m,t} \qquad \forall m, n, t \tag{4.22k}$$

In problem *P3*, there are now separate optimization problems for each zonal TSO (indicated by Y_m), with the objective to minimize costs from zonal grid and cross-border capacity extensions as well as from zonal redispatch measures (Equation (4.22e)). For the redispatch, TSOs have to consider the same restrictions as in the previous setting (Equations (4.22h) and (4.22i)). TSOs are assumed to negotiate about the extension of cross-border capacities according to some regulatory rule that ensures the acceptance of a unique price for each cross-border line by both of the neighboring TSOs. For instance, the regulatory rule may be specified such that both TSOs are obliged to accept the higher price offer, or, equivalently, the lower of the two capacities offered for the specific cross-border line. Corresponding costs from inter-zonal grid extensions are assumed to be shared among the TSOs according to the cost allocation key $\sigma_{i,j}$. According to Equation (4.22j), prices for transmission between zones that are provided to the generation stage ($\kappa_{m,n,t}$) are determined just as in the previous *Setting II* with only one TSO, depending on the type of market coupling, i.e., the specification of function g. The only difference is that line-specific prices $\kappa_{i,j,t}$ may now deviate from *Setting II* as they result from the separated activities of each zonal TSO (specifically, from Equation (4.22g), i.e., the restriction of flows on intra-zonal and cross-border lines).

Model of Setting IV: coupled zonal markets with zonal TSOs and g-component

Mathematically, the model of *Setting IV*, representing coupled zonal markets with zonal TSOs and g-component, is formulated as follows:

P4a *Generation*

$$\min_{\overline{G}_i, G_{i,t}, T_{m,n,t}} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \sum_{i,j,t} \kappa_{i,j,t} T_{i,j,t} \tag{4.23a}$$

$$\text{s.t.} \quad \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_n T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \qquad \forall m,t \qquad | \, \lambda_m \tag{4.23b}$$

$$G_{i,t} \leq \overline{G}_i \qquad \forall i,t \tag{4.23c}$$

$$T_{m,n,t} = -T_{n,m,t} \qquad \forall m,n,t \tag{4.23d}$$

P4b *Transmission*

$$\min_{\overline{P}_{i,j \in \mathbf{I_m}, \mathbf{I_{m,cb}}}} \quad Y_m = \sum_{i,j \in \mathbf{I_m}} \mu_{i,j} \overline{P}_{i,j} + \sum_{i,j \in \mathbf{I_{m,cb}}} \sigma_{i,j} \mu_{i,j} \overline{P}_{i,j} \qquad \forall m \tag{4.23e}$$

$$\text{s.t.} \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_n T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \qquad \forall m,t \tag{4.23f}$$

$$|T_{i,j,t}| = |P_{i,j,t}(\overline{P}_{k,l}, G_{k,t}, d_{k,t})| \leq \overline{P}_{i,j} \quad \forall t, i, j \in \mathbf{I_m}, \mathbf{I_{m,cb}} \quad | \, \kappa_{i,j \in \mathbf{I_m}, \mathbf{I_{m,cb}}, t} \tag{4.23g}$$

$$T_{m,n,t} = -T_{n,m,t} \qquad \forall m,n,t \tag{4.23h}$$

Problem *P4a* is almost identical to *P2a* (and *P3a*), with the exception of one term in the objective function (4.23a). With a g-component, generators pay nodal instead of zonal prices for transmission ($\kappa_{i,j,t}$ instead of $\kappa_{m,n,t}$), depending on the impact of their nodal generation level on the grid infrastructure (by means of $T_{i,j,t} = T_{i,j,t}(G_{k,t}, d_{k,t})$). These prices are determined by the zonal TSOs via their flow-restriction (4.23g).

Numerical algorithm for NTC-coupled zonal markets, zonal TSOs, and zonal redispatch

In Section 4.3, we have shown the numerical implementation of the nodal pricing regime. For the sake of clarifying the major changes needed to represent the alternative Settings *II-IV*, we here present the model for m zonal (instead of nodal) markets that are coupled via NTC-based capacity restrictions, along with multiple zonal TSOs (instead of only one), all having the possibility to deploy zonal redispatch as an alternative to grid expansion. Hence, the model corresponds to *Setting III* with NTC-based market coupling. Compared to nodal pricing, no more nodal or time-specific information about grid costs is provided. Instead, an aggregated price $\kappa^{(v)}_{m,n,t}$ for each border is calculated via a function g_{NTC} and passed on to generation level. The model with flow-based market coupling works in the same way, only that the price $\kappa^{(v)}_{m,n,t}$ is calculated via a different function g_{FB}.

$v = 1$; convergence=false

While(convergence=false) {

Master problem: generation

$$\min_{\overline{G}_i, G_{i,t}, T_{m,n,t}, \alpha} \quad X = \sum_i \delta_i \overline{G}_i + \sum_{i,t} \gamma_{i,t} G_{i,t} + \alpha \tag{4.24a}$$

$$\text{s.t.} \quad \sum_{i \in \mathbf{I_m}} G_{i,t} - \sum_n T_{m,n,t} = \sum_{i \in \mathbf{I_m}} d_{i,t} \quad \forall m, t \tag{4.24b}$$

$$G_{i,t} \leq \overline{G}_i \quad \forall i, t \tag{4.24c}$$

$$T_{m,n,t} = -T_{n,m,t} \quad \forall m, n, t \tag{4.24d}$$

$$\sum_m Y^{(u)}_m + \sum_{m,n,t} \kappa^{(u)}_{m,n,t} \cdot (T_{m,n,t} - T^{(u)}_{m,n,t}) \leq \alpha \quad \forall u = 1, ..., v-1 | v > 1 \tag{4.24e}$$

$$G^{(v)}_{i,t} = \text{Optimal value of } G_{i,t} \quad \forall i, t \tag{4.24f}$$

Sub-problem: transmission

$$\min_{\overline{P}_{i,j\in \mathbf{I_m},\mathbf{I_{m,cb}}},R_{i\in \mathbf{I_m},t},T_{m,n,t}} Y_m = \sum_{i,j\in \mathbf{I_m}} \mu_{i,j}\overline{P}_{i,j} + \frac{1}{2}\sum_{i,j\in \mathbf{I_{m,cb}}} \mu_{i,j}\overline{P}_{i,j} + \sum_{i\in \mathbf{I_m},t} R_{i,t}\gamma_{i,t} \quad \forall m \tag{4.24g}$$

$$\text{s.t.} \quad \left|P_{i,j,t}\right| = \left|\sum_k PTDF_{k,i,j}\cdot(G_{k,t}^{(v)} + R_{k,t} - d_{k,t})\right| \leq \overline{P}_{i,j} \quad \forall i,j,t \quad |\,\kappa_{i,j,t}^{(v)} \tag{4.24h}$$

$$0 \leq R_{i,t} + G_{i,t}^{(v)} \leq \overline{G}_i \quad \forall i,t \in \mathbf{I_m} \tag{4.24i}$$

$$\sum_{i\in \mathbf{I_m}} R_{i,t} = 0 \tag{4.24j}$$

−−−−−−−−−−−−−−−−−−−−−−−

$$Y_m^* = \text{Optimal value of } Y_m \tag{4.24k}$$

$$PTDF^{(v)} = \text{New PTDF matrix calculated based on } \overline{P}_{i,j}\kappa_{m,n,t} = g_{NTC}(\kappa_{i,j,t}^{(v)}) \tag{4.24l}$$

$$\overline{P}_{i,j\in \mathbf{I_{m,cb}}} = \overline{P}_{i,j\in \mathbf{I_{n,cb}}} = \min\left\{\overline{P}_{i,j\in \mathbf{I_{m,cb}}};\overline{P}_{i,j\in \mathbf{I_{n,cb}}}\right\} \tag{4.24m}$$

if(convergence criterion < threshold; convergence=true)

$v = v + 1$

};

Numerical assumptions for the large-scale application

Country	2011	2020	2030
Belgium	87	98	105
Germany	573	612	629
France	466	524	559
Luxembourg	7	8	8
Netherlands	113	128	137
Eastern	276	328	366
Northern	387	436	465
Southern	450	528	594
Southwest	317	378	433
United Kingdom	400	450	481

Table 4.3: Assumptions for the gross electricity demand [TWh]

To depict the CWE region in a high spatial resolution, we split the gross electricity demand per country among the nodes belonging to this country according to the percentage of population living in that region.

Technology	**2020**	**2030**
Wind Onshore	1253	1188
Wind Offshore (<20m depth)	2800	2350
Wind Offshore (>20m depth)	3080	2585
Photovoltaics (roof)	1260	935
Photovoltaics (ground)	1110	785
Biomass gas	2398	2395
Biomass solid	3297	3295
Biomass gas, CHP	2597	2595
Biomass solid, CHP	3497	3493
Geothermal	10504	9500
Compressed Air Storage	1100	1100
Pump Storage	1200	1200
Lignite	1500	1500
Lignite Innovative	1600	1600
Coal	1200	1200
Coal Innovative	2025	1800
IGCC	1700	1700
CCGT	711	711
OCGT	400	400
Nuclear	3157	3157

Table 4.4: Assumptions for the generation technology investment costs [€ /kW]

Fuel type	**2011**	**2020**	**2030**
Nuclear	3.6	3.3	3.3
Lignite	1.4	1.4	2.7
Oil	39.0	47.6	58.0
Coal	9.6	10.1	10.9
Gas	14.0	23.1	25.9

Table 4.5: Assumptions for the gross fuel prices [€ /MWh_{th}]

Grid Technology	**Extension costs**	**FOM costs**
AC overhead line incl. compensation	445 € /(MVA*km)	2.2 € /(MVA*km)
DC overhead line	400 € /(MW*km)	2.0 € /(MW*km)
DC underground	1250 € /(MW*km)	6.3 € /(MW*km)
DC submarine	1100 € /(MW*km)	5.5 € /(MW*km)
DC converter pair	150000 € /MW	750.0 € /MW

Table 4.6: Assumptions for the grid extension and FOM costs

5 Regulation of non-marketed outputs and substitutable inputs

We study the regulation of a monopolistic firm that provides a non-marketed output based on multiple substitutable inputs. The regulator is able to observe the effectiveness of the provision, but faces information asymmetries with respect to the efficiency of the firm's activities. Motivated by the example of electricity transmission services, we consider a setting where one input (grid expansion) and the output (uninterrupted electricity transmission) are observable, while another input (sophisticated grid operation) and related costs are not. Multi-dimensional information asymmetries are introduced by discrete distributions for the functional form of the marginal rate of substitution between the inputs as well as for the input costs. For this novel setting, we investigate the theoretically optimal Bayesian regulation mechanism. We find that the first-best solution cannot be obtained in case of shadow costs of public funding. The second-best solution implies separation of the most efficient type with first-best input levels, and upwards distorted (potentially bunched) observable input levels for all other types. Moreover, we compare these results to a simpler non-Bayesian approach and hence, bridge the gap between the academic discussion and regulatory practice. We provide evidence that under certain conditions, a single contract non-Bayesian regulation can indeed get close to the second-best of the Bayesian menu of contracts regulation.

5.1 Introduction

Numerous goods and services are provided by regulated firms with a monopolistic status. For instance, uninterrupted electricity transmission services - being a textbook example of a natural monopoly - are usually provided by a single firm. Currently, an increasing deployment of renewable energy sources leads to substantially changing requirements to secure an uninterrupted electricity transmission, while multiple substitutable measures may exist to cope with it, such as grid expansion or sophisticated grid operation.[1] Due to the fact that an uninterrupted electricity transmission is crucial for society, the regulator will be well aware of whether or not it has been provided *effectively*.[2] In contrast, however, electricity systems are highly complex, such that interdependent activity levels as well as related cost figures are hard to assess. Hence, it may be difficult for the regulator to judge the *efficiency* of the firm's underlying measures. Technically speaking, this situation may be seen as a production process involving multiple substitutable inputs, incorporating two adverse selection problems: First, the regulator may have a hard time estimating the necessary overall level of the firm's activity, determined by the marginal rate of technical substitution (MRTS), i.e., the isoquant function describing the relation of inputs needed to produce the requested output. Second, the regulator may have difficulties verifying the unit costs of one or multiple inputs. This multi-dimensional asymmetric information increases the complexity of finding an adequate regulation.

In theory as well as in practice, problems of information asymmetry between the regulator and the firm have been tackled by different forms of regulation. Typical approaches in regulatory practice range from cost-based regulation to widely applied incentive regulation (discussed, e.g., in Joskow (2014)), or a linear combination of those two extremes (e.g., Schmalensee (1989)). For instance, the German regulator offers one single contract to electricity transmission firms, dependent on grid expansion, which corresponds to a cost-based regulation of capital.[3] The academic discussion has not yet fully covered the specific multi-dimensional problems of asymmetric information regarding the level and mix of inputs, but more recent theoretical approaches suggest that the best theoretical solution consists of the regulator offering the firm a *menu* of contracts, such that the firm reveals her private information (e.g., Laffont and Tirole (1993)). Even though the dichotomy between such Bayesian models of regulation (which tend to dominate the academic discussion) and simpler non-Bayesian models (which are closer to regulatory practice) is well perceived, corresponding explanations are rather vague. For instance, as Arm-

[1]The German Transmission System Operators estimate the necessary investments into grid reinforcements and expansion to be around 22 bn. for the period 2013-2022 (Netzentwicklungsplan (2013)), which doubles the annual figures for 2012 and quadruples the value for 2006 (Monitoringbericht (2013)).

[2]For instance, in Germany the regulator has defined five observable, quantifiable dimensions for measuring grid quality.

[3]In Germany, transmission system operators formulate a network expansion plan for which they get an allowed investment. In line with economic theory, the chosen levels may be suspected to be inefficiently high (see Footnote 1 for related cost figures). This regulation corresponds to a cost-based regulation for the input factor grid expansion, while neglecting any other possible input, such as better operational measures. Obviously, this triggers some sort of Averch-Johnson-effect and leads to suboptimal distortions of the input levels.

strong and Sappington (2007) note, [...] *regulatory plans that encompass options are 'complicated', and therefore prohibitively costly to implement*.

The goal of this paper is twofold: First, to identify and investigate the optimal Bayesian regulation for the multi-dimensional problem at hand, and second, to bridge the gap between the theoretically optimal solution and simpler regimes applied in regulatory practice.

To derive an optimal regulation strategy, we build on the theory of incentives and contract menus. It is well known that in a simple setting with two types of the firm, the efficient type is incentivized via a contract with first-best (price) levels along with some positive rent, while the inefficient type's contract includes prices below the first-best and no rent (e.g., Laffont and Tirole (1993)). This analysis has been extended to represent multiple dimensions of information asymmetry in terms of adverse selection, e.g., by Lewis and Sappington (1988b), Dana (1993), Armstrong (1999) or Aguirre and Beitia (2004). While Dana (1993) analyzes a multi-product environment, Lewis and Sappington (1988b), Armstrong (1999) and Aguirre and Beitia (2004) consider two-dimensional adverse selection with only one screening variable. Specifically, the latter three derive optimal regulation strategies in a marketed-good environment (in the sense of Caillaud et al. (1988)) with unknown cost and demand functions. In our paper, unlike Lewis and Sappington (1988b) and Armstrong (1999), we consider shadow costs of public funding instead of distributional welfare preferences. Despite technical differences, this is largely in line with the analysis of Aguirre and Beitia (2004).[4] However, in contrast to all these papers, we solve the two-dimensional adverse selection problem for a non-marketed good environment and a production process that involves two substitutable inputs with an uncertain isoquant and input factor costs.[5]

For the novel setting of multi-dimensional inputs and a non-marketed output, we are able to confirm the general insights from the above literature. We find that expected social welfare necessarily includes positive rents for some types of the firm, such that the first-best solution cannot be achieved. While the efficient type is always set to first-best input levels, the other contracts' (observable) input levels are dis-

[4]Aguirre and Beitia (2004) show the difference between shadow costs of public funding and distributional welfare preferences based on a model with continuous probability distribution, while we assume a discrete distribution.

[5]Noticeably, with the (discrete) two-dimensional adverse selection problem, our problem setting is technically closest to the model discussed by Armstrong (1999).

torted upwards.[6] Separation of at least three types is always possible, while bunching of two types may be unavoidable in case of a very asymmetric distribution of costs or very flat isoquants.

We compare the obtained optimal Bayesian regulation to the results of a non-Bayesian regulation that we obtain by restricting our regulation problem to one single contract. We find that despite the general inferiority a non-Bayesian cost-based regulatory regime may indeed be close to the optimal Bayesian solution for specific circumstances. This especially holds true if the overall input level probably needs to be high, and shadow costs of public funding are large. Considering current circumstances observed in the electricity sector, i.e., substantial changes in the supply structure and ongoing intense discussions about grid tariffs, these conditions may indeed prevail.

The paper is organized as follows: Section 5.2 introduces the model, Section 5.3 presents the optimal regulation strategy, Section 5.4 compares the optimal regulation to simpler regimes, and Section 5.5 concludes.

5.2 The model

Consider a single firm that is controlled by a regulator. The firm uses two inputs to provide an output in terms of a good or service level q that is requested by the regulator. The regulator's choice of q could, for instance, result from counterbalancing the economic value of the provided with the related social costs. For simplicity, however, we assume q to be invariant throughout the paper. Although this assumption might seem restrictive at first sight, it may indeed fit a number of relevant cases very well. For instance, due to the very high societal value of uninterrupted electricity transmission, changes in costs will hardly affect the desired level of the transmission service quality q.

In our model, probability μ (respectively $1-\mu$) leads to a low (high) aggregated input that is necessary to reach the same requested output q. This could, e.g., be an exogenous shock induced by the increased deployment of renewable energies, triggering a changing spatial distribution of supply and hence impacting the neces-

[6] Upwards distorted observable input levels coincide with upwards distorted prices for the inefficient type as shown in Laffont and Tirole (1993). They also agree with the results in a setting with unknown cost and demand functions as long as shadow costs of public funding are considered (Aguirre and Beitia (2004)). Noticeably, the case of prices below marginal costs, as found in Lewis and Sappington (1988b) and Armstrong (1999), is mainly triggered by using a distributive social welfare function instead of shadow costs of public funding.

sary overall activity level in the grid sector to achieve secure electricity transmission. From the firm's perspective, an output level q can be provided by means of two different inputs, one of which is observable (x) and one non-observable (y) by the regulator. Stressing again our introductory example of electricity transmission services, x could be the level of grid expansion that is easily observable by the regulator, even by people unfamiliar with the details of electricity transmission. Utilization and measures of sophisticated grid operation, especially as a partial substitute to grid expansion, however, are hardly observable. The tradeoff between those two inputs needed to reach output q is commonly described by a production function $q = f(x, y)$ which can be illustrated by means of isoquants. We assume smooth and decreasing marginal returns of both inputs, such that the isoquants are downward sloping, convex and differentiable. Noticeably, two different isoquants can never cross. An example fulfilling these requirements is a Cobb-Douglas-type production function. The inverse production function $g(q, x)$ reflects the necessary level of the non-observable input y needed to reach output q, given a level of x. We will mostly use this inverse function hereafter. Due to the exogenous shock leading to a low (l) or high (h) aggregated input necessary for the envisaged output level q, the inverse function takes one of two possible functional forms, i.e. $g_i(q, x)$, with $i \in [l, h]$ and $g_l(q, x) < g_h(q, x)$.

The optimal rate of substitution between the two inputs minimizing total costs for reaching the requested output depends on the cost functions of the inputs. We consider the cost function $c^x(x)$ of the the observable input to be fixed and common knowledge, while the cost function of the non-observable input $c_j^y(y)$ is subject to a nature draw, which leads with probability ν (respectively $1 - \nu$) to a low (high) cost function (i.e., $j \in [l, h]$). For simplicity, we assume constant factor costs of both inputs, i.e., $c^x(x) = c^x$ and $c_j^y(y) = c_j^y$. The realization of c_j^y influences the isocost line of the two inputs and hence, the optimal rate of substitution.[7] Hence, depending on the two random draws for the isoquant and the costs of the non-observable input, there are four possible firstbest bundles of inputs, which we denote by $\{x_{ll}^{fb}, y_{ll}^{fb}\}$, $\{x_{lh}^{fb}, y_{lh}^{fb}\}$, $\{x_{hl}^{fb}, y_{hl}^{fb}\}$ and $\{x_{hh}^{fb}, y_{hh}^{fb}\}$. As a last precondition, we require the expansion path, i.e. the curve connecting the optimal input combinations of the different isoquants, to be pointing rightwards as the necessary aggregated

[7] As it is well known from production theory, the optimal rate of substitution is determined by equating the marginal rate of technical substitution between the factors (i.e., the slope of the isoquant) with the relative factor costs (i.e., the slope of the isocost line).

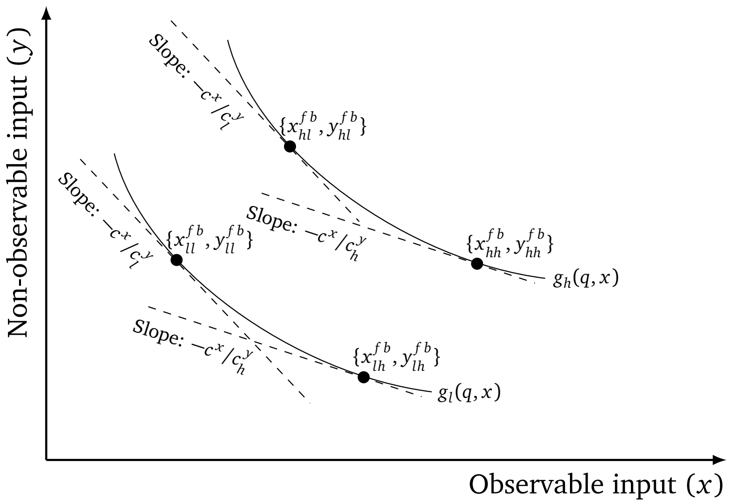


Figure 5.1: Problem setting with double adverse selection

input increases.[8] In terms of the first-best input levels, this requires $x_{ll}^{fb} > x_{hl}^{fb}$ and $x_{lh}^{fb} > x_{hh}^{fb}$, which again holds true for a wide range of possible production function specifications, including the above mentioned Cobb-Douglas type.

Under optimal Bayesian regulation, the goal of the regulator is to incentivize the firm via a suitable contract framework to choose the welfare-optimizing bundle of inputs, which we will derive based on classic mechanism design entailing truthful direct revelation. Contrary to the firm, the regulator cannot observe the realizations of the two random draws, although the possible realizations as well as the occurrence probabilities are common knowledge. She knows the cost function of the observable input and can observe the corresponding input level. The output is also observable and verifiable.[9] For an optimal regulation, the regulator offers the firm a menu of four contracts, each with a level of the observable input x_{ij} and a corresponding transfer T_{ij}. Naturally, the contracts can be conditioned on observable parameters only, i.e., the output as well as the amount of the observable input used. Both are enforceable by means of suitably high penalties in case the firm deviates from the requested/contracted level.

The timing – as shown in Figure 5.2 – is as follows. First, the random draws are realized and the cost function of the non-observable input and the necessary

[8]For an analysis involving continuous variables, this would require the expansion path to behave like a function with a unique function value y for each x, or, in other words, an expansion path that is not bending backwards.

[9]Stochastic deviations due to force majeure are supposed to be detectable and excludable from the contract framework.

aggregated input relation (isoquant) are observed by the firm. The firm then chooses between several (in our case, four) contracts offered by the regulator. She then realizes the input levels to produce the requested output. The regulator observes one input level (x) and whether the output is as requested; if those are as agreed upon, the contract is executed and the transfer realized.

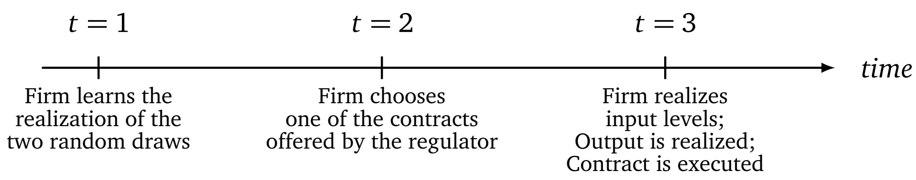


Figure 5.2: Timing

The rent of the firm R_{ij} given a realization $i \in [l, h]$ and $j \in [l, h]$, results from the transfer T_{ij} minus the private cost of the firm's activities:[10]

$$R_{ij} = T_{ij} - c^x x_{ij} - c_j^y g_i(q, x_{ij}) \tag{5.1}$$

The regulator maximizes expected social welfare, defined as the sum of expected social utility and firm surplus, by adjusting the observables, i.e.:

$$\max_{x_{ij}, T_{ij}} W = \mathbb{E}\left[\underbrace{S_q - (1+\lambda)T_{ij}}_{\text{Net social utility}} + \underbrace{(T_{ij} - c^x x_{ij} - c_j^y g_i(q, x_{ij}))}_{\text{Firm's rent } (R_{ij})}\right] \tag{5.2}$$

where S_q is the gross social utility from reaching output q, and λ denotes the shadow costs of public funding, i.e., the costs due to raising and transferring finances through public channels (for a discussion, see, e.g., Laffont and Tirole (1993)). As discussed previously, we assume q – and hence also gross social utility S_q – to be invariant and independent of the random draws, yielding[11]

$$\max_{x_{ij}, T_{ij}} W = S_q - \mathbb{E}\left[\underbrace{(1+\lambda)T_{ij}}_{\text{Transfer costs}} - \underbrace{(T_{ij} - c^x x_{ij} - c_j^y g_i(q, x_{ij}))}_{\text{Firm's rent } (R_{ij})}\right] \tag{5.3}$$

[10]It goes without saying here that the firm is characterized such that she tries to maximize her rent.

[11]This is the reason why q appears as a subscript here. In case of a more complex analysis involving q as a variable, S_q would be replaced by $S(q, x)$ to reflect the counterbalancing of the economic value of the provided output with the related social costs.

As an important consequence of Equation (5.3), we see that the optimization problem of the regulator can be reformulated in terms of a cost-minimization problem, essentially stating that the desired output shall be reached at minimal expected social costs:

$$\min_{x_{ij},T_{ij}} C = \mathbb{E}\left[C_{ij}\right] = \mathbb{E}\left[\lambda \underbrace{R_{ij}}_{\text{Firm's rent}} + (1+\lambda)(\underbrace{c^x x_{ij}}_{\substack{\text{Costs of}\\ \text{observable input}}} + \underbrace{c_j^y g_i(q, x_{ij})}_{\substack{\text{Costs of}\\ \text{non-observable input}}})\right] \tag{5.4}$$

While choosing x_{ij} and T_{ij} such that social costs are minimized, the regulator is restricted by several participation and incentive constraints for the firm's rent:

$$R_{ij} \geq 0 \qquad \forall i, j \tag{5.5}$$

$$R_{ij} \geq R_{i'j'} + c_{j'}^y g_{i'}(q, x_{i'j'}) - c_j^y g_i(q, x_{i'j'}) \qquad \forall \text{ pairs } i, j \text{ and } i', j' \tag{5.6}$$

Equation (5.5) ensures that all types of firms have a non-negative profit and therefore participate.[12] In line with the revelation principle, Equation (5.6) provides the firm with the incentive to truthfully report the realized isoquant and non-observable input costs.

Written explicitly, the four participation constraints for the four possible firm types become

$$R_{ll} \geq 0 \tag{5.7a}$$

$$R_{lh} \geq 0 \tag{5.7b}$$

$$R_{hl} \geq 0 \tag{5.7c}$$

$$R_{hh} \geq 0, \tag{5.7d}$$

and the twelve incentive constraints (each of the four types might be tempted to

[12] Hence, we implicitly assume zero liability for the firm.

choose a contract of one of the other three types)

$$R_{ll} \geq R_{lh} + c_h^y g_l(q, x_{lh}) - c_l^y g_l(q, x_{lh}) \tag{5.8a}$$
$$R_{ll} \geq R_{hl} + c_l^y g_h(q, x_{hl}) - c_l^y g_l(q, x_{hl}) \tag{5.8b}$$
$$R_{ll} \geq R_{hh} + c_h^y g_h(q, x_{hh}) - c_l^y g_l(q, x_{hh}) \tag{5.8c}$$
$$R_{lh} \geq R_{ll} + c_l^y g_l(q, x_{ll}) - c_h^y g_l(q, x_{ll}) \tag{5.8d}$$
$$R_{lh} \geq R_{hl} + c_l^y g_h(q, x_{hl}) - c_h^y g_l(q, x_{hl}) \tag{5.8e}$$
$$R_{lh} \geq R_{hh} + c_h^y g_h(q, x_{hh}) - c_h^y g_l(q, x_{hh}) \tag{5.8f}$$
$$R_{hl} \geq R_{ll} + c_l^y g_l(q, x_{ll}) - c_l^y g_h(q, x_{ll}) \tag{5.8g}$$
$$R_{hl} \geq R_{lh} + c_h^y g_l(q, x_{lh}) - c_l^y g_h(q, x_{lh}) \tag{5.8h}$$
$$R_{hl} \geq R_{hh} + c_h^y g_h(q, x_{hh}) - c_l^y g_h(q, x_{hh}) \tag{5.8i}$$
$$R_{hh} \geq R_{ll} + c_l^y g_l(q, x_{ll}) - c_h^y g_h(q, x_{ll}) \tag{5.8j}$$
$$R_{hh} \geq R_{lh} + c_h^y g_l(q, x_{lh}) - c_h^y g_h(q, x_{lh}) \tag{5.8k}$$
$$R_{hh} \geq R_{hl} + c_l^y g_h(q, x_{hl}) - c_h^y g_h(q, x_{hl}). \tag{5.8l}$$

5.3 Optimal regulation

5.3.1 Preparatory analysis

As a first preparatory step in the analysis we shall check whether the contract variable x is actually suitable to provide incentives to the firm to reveal her true type. To this end, we investigate whether the incentive to choose another type's contract (motivated by a potential increase in rent) regarding one of the two random draws is impacted by an adjustment of x. This is often referred to as "single crossing" conditions. For the incentive to choose another type's contract regarding the realized input cost, we find that[13]

$$\frac{\partial}{\partial x}(R_{ih}(x) - R_{il}(x)) = (c_l^y - c_h^y) g_i'(q, x) \qquad \text{for} \qquad i = l, h, \tag{5.9}$$

which is clearly greater than zero due to $c_h > c_l$ and $g_i'(q, x) < 0$. Hence, by an upwards distortion of x, we are able to reduce the incentive for the firm to choose the contract of a high cost type instead of truly revealing the realized low cost type.

[13]Here and in the following, a prime denotes derivation with respect to x.

Similarly, for the incentive to choose a contract for an isoquant different from the realized one, we find that

$$\frac{\partial}{\partial x}(R_{hj}(x) - R_{lj}(x)) = c_j^y (g_l'(q,x) - g_h'(q,x)) \qquad \text{for} \qquad j = l,h \tag{5.10}$$

which is greater than zero as long as $g_h'(q,x) < g_l'(q,x)$. Recalling from Section 5.2 that we have assumed rightwards pointing expansion paths (a property exhibited by a wide range of possible production function specifications, including the Cobb-Douglas type), this condition will always hold true. Hence, upwards distorting x will provide a possibility to reduce the incentive for the firm to choose the contract with a high isoquant instead of truly revealing the realized low isoquant.

The effect of changing incentives following a distortion of x helps us to derive a first characterization of the optimal solution of our regulatory problem. In fact, in order to comply with the incentive constraints (5.8a)-(5.8l) (which need to be fulfilled for the optimal solution anyway), input levels x_{ij} need to follow a certain ordering. Note that for each pair of types there are two relevant incentive constraints (e.g., Equations (5.8a) and (5.8d) for the types ll and lh). Adding those and using the above single crossing conditions, the necessary ordering can be obtained as follows:[14]

$$x_{ll} \leq x_{lh} \leq x_{hh} \tag{5.11}$$

$$x_{ll} \leq x_{hl} \leq x_{hh} \tag{5.12}$$

Moreover, from the incentive constraints (5.8a) and (5.8i) it follows that only the participation constraints (5.7b) and (5.7d) (i.e., limited liability of the lh and the hh-type) remain relevant for further analyses. In contrast, the other two participation constraints (those of the low-cost types) are implicitly fulfilled if these two incentive constraints hold.

So far unclear from the above analysis, however, is the ordering of the intermediate cases x_{lh} and x_{hl}, which depends on whether the term $R_{hl}(x) - R_{lh}(x)$ is increasing or decreasing in x. Differentiating with respect to x yields

$$\frac{\partial}{\partial x}(R_{hl}(x) - R_{lh}(x)) = (c_h^y g_l'(q,x) - c_l^y g_h'(q,x)) \tag{5.13}$$

[14]For instance, adding Equations (5.8a) and (5.8d) yields $(c_l^y - c_h^y) g_l'(q, x_{lh}) \geq (c_l^y - c_h^y) g_l'(q, x_{ll})$, which, together with (5.9), implies that $x_{lh} \geq x_{ll}$.

which is increasing in x as long as

$$\frac{c_h^y}{c_l^y} < \frac{g_h(q,x)}{g_l(q,x)}, \tag{5.14}$$

and decreasing in x otherwise. Together with incentive constraints (5.8e) and (5.8h) we infer that if the cost variation is small compared to the isoquant variation, then $x_{lh} \leq x_{hl}$. If the aggregated input level variation is small compared to the cost variation, then $x_{lh} \geq x_{hl}$. For an intuition, recall Figure 5.1. If the aggregated input level variation and hence the distance between the isoquants is large, x_{hl}^{fb} is larger than x_{lh}^{fb}. If the cost variation, and hence, the vertical distance between the corresponding first-best solutions is large, x_{lh}^{fb} is larger than x_{hl}^{fb}.

The results of our preparatory analysis are summarized in the following two Lemmas.

Lemma 1. *Limited liability is only an issue for the high-cost types. Hence, the only relevant participation constraints are* (5.7b) *and* (5.7d)*, whereas* (5.7a) *and* (5.7c) *are implicitly fulfilled.*

Lemma 2. *In order to reach incentive compatibility, input levels x_{ij} must be ordered as follows:*

(A) If the cost variation is small compared to the isoquant variation, then $R_{hl}(x) - R_{lh}(x)$ is increasing in x and requires

$$x_{ll} \leq x_{lh} \leq x_{hl} \leq x_{hh}. \tag{5.15}$$

(B) If the cost variation is large compared to the isoquant variation, then $R_{hl}(x) - R_{lh}(x)$ is decreasing in x and requires

$$x_{ll} \leq x_{hl} \leq x_{lh} \leq x_{hh}. \tag{5.16}$$

5.3.2 Full information benchmark

If the regulator had no information deficit, she would observe the realized isoquant as well as the realized isocost line. Differentiating all possible realizations of the social cost function C_{ij} with respect to the observable input levels x_{ij} shows that all of them are single-peaked with a unique minimum at $g_i^{'}(q, x_{ij}) = -\frac{c^x}{c_j^y}$, which is necessarily realized at $x_{ij} = x_{ij}^{fb}$. The regulator would easily derive the first-best levels of inputs to supply the requested output at minimal social costs, i.e., $\{x_{ij}^{fb}, y_{ij}^{fb}\}$, by equating the known realized marginal rate of technical substitution of the inputs with the realized isocost line. Moreover, she would be able to enforce the implementation of the first-best due to the full observability. The corresponding optimal transfers would be $T_{ij}^{fb} = c^x x_{ij}^{fb} + c_j^y y_{ij}^{fb}$, leaving all types of the firm with zero rent. In the case of full information, social costs amount to $C_{ij}^{fb} = (1+\lambda)T_{ij}^{fb} = (1+\lambda)(c^x x_{ij}^{fb} + c_j^y y_{ij}^{fb})$, corresponding to the welfare-optimizing first-best solution that could thus be obtained.

5.3.3 Asymmetric information

In the case of asymmetric information, the only two observables for the regulator are the output q and the observable input x. In addition, she can choose an appropriate level of transfer payment T. As q is invariable and observable, its implementation can be enforced by means of suitably high penalties in case the firm deviates. Hence, x and T are the two variables the regulator will condition her contracts on. The general idea for the regulator's optimal regulation strategy is to offer a menu of contracts with optimized variables $\{x_{ij}^*, T_{ij}^*\}$, such that expected social costs are minimized (as stated in Equation (5.4)), and participation (Equation (5.5)) and incentive constraints (Equation 5.6) fulfilled. Hence, we restrict our attention to incentive compatible contracts ensuring that the firm always reveals her true type. Under these conditions, the revelation principle requires that the solution found (if any) is a Bayesian-Nash equilibrium (Myerson (1979), Laffont and Martimort (2002)).

One-dimensional asymmetric information

We shall first investigate a simplified problem with one-dimensional asymmetric information only, i.e., isoquant *or* cost uncertainty. Eliminating the isoquant uncertainty (by setting $\mu = 0$, $\mu = 1$ or $g_{lj} = g_{hj}$), we are left with two constraints binding: the participation constraint of the high cost type (5.7b or 5.7d) and the incentive constraint from the low to the high cost type (5.8a or 5.8i). This leads to the simplified cost function:

$$\begin{aligned} C = &\nu\big[\lambda\left(g_i(x_{ih})(c_h^y - c_l^y)\right) + (1+\lambda)\left(c^x x_{il} + c_l^y g_i(x_{il})\right)\big] \\ &+ (1-\nu)\big[(1+\lambda)\left(c^x x_{ih} + c_h^y g_i(x_{ih})\right)\big] \end{aligned} \tag{5.17}$$

Derivating with respect to $x_{ij}, j \in l, h$ yields the following first order conditions:

$$\frac{\partial C}{\partial x_{il}} = 0 \Leftrightarrow g_i'(x_{il}^*) = -\frac{c^x}{c_l^y}, \tag{5.18}$$

$$\frac{\partial C}{\partial x_{ih}} = 0 \Leftrightarrow \underbrace{\nu\lambda(c_h^y - c_l^y)g_i'(x_{ih}^*)}_{<0} + \underbrace{(1-\nu)(1+\lambda)(c^x + c_h^y g_i'(x_{ih}^*))}_{\substack{=0 \text{ for } x_{ih}=x_{ih}^{fb} \\ <0 \text{ for } x_{ih}<x_{ih}^{fb} \\ >0 \text{ for } x_{ih}>x_{ih}^{fb}}} = 0 \tag{5.19}$$

Similarly, in case of no cost uncertainty, the observable input levels of the low isoquant types are first-best, whereas the high isoquant types are distorted upwards:[15]

Lemma 3. *In case of asymmetric information about either costs or isoquants, the respective l-type is set to first-best, while the h-type is distorted upwards compared to its first-best.*

Note that the result of an adverse selection problem with one-dimensional information asymmetry on costs is well-known from the literature (e.g., Baron and Myerson (1982) or Sappington (1983)). Also note that the results concerning isoquant uncertainty are strikingly different compared to the one-dimensional demand uncertainty (which essentially corresponds to the isoquant in our setting) studied by Lewis and Sappington (1988a) or Armstrong (1999). In contrast to our model – due to neglecting shadow costs of public funding – they find that the first-best can be achieved in the one-dimensional case of demand uncertainty.

Two-dimensional asymmetric information

Solving the full optimal regulation problem requires minimization of social costs, subject to all imposed four participation and twelve incentive constraints. Due to the large number of constraints, we approach the optimization by solving a relaxed problem where only a subset of the constraints is considered. To this end, we need to come up with an educated guess about the binding constraints in the optimum. If

[15]Due to the obvious symmetry of the problem, we omit the detailed calculation here.

we can later show that the remaining constraints are fulfilled at the solution of the relaxed problem, we will have obtained the solution of the full problem.

We already know from Lemma 1 that the participation constraints of the high-cost types are the only relevant ones. Furthermore, it generally seems to be a good approach to assume the "upwards" incentive constraints, i.e., from low to high isoquant, and from low to high costs, to be binding. Moreover, it seems plausible to assume binding incentive constraints from the most efficient to an intermediate type (i.e., lh or hl), and from an intermediate type to the least efficient type. If we consider the isoquant variation more relevant than the cost variation, assuming the incentive constraints according to the ordering shown in Lemma 2, Case *(A)*, to be binding appears to be the most educated guess we can come up with.[16] Hence, we assume that incentive constraints (5.8a) ($ll \to lh$), (5.8e) ($lh \to hl$), and (5.8i) ($hl \to hh$) are fulfilled with equality. In addition, we assume the participation constraint of the hh-type to bind since this is the only type remaining that is not attracted by any other type. Figure 5.3 illustrates with arrows the binding incentive constraints, such that the former type is not attracted by the latter type-contract. Diamonds mark the binding participation constraints.

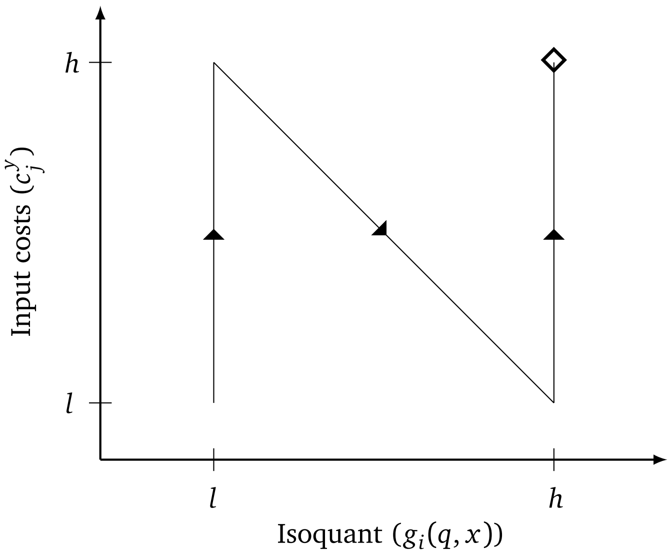


Figure 5.3: Constraints considered binding for Case *(A)*

We find that this set of assumptions does indeed lead us to the optimal regulation strategy. The results are summarized in the following Proposition 1.

[16]The ordering and solution of Case *(B)* is reversed, but similar. The corresponding discussion can be found in the appendix.

Proposition 1. *For Case (A),*

(i) Optimal regulation is achieved under the following set of observable input levels:

$$x^*_{ll} = x^{fb}_{ll} \tag{5.20}$$
$$x^*_{lh} \geq x^{fb}_{lh} \tag{5.21}$$
$$x^*_{hl} \geq x^{fb}_{hl} \tag{5.22}$$
$$x^*_{hh} \geq x^{fb}_{hh}, \tag{5.23}$$

while respecting $x^*_{ll} < x^*_{lh} \leq x^*_{hl} \leq x^*_{hh}$.

(ii) The most efficient (ll*) type can always be separated. Moreover, separation of at least three types is always possible, while bunching of the* lh *and* hl *types is unavoidable in case of* $\nu \to 1$. *The* hl *and* hh *types may need to be bunched in case of* $g'_l(q,x) \to 0$ *together with* c^y_l *being large.*

Proof. See Appendix. □

Corollary 1. *For* $\lambda = 0$*, the optimal solution is first-best. All input levels amount to* $x^*_{ij} = x^{fb}_{ij}$*, and expected social costs to* $C = C^{fb}$.

Proof. Follows immediately from the solution of Case *(A)* when setting $\lambda = 0$. □

According to Corollary 1, with no shadow costs of public funding, all input levels x^*_{ij} are first-best. The regulator optimizes overall welfare, but has no preference regarding the distribution of social surplus. Hence, she can give the firm an arbitrarily high budget at no social costs, and the firm maximizes her rent by setting efficient input levels. In this case, the maximization of the firm and the maximization of social welfare coincide, i.e., there is no problem of aligning the activities of the firm with social interests. Of course, larger parts of the welfare are then given to the firm.

For the general case of $\lambda > 0$, observable input levels of all types besides the ll-one are distorted above first-best levels, leading to a second-best solution only. Naturally, the overall level of inefficiency increases in λ, but also for decreasing μ and ν (i.e., when there is a high probability for "costly" outcomes of the random draws) as well as for $c^y_h - c^y_l$ and $g_h(q,x) - g_l(q,x)$ getting large. In contrast, however, the less significant the cost variation becomes compared to the isoquant variation, the more efficient the solution will be.

Due to keeping the most efficient (ll) type at first-best level combined with the ordering according to Lemma 2, the type can always be separated in the contract framework. Moreover, we find that at least three types can always be separated, while bunching of two types may be unavoidable in case of vanishing isoquant or cost uncertainties, or if the isoquant variation becomes extremely large. As a last remark, it is worth mentioning that the ordering of rents is (and must be) as depicted in Figure 5.3, i.e. $0 = R^*_{hh} < R^*_{hl} < R^*_{lh} < R^*_{ll}$.

The results for Case *(B)* are symmetric but structurally identical to Case *(A)*, i.e., the ll-type is incentivized to first-best input levels while the other types show upwards distortions of x_{ij}. However, roles of isoquants and costs are interchanged, reflected in the inverse occurrence of the terms $g_i \longleftrightarrow c^y_j$ and $\mu \longleftrightarrow \nu$. At the same time, as imposed by Lemma 2, Case *(B)*, the sequence of the "intermediate" types is now $hl \rightarrow lh$. Hence, the ordering of observable input levels x_{lh} and x_{hl} as well as rents R_{lh} and R_{hl} need to be reversed to obtain an optimal regulatory contract framework.[17]

5.4 Comparing the optimal regulation to simpler regimes

In contrast to the optimal Bayesian menu of contracts studied in the previous section, regulatory authorities often apply alternative, simpler approaches. In fact, in the case of electricity transmission grids, it appears that they mostly offer a non-Bayesian, i.e., *single*, contract, while the application of Bayesian contracts in terms of *menus* of contracts, has been very rare.[18] For instance, regulatory practice in Germany is such that TSOs formulate a grid expansion plan, which is then reviewed and approved by the regulator. For the approved measures, the TSOs get their costs reimbursed. This corresponds to a cost-based regulation for the input factor grid expansion, while neglecting any other possible input, such as better operational measures. Meanwhile, driven e.g. by social acceptance issues, the regulator is expected to limit the approval of extensive grid expansion to some reasonable level.

Transferring such a simple non-Bayesian approach into our model, we need to limit the set of regulatory choice variables to one single contract with contract variables $\bar{x}$ and $\bar{T}$, such that the objective function of the regulator (in contrast to Equa-

[17] See the appendix for a detailed discussion and the corresponding proposition and proof.
[18] The system operator for England and Wales and the electric distribution companies in the UK are the only two examples for menus of contracts being applied in regulatory practice Joskow (2014).

tion (5.4) as in the case of optimal regulation) becomes:

$$\min_{\bar{x},\bar{T}} \bar{C} = \mathbb{E}\left[\lambda \underbrace{\bar{R}_{ij}}_{\text{Firm's rent}} + (1+\lambda)\Big(\underbrace{c^x \bar{x}}_{\substack{\text{Costs of}\\ \text{observable input}}} + \underbrace{c_j^y g_i(q,\bar{x})}_{\substack{\text{Costs of}\\ \text{non-observable input}}} \Big)\right] \tag{5.24}$$

In contrast to the solution of the optimal regulation, this minimization is only subject to the participation constraints (5.5). With a sole contract and hence, only one observable input $\bar{x}$ for all types, the regulator has no possibility to separate types, which makes the incentive constraints obsolete. As before, the only participation constraint holding with equality is the one of the *hh*-type. Considering that this type gets full cost reimbursement but cannot be distinguished from the other types, it becomes clear that all other types must then necessarily receive a positive rent. The following proposition summarizes the solution of this non-Bayesian regulatory approach.[19]

Proposition 2. *Under a single contract cost-based regulation with quantity restriction, the optimal input level $\bar{x}^*$ represents an expected average of the first-best solutions of the four possible types, adjusted by some upwards distortion in case of $\lambda > 0$. As an expected average, it lies between the extreme types' first-best input levels, i.e. $x_{ll}^{fb} < \bar{x}^* < x_{hh}^{fb}$.*

Proof. See Appendix. □

Proposition 3. *Compared to a single contract, the regulatory approach based on a menu of contracts is superior with respect to expected social welfare.*

Proof. It is easy to show that the optimal solution of the single contract is a feasible solution of the menu of contracts problem. Due to the fact that the solution for the menu of contracts, as stated in Proposition 1, is both optimal and different from the one in Proposition 2, it must necessarily be superior. □

As stated in Proposition 3, the solution of the single contract regime is always inferior to the one obtained with the menu of contracts. Nevertheless, the characteristics of the different regimes can be compared and deserve a closer look. We contrast the

[19]Note that the solution for a pure cost-based regulation without quantity restriction would simply reimburse the costs of the observable input. This would incentivize the firm to choose infinitely high values of x (known as the gold-plating effect). Assuming that the regulator restricts her set of choices by an upper level of $\bar{x} = x_{hh}^{fb}$ in order to limit excessive (socially costly) rents, all types would then choose this level. In contrast to this *very* simple approach, the regulatory regime considered in this section makes use of being able to use the observable input x as a contracting variable.

outcome of the optimal menu of contracts with the one of the single contract regime considering three aspects: input levels, cost-efficiency of the input levels, and rents of the firm.

Input levels for the different types have been characterized in Proposition 1 for the menu of contract, stating that all types besides the ll-one are distorted above first-best levels. According to Proposition 3, the optimal input level for the single contract regime, $\bar{x}^*$, represents an expected average of the first-best solution of the four possible types, adjusted by some upwards distortion in case of $\lambda > 0$. Hence, chosen input levels are generally different. However, $\bar{x}^*$ may get close to x^*_{hl} in case of λ being large and μ small (i.e., for a high probability of realizing a high isoquant). At the same time, it will never be as high as x^*_{hh}, due to $\bar{x}^* < x^{fb}_{hh} < x^*_{hh}$.

Cost-efficiency of the input levels is closely connected to the input levels and their deviation from the first-best optimal solution. The optimal menu of contracts approaches first-best cost-efficiency of the input levels for $\lambda \to 0$, as input levels then converge towards first-best levels, i.e., $\{x^*_{ij}, y^*_{ij}\} \to \{x^{fb}_{ij}, y^{fb}_{ij}\}$. In contrast, cost-efficiency is poor for the single contract regime under this condition. However, first-best input levels may also be reached, but only under very restrictive conditions, namely if $\lambda \to 0$ *and* the occurrence probability for one specific type is particularly large (e.g., if $\mu, \nu \to 1$). Type-specific as well as expected cost-efficiency of input levels is (only) then approaching first-best optimality for both contracting frameworks. For the general case of $\lambda \geq 0$, it is clear that cost-efficiency of the input levels is inferior for the ll type in the single contract regime, while the ordering is ambiguous for all other types, depending on the optimal choice of $\bar{x}^*$ in comparison to x^*_{ij}.

Regarding **rents of the firm**, remember that they are only an issue for social welfare if there are shadow costs of public funding, i.e., if $\lambda > 0$. Then, however, the well known trade-off for rent-extraction and efficiency becomes relevant. For both contracting regimes, the rent of the inefficient hh type is set to zero. Moreover, for both regimes it holds true that $0 = R^*_{hh} < R^*_{hl} \ll R^*_{lh} < R^*_{ll}$ (respectively, $0 = \bar{R}^*_{hh} < \bar{R}^*_{hl} \ll \bar{R}^*_{lh} < \bar{R}^*_{ll}$), if isoquant variation is more relevant than cost variation. For the rent of specific types, we find that $R^*_{hl} < \bar{R}^*_{hl}$, while the ordering of other types' rent is generally ambiguous. Interestingly, however, if $g'_i(x)$ is small in the relevant range, $R^*_{ij} < \bar{R}^*_{ij}$ for all i, j.

Based on the above comparative statics, a singular interesting constellation can be identified for which the two contracting frameworks effectively approach each other.

Proposition 4. *For λ being large and μ small, the performance of the single contract is close to the one of the menu of contracts.*

Proof. For λ being large, $\bar{x}^*$ is distorted upwards (see Proposition 2), while $x^*_{hl} \approx x^{fb}_{hl}$ for μ small. Hence, in this case, $\bar{x}^* \approx x^*_{hl}$. Moreover, due to the fact that we consider Case (A) where cost uncertainty is relatively low, we know that the upwards distortion of x^*_{hh} is low (see Equation (5.29)), such that x^*_{hl} is not far from x^*_{hh}. Under these conditions, $\bar{C}^* \approx C^*$. □

Transferring Proposition 4 to our example of electricity transmission services and the regulation of the German TSOs, one may indeed come to the conclusion that the practically applied non-Bayesian regulatory approach could be close to the optimal second-best strategy. In fact, a high overall input level appears to be likely due to the strongly changing supply infrastructure, while ongoing intense discussions about the burden of electricity costs and grid tariffs for consumers could indicate high shadow costs of public funding. In the end, however, reasons for the chosen regulation are probably manifold, and might also include an explicit disutility of grid expansion, a commitment problem,[20] or the prohibitively high costs of implementing a 'complicated' regulatory regime (Armstrong and Sappington, 2007).

5.5 Conclusion

We considered a regulated firm providing a non-marketed output with substitutable inputs. We presented the optimal Bayesian regulation in terms of a menu of contracts when the regulator faces information asymmetries regarding the aggregated input level needed to provide the output as well as the realized optimal marginal rate of substitution between the inputs. Finally, the optimal Bayesian regulation was compared to a simpler non-Bayesian approach which appears to be closer to regulatory practice.

We found that in the optimal Bayesian regulation, the first-best solution cannot be

[20]Noticeably, a commitment problem of the regulator might impede the implementation of an incentive-based approach, which would be welfare-superior compared to a cost-based regulation. If the firm gets an unconditional payment representing the pay-off of the *hh*-type, i.e., $\tilde{T} = c^x x^{fb}_{hh} + c^y_h g_h(q, x^{fb}_{hh})$, she will realize first-best input quantities $\{x^{fb}_{ij}, y^{fb}_{ij}\}$. In this case, the realized rent of the firm becomes $R_{ij} = c^x x^{fb}_{hh} + c^y_h g_h(q, x^{fb}_{hh}) - c^x x^{fb}_{ij} - c^y_j g_i(q, x^{fb}_{ij})$. However, due to the (observable) separation of types via the realized input x, the regulator might be tempted to adjust the regulatory contract ex-post, and hence, jeopardize the regulatory success if the firm anticipates this behavior.

achieved under the considered information asymmetries and shadow costs of public funding. This implies a strictly positive rent for the firm. The second-best solution that we then characterized depends on the relative importance of the information asymmetries. However, the most efficient type is always set to first-best, while the levels of the observable input are distorted upwards for all other types. At least three types can always be separated, while bunching of two types may be unavoidable in case of a very asymmetric distribution of costs or very flat isoquants. These results are structurally similar to the solutions for multi-dimensional adverse selection problems in the literature (e.g. Lewis and Sappington (1988b), Armstrong (1999) or Aguirre and Beitia (2004)). However, in contrast to existing results, our model explains upwards distortions of input levels rather than prices. Hence, we obtained important insights regarding the optimal mechanism design in the context of a regulated monopolistic firm producing a non-marketed good with multi-dimensional inputs.

The comparison to a single contract cost-based approach, as it is often applied in regulatory practice, showed that the menu of contracts is welfare superior. However, there are situations in which the performance of the approaches converge, namely if the overall input level probably needs to be high, and shadow costs of public funding are large. Given our motivating example of electricity transmission services and the current situation, e.g., in Germany, these circumstances may indeed prevail, possibly explaining the gap between the theoretically optimal Bayesian approach and the simpler non-Bayesian regulation applied in practice.

Lastly, we note that our general approach as well as our insights might also be applicable to other industries that show similar characteristics, such as public works or administrative services. Besides investigating such areas of application, future research could relax the limited liability assumption and hence, allow for a shut down of firms. Another expansion could allow the good to be marketed, which would trigger a demand reaction of the regulator (or consumers) and possibly lead to interesting variations of the conclusions derived in this paper.

Acknowledgments

We thank Felix Höffler and Christian Tode for helpful comments, Clara Dewes for support, and the participants of the IAEE 2014 in Rome and the SMYE 2015 in Ghent for valuable discussions. Funding of the German research society DFG through research grant HO5108/2-1 is gratefully acknowledged.

5.6 Appendix

Proof of Proposition 1

Proof. (i) Under the constraints considered binding for Case *(A)* – as discussed and shown in Figure 5.3 – the social cost function (5.4) becomes

$$
\begin{aligned}
&C = \\
&\mu\nu\Big[\lambda\left(g_h(x_{hh})(c_h^y - c_l^y) + c_l^y g_h(x_{hl}) - c_h^y g_l(x_{hl}) + g_l(x_{lh})(c_h^y - c_l^y)\right) \\
&+(1+\lambda)\left(c^x x_{ll} + c_l^y g_l(x_{ll})\right)\Big] \\
&+\mu(1-\nu)\Big[\lambda\left(g_h(x_{hh})(c_h^y - c_l^y) + c_l^y g_h(x_{hl}) - c_h^y g_l(x_{hl})\right) \\
&+(1+\lambda)\left(c^x x_{lh} + c_h^y g_l(x_{lh})\right)\Big] \\
&+(1-\mu)\nu\Big[\lambda\left(g_h(x_{hh})(c_h^y - c_l^y)\right) + (1+\lambda)\left(c^x x_{hl} + c_l^y g_h(x_{hl})\right)\Big] \\
&+(1-\mu)(1-\nu)\Big[(1+\lambda)\left(c^x x_{hh} + c_h^y g_h(x_{hh})\right)\Big]. \qquad (5.25)
\end{aligned}
$$

To derive the optimal observable input levels, we need to derive the above equation with respect to each of the four possible x_{ij}. Minimizing C with respect to x_{ll} yields

$$g_l'(x_{ll}^*) = -\frac{c^x}{c_l^y}, \qquad (5.26)$$

which implies that $x_{ll}^* = x_{ll}^{fb}$. Derivations of C with respect to x_{lh}, x_{hl} and x_{hh} take the following forms:

$$\frac{\partial C}{\partial x_{lh}} = \underbrace{\mu\nu\lambda(c_h^y - c_l^y)g_l'(x_{lh})}_{<0} + \underbrace{\mu(1-\nu)(1+\lambda)(c^x + c_h^y g_l'(x_{lh}))}_{\substack{=0 \text{ for } x_{lh}=x_{lh}^{fb} \\ <0 \text{ for } x_{lh}<x_{lh}^{fb} \\ >0 \text{ for } x_{lh}>x_{lh}^{fb}}} \qquad (5.27)$$

$$\frac{\partial C}{\partial x_{hl}} = \underbrace{\mu\lambda(c_l^y g_h'(x_{hl}) - c_h^y g_l'(x_{hl}))}_{<0} + \underbrace{(1-\mu)\nu(1+\lambda)(c^x + c_l^y g_h'(x_{hl}))}_{\substack{=0 \text{ for } x_{hl}=x_{hl}^{fb} \\ <0 \text{ for } x_{hl}<x_{hl}^{fb} \\ >0 \text{ for } x_{hl}>x_{hl}^{fb}}} \qquad (5.28)$$

$$\frac{\partial C}{\partial x_{hh}} = \underbrace{(\mu + (1-\mu)\nu)\lambda g_h'(x_{hh})(c_h^y - c_l^y)}_{<0}$$
$$+ \underbrace{(1-\mu)(1-\nu)(1+\lambda)(c^x + c_h^y g_h'(x_{hh}))}_{\substack{=0 \text{ for } x_{hh}=x_{hh}^{fb} \\ <0 \text{ for } x_{hh}<x_{hh}^{fb} \\ >0 \text{ for } x_{hh}>x_{hh}^{fb}}}. \tag{5.29}$$

From Equation (5.27), we see that $\frac{\partial C}{\partial x_{lh}}$ is strictly smaller than 0 for $x_{lh} = x_{lh}^{fb}$ and monotonically increasing in x_{lh}, which implies that $x_{lh}^* > x_{lh}^{fb}$ must always hold. The same logic applies for x_{hl}^* and x_{hh}^*.

(ii) From the fact that $x_{ll}^{fb} < x_{lh}^{fb}$ and the strict upwards distortion of all other types, it follows that the ll-type can always be separated. In order to investigate whether the types lh, hl and hh can be separated or need to be bunched, we proceed as follows: For each of the possible pairs $lh - hl$, $hl - hh$ and $lh - hh$, we check the derivative of C with respect to the former type at the optimal level of x^* of the latter type (derived from the first order condition). If the change in C is greater than 0 we can conclude that we have already surpassed the optimal level of the former type, which then must be smaller than the optimal level of the latter type. In other words, we check the level of upwards distortion for the lh, hl and hh types while considering the necessary ordering of the types according to Lemma 2. For the pair lh-hl, we find that x_{lh}^* may surpass x_{hl}^* in case of $\nu \to 1$, while they are otherwise clearly separated from each. For the pair hl-hh, bunching may occur for $g_l'(q,x) \to 0$ together with c_l^y being large. Furthermore, we find that lh-hh can always be separated, implying that at most two types (i.e., either lh-hl or hl-hh) may be bunched under certain parameter constellations.

Lastly, it is straightforward to check that the remaining constraints are satisfied under the obtained solution of the relaxed problem. Hence, we have indeed obtained to optimal solution for the full regulatory problem we are facing in Case *(A)*. □

Proof of Proposition 2

Proof. Written explicitly, Equation (5.24) becomes

$$\begin{aligned}\bar{C} =& \mu\nu\left[\lambda\left(c_h^y g_h(\bar{x}) - c_l^y g_l(\bar{x})\right) + (1+\lambda)\left(c^x\bar{x} + c_l^y g_l(\bar{x})\right)\right]\\ &+ \mu(1-\nu)\left[\lambda\left(c_h^y g_h(\bar{x}) - c_h^y g_l(\bar{x})\right) + (1+\lambda)\left(c^x\bar{x} + c_h^y g_l(\bar{x})\right)\right]\\ &+ (1-\mu)\nu\left[\lambda\left(c_h^y g_h(\bar{x}) - c_l^y g_h(\bar{x})\right) + (1+\lambda)\left(c^x\bar{x} + c_l^y g_h(\bar{x})\right)\right]\\ &+ (1-\mu)(1-\nu)\left[(1+\lambda)\left(c^x\bar{x} + c_h^y g_h(\bar{x})\right)\right]. \end{aligned} \tag{5.30}$$

Deriving the above with respect to $\bar{x}$ yields, after a few calculations, $\mathbb{E}(g_i'(\bar{x}^*))\mathbb{E}(c_j^y) + c^x + \lambda(c_h^y g_h'(\bar{x}^*) + c^x) = 0$. Hence, for $\lambda = 0$, $\mathbb{E}(g_h'(\bar{x}^*)) = -\frac{c^x}{\mathbb{E}(c_j^y)}$. □

Two-dimensional asymmetric information, Case *(B)*: Cost variation large compared to isoquant variation

To solve the second case following from Lemma 2, we need to apply a different educated guess with respect to the binding constraints. However, we apply a similar reasoning as in Case *(A)*, but take account of the fact that now, cost variation is more relevant than isoquant variation. Hence, we choose a symmetric setting and imply incentive constraints (5.8b) ($ll \rightarrow hl$), (5.8h) ($hl \rightarrow lh$) and (5.8f) ($lh \rightarrow hh$) to be binding. Again, we assume the participation constraint of the hh-type to be binding. Figure 5.4 illustrates this setting.

After having determined the results and checked all remaining constraints, we find the setting of binding constraints as in Figure 5.4 indeed to be optimal for Case *(B)*. Results are summarized in the following Proposition 5.

Proposition 5. *For case (B),*

(i) Optimal regulation is achieved under the following set of observable input levels:

$$x_{ll}^* = x_{ll}^{fb} \tag{5.31}$$
$$x_{lh}^* \geq x_{lh}^{fb} \tag{5.32}$$
$$x_{hl}^* \geq x_{hl}^{fb} \tag{5.33}$$
$$x_{hh}^* \geq x_{hh}^{fb}, \tag{5.34}$$

while respecting $x_{ll}^* < x_{hl}^* \leq x_{lh}^* \leq x_{hh}^*$.

(ii) The most efficient (ll) type can always be separated. Moreover, separation of at least three types is always possible, while bunching of the hl and lh types is unavoidable in case of $\mu \to 1$.The lh and hh types may be bunched in case of c_l^y being small and $g_h'(q,x)$ large.

Proof. (i) Under the constraints considered binding for Case *(B)* – as discussed and shown in Figure 5.4 – the social cost function (5.4) becomes

$$\begin{aligned}
&C = \\
&\mu\nu\big[\lambda\big(c_h^y(g_h(x_{hh})-g_l(x_{hh}))+c_h^y g_l(x_{lh})-c_l^y g_h(x_{lh})+c_l^y(g_h(x_{hl})-g_l(x_{hl}))\big) \\
&+(1+\lambda)\big(c^x x_{ll}+c_l^y g_l(x_{ll})\big)\big] \\
&+\mu(1-\nu)\big[\lambda\big(c_h^y(g_h(x_{hh})-g_l(x_{hh}))\big)+(1+\lambda)\big(c^x x_{lh}+c_h^y g_l(x_{lh})\big)\big] \\
&+(1-\mu)\nu\big[\lambda\big(c_h^y(g_h(x_{hh})-g_l(x_{hh}))\big)+c_h^y g_l(x_{lh})-c_l^y g_h(x_{lh}) \\
&+(1+\lambda)\big(c^x x_{hl}+c_l^y g_h(x_{hl})\big)\big] \\
&+(1-\mu)(1-\nu)\big[(1+\lambda)\big(c^x x_{hh}+c_h^y g_h(x_{hh})\big)\big].
\end{aligned} \tag{5.35}$$

Minimizing C with respect to x_{ll} yields

$$g_l'(x_{ll}^*) = -\frac{c^x}{c_l^y}, \tag{5.36}$$

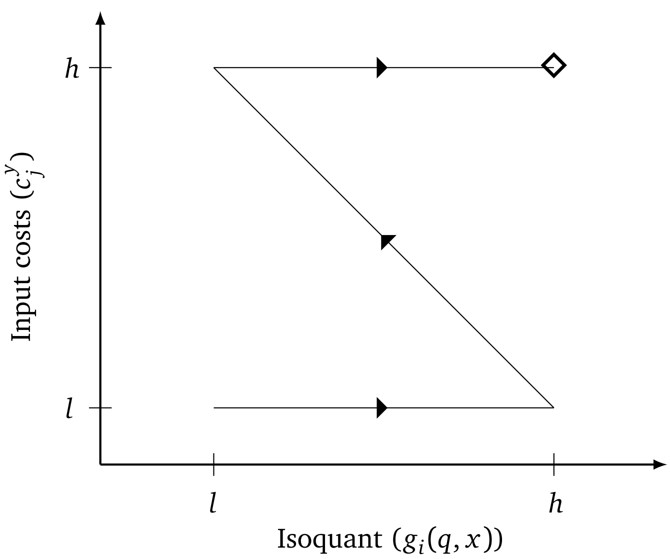


Figure 5.4: Constraints considered binding for Case *(B)*

which implies that $x_{ll}^{*} = x_{ll}^{fb}$. Derivation of C with respect to x_{lh}, x_{hl} and x_{hh} yields:

$$\frac{\partial C}{\partial x_{lh}} = \underbrace{\mu\lambda(c_h^y g_l'(x_{lh}) - c_l^y g_h'(x_{lh})}_{<0} + \underbrace{\mu(1-\nu)(1+\lambda)(c^x + c_h^y g_l'(x_{lh}))}_{\substack{=0 \text{ for } x_{lh}=x_{lh}^{fb} \\ <0 \text{ for } x_{lh}<x_{lh}^{fb} \\ >0 \text{ for } x_{lh}>x_{lh}^{fb}}} \tag{5.37}$$

$$\frac{\partial C}{\partial x_{hl}} = \underbrace{\mu\nu\lambda(c_l^y g_h'(x_{hl}) - c_l^y g_l'(x_{hl}))}_{<0} + \underbrace{(1-\mu)\nu(1+\lambda)(c^x + c_l^y g_h'(x_{hl}))}_{\substack{=0 \text{ for } x_{hl}=x_{hl}^{fb} \\ <0 \text{ for } x_{hl}<x_{hl}^{fb} \\ >0 \text{ for } x_{hl}>x_{hl}^{fb}}} \tag{5.38}$$

$$\frac{\partial C}{\partial x_{hh}} = \underbrace{(\mu + (1-\mu)\nu)\lambda c_h^y (g_h'(x_{hh}) - g_l'(x_{hh}))}_{<0} \tag{5.39}$$

$$+ \underbrace{(1-\mu)(1-\nu)(1+\lambda)(c^x + c_h^y g_h'(x_{hh}))}_{\substack{=0 \text{ for } x_{hh}=x_{hh}^{fb} \\ <0 \text{ for } x_{hh}<x_{hh}^{fb} \\ >0 \text{ for } x_{hh}>x_{hh}^{fb}}}. \tag{5.40}$$

From Equation (5.37), we see that $\frac{\partial C}{\partial x_{lh}}$ is strictly smaller than 0 for $x_{lh} = x_{lh}^{fb}$ and monotonically increasing in x_{lh}, which implies that $x_{lh}^{*} > x_{lh}^{fb}$ must always hold. The same logic applies for x_{hl}^{*} and x_{hh}^{*}.

(ii) From $x_{ll}^{fb} < x_{lh}^{fb}$ and the strict upwards distortion of all other types, it follows that the ll-type can always be separated. x_{hl}^{*} may surpass x_{lh}^{*} in case of $\mu \to 1$. If the low costs c_l^y are small and $g_h'(q,x)$ becomes large, lh and hh types may need to be bunched, without impacting the separation of the other types.

The remaining constraints are satisfied under the obtained solution. □

As in Case (A), the first-best solution can be obtained for $\lambda = 0$, while the solution is second-best and incurring an increasing level of inefficiency for increasing levels of λ. Also again, the most efficient type can always be separated, while bunching of the hl and lh types (lh and hh types) may occur for very high occurrence probability of low isoquants, or if $g_h(q,x)$ is very steep and c_l^y small.

Bibliography

ACER, March 2014. Report on the influence of existing bidding zones on electricity markets.

Ackermann, T., Cherevatskiy, S., Brown, T., Eriksson, R., Samadi, A., Ghandhari, M., Söder, L., Lindenberger, D., Jägemann, C., Hagspiel, S., Cuk, V., Ribeiro, P. F., Cobben, S., Bindner, H., Isleifsson, F. R., Mihet-Popa, L., May 2013. Smart Modeling of Optimal Integration of High Penetration of PV - Smooth PV. Final Report.

Agentur für erneuerbare Energien, 2014. Yearly power generation from wind energy per federal state.
URL `http://www.foederal-erneuerbar.de/uebersicht/kategorie/wind/bundeslaender/BW|BY|B|BB|HB|HH|HE|MV|NI|NRW|RLP|SL|SN|ST|SH|TH|D`

Aguirre, I., Beitia, A., 2004. Regulating a monopolist with unknown demand: Costly public funds and the value of private information. Journal of Public Economic Theory 6 (5), 693–706.

Alexander, C., 2008. Market Risk Analysis Volume II: Practical Financial Econometrics. John Wiley & Sons Ltd.

Amelin, M., 2009. Comparison of capacity credit calculation methods for conventional power plants and wind power. IEEE Transactions on Power Systems 24 (2), 685–691.

Anderson, R., Taylor, L., 1986. The social cost of unsupplied electricity: A critical review. Energy Economics 8 (3), 139 – 146.

Andersson, G., 2011. Power System Analysis. Eidgenössische Technische Hochschule Zürich (ETH).

Armstrong, M., 1999. Optimal regulation with unknown demand and cost functions. Journal of Economic Theory 84, 196–215.

Armstrong, M., Sappington, D., 2007. Chapter 27 recent developments in the theory of regulation. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization. Elsevier, pp. 1557–1700.

Babich, V., Burnetas, A. N., Ritchken, P. H., 2007. Competition and diversification effects in supply chains with supplier default risk. Manufacturing & Service Operations Management 9 (2), 123–146.

Baron, D. P., Myerson, R. B., 1982. Regulating a monopolist with unknown costs. Econometrica: Journal of the Econometric Society 50 (4), 911–930.

Bazaraa, M. S., Sherali, H. D., Shetty, C. M., 2006. Nonlinear Programming - Theory and Algorithms. John Wiley & Sons.

Benders, J. F., 1962. Partitioning procedures for solving mixed-variables programming problems. Numerische Mathematik 4, 238–252.

Bessembinder, H., Lemmon, M. L., 2002. Equilibrium pricing and optimal hedging in electricity forward markets. The Journal of Finance 57, 1347–1382.

Billinton, R., 1970. Power system reliability evaluation. Taylor & Francis.

Billinton, R., Allan, R. N., 1996. Reliability evaluation of power systems, 2nd ed. New York: Plenum Publishing Corporation.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Boyd, S., Xiao, L., Mutapcic, A., Mattingley, J., 2008. Notes on Decomposition Methods.
URL `http://see.stanford.edu/materials/lsocoee364b/08-decomposition_notes.pdf`

Brandenburger, A., Stuart, H., 2007. Biform games. Management Science 53 (4), 537–549.

Brunekreeft, G., Neuhoff, K., Newbery, D., 2005. Electricity transmission: An overview of the current debate. Utilities Policy 13 (2), 73–93.

Bundesnetzagentur, 2014. Kraftwerksliste Bundesnetzagentur (Bundesweit; alle Netz- und Umspannebenen); Stand 16.07.2014.
URL `http://www.bundesnetzagentur.de/cln_1432/EN/Areas/Energy/Companies/SpecialTopics/PowerPlantList/PubliPowerPlantList_node.html`

Burger, M., Klar, B., Müller, A., Schindlmayr, G., 2006. A spot market model for pricing derivatives in electricity markets. Quantitative Finance 4, 109–122.

Burstedde, B., 2012. Essays on the economic of congestion management - theory and model-based analysis for Central Western Europe. Ph.D. thesis, Universität zu Köln.

Caillaud, B., Guesnerie, R., Rey, P., Tirole, J., 1988. Government intervention in production and incentives theory: A review of recent contributions. The RAND Journal of Economics 29 (1), 1–26.

Cambridge Energy Research Associates, 2001. Short circuit: will the California energy crisis derail the state's economy? The UCLA Anderson forecast for the nation and California. Los Angeles, California.

Capacity Allocating Service Company, May 2014. Documentation of the CWE FB MC solution - As basis for the formal approval-request.

Caramanis, M., December 1982. Investment decisions and long-term planning under electricity spot pricing. IEEE Transactions on Power Apparatus and Systems PAS-101 (12), 4640–4648.

Chao, H.-p., Peck, S., Oren, S., Wilson, R., 2000. Flow-based transmission rights and congestion management. The Electricity Journal 13 (8), 38–58.

Conejo, A. J., Castillo, E., Minguez, R., Garcia-Bertrand, R., 2006. Decomposition Techniques in Mathematical Programming - Engineering and Science Applications. Springer.

Council of European Energy Regulators, 2014. Recommendations for the assessment of electricity generation adequacy.

Cramton, P., Ockenfels, A., Stoft, S., 2013. Capacity market fundamentals. Economics of Energy & Environmental Policy 2 (2), 27–46.

Cutler, N. J., Boerema, N. D., MacGill, I. F., Outhred, H., 2011. High penetration wind generation impacts on spot prices in the australian national electricity market. Energy Policy 39, 5939–5949.

Dada, M., Petruzzi, N. C., Schwarz, L. B., 2007. A newsvendor's procurement problem when suppliers are unreliable. Manufacturing & Service Operations Management 9 (1), 9–32.

Dana, J. D., 1993. The organization and scope of agents: Regulating multiproduct industries. Journal of Economic Theory 59, 288–310.

Daxhelet, O., Smeers, Y., 2007. The EU regulation on cross-border trade of electricity: A two-stage equilibrium model. European Journal of Operations Research 181, 1396–1412.

dena, 2010. Kurzanalyse der Kraftwerksplanung in Deutschland bis 2020 (Aktualisierung).

Densing, M., 2013. Dispatch planning using newsvendor dual problems and occupation times: Application to hydropower. European Journal of Operational Research 228, 321–330.

Dragoon, K., Dvortsov, V., 2006. Z-method for power system resource adequacy applications. Power Systems, IEEE Transactions on 21 (2), 982–988.

Durante, F., Sempi, C., 2010. Copula theory: An introduction. In: Jaworski, P., Durante, F., Hardle, W. K., Rychlik, T. (Eds.), Copula Theory and Its Applications. Lecture Notes in Statistics. Springer Berlin Heidelberg, pp. 3–31.

DWD, 2014. Hourly mean values of wind speeds.
URL ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/wind/historical/

EEX Transparency Platform, 2012. Day-ahead prognosis of German wind power.
URL http://www.transparency.eex.com/en/Statutory%20Publication%20Requirements%20of%20the%20Transmission%20System%20Operators/Power%20generation/Expected%20wind%20power%20generation

Ehrenmann, A., Smeers, Y., 2005. Inefficiencies in European congestion management proposals. Utilities policy 13 (2), 135–152.

Elberg, C., Hagspiel, S., 2015. Spatial dependencies of wind power and interrelations with spot price dynamics. European Journal of Operational Research 241 (1), 260–272.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation. Econometrica 50, 987–1008.

ENTSO-E, 2012. Hourly load levels for Germany.
URL https://www.entsoe.eu/resources/data-portal/consumption/

European Commission, March 2013. Green Paper - A 2030 framework for climate and energy policies. COM(2013) 169 final.
URL http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52013DC0169

European Commission, January 2014. Impact Assessment - A 2030 framework for climate and energy policies. SWD(2014) 15 final.
URL http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52013DC0169

Eurostat, 2012. Electricity production from wind energy.
URL http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

Eurostat, 2015. Energy statistics (nrg).
URL http://ec.europa.eu/eurostat/data/database

Falbo, P., Felletti, D., Stefani, S., 2010. Integrated risk management for an electricity producer. European Journal of Operational Research 207, 1620–1627.

Firtin, E., Önder Güler, Akdag, S. A., 2011. Investigation of wind shear coefficients and their effect on electrical energy generation. Applied Energy 88 (11), 4097–4105.

Fürsch, M., Hagspiel, S., Jägemann, C., Nagl, S., Lindenberger, D., Tröster, E., 2013. The role of grid extensions in a cost-effcient transformation of the European electricity system until 2050. Applied Energy 104, 642–652.

García-González, J., Parrilla, E., Mateo, A., 2007. Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. European Journal of Operational Research 181 181, 1354–1369.

Garver, L. L., 1966. Effective load carrying capability of generating units. IEEE Transactions on Power Appartus and Systems 85 (8), 910–919.

Gelabert, L., Labandeira, X., Linares, P., 2011. An ex-post analysis of the effect of renewables and cogeneration on spanish electricity prices. Energy Economics 33, S59–S65.

General Electric, 2010. 2.5 mw wind turbine series. Tech. rep.
URL http://site.ge-energy.com/prod_serv/products/wind_turbines/en/downloads/GEA17007A-Wind25Brochure.pdf

Genest, C., Gendron, M., Bourdeau-Brien, M., 2009. The advent of copulas in finance. The European Journal of Finance 15 (7-8), 609–618.

German Wind Energy Association (BWE), 2012. Installed wind power capacities in the German federal states.
URL http://www.wind-energie.de/en/infocenter/statistiken/deutschland/installed-wind-power-capacity-germany

Girard, R., Laquaine, K., Kariniotakis, G., 2013. Assessment of wind power predictability as a decision factor in the investment phase of wind farms. Applied Energy 101 (0), 609 – 617.

Glachant, J.-M., 2010. The achievement of the EU electricity internal market through market coupling. EUI Working Papers, RSCAS 2010/87.

Green, R., 2007. Nodal pricing of electricity: how much does it cost to get it wrong? Journal of Regulatory Economics 31 (2), 125–149.

Green, R., Vasilakos, N., 2012. Storing wind for a rainy day: What kind of electricity does denmark export? Energy Journal 33, 1–22.

Grothe, O., Schnieders, J., 2011. Spatial dependence in wind and optimal wind power allocation: A copula-based analysis. Energy Policy 39, 4742–4754.

Growitsch, C., Malischek, R., Nick, S., Wetzel, H., 2014. The costs of power interruptions in Germany: A regional and sectoral analysis. German Economic Review.

Gürler, Ü., Parlar, M., 1997. An inventory problem with two randomly available suppliers. Operations Research 45 (6), 904–918.

Haghi, H. V., Bina, M. T., Golkar, M., Moghaddas-Tafreshi, S., 2010. Using copulas for analysis of large datasets in renewable distributed generation: Pv and wind power integration in iran. Renewable Energy 35 (9), 1991–2000.

Hagspiel, S., Jägemann, C., Lindenberger, D., Brown, T., Cherevatskiy, S., Tröster, E., 2014. Cost-optimal power system extension under flow-based market coupling. Energy 66, 654–666.

Hagspiel, S., Papaemannouil, A., Schmid, M., Andersson, G., 2012. Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid. Applied Energy 96, 33–44.

Hasche, B., Keane, A., O'Malley, M., 2011. Capacity value of wind power, calculation, and data requirements: the irish power system case. IEEE Transactions on Power Systems 26 (1), 420–430.

Höffler, F., Wambach, A., 2013. Investment coordination in network industries: The case of electricity grid and electricity. Journal of Regulatory Economics 44 (3), 287–307.

Hogan, W., Rosellón, J., Vogelsang, I., 2010. Toward a combined merchant-regulatory mechanism for electricity transmission expansion. Journal of Regulatory Economics 38, 113–143.

Hogan, W. W., 1992. Contract networks for electric power transmission. Journal of Regulatory Economics 4 (2), 211–242.

Howison, S., Coulon, M. C., 2009. Stochastic behaviour of the electricity bid stack: From fundamental drivers to power prices. The Journal of Energy Markets 2, 29–69.

Huppmann, D., Egerer, J., 2014. National-strategic investment in European power transmission capacity. DIW Discussion Papers, No. 1379.

IPCC, 2014. Climate Change 2014: Mitigation of Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, Ch. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)].

Jägemann, C., Fürsch, M., Hagspiel, S., Nagl, S., 2013. Decarbonizing Europe's power sector by 2050 - analyzing the implications of alternative decarbonization pathways. Energy Economics 40, 622–636.

Jónsson, T., Pinson, P., Madsen, H., 2010. On the market impact of wind energy forecasts. Energy Economics 32, 313–320.

Joe, H., 1997. Multivariate models and dependence concepts. Chapman & Hall, London.

Jorion, P., 2007. Value at risk: the new benchmark for managing financial risk. Vol. 3. McGraw-Hill New York.

Joskow, P., Tirole, J., June 2005. Merchant transmission investment. The Journal of Industrial Economics LIII (2), 233–264.

Joskow, P. L., 2008a. Capacity payments in imperfect electricity markets: Need and design. Utilities Policy 16 (3), 159–170.

Joskow, P. L., 2008b. Lessons learned from electricity market liberalization. The Energy Journal 29 (2), 9–42.

Joskow, P. L., 2014. Incentive Regulation in Theory and Practice: Electric Transmission and Distribution Networks. University of Chicago Press, Ch. 5.

Keane, A., Milligan, M., Dent, C., Hasche, B., D'Annunzio, C., Dragoon, K., Holttinen, H., Samaan, N., Soder, L., O'Malley, M., 2011. Capacity value of wind power. IEEE Transactions on Power Systems 26 (2), 564–572.

Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the copula r package. Journal of Statistical Software 34 (9), 1–20.

Kunz, F., 2013. Improving congestion management: How to facilitate the integration of renewable generation in Germany. The Energy Journal 34 (4), 55–78.

Kurzidem, M. J., 2010. Analysis of flow-based market coupling in oligopolistic power markets. Ph.D. thesis, ETH Zurich.

Laffont, J.-J., Martimort, D., 2002. The theory of incentives - the principial-agent model. Princeton University Press.

Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press.

Leuthold, F., Weigt, H., von Hirschhausen, C., 2008. Efficient pricing for European electricity networks - the theory of nodal pricing applied to feeding-in wind in Germany. Utilities Policy 16, 284–291.

Lewis, T. R., Sappington, D. E., 1988a. Regulating a monopolist with unknown demand. The American Economic Review 78 (5), 986–998.

Lewis, T. R., Sappington, D. E., 1988b. Regulating a monopolist with unknown demand and cost functions. The RAND Journal of Economics 19 (3), 438–457.

Li, J., Wang, S., Cheng, T. E., 2010. Competition and cooperation in a single-retailer two-supplier supply chain with supply disruption. International Journal of Production Economics 124 (1), 137–150.

Li, Z., Xu, S. H., Hayya, J., 2004. A periodic-review inventory system with supply interruptions. Probability in the Engineering and Informational Sciences 18 (01), 33–53.

Louie, H., 2014. Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures. Wind Energy 17 (2), 225–240.

Mas-Colell, A., Whinston, M. D., Green, J. R., et al., 1995. Microeconomic theory. Vol. 1. Oxford university press New York.

Masih-Tehrani, B., Xu, S. H., Kumara, S., Li, H., 2011. A single-period analysis of a two-echelon inventory system with dependent supply uncertainty. Transportation Research Part B: Methodological 45 (8), 1128–1151.

Minner, S., 2003. Multiple-supplier inventory models in supply chain management: A review. International Journal of Production Economics 81, 265–279.

Monitoringbericht, 2013. Bundesnetzagentur, Bundeskartellamt.

Murty, K. G., 1983. Linear Programming. John Wiley & Sons.

Myerson, R. B., 1979. Incentive compatibility and the bargaining problem. Econometrica: Journal of the Econometric Society 47 (1), 61–73.

Nelsen, R., 2006. An Introduction to Copulas. Springer, New York.

Netzentwicklungsplan, 2013. Netzentwicklungsplan Strom 2013 - Zweiter Entwurf der Übertragungsnetzbetreiber. 50Hertz Transmission, Amprion, TenneT TSO, TransnetBW.

Neuhoff, K., Boyd, R., Grau, T., Barquin, J., Echabarren, F., Bialek, J., Dent, C., von Hirschhausen, C., Hobbs, B. F., Kunz, F., Weigt, H., Nabe, C., Papaefthymiou, G., Weber, C., 2013. Renewable electric energy integration: Quantifying the value of design of markets for international transmission capacity. Energy Economics 40, 760–772.

Nordex, 2007. Nordex s70/s77.
URL http://www.nordex-online.com/fileadmin/MEDIA/Produktinfos/Nordex_S70-S77_D.pdf

Oggioni, G., Allevi, Y. S. E., Schaible, S., 2012. A generalized nash equilibrium model of market coupling in the European power system. Networks & Spatial Economics 12, 503–560.

Oggioni, G., Smeers, Y., 2012. Degree of coordination in market coupling and counter-trading. The Energy Journal 33 (3), 39–90.

Oggioni, G., Smeers, Y., 2013. Market failures of market coupling and counter-trading in Europe: An illustrative model based discussion. Energy Economics 35, 74–87.

Ozdemir, O., Munoz, F. D., Ho, J. L., Hobbs, B. F., 2015. Economic analysis of transmission expansion planning with price-responsive demand and quadratic losses by successive lp. IEEE Transactions on Power Systems PP (99), 1–12.

Papaefthymiou, G., 2006. Integration of stochastic generation in power systems. Ph.D. thesis, Technische Universiteit Delft.

Parlar, M., Perry, D., 1996. Inventory models of future supply uncertainty with single and multiple suppliers. Naval Research Logistics (NRL) 43 (2), 191–210.

Patton, A. J., 2012. A review of copula models for economic time series. Journal of Multivariate Analysis 110, 4–18.

Pérez-Arriaga, I. J., 2013. Regulation of the power sector. Springer Verlag London.

Pinson, P., Girard, R., 2012. Evaluating the quality of scenarios of short-term wind power generation. Applied Energy 96 (0), 12–20.

Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12, 51–62.

PJM, 2015. PJM Manual 18: PJM Capacity Market.

Platts, December 2009. UDI World Electric Power Plants Data Base (WEPP).

Porchet, A., Touzi, N., Warin, X., 2009. Valuation of power plants by utility indifference and numerical computation. Mathematical Methods of Operations Research 70 (1), 47–75.

Richter, J., 2011. DIMENSION - a dispatch and investment model for European electricity markets. EWI WP 11/03.

Rious, V., Dessante, P., Perez, Y., 2009. Is combination of nodal pricing and average participation tariff the best solution to coordinate the location of power plants with lumpy transmission investment? EUI Working Papers, RSCAS 2009/14.

Sappington, D., 1983. Optimal regulation of a multiproduct monopoly with unknown technological capabilities. The Bell Journal of Economics 14 (2), 453–463.

Sauma, E. E., Oren, S. S., 2006. Proactive planning and valuation of transmission investment in restructured electricity markets. Journal of Regulatory Economics 30, 261–290.

Schmalensee, R., 1989. Good regulatory regimes. The RAND Journal of Economics 20 (3), 417–436.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., Bohn, R. E., 1988. Spot Pricing of Electricity. Norwell, MA: Kluwer.

Shapley, L., 1953. A value for n-person games. In: Kuhn, H., Tucker, W. (Eds.), Annals of Mathematical Studies. Vol. 28. Princeton University Press, Ch. Contributions to the Theory of Games II.

Sklar, A., 1959. Fonctions de repartition a n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Universite de Paris.

Snyder, L. V., Atan, Z., Peng, P., Rong, Y., Schmitt, A. J., Sinsoysal, B., 2014. OR/MS models for supply chain disruptions: A review. Available at SSRN 1689882.

Stoft, S., June 2002. Power System Economics: Designing Markets for Electricity. Wiley-IEEE Press.

Tang, C. S., 2006. Perspectives in supply chain risk management. International Journal of Production Economics 103 (2), 451–488.

Tasche, D., Tibiletti, L., 2003. A shortcut to sign incremental value at risk for risk allocation. The Journal of Risk Finance 4 (2), 43–46.

Telson, M. L., 1975. The economics of alternative levels of reliability for electric power generation systems. The Bell Journal of Economics 6 (2), 679–694.

Thompson, M., Davison, M., Rasmussen, H., 2004. Valuation and optimal operation of electric power plants in competitive markets. Operations Research 52 (4), 546–562.

Tomlin, B., 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. Management Science 52 (5), 639–657.

van der Weijde, A. H., Hobbs, B. F., 2011. Locational-based coupling of electricity markets: benefits from coordinating unit commitment and balancing markets. Journal of Regulatory Economics 39, 223–251.

VGB and Eurelectric, 2012. Availability of Thermal Power Plants 2002 - 2011, Edition 2012.

Wagner, S. M., Bode, C., Koziol, P., 2009. Supplier default dependencies: Empirical evidence from the automotive industry. European Journal of Operational Research 199 (1), 150–161.

Wang, Z., 2002. The properties of incremental var in monte carlo simulations. The Journal of Risk Finance 3 (3), 14–23.

Weare, C., 2003. The California Electricity Crisis: Causes and Policy Options. Public Policy Institute of California.

Woo, C., Horowitz, I., Moore, J., Pacheco, A., 2011. The impact of wind generation on the electricity spot-market price level and variance: The texas experience. Energy Policy 39, 3939–3944.

Zachary, S., Dent, C. J., 2011. Probability theory of capacity value of additional generation. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 226 (1), 33–43.

Zhou, Z., Botterud, A., Wang, J., Bessa, R., Keko, H., Sumaili, J., Miranda, V., 2013. Application of probabilistic wind power forecasting in electricity markets. Wind Energy 16, 321–338.