Competition and Optimization in Electricity Systems

by

Majid M. Al-Gwaiz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2014

Doctoral Committee:

Professor Xiuli Chao, Chair
Professor Katta G Murty
Professor Romesh Saigal
Assistant Professor Owen Wu

ACKNOWLEDGEMENTS

I would like express my greatest appreciation to my advisor, Professor Xiuli Chao, for his mentoring and guidance throughout my PhD. program. I would also like to acknowledge my dissertation committee for their valuable feedback on my research. I am particularly grateful to Professor Own Wu for his help in the renewable curtailment and supply function equilibria research. I am grateful to all the IOE staff and faculty who have helped me throughout my PhD. program.

I would also like to thank my sponsoring company, Saudi Aramco, for funding my PhD. program. I would like to especially thank Tareq Alnuaim, Dr. Mahmoud Bahy Noureldin, and Omar Bazuhair for their efforts in approving my sponsorship. Lastly, I am especially grateful to my parents and wife for their support and encouragement throughout my studies.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

CSP concentrating solar thermal power

EC economic curtailment

FG flexible generators

FIT Feed-In-Tariffs

FTR financial transmission rights

IG inflexible generators

ISO Independent System Operator

ITC investment tax credits

LMP locational marginal price

MISO Midcontinent Independent System Operator

MSNE Mixed Strategy Nash Equilibrium

ODE Oridinary Differential Equations

PBS production based subsidies

PSNE Pure Strategy Nash Equilibrium

PTDF power transfer distribution factors

PTC production tax credits

PV photovoltaic

RPS Renewable Portfolio Standards

SFE Supply Function Equilibrium

KKT Karush-Kuhn-Tucker

VG variable generation

ABSTRACT

Competition and Optimization in Electricity Systems

by

Majid Al-Gwaiz

Chair: Xiuli Chao

Electricity prices are characterized by high volatility and severe price spikes. At the root of these phenomena is the strategic behavior of market participants. A good understanding of the market competition is key to making better regulation, contract, and investment decisions. The goal of this thesis is to study the following market competition problems: (1) the competition between flexible generators with fast ramping rates and inflexible generators with constant production rates, (2) the effect of the renewable generation penetration and production based subsidies on the competition and operating efficiency, (3) generation competition in transmission constrained networks, and (4) competition in the capacity expansion of electricity networks.

We first consider a centralized electricity model and find that reducing the production based subsidies to renewable plants dampens their intermittency effect through controlled curtailment, cuts operational cost, and improves the system's balance. We then consider an oligopoly in which generators submit supply function bids and analyze a Supply Function Equilibrium (SFE) model with generators that have different ramping rates. We find that the controlled curtailment of renewables has an additional benefit in oligopolistic markets as it can reduce generator market power, which has favorable operational efficiency and electricity price ramifications. We also find that the classical SFE model is inadequate for modeling renewables and inflexible generators, and can grossly overestimate the competition intensity. We modify the SFE model to account for these issues. Afterwards, a Bertrand model is used to study the duopoly competition in a transmission constrained network. We find that adding transmission constraints in this model does not change the bidding policy, instead it changes the critical demand levels at which firms revise their position from competitive to aggressive bidding. We also solve the symmetric mixed strategy Nash equilibrium problem for multiple generators in a Bertrand electricity auction. Finally, we study several transmission expansion schemes and devise two investment mechanisms that achieve near social optimality.

CHAPTER I

Introduction

Four decades ago, electricity markets around the globe were either government owned and operated or operated as a regulated monopoly. After the oil crisis of the 70's, many governments saw the need to shift to a decentralized competitive electricity market structure. The earliest market decentralization took place in Chile in the 1980s, followed by Argentina and other Latin American countries. It was not until the early 1990's that the UK and other commonwealth countries started the deregulation process, followed by other European countries. Some US states started decentralizing their electricity markets in the late 1990's, but others had reservations especially after the California electricity crisis in the turn of the century. The decentralization mechanisms varied by region and were influenced by many factors such as the market and the power grid structures.

An electricity system consists of generators that consume energy and incur other costs to produce electricity, and a network that transmits the electrical power to the end users. In a centralized electricity system, the government operates or heavily regulates the network and generation sectors. In such a setting, a single entity has complete control over all generation and network assets, and operates the system to maximize the overall benefit to society. In a decentralized electricity market, multiple profit maximizing firms own the electricity generators. However, the power transmission and distribution network is recognized as a natural monopoly (*Hogan* (1992)) and continues to be managed by a single independent agent. This is partly due to the nature of power networks, where hard network constraints must hold for the effective and healthy operation of the system.

It is the responsibility of an Independent System Operator (ISO) to coordinate electricity dispatch while satisfying the network's constraints as well as maintaining other reliability measures. Network constraints entail achieving electricity balance and abiding by capacity limits while accounting for power losses in the network. ISOs are also responsible for satisfying the random demand fluctuation and are required to sustain excess electricity reserves for reliable operations. In centralized electricity systems, the ISO is also responsible for the efficient dispatch of electricity from generators to users. *Schweppe et al.* (1988) classify the different economic factors that affect electricity dispatch and derive *spot prices* for electricity at every node in the network that satisfy necessary and sufficient conditions for electric dispatch optimality. The spot prices are determined by generator marginal costs, network congestions, and transmission power losses.

With the decentralization of electricity systems, markets need to adopt electricity contracts and reflect accurate locational prices. The *loop flow* phenomenon, electricity flows through the paths of least resistance in the network, makes it impossible to specify a contract path between two nodes in a complex network. To resolve this issue, *Hogan* (1992) uses the spot pricing idea to introduce *contract networks* and *transmission rights*. *Chao and Peck* (1996) extends the concept of transmission rights to account for network externalities in electricity contracts. Variants of these decentralization mechanisms have been widely implemented. Today, most of the electricity trades in decentralized markets either take place over contract agreements or through the electricity spot market.

Private generating firms participating in the electricity market submit supply function bids to the ISO. These supply functions convey the electricity rate the firms are willing to produce for every market price. The ISO essentially treats these bids as generator cost functions and sets electricity dispatch schedules with the minimum costs to customers. To maximize their profits, generating firms typically submit bids that exceed their true marginal operating costs. Generating firms may also collude or exercise market power to raise electricity prices.

To protect customers from high and volatile prices, the ISO may exercise a variety of regulatory actions such as price ceilings and other bidding restrictions. Large customers, electricity retailers, and generating firms can also purchase a portfolio of electricity contracts and options to hedge against the highly volatile electricity prices. Although high prices can be attributed to a variety of sources, such as large capital investments and expensive fuel costs, the high price volatility and price spikes are often caused by generators exercising market power when strategic opportunities arise, such as demand peaks, supply shortages, or network congestion. To make better regulation and contract decisions, regulators and other entities need to have a good understanding of the market competition.

Most of the generating firms' strategic behavior models fall under three categories: Cournot competition, Bertrand competition, and SFE. Generating firms are assumed to bid production quantity offers in a Cournot model, price bid offers in a Bertrand model, and production-price pair offers in a SFE model. The Cournot model is typically the easiest to solve among the three, but can be inaccurate, while the SFE model is the most difficult to handle but the most representative among the three. In this thesis, the Bertrand and the SFE models will be used to study the competition in the electricity market.

The goal of this thesis is to study the following market competition problems: (1) the competition between flexible generators with fast ramping rates and inflexible generators with constant production rates, (2) the effect of the renewable generation penetration and production based subsidies on the competition and operating efficiency, (3) generation competition in transmission constrained networks, and (4) competition in the capacity expansion of electricity networks.

The impact of the renewable generation on centralized electricity systems is studied in Chapter II, and their combined impact along with generation inflexibility on the market competition is then studied in Chapter III using a SFE model. In this chapter, a Nash equilibrium for the competition is found using a linear SFE model. This problem is studied again in Chapter IV but with a focus on the high demand daytime market. Under such a scenario, the set of differential equations that characterize the SFE become easier to solve, and a more general solution can be attained.

In Chapter V, a Bertrand model is used to study the duopoly competition in a transmission constrained network. A symmetric Mixed Strategy Nash Equilibrium (MSNE) for the multiple generator problem is then solved using the Bertrand competition model. Afterwards, network transmission investment models are studied in Chapter VI. In this chapter, several transmission expansion schemes are considered in which profit maximizing firms expand the network to gain transmission rent. Two mechanisms that achieve social optimality and near social optimality are presented in this chapter. Finally, Chapter VII concludes this thesis followed by Appendices with supporting material.

For convenience, the following operators will be used throughout the chapters: $a^+ = \max\{a, 0\}$, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, and the indicator function $\mathbf{1}_{\{A\}} = 1$ if event A is true and 0 if A is false. In addition, the convention $\sum_{i=a}^{b} x_i = 0$ for integers $a > b$ is used.

CHAPTER II

Using Controlled Curtailment for Integrating Intermittent Renewable Generation in Electricity Systems

2.1 Introduction

The rapidly expanding renewable technology capacity in electricity systems is motivated by environmental, sustainability, and independence considerations. About 20% of the world's electricity production in 2011 came from renewable energy sources, and about 5% of which from non-hydroelectric plants[1]. According to *Renewable Energy Policy Network for the 21st Century (REN21)* (2013), wind, solar photovoltaic (PV), and concentrating solar thermal power (CSP) generation are the fastest growing renewable technologies with capacity growth rates averaging 26%, 58%, and 37% from 2006 through 2011 to reach a total of 310 GW by the end of 2011. In the United State, non-hydroelectric renewable generation has increased by 110% from 1999 to 2010, with an installed capacity expansion from about 16 GW to about 54 GW in the same period (*EIA*, 2010). In fact, in the 2012 renewable electricity futures study, *Hand et al.* (2012) concluded that 80% of the United State's electricity demand can be supplied by renewable generation by 2050. Many other countries have set ambitious

[1] *Renewable Energy Policy Network for the 21st Century (REN21)* (2013).

renewable portfolio targets according to *REN21* (2013). The report predicts that renewables would make 50-80% of the global electricity generation by 2050.

To meet these targets and promote their technological development, many countries have introduced various forms of governmental subsidies and tax incentives, such as direct grants, production tax credits (PTC) that entitle renewable plants to reduced federal income taxes based on their production rates, investment tax credits (ITC) that entitle renewable plants to reduced federal income taxes based on their capital investments, Feed-In-Tariffs (FIT) that guarantee electricity purchase from the renewable plants at fixed prices, Renewable Portfolio Standards (RPS) that require customers to purchase a portion of their electricity demand from renewable plants, and carbon cap-and-trade that prompts customers to purchase electricity from clean sources because of emission allowances. There have been numerous studies comparing the effectiveness of the renewable subsidy schemes (*Fischer and Newell* (2008); *Menanteau et al.* (2003); *Neuhoff* (2005); *Palmer and Burtraw* (2005)). Despite the renewable generation virtues and governmental promotional efforts, there have been some obstacles to their widespread. The main impediment to the renewable integration initiative is the intermittent nature of some renewable technologies, such as wind, solar, and ocean tidal power. To distinguish between these technologies and other non-intermittent renewable technologies, such as hydroelectric and biomass, we refer to the intermittent technologies as variable generation (VG) technologies.

Intermittent generation can cause scheduling and balancing problems in electricity networks. Scheduling of electricity generation is made over varying time durations. Unit commitment schedules are set a day or more in advance and account for seasonal trends, while economic dispatch and load following schedules are done on the day of operation and occur between 10 minutes to a few hours in advance to account for the daily consumption trends. Because electricity networks have very little tolerance to system imbalance, sufficient capacity must be ensured to account for uncertainty.

This safety capacity is categorized into contingency and operating reserves. Contingency reserves, in the form of spinning and non-spinning reserves, are used to counter discrete imbalance events such as generation outages, while operating reserves are used for continuous imbalances to correct forecast mismatches and to fine tune the electricity supply and demand balance. The types of generators used for the different dispatch schedules and contingency categories vary based on their operating costs (including fuel costs, emissions, wear and tare, maintenance, etc.), ramp rates (how fast they can change their production levels), and how fast they can be shut down and turned on. The inflexible generators (IG) with slow ramp rates and long up and down times usually have narrow operating ranges throughout the day that are set in the unit commitment phase, while flexible generators (FG) that have fast ramp rates and short up and down times may have a more varying production schedule to follow the demand forecast and balance the network.

The growing VG penetration trend can increase the forecast variance and make balancing the network more costly. Several studies have confirmed the additional operating reserve requirements as VG installations continue to increase. This may even result in more fossil fuel consumption, increased environmental harm, higher electricity prices for consumers, and substantial investments in FG plants. While most studies conclude that VGs do not affect the network contingency requirements, there have been mixed findings on the impact of VGs on load following requirements. *Ela et al.* (2011) summarize the findings of several renewable integration studies.

Renewable integration has received considerable attention over the past decade, and there have been a variety of technology and market based solution proposals for achieving high VG penetration levels. In fact, renewable enabling technologies have been the subject of numerous smart grid applications, and several approaches have been considered to counter the renewable intermittency. One such approach is the Demand Response (DR), which shifts some of the balancing burden to end users that willingly curtail their load to help balance the network (*Albadi and El-Saadany*, 2007). Load curtailments can either be contracted in advance with customers or may be prompted using real time price signals that incent some of the flexible electricity consumers to adjust their consumption. Another approach is to use storage buffers to reduce the intermittency *Denholm et al.* (2010). Even with the recent advancements in energy storage technologies, their relatively low efficiencies and high costs limit their wide spread implementation for the time being. One of the operational modifications that can be used to lower balancing costs is to use controlled curtailment of renewables (*Wu and Kapuscinski* (2013)). In this approach, the electricity generation from renewable plants is voluntarily cutback in order to maintain less variable output from renewable plants. This in turn reduces the forecast variance and the balancing load. In fact, VG curtailment can be part of the solution to reduce the variability caused by the customer demand (*Kirby et al.*, 2010; *Miller et al.*, 2011). We will focus in this chapter on the controlled curtailment approach as a means to integrate variable renewable generation.

The main impediment to implementing controlled curtailment in many of the existing markets is ironically the very incentive used to promote their widespread. Mechanisms that subsidize renewable generation through operational rewards, such as the PTC and FIT schemes, encourage VGs to maximize their production regardless of the operational consequences, which makes VGs resist their curtailment. In fact, some system operators are contractually obligated to accept all production from renewable plants with few exceptions. Although such incentive schemes have helped spread renewable technologies, giving VGs production priority over other generators may discard some of the technically feasible and economically superior operating strategies. We show in this chapter that the reduction in the overall generation efficiency due to VG priority can become significant when VGs make up a substantial part of the power supply portfolio, which implies that operational policies that give VGs production priority may not be sustainable in systems with high VG penetration levels.

Some renewable integration initiatives redefine the role VGs can play in the network. Conventionally, gas turbine and some combined cycle plants are used for operating reserves because of their fast ramp rates and wide operating ranges, while VGs are merely used as energy providers. However, under certain control schemes some renewable plants can produce similar ramp rates and operating range requirements for regulation reserves. In fact, some recent studies identify renewable plants as candidates for regulation reserves. *Erlich and Wilch* (2010) give wind generator control concepts that can be used in the grid's frequency control. *Rivier Abbad* (2010) conjectures that conventional ancillary service providers may fail to provide sufficient operational reserves in high VG penetration scenarios, and suggests that VGs provide such services. *Liang, Grijalva, and Harley* (2011) consider a market where wind generators can bid in both the energy and the regulation markets. *Ela et al.* (2011) discuss the possible restructuring of ancillary services markets to allow for higher VG penetration, including the option of VG supplied ancillary services.

In this chapter, we build on the idea of using VGs as energy suppliers as well as regulation reserves to balance the network. Having the controlled curtailment of renewables as a viable option to the system operator is central to this operating policy. We consider in this chapter a simplified electricity system with three types of generators; IG, FG, and VG. We first study the electricity system balancing problem in §2.2 and characterize the optimal operational policy in §2.3. In §2.4 we consider a variant of the model in which the system could not tolerate any supply and demand discrepancy. We also compare in §2.5 the option of giving VGs priority over other generators and the option of curtailing VGs for economic gain under different VG penetration levels. The concepts introduced in this chapter are illustrated using a numerical example in §2.6 and a brief conclusion is given in §2.7.

2.2 The Electricity System Model

In this section, we consider a centralized electrical network balancing problem where an ISO sets generation schedules to balance customer demand at the lowest possible cost. We first describe our model for the generators in the electricity system and then formulate the system operator's problem. The length of the operating horizon is T, which can be several hours to one day.

2.2.1 Generator Types and Costs

To approximate a fleet of power generators, we assume that the system consists of three types of generators: *inflexible*, *flexible*, and *variable* generators. We assume that each firm owns one generator, thus we use "firm" and "generator" interchangeably throughout the chapter.

Inflexible generators (IG), indexed by $i \in G^I$, cannot adjust their output rates during $[0, T]$. The output rate of generator $i \in G^I$, denoted as $q_i \geq 0$, is determined by the system operator prior to $t = 0$ and stays constant over $[0, T]$. Let $C_i(q_i)$ denote generator i's operating cost per unit of time.

Flexible generators (FG), indexed by $j \in G^F$, can adjust their output rates instantaneously. Let $q_{jt} \geq 0$ denote the output rate of generator $j \in G^F$ at time $t \in [0, T]$, and let $C_j(q_{jt})$ denote its operating cost rate at time t.

The output rates, q_i and q_{jt}, are implicitly within generators' capacities. For the purpose of the analysis in this chapter, we assume IGs and FGs' capacities are not binding constraints.

Variable generators (VG) have time-varying potential outputs, which depend on factors such as wind speed or solar radiation. Let K denote the total installed VG capacity, and $W_t \in [0, K]$ denote the *potential output* of VGs at time $t \in [0, T]$, which is the maximum possible total VG production level in time t. VGs may adjust their actual output below W_t, which is known as curtailment. Curtailment is achieved, for example, by pitching the blades of wind power generators or rotating solar panels to reduce power output.

The costs of the generators satisfy the following assumption.

Assumption II.1. *(i) For any generator $k \in G^I \cup G^F$, the cost rate function $C_k(q)$ is convex, strictly increasing, and continuously differentiable in q, and $C_k(0) = 0$; (ii) VGs produce energy at negligible operating cost and receive a subsidy of $r \geq 0$ per unit of output that is not curtailed; (iii) FGs and VGs output can be adjusted instantaneously at negligible cost.*

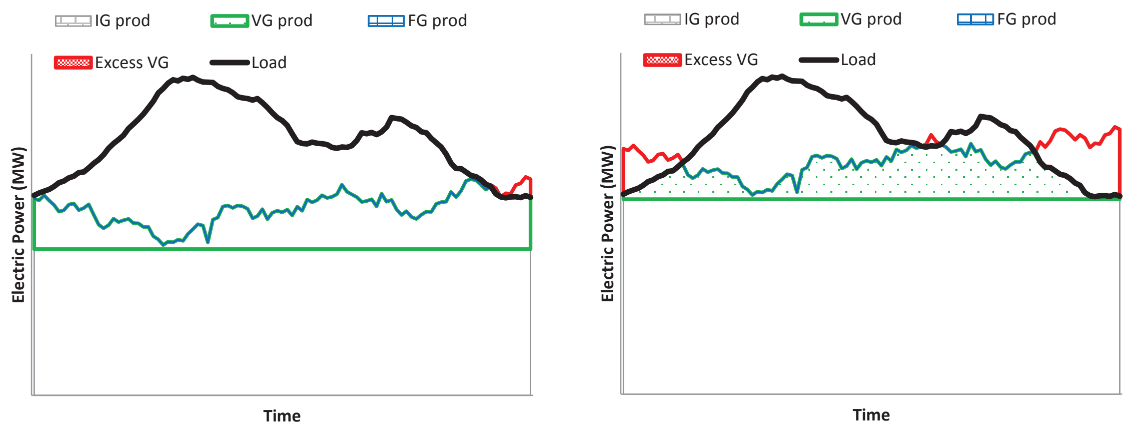
The convexity and monotonicity assumption in part (i) matches with the reality. Part (ii) states that VGs receive production-based subsidy and implies that the marginal cost of VGs is $-r$. Although there can be multiple VGs in the system, the identical production cost assumption for VGs allows us, without loss of generality, to aggregate all VGs into a single VG plant. Part (iii) assumes that FGs have fast ramp rates and ramping is costless. We refer the reader to *Wu and Kapuscinski* (2013) for a model that includes ramping speed and cost.

2.2.2 Problem Description

Let $L_t \in [\underline{L}, \overline{L}]$ denote the price-insensitive load[2] at time $t \in [0, T]$, where $\overline{L} > \underline{L} \geq 0$. The load L_t must be satisfied instantaneously for all $t \in [0, T]$. The load L_t and the VG potential output W_t are the two sources of uncertainties in our model and they can be correlated. We consider two ways to balance the load in real time: (1) vary the production of FG plants, and (2) the curtailment of VG plants. Although the production of IG plants must be fixed before L_t and W_t are realized, the choice of total IG production is critical to the operational cost as it sets a tradeoff between VG curtailment and FG production. While increasing the total IG production level replaces the expensive FG production it may also offset some of the free VG capacity.

[2]Load is the term often used for the customer demand in electricity systems.

This tradeoff is illustrated in Figure 2.1 for a given electricity load and VG capacity realization. In Figure 2.1(a) the system operator uses a low IG production level, in which case the FG production is high, but very little VG is curtailed. The IG production is increased in Figure 2.1(b) and the total required FG generation is reduced, but a larger portion of the free VG production capacity is lost.



(a) System operation under low IG production allocation.

(b) System opearation under high IG production allocation.

Figure 2.1: Production levels of each generator for a given load and VG capacity sample path.

VG Production Subsidies

As discussed in the introduction, the VG producers are often promoted by production based subsidies (PBS) in the form of production tax credits and renewable credits. These subsidies are an additional stream of revenue to renewable plants that may give incentives for VGs to produce even at a loss, and hence directly affect the production schedule. To model the PBS effect on the system we denote r as the monetary value paid to VGs for producing 1 MWh of electricity. In other words, a VG is paid the market price in addition to r for every MWh of electricity it produces. The value of r essentially sets a VG priority level over other generators. VGs would find it economical to produce as long as the market price is above $-r$. This alters the tradeoff between the VG curtailment and the FG production which in turn biases the operating policy towards less IG production as we will see later in this chapter.

Oversupply Penalty

Although electrical networks have little tolerance for violating the supply and demand balance in real time, they do have a narrow imbalance allowance. A mismatch between power supply and load causes the electricity's frequency to diverge from its nominal value, which may harm some of the equipment and appliances if the frequency offset becomes significant. Some generators are unwilling to reduce their production levels even at a zero electricity market price, which is typical for baseload plans because of their inflexible nature, as well as renewable plants that are given production based subsidies. This may cause electricity prices to become negative, which results in a situations where generators are charged for producing electricity and customers are paid for their consumption (*Fink et al.*, 2009).

In Figure 2.1b, the total IG production is close to the minimum load. If the total IG production were a little higher it could have exceeded the minimum load. Although a small mismatch between supply and demand can be tolerated for short durations, the system operator is required to take action to avoid prolonged or significant supply and demand mismatches. Such actions include:

- *Invoking responsive loads to consume extra electricity:* Such loads may require monetary incentives to deviate from their preferred consumption patterns.

- *Shutting down IGs:* This may not be an option for all IG plants as some can take several hours or even days to shutdown. Even if an imbalance could be anticipated early enough to schedule IG shutdowns, the shutdown and start up of these plants incurs extra fuel and significant wear-and-tear costs. Consequently, the system wide operating costs could increase over the start up and shut down durations of these IGs.

- *Operating generators at emergency minimum level:* Generators typically operate within best efficiency ranges. Some generators can run below their nominal operating ranges, but at a severe drop in efficiency and possible wear-and-tear costs.

- *Curtailing VGs:* Because of the PBS, VGs are often willing to pay the network for their production. In order to curtail VGs the system operator may have to set the electricity price below $-r$.

- *Storing the excess supply:* For systems that have available storage capacity (such as batteries, pumped hydro, and compressed air storage) some of the excess electricity may be stored. This storage comes at a cost due to the energy charging, discharging, and decay losses.

We do not intend to accurately model the system operator action when the supply exceeds the demand. Instead, we model the extra costs for handling oversupply situations using a penalty function $h(e)$, which represents the extra cost rate when the excess supply rate (total output minus the load) is $e \geq 0$. A similar approach is used in practice. For example, in the Texas electricity system, a penalty for violating the power balance constraint is included in the objective function of the security-constrained economic dispatch problem (ERCOT 2012, p. 24).

Assumption II.2. *The oversupply penalty rate function $h(e)$ is strictly convex, strictly increasing, and continuously differentiable in e for $e \geq 0$, and $h(0) = 0$.*

Our model does not involve undersupply, because FGs are flexible enough to ensure all demand is met. With the introduction of h the tradeoffs between low cost production using IGs, the tolerance for excess supply, and the curtailment of VGs become clear. The system has the following cost components:

- The *generation costs* for IGs and FGs.

- The VG *production subsidy.* This cost term is given by $-rq$ when the total VG production is q

- The *oversupply penalty.* This cost is only incurred when the supply exceeds the demand.

The system operator selects the production level for each generator to minimize the sum of these three cost components over the time horizon. Although L_t and W_t are treated as random variables, the model studied in this chapter extends to the deterministic case when L_t and W_t represent the (known) variable load and capacity duration correspondences over the time horizon.

2.2.3 The System Operator's Problem

Let q_i and q_{jt} respectively denote the production levels of generator $i \in G^I$, $j \in G^F$, and q_t^V denote the total VG production at time t. Note that q_i is time-invariant because once an IG's production level is set, it cannot be changed throughout the planning horizon $[0, T]$. The system operator's problem is to decide q_i prior to $t = 0$ and decide q_{jt} and q_t^V in real time in response to the load and wind power realizations, with the objective of minimizing the total expected cost of serving the demand over $[0, T]$.

The system operator's problem can be formulated as first deciding the aggregate production level for each type of generators and then allocate the aggregate production to individual generators. Let q^I and q_t^F denote the aggregate production at time t for IGs and FGs, respectively. The cost-minimizing allocations of q^I and q_t^F are determined by solving the following problems:

$$C^I(q^I) \stackrel{\text{def}}{=} \min_{q_i \geq 0} \Big\{ \sum_{i \in G^I} C_i(q_i) : \sum_{i \in G^I} q_i = q^I \Big\}, \tag{2.1}$$

$$C^F(q_t^F) \stackrel{\text{def}}{=} \min_{q_{jt} \geq 0} \Big\{ \sum_{j \in G^F} C_j(q_{jt}) : \sum_{j \in G^F} q_{jt} = q_t^F \Big\}. \tag{2.2}$$

The following lemma summarizes the properties of the aggregate production cost functions.

Lemma II.3. *The total IG and FG cost functions $C^I(q)$ and $C^F(q)$ are continuously differentiable, convex, and strictly increasing in q.*

Proof of this lemma and other results in this chapter can be found in Appendix B.

Because the per-MWh subsidy r applies to all VGs, the total subsidies depend only on the aggregate VG production q_t^V. The allocation for q_t^V among VGs can be arbitrary as long as it satisfies the capacity constraint for each VG. With the optimal allocation of the aggregate production in (2.1)-(2.2), the system operator's problem can be written as

$$\min_{\{q^I, q_t^F, q_t^V\}} TC^I(q^I) + \mathsf{E}\Big[\int_0^T \Big(C^F(q_t^F) - r\, q_t^V + h(e_t)\Big)\, dt\Big] \tag{2.3}$$

$$\text{s.to} \quad e_t \equiv q^I + q_t^F + q_t^V - L_t \geq 0, \qquad \forall\, t \in [0, T], \tag{2.4}$$

$$q_t^V \leq W_t, \qquad \forall\, t \in [0, T], \tag{2.5}$$

$$q^I,\, q_t^F,\, q_t^V \geq 0, \qquad \forall\, t \in [0, T]. \tag{2.6}$$

This problem contains two stochastic processes: the electric load $\{L_t : 0 \leq t \leq T\}$ and the maximum wind power generation $\{W_t : 0 \leq t \leq T\}$. The two processes can be correlated and we assume that their (joint) probability distributions are known to all players. The expectation is taken with respect to these two processes. The inequality in (2.4) ensures sufficient supply to meet the load, whereas excess supply (if $e_t > 0$) is penalized in the objective (2.3).

2.3 Optimal Dispatch

To solve the system optimization problem in (2.3)-(2.6), we first fix IGs' output rate q^I and solve for the optimal q_t^F and q_t^V in response to the realizations of L_t and W_t. Then we decide the optimal q^I prior to $t = 0$. These two steps are analyzed in §2.3.1 and §2.3.2, respectively.

2.3.1 Optimal Flexible and Variable Generation Under Given q^I

Suppose IG output rate $q^I \geq 0$ is given. At time t, knowing the realized load L_t and VG potential output W_t, we decide the optimal FG and VG output by the following convex program:

$$\widetilde{C}(q^I, L_t, W_t) \stackrel{\text{def}}{=} \min_{\{q_t^F, q_t^V\}} C^F(q_t^F) - r\, q_t^V + h(e_t) \tag{2.7}$$

$$\text{s.t.} \quad e_t \equiv q^I + q_t^F + q_t^V - L_t \geq 0, \tag{2.8}$$

$$q_t^V \leq W_t, \qquad q_t^F, q_t^V \geq 0. \tag{2.9}$$

Recall the marginal penalty $h'(e)$ strictly increases in e with minimum value $h'(0) \geq 0$. We define

$$\mu(r) \stackrel{\text{def}}{=} \begin{cases} (h')^{-1}(r), & \text{if } r \geq h'(0), \\ 0, & \text{if } r < h'(0). \end{cases} \tag{2.10}$$

When $r \geq h'(0)$, $\mu(r)$ gives the oversupply level at which the marginal oversupply penalty equals r. The following theorem provides an explicit solution to the problem in (2.7)-(2.9).

Theorem II.4. *For a given IG production level $q^I \geq 0$, under the realized VG potential output W_t and the load L_t, the optimal FG and VG production rates at time t*

are

$$q_t^{F*} = (L_t - q^I - W_t)^+ \qquad \textit{and} \qquad q_t^{V*} = \min\{W_t,\, (L_t - q^I + \mu(r))^+\}. \qquad (2.11)$$

Furthermore, the real-time cost rate $\widetilde{C}(q^I, L_t, W_t)$ *in* (2.7) *is jointly convex in* (q^I, L_t, W_t).

The solutions in (2.11) under various realized values of L_t and W_t are illustrated in Figure 2.2. If the load L_t drops below the IG output q^I to such an extent that the marginal oversupply penalty exceeds the subsidy, $h'(q^I - L_t) > r$ (or $q^I - L_t \geq \mu(r)$, i.e., event A_1), then the VG output does not bring net benefit to the system and is completely curtailed. However, if the subsidy r exceeds the marginal oversupply penalty $h'(q^I - L_t)$, then some or all of the VG potential output is used, corresponding to the next three cases.

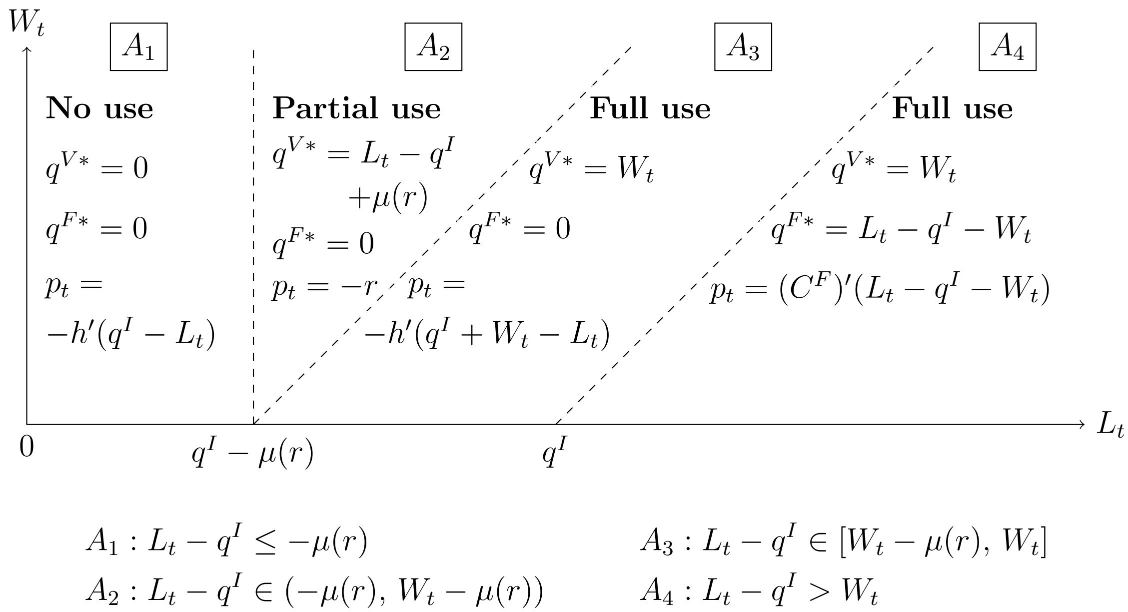


Figure 2.2: Real-Time Operating Policy and Price.

In event A_2, VGs are partially curtailed and the VG output is such that the per-unit subsidy equals the marginal oversupply penalty, $r = h'(e_t)$ or $\mu(r) = e_t = q^I + q_t^{V*} - L_t$. In event A_3, the per-unit subsidy outweighs the marginal oversupply penalty when all the VG potential output is used, $r \geq h'(q^I + W_t - L_t)$, and thus, no curtailment occurs. In event A_4, IGs and VGs cannot meet the entire load, and FGs serve the remaining load.

The four events together imply that FGs produce if and only if the load cannot be satisfied by IGs and VGs. In other words, FGs only produce to make up the supply shortage. The above discussions lead to a complementary property of the optimal operating policy:

$$q_t^{F*}\left(q_t^{V*} - W_t\right) \;=\; 0. \tag{2.12}$$

That is, when FGs produce, VGs' potential output is fully used. When VG curtailment occurs, FGs do not produce.

Figure 2.2 also describes the real-time price, p_t, which equals the system marginal cost of serving the load at time t. When the load exceeds the combined output of IGs and VGs (event A_4), FGs must produce and the real-time price is $p_t = (C^F)'(L_t - W_t - q^I) > 0$. When the load can be met by IGs and VGs, the real-time price becomes zero or negative:

a) The real-time price is zero when VG output is partially curtailed (event A_2 occurs) and no subsidy is provided ($r = 0$). A small incremental load can be served by VG at zero cost.
b) The real-time price is negative when additional load lowers the total revealed cost by either reducing the oversupply penalty or increasing VG output (when $r > 0$). In event A_1, all VG output is curtailed, oversupply is $q^I - L_t$, and price is $p_t = -h'(q^I - L_t) < -r$. In event A_2, VG is partially curtailed and $p_t = -r$. In event A_3, $p_t = -h'(q^I + W_t - L_t) \in (-r, 0)$.

Summarizing the above discussion, we can express the real-time price p_t as a function of q^I, L_t, and W_t as in (2.13) below. In this expression, the dependence on

FG cost function $C^F(\cdot)$ is also emphasized.

$$\begin{aligned} P(q^I, L_t, W_t, C^F) \overset{\text{def}}{=} & (C^F)'(L_t - W_t - q^I)\mathbf{1}_{A_4} - h'(q^I + W_t - L_t)\mathbf{1}_{A_3} \\ & - r\mathbf{1}_{A_2} - h'(q^I - L_t)\mathbf{1}_{A_1}, \end{aligned} \tag{2.13}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function taking value 1 if statement in $\{\cdot\}$ is true and 0 otherwise.

Using (2.13), the time-average of the expected real-time price can be written as

$$\overline{P}(q^I, C^F) \overset{\text{def}}{=} \frac{1}{T} \int_0^T \mathsf{E}\big[P(q^I, L_t, W_t, C^F)\big]\, dt, \tag{2.14}$$

where the expectation is taken prior to $t = 0$. The function $\overline{P}(q^I, C^F)$ relates the average price to the IG output q^I under a given aggregate FG cost function, thus $\overline{P}(q^I, C^F)$ is the IGs' inverse residual demand function. Note that $\overline{P}(q^I, C^F)$ decreases in q^I, because $P(q^I, L_t, W_t, C^F)$ decreases in q^I due to the monotonicity of $C^F(\cdot)$ and $h'(\cdot)$.

2.3.2 Optimal Inflexible Generation

With the real-time problem solved in (2.7)-(2.9), the problem in (2.3)-(2.6) can be written as

$$\min_{q^I \geq 0} \; TC^I(q^I) + \mathsf{E}\left[\int_0^T \widetilde{C}(q^I, L_t, W_t)\, dt\right]. \tag{2.15}$$

The convexity of $C^I(q^I)$ and $\widetilde{C}(q^I, L_t, W_t)$ in Theorem II.4 implies that the objective in (2.15) is convex in q^I. This property allows us to characterize the optimal IG production q^{I*}. For most of the realistic situations, inflexible generators do produce energy, and thus for the rest of this chapter, we will focus our attention to $q^{I*} > 0$. The optimal IG production is given by the following theorem.

Theorem II.5. *The optimal IG output q^{I*} is the unique solution to*

$$(C^I)'(q^{I*}) = \overline{P}(q^{I*}, C^F). \tag{2.16}$$

This Theorem states that at the optimal IG production level, the marginal production cost of IGs equals the *time-average* expected system marginal cost over the planning horizon. As q^I increases from zero, the marginal production cost of IGs increases while the time-average system marginal cost decreases; the optimal q^{I*} is the unique intersection of these two curves. The system marginal cost can be interpreted as the the spot price under perfect competition. Thus, Theorem II.5 implies that under perfect competition, IGs equate their marginal cost with the time-average spot price, so that their expected marginal profit is zero.

2.4 Robust Balancing Requirement

We consider in this section the policy when the system is required to be balanced at all times. The problem (2.3)-(2.6) can be modified to accommodate this requirement by dropping the overproduction penalty term $h(\cdot)$ from (2.3) and changing the "$\geq$" in (2.4) to "$=$". When there is no overproduction penalty the real-time operating policy becomes trivial: maximize the generation from VGs as long as it does not exceed $L_t - q^I$ and make up the balance using FGs. The optimal VG production becomes

$$q_t^{V*} = \min\{W_t, L_t - q^I\}$$

and the total FG production has the same solution as in (2.11). The real-time cost function becomes

$$\widetilde{C}(q^I, L_t, W_t) = C^F((L_t - q^I - W_t)^+) - r\min\{W_t, L_t - q^I\},$$

which is clearly jointly convex in (q^I, L_t, W_t). This cost function shows that FGs are the marginal generators when $L_t - q^I > W_t$ and VGs become marginal when $L_t - q^I \leq W_t$. When FGs set the system's marginal cost, the spot price becomes $p_t = C^{F\prime}(L_t - q^I - W_t)$, and when VGs are marginal the spot price is $p_t = -r$. Therefore, the average price over the time horizon is

$$\begin{aligned}\overline{P}(q^I, C^F) =& \frac{1}{T}\int_0^T \mathsf{E}\left[C^{F\prime}(L_t - q^I - W_t)\mathbf{1}_{\{L_t - q^I > W_t\}} - r\mathbf{1}_{\{L_t - q^I \leq W_t\}}\right] dt \\ =& \frac{1}{T}\int_0^T \left(\mathsf{E}\left[C^{F\prime}(L_t - q^I - W_t)\mathbf{1}_{\{q^I < L_t - W_t\}}\right] - r\mathbb{P}\{L_t \leq W_t + q^I\}\right) dt.\end{aligned}$$

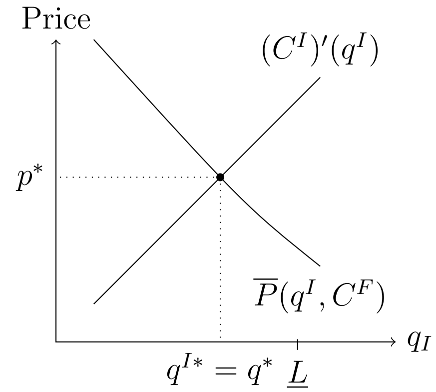
The optimal IG production problem can be solved using the same approach as in §2.3.2 with the additional requirement that the total IG production not exceed the total demand to prevent excess supply, which is given by the constraint $q^I \leq \underline{L}$. If we use the dual variable $\nu \geq 0$ for this constraint then, from (2.16), q^I's first order optimality condition becomes

$$C^{I\prime}(q^I) = \frac{1}{T}\int_0^T \left(\mathsf{E}\left[C^{F\prime}(L_t - q^I - W_t)\mathbf{1}_{\{q^I < L_t - W_t\}}\right] - r\mathbb{P}\{L_t \leq W_t + q^I\}\right) dt - \nu. \tag{2.17}$$

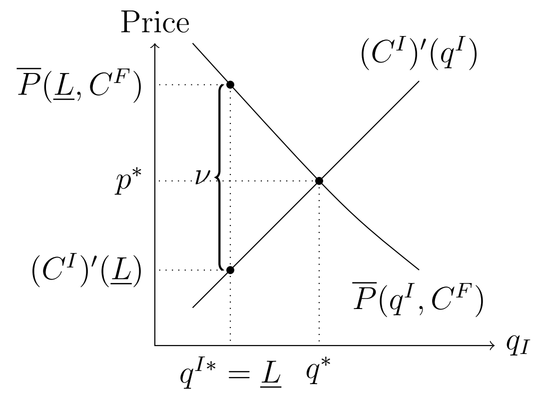
By complimentary slackness, ν is 0 when $q^I < \underline{L}$ but can be positive when $q^I = \underline{L}$. Equivalently, to get the optimal q^I the system operator can find q^* that solves

$$C^{I\prime}(q^*) = \frac{1}{T}\int_0^T \left(\mathsf{E}\left[C^{F\prime}(L_t - q^* - W_t)\mathbf{1}_{\{q^* < L_t - W_t\}}\right] - r\mathbb{P}\{L_t \leq W_t + q^*\}\right) dt$$

and then set $q^{I*} = \min\{q^*, \underline{L}\}$. When $\nu > 0$ the price IGs get for their electricity becomes less than the average spot price over the time horizon in a decentralized competitive market. Figure 2.3 illustrates this phenomenon.



(a) Small q^* scenario: The robust model gives the same solution as the original model.



(b) Large q^* scenario: The robust model gives a lower q^{I*} and a higher average spot price than the original model.

Figure 2.3: IG production under the robust balancing model.

In Figure 2.3(a), the IG marginal cost and average price curves intersect at $q^{I*} < \underline{L}$, which gives a solution that is similar to the original problem's solution shown in (2.16). However, when the two curves intersect in $q^* > \underline{L}$ as shown in Figure 2.3(b), then the system operator sets q^{I*} to the maximum feasible production level $\underline{L}$. ν in in Figure 2.3(b) represents the difference between the average market price and the marginal IG cost. This implies that the optimal solution may not be implementable in a uniform price market mechanism and discriminatory pricing may be more applicable, otherwise the higher market price may signal IGs to overproduce and violate the network balance.

Another approach to solve the balanced system problem is to allow an imbalance to occur but charge a large overproduction penalty that never makes it economical to violate the system balance. Consider a penalty function h with an initial marginal cost $h'(0) > r$, which implies that $\mu(r) = 0$. Equation (2.16) becomes

$$C^{I'}(q^I) = \frac{1}{T}\int_0^T \Big(\mathsf{E}\left[C^{F'}(L_t - q^I - W_t)\mathbf{1}_{\{q^I < L_t - W_t\}} - h'(q^I - L_t)\mathbf{1}_{\{q^I \geq L_t\}} \right]$$

$$-r\mathbb{P}\{q^I < L_t \leq q^I + W_t\} \Big)\, dt.$$

This confirms that if the optimal q^I from Equation (2.16) does not exceed $\underline{L}$ then the solution from (2.16) and (2.17) agree. However, if the optimal q^I from (2.16) exceeds $\underline{L}$ then it must solve

$$C^{I\prime}(q^I) = \frac{1}{T}\int_0^T \big(\mathsf{E}\big[C^{F\prime}(L_t - q^I - W^t)\mathbf{1}_{\{q^I < L_t - W_t\}}\big] - r\mathbb{P}\{q^I < L_t \leq q^I + W_t\} \\ -\mathbb{P}\{L_t \leq q^I\}\mathsf{E}[h'(q^I - L_t)|L_t \leq q^I]\big)\, dt. \tag{2.18}$$

Since the LHS is increasing and the RHS is decreasing in q^I we get

$$C^{I\prime}(\underline{L}) < \frac{1}{T}\int_0^T \left(\mathsf{E}\left[C^{F\prime}(L_t - \underline{L} - W_t)\mathbf{1}_{\{\underline{L} < L_t - W_t\}}\right] - h'(0)\mathbb{P}\{L_t \leq q^I\}\right) dt$$

$$\Rightarrow \int_0^T \mathbb{P}\{L_t \leq q^I\}\, dt < \frac{\int_0^T \mathsf{E}\left[C^{F\prime}(L_t - \underline{L} - W_t)\mathbf{1}_{\{\underline{L} < L_t - W_t\}}\right] dt - TC^{I\prime}(\underline{L})}{h'(0)},$$

and hence $q^I \to \underline{L}$ as $h'(0) \to \infty$. The multiplication of $h'(0)$ and $\int_0^T \mathbb{P}\{L_t \leq q^I\}\, dt$ becomes indeterminate as $h'(0) \to \infty$, but we know that it must close the balance in (2.18), and hence

$$\lim_{h'(0)\to\infty} h'(0)\int_0^T \mathbb{P}\{L_t \leq q^I\} =$$

$$\int_0^T \left(\mathsf{E}[C^{F\prime}(L_t - \underline{L} - W_t)\mathbf{1}_{\{\underline{L} < L_t - W_t\}}] - r\mathbb{P}\{L_t \leq W_t + \underline{L}\}\right) dt - TC^{I\prime}(\underline{L}),$$

which is precisely $T\nu$ from (2.17) when $q^I = \underline{L}$. Therefore, the optimal q^I from (2.16) converges to the same solution as the optimal q^I from (2.17) when the initial marginal penalty $h'(0) \to \infty$.

This shows that requiring the network to be balanced is a limiting case of the original model with the overproduction penalty function. However, solving this problem using the infinite marginal penalty approach gives another explanation to the price discrimination based on generator flexibility. This approach shows that the IG marginal cost is indeed equal to the average spot market price, but the average spot market price becomes infinitely negative when $L_t = q^I$ which almost surely never happens, but the product of this infinitely negative price and its infinitesimally short duration amounts to a constant that equals this price difference. It just happens that all other generators don't produce during this infinitesimally short duration and hence aren't subject to this negative price. In fact, the average prices other generators see may be different from the IG prices even when overproduction is allowed. For example, when $L_t \leq q^I - \mu(r)$ the price becomes $-h'(q^I - L_t) < 0$ and neither FGs nor VGs produce in this case, and if this event occurs with positive probability then the FGs and VGs would sell their products at larger average prices than IGs.

2.5 The Impact of VG Subsidy and Penetration

In this section, we study the effects of subsidy r and total VG capacity K on system operations. In particular, we focus on their effects on the optimal inflexible generation q^{I*} and the system's average operating cost excluding the subsidies:

$$C^a(r, K) \stackrel{\text{def}}{=} TC^I(q^{I*}) + \int_0^T \mathsf{E}\Big[C^F(q_t^{F*}) + h(e_t^*)\Big]\, dt,$$

where the optimal generation levels q^{I*} and q_t^{F*} and the excess supply e_t^* are given by Theorems II.4 and II.5.

To disaggregate the impact of the installed VG capacity K and the natural energy sources that drive the VG units, such as the wind speed and solar illumination, we define the VG capacity factor $\rho_t = W_t/K$ and we assume that the distribution of ρ_t is independent of K. The total realized VG capacity in time t, $W_t = \rho_t K$, is increasing in both ρ_t and the total installed VG capacity K.

Theorem II.6. *The production based subsidy r and VG penetration K have the following impacts on the solution:*

(i) The optimal IG production q^{I*} is decreasing in r and K.

(ii) The average production cost C^a is increasing in r.

(iii) If $r = 0$ then C^a decreases in K.

(iv) C^a may increase in K when $r > 0$.

The willingness to sell power in spite of the negative price signals gives priority to VGs over other generators, and the value of r can be regarded as a priority level for VGs over other generators or imbalance response actions. Specifically, a larger r gives a larger VG priority over IGs, which in turn reduces q^I as shown in Theorem II.6 (i). Additionally, the subsidies skew the VG profits from their true values and raising r increases the distortion in the optimal generation allocation problem, causing C^a to increase as pointed out in Theorem II.6 (ii). K's impact on the cost is more complicated as it depends on two counteractive factors; the reduced cost due to free energy from VG plants and the increased balancing costs that result from the VG priority. In the special case $r = 0$ the subsidy term vanishes from the system operator's problem, and therefore increasing K necessarily improves the non-subsidized solution as indicated in Theorem II.6 (iii). Furthermore, the VG dispatch priority that comes with r raises the VG capacity utilization $\int_0^T \mathsf{E}[q_t^V/W_t|W_t > 0]\mathbb{P}\{W_t > 0\}\, dt$ as evident by Equation (2.11), making the VGs less susceptible to curtailment. In particular, when $r = 0$ VGs lose their preferential treatment and the VG curtailment is at its highest. We refer to this maximal VG curtailment scenario when $r = 0$ as the economic curtailment (EC) scenario.

2.6 Numerical Example

To illustrate the impact of r and K on the system we will consider a simple example of a system with truncated time invariant normally distributed $L \in [80, 100]$ MW and $\rho \in [0, 1]$ with means 90 MW and 0.5, and standard deviations 10/3 MW and 1/6 respectively. We will consider a unity time horizon and the aggregate generation cost functions $C^I(q) = 2q + 0.05q^2$ and $C^F(q) = 5q + q^2$, and the penalty function $h(q) = 10q + 2.5q^2$. q^I and C^a are plotted against $K \in [0, 120]$ MW for various r values in figures 2.4 and 2.5. These figures illustrates the declining trend of q^I with respect to r and K and the increasing trend of C^a with respect to r. Figure 2.5 shows that the inefficiency due to r is minuscule for low VG penetrations but can become significant for systems with high VG penetration levels. Figure 2.5 also proves that C^a can increase with K if $r > 0$. The relationship between the VG curtailment as a fraction of the available capacity and $r \in [0, 60]$ \$/MWh is shown in Figure 2.6 for several K values. As expected, the VG curtailment percentage decreases with r. Although the VG curtailment value in MW increases with K, this result cannot be extended to the % of curtailed VG capacity as demonstrated by the intersections and non-monotonicity of the W = 5, 10, 20, and 40 MW curves in Figure 2.6. The non-monotonicity is caused by the rapid rate of change of q^I with respect to K for small values of K that can make $\frac{d}{dW}(q_t^V/K\rho_t) > 0$.

2.7 Conclusion

We have developed in this chapter a simplified model to study the impact of of the renewable penetration and production based subsidy on the power system, and quantified the short term benefits of curtailing VG capacity. Our results show that the subsidies increase production costs, and can make additional VG installations harmful. The intuition for these results is that subsidies enable VGs to produce in

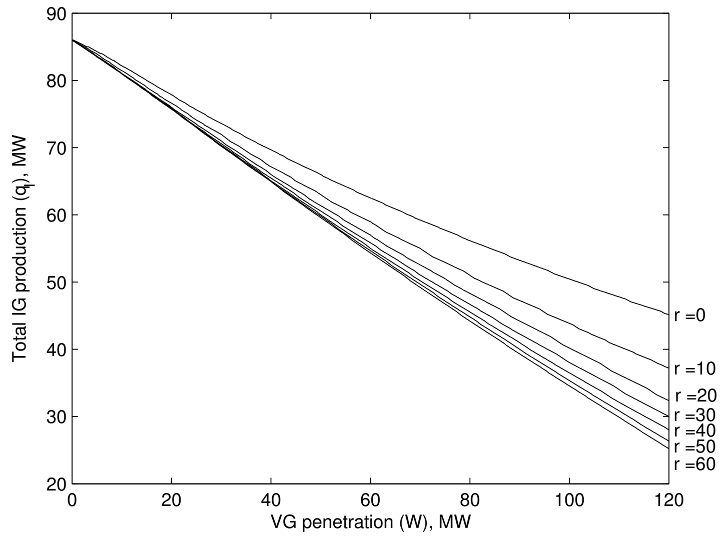


Figure 2.4: The impact of r and K on the total IG production.

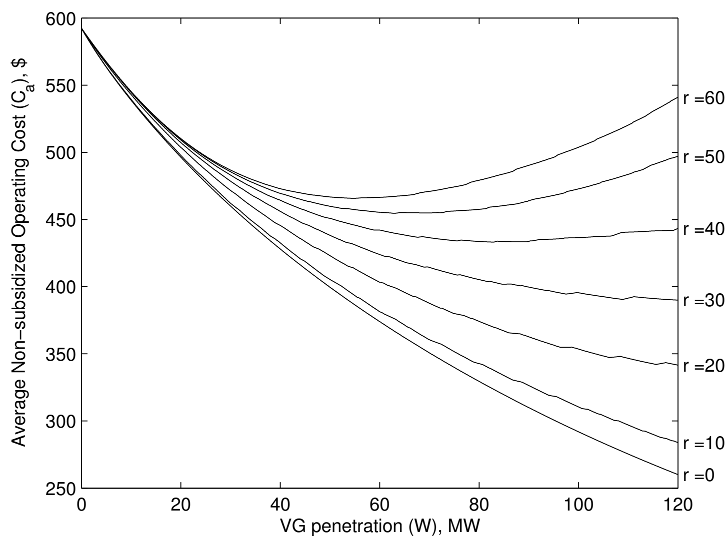


Figure 2.5: The impact of r and K on the average non-subsidized cost.

otherwise uneconomical circumstances, which can cause inefficient generation allocation. On the other hand, the additional revenue from the PBS mechanism promotes the renewable expansion and potentially speeds up technology advancements that can improve the renewable generation efficiency and reduce its costs. We did not intend in this chapter to find the best subsidies that give the optimal tradeoff between VG expansion and operational efficiency, nor did we intend to compare the subsidies

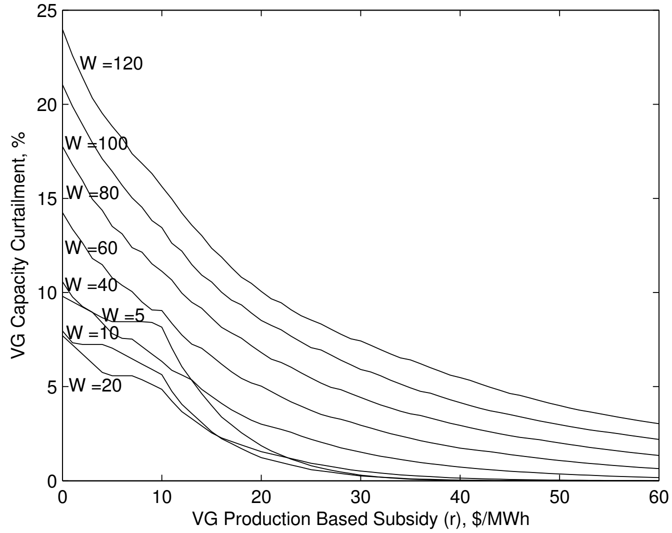


Figure 2.6: The impact of r and K on the VG curtailment.

with other renewable promotional mechanisms. Our aim was to study the short term implication of the production based subsidies on the system's operating policy and production costs for various renewable penetration levels. As it turns out, the VG subsidies and penetration can influence the generator competitive behavior as well. We will consider the impact of these factors on the market competition in the next chapter.

CHAPTER III

Supply Function Competition in Electricity Markets with Flexible, Inflexible, and Variable Generation

3.1 Introduction

In this chapter we study the supply function competition between power-generation firms with different levels of flexibility. Inflexible firms produce power at a constant rate over an operating horizon, while flexible firms can adjust their output to meet the fluctuations in electricity demand. Both types of firms compete in an electricity market by submitting supply functions to a system operator, who solves an optimal dispatch problem to determine the production level for each firm and the corresponding market price. We study how firms' (in)flexibility affects their equilibrium behavior and the market price. We also analyze the impact of variable generation (such as wind and solar power) on the equilibrium, with the focus on the effects of the amount of variable generation, its priority in dispatch, and the production-based subsidies. We find that the classical supply function equilibrium model tends to overestimate the intensity of the market competition, and even more so as more variable generation is introduced into the system. The policy of economically curtailing variable generation intensifies the market competition, reduce price volatility, and improve the system's overall efficiency. Moreover, we show that these benefits are most significant in the absence of the production-based subsidies.

The special nature of the electricity industry (quick and random fluctuations of demand, limited storage capability) requires production decisions to be automated and coordinated instantaneously. Thus, in an electricity market, the instruments of competition are supply functions, which specify the amount of electricity each firm is willing to generate at every market price. Based on the submitted supply functions, a system operator finds the most economical production schedule to meet the electricity demand and determines the payment to each firm. A set of supply functions from which no firm would benefit by unilaterally altering its supply function is known as a supply function equilibrium (SFE).

Klemperer and Meyer (1989) pioneered the effort in analyzing the SFE in general industrial contexts. *Green and Newbery* (1992) and *Bolle* (1992) are the first to employ the SFE framework to analyze electricity markets. These works and the following stream of research provide important economic insights and policy recommendations, which we will review in §3.2.

Most SFE models for electricity systems assume that all firms have the flexibility to adjust their power output at different prices and do not explicitly consider ramping constraints. This assumption can be justified in two situations. First, each firm owns a portfolio of power generators and offers a supply curve representing the aggregate output as a function of the market price. The portfolio is not dominated by inflexible generators (e.g., nuclear and some coal-fired generators), and the aggregate output can be adjusted in response to the price changes throughout the day. This situation is studied by *Green and Newbery* (1992), *Green* (1996), *Rudkevich* (1999), *Baldick, Grant, and Kahn* (2004), among others. Second, in the real-time market that runs and clears every hour (or half-hour in some markets), firms with flexible generators submit real-time supply offers to meet the energy imbalance (the energy that deviates from the day-ahead schedule). This situation is considered by, for example, *Holmberg* (2007, 2008) and *Sioshansi and Oren* (2007). The theoretical framework of SFE is applicable to both situations, as discussed in *Anderson and Philpott* (2002) and *Holmberg and Newbery* (2010).

As industry deregulation continues, firms downsize their portfolios by selling off parts of their generation assets, and independent power producers emerge and participate in the power markets. In many of the current markets, firms exhibit different levels of flexibility: Firms that own mainly inflexible generators cannot change their power output in a short time, whereas firms owning flexible generators can quickly ramp up or down their production. All firms engage in a supply function competition in the day-ahead market and the system operator determines the production schedule taking into account the firms' different levels of flexibility.

Because firms' flexibility directly affects their production and revenue, it is natural to ask the following questions: *How do firms with different levels of flexibility behave in a supply function competition? How does the presence of inflexibility affect the equilibrium market price?* Answers to these questions will help policy makers understand whether the classical SFE model may over- or underestimate the intensity of the market competition. Understanding the effect of generation flexibility/inflexibility on competition is also important in this era of generation technology evolution, with coal-fired generation shifting toward more flexible generation fueled by natural gas.

An integral part of this technology evolution is the increasing variable generation from renewable sources. According to the Renewable Energy Policy Network (2013), globally the fastest growing renewable energy technologies from 2008 to 2012 are solar photovoltaic, concentrating solar thermal power, and wind power, with average annual capacity growth rates of 60%, 43%, and 25%, respectively. Variable generation from renewable sources displaces conventional flexible and inflexible generation, and thus changes the competition between them, but the classical SFE model does not address competitions involving inflexible firms. This raises another question: *How does variable generation impact the competition between flexible and inflexible firms?*

The answer to this question depends on the priority of variable generation. Due to its environmental and economic benefits, variable generation is often given the highest priority in dispatch, i.e., it is curtailed only when the excessive energy from variable generation threatens system reliability. However, curtailing variable generation may also provide economic benefits, as shown by *Ela* (2009), *Ela and Edelson* (2012), and *Wu and Kapuscinski* (2013). Hence, a relevant question is: *How does the economic curtailment policy affect the competition between flexible and inflexible firms?* A caveat is that even if economic curtailment policy is in effect, the production-based subsidies for renewable energy may lead to partial economic curtailment. Therefore, in addressing the last question, we also examine the case of partial economic curtailment.

This chapter aims to address the several questions raised above through theoretical and computational analysis. Our model is not intended to be a comprehensive depiction of the electricity industry, but to be a stylized model that captures the relevant tradeoffs. We assume each firm's generators are either fully flexible or inflexible. Inflexible generators (IGs) produce power at a constant rate over an operating horizon (e.g., several hours to one day), while flexible generators (FGs) can adjust their output to meet the load fluctuations. We formulate the system operator's optimal dispatch problem and derive the market clearing conditions. We then characterize and compute the SFE with linear supply functions, which is commonly adopted both in practice and in the research literature, assuming variable generators are price-takers.

The main insights from this chapter are summarized as follows. First, the classical SFE model tends to overestimate the intensity of the supply function competition. IGs do not compete with FGs in matching production with uncertain demand and, in equilibrium, FGs offer significantly lower output at each price than predicted by the classical SFE model. IGs compete with all other generators for market share and in equilibrium offer slightly less output than in the classical model. Our equilibrium model with flexibility consideration also leads to a higher average price and a higher price volatility than predicted by the classical SFE model.

Second, when variable generation receives absolute priority in dispatch, the additional variability introduced into the system must be balanced by FGs, leading to more market share for FGs, as well as less intense market competition. As more variable generation displaces the conventional generation, the average market price drops, but the price volatility increases significantly.

Third, economic curtailment of variable generation is a partial substitute for FGs in balancing against variability. Thus, economic curtailment intensifies the market competition: All IGs and FGs offer more competitive supply functions than under priority dispatch, and IGs' supply functions may be even more competitive than predicted by the classical SFE model. Economic curtailment has little impact on the average price, but substantially reduces the price volatility. The overall operating cost of the system is also reduced by the economic curtailment, but the emission reduction depends on the types of the generators in the electricity system.

Finally, production-based subsidies increase the priority of variable generation and reduce the amount of VG curtailment. With the presence of the subsidies, the economic curtailment policy does not achieve its full benefit to encourage competition and improve the system's efficiency.

3.2 Literature Review

In their original work, *Klemperer and Meyer* (1989) showed the existence of a family of SFE for competing firms with identical cost functions and without capacity constraints. They characterized the SFE by differential equations and show that, given the support of the uncertainty, the equilibria are independent of the distribution of the uncertainty. Since this seminal work, the SFE framework has been applied extensively to the research in electricity markets. Comprehensive reviews of this area are provided by *Ventosa et al.* (2005), *Holmberg and Newbery* (2010), and *Li, Shi, and Qu* (2011). Thus, we review below only the works most relevant to ours.

Green and Newbery (1992) consider the effect of capacity constraints on SFE and calibrate the model for the British electricity industry. Their results suggest that the market power had been seriously underestimated by the policy makers. *Rudkevich, Duckworth, and Rosen* (1998) study symmetric SFE with inelastic demand, and find that even with a relatively high number of competing firms, the market clearing prices are still significantly higher than perfectly competitive prices. *Anderson and Philpott* (2002) derive the conditions under which a supply function can represent a firm's optimal response to the offers of other firms and show that their model admits symmetric SFE. *Holmberg* (2008) proves the SFE is unique when power shortage occurs with positive probability and a price cap exists. All these studies focus on the case of symmetric equilibria in which firms offer identical supply functions.

When firms differ in costs, the general asymmetric SFE is difficult to find and, thus, linear supply functions are often used to simplify the analysis. *Green* (1996) obtains the linear supply function equilibrium for the asymmetric case and studies the effects of various policies that could increase the competition in the electricity market. *Rudkevich* (1999) provides a more explicit solution to the SFE with linear supply functions and further finds that this equilibrium could be reached by a learning process.

Several studies analyze firms with identical cost but asymmetric capacities. *Genc and Reynolds* (2011) consider SFE when some firms are pivotal, i.e., they can move the market price to the price cap with positive probability. *Holmberg* (2007) establishes the uniqueness of SFE in a real-time market under certain conditions.

It is also possible to consider asymmetries in both costs and capacities. *Baldick, Grant, and Kahn* (2004) focus on SFE with piecewise linear supply functions and point out that linear supply functions are useful for practical applications. *Anderson and Hu* (2008) consider more general SFE and analyze situations when supply functions may have jumps. They also develop numerical methods for calculating asymmetric SFE. *Anderson* (2013) establishes the existence of an SFE under more general conditions.

Incorporating physical constraints, especially the network transmission constraints, into the SFE model is also an important research direction. *Berry et al.* (1999) find that the strategic behaviors on networks may lead to results that differ from those predicted by the traditional models. *Wilson* (2008) characterizes the necessary conditions for an equilibrium when transmission capacity is uncertain and transmission constraints may be binding. *Holmberg and Philpott* (2012) consider a situation where demand shocks exist at all nodes and transmission capacities are known.

This chapter contributes to the literature by analyzing supply function competition between firms with different levels of flexibility. To the best of our knowledge, the impact of flexibility on firms' strategic interactions has not been analyzed, and this chapter tries to begin filling this gap. The key feature of our model is that the system operator takes into account the firms' different levels of flexibility when determining the production schedule. The outputs of inflexible firms, once determined, stay constant throughout the operating horizon. The optimality condition of the system operator's problem serves as a constraint in the firm-level profit-maximization problem. Our approach shares similar features with the bi-level optimization procedure by *Hobbs, Metzler, and Pang* (2000).

Integration of variable generation into electricity systems has received substantial research attention over the past decade. The National Renewable Energy Laboratory recently completed two large variable generation integration studies: the Western Wind and Solar Integration Study (WWSIS) (*GE Energy*, 2010) and the Eastern Wind Integration and Transmission Study (EWITS) (*EnerNex*, 2011). Excellent reviews of these and earlier variable generation studies are provided by *Smith et al.* (2007), *Ela et al.* (2009), and *Hart et al.* (2012). Most of these integration studies focus on quantifying system cost reduction due to variable generation, as well as the integration cost, i.e., the incremental cost in balancing against variable generation.

The impact of variable generation on the SFE in electricity markets has not been considered until recently. *Sioshansi* (2011) recognizes the difficulty in modeling simultaneous-move of wind and conventional generators and analyzes a Stackelberg game with embedded supply function competition among conventional generators. Assuming wind-power generators are price-takers and have priority in dispatch, *Buygi, Zareipour, and Rosehart* (2012) analyze an SFE with linear supply functions and find that although the intermittency of wind power tends to increase the market price, the net impact of wind power is reduced market prices. In this chapter we also treat variable generation as price-takers and study its impact on both average price and price volatility. We further consider the impact of dispatch policies (priority dispatch vs. economic curtailment) on SFE and market prices.

The role of economic curtailment policy has been investigated in several studies. *Ela* (2009) explores the network effects of economic curtailment. *Ela and Edelson* (2012) analyze the benefit of curtailment on relieving physical constraints of generation resources, thereby bringing substantial cost savings. *Bentek Energy* (2010) points out that accommodating variable generation may lead to increased cycling cost of conventional generators and increased emissions. *Katzenstein and Apt* (2009) also find that, due to extra emissions from cycling, the emission reductions are likely to be significantly less than those assumed by policy makers. *Wu and Kapuscinski* (2013) analyze the impact of economic curtailment on cycling cost and peaking cost, and find that curtailing wind power can be both economically and environmentally beneficial under certain situations. This chapter complements the above body of work by studying the impact of economic curtailment on market competition. We find an additional benefit of economic curtailment—economic curtailment intensifies market competition.

3.3 The Model

In a decentralized electricity system, an ISO receives supply function offers from electricity suppliers, then sets dispatch schedules to satisfy demand with the minimum cost to consumers. Upon receiving the supply offers, the ISO can convert them into *revealed cost* functions and use the centralized electricity system model studied in Chapter II. We will go over the supply offers and show how they are converted into revealed costs next.

3.3.1 Supply Offers and Revealed Costs

The supply offers from IGs and FGs are characterized by supply functions. Prior to $t = 0$, FG $j \in G^F$ submits a supply function $S_j(p)$, $p \in \Re$, which specifies the output rate it is willing to produce when the real-time price is p. Also prior to $t = 0$, IG $i \in G^I$ submits a supply function $S_i(p)$, $p \in \Re$, which specifies the fixed output rate it is willing to set over $[0, T]$ if the average price over $[0, T]$ is p. We assume generators are risk-neutral, so that IGs care only about the average price. The aggregate supply functions are defined as

$$S^I(p) \stackrel{\text{def}}{=} \sum_{i \in G^I} S_i(p) \qquad \text{and} \qquad S^F(p) \stackrel{\text{def}}{=} \sum_{j \in G^F} S_j(p). \tag{3.1}$$

To fix ideas, we can set the operating time horizon $T = 1$ day and assume that each generator submits a supply function for the entire day (i.e., the supply offer is "long-lived"). In many markets, although generators are allowed to submit hourly offers, most generators submit the same supply functions for the entire day. This is because they do not anticipate status changes (e.g., maintenance) of their own or other generators, and thus, a single supply function describes the generators preferred output at different prices during the day. Indeed, using the historical generator offer data from Midcontinent Independent System Operator (MISO)[1], we find that about 90% of the generators submit the same supply offers for the entire day.

The supply functions satisfy the following assumption.

Assumption III.1. *For any $k \in G^I \cup G^F$: (i) There exists $p_k^{\min} \geq 0$, such that $S_k(p) = 0$ for $p \leq p_k^{\min}$; (ii) $S_k(p)$ strictly increases in p for $p \geq p_k^{\min}$; (iii) $\lim_{p \to 0} S_k(p) = 0$.*

Assumption III.1(i) implies that no generator is willing to produce when the price (or average price in the case of IGs) is negative. Part (ii) is consistent with practice[2]. Part (iii) is automatically satisfied if $p_k^{\min} > 0$ due to part (i); when $p_k^{\min} = 0$, part (iii) states that no generator is willing to produce when the price drops to nearly zero. All these assumptions are mild. The commonly used affine supply function $S_k(p) = \beta_k(p - p_k^{\min})^+$, where $\beta_k > 0$ is a constant, satisfies Assumption III.1. (Throughout the chapter, we use notation $x^+ = \max\{x, 0\}$ for any real number x.)

The system operator computes generators' revealed cost functions based on their submitted supply functions. The revealed cost function of generator k is defined as

$$\widehat{C}_k(q) \overset{\text{def}}{=} \int_0^q S_k^{-1}(x)dx, \qquad \forall\, k \in G^I \cup G^F, \tag{3.2}$$

where the inverse supply function is $S_k^{-1}(q) \overset{\text{def}}{=} \inf\{p : S_k(p) > q\}$. If generator k submits its inverse marginal cost function as its supply function, then the revealed cost is its true cost function. Assumption III.1 ensures that the revealed costs have the same properties of the true costs (increasing, convex, and continuously differentiable),

[1] Available at https://www.misoenergy.org/Library/ MarketReports.

[2] E.g. MISO's Business Practice Manual states that the price-quantity pairs that form a supply function must be increasing for price and strictly increasing for quantity (MISO, 2013, p. 92).

and hence all the results based on these properties for the cost functions in Chapter II hold.

Unlike IGs and FGs, VGs are unable to guarantee an output rate because of their inherent intermittency. Therefore, we assume each VG submits a price offer for its potential output. To focus on analyzing the strategic interactions between IGs and FGs, we assume that VGs submit their marginal cost $-r$ as the offer price, where r is the subsidy per unit of output. This means that VGs produce W_t when the price exceeds $-r$, completely curtail output when the price drops below $-r$, and are willing to produce any amount in $[0, W_t]$ when the price is $-r$.

3.3.2 System Operator's Problem

The system operator's problem can be divided into two parts. First, the ISO chooses the aggregate production levels form the different generator types, and then allocates the production from each type to the individual generator units. The second problem is given in Chapter II by Equaitons (2.1) and (2.2). Since the revealed cost is used in place of the true cost in the decentralized problem, we slightly modify these equations to

$$C^I(q^I) \overset{\text{def}}{=} \min_{q_i \geq 0} \left\{ \sum_{i \in G^I} \widehat{C}_i(q_i) : \sum_{i \in G^I} q_i = q^I \right\}, \tag{3.3}$$

$$C^F(q_t^F) \overset{\text{def}}{=} \min_{q_{jt} \geq 0} \left\{ \sum_{j \in G^F} \widehat{C}_j(q_{jt}) : \sum_{j \in G^F} q_{jt} = q_t^F \right\}. \tag{3.4}$$

The fact that $\widehat{C}_i$ and $\widehat{C}_j$ have the same properties as the true costs implies that Lemma II.3 holds. The following result shows the relation between the aggregate supply functions and the aggregate revealed costs.

Lemma III.2. *The aggregate revealed cost functions satisfy*

$$(C^I)'(q) = (S^I)^{-1}(q) \qquad \text{and} \qquad (C^F)'(q) = (S^F)^{-1}(q). \tag{3.5}$$

Proof of this lemma and other results in this chapter are shown in Appendix C.

With this modification, the problem of determining the aggregate production levels q^I, q_t^F, and q_t^V is identical to the Problem (2.3)-(2.6). Consequently, Theorems II.4 and II.5 hold, and the FG and FG real-time production levels are given by (2.11) and the optimal IG production level solves (2.16).

To express the spot and average prices in terms of S^F instead of C^F, we will modify the definitions of P and $\overline{P}$ from Equations (2.13) and (2.14) in Chapter II to

$$\begin{aligned} P(q^I, L_t, W_t, S^F) \stackrel{\text{def}}{=} &(S^F)^{-1}(L_t - W_t - q^I)\mathbf{1}_{A_4} - h'(q^I + W_t - L_t)\mathbf{1}_{A_3} \\ &- r\mathbf{1}_{A_2} - h'(q^I - L_t)\mathbf{1}_{A_1}, \end{aligned} \tag{3.6}$$

$$\overline{P}(q^I, S^F) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T \mathsf{E}\big[P(q^I, L_t, W_t, S^F)\big]\,dt. \tag{3.7}$$

The events A_1, A_2, A_3, and A_4 are defined in Figure 2.2 in Chapter II. Note that for a given q^I, the real-time price does not depend on IGs' supply function $S^I(\cdot)$. Also note that $P(q^I, L_t, W_t, S^F)$ decreases in q^I due to the monotonicity of $S^F(\cdot)$ and $h'(\cdot)$, and hence $\overline{P}(q^I, S^F)$ also decreases in q^I. The supply function form of Lemma III.2 also gives an alternative formula for the spot price as shown in the following result.

Corollary III.3. *The real-time price function in* (3.6) *can be expressed as*

$$P(q^I, L_t, W_t; S^F) = \inf\big\{p : S^F(p) + W_t\mathbf{1}_{\{p\geq -r\}} - \mu(-p) \geq L_t - q^I\big\}. \tag{3.8}$$

This Corollary gives a supply function-based method for calculating the real-time price. In (3.8), $S^F(p)$ is FGs' supply function, $W_t\mathbf{1}_{\{p\geq -r\}}$ is VGs' supply function (VGs offer the entire potential output whenever the price is at least $-r$), and the oversupply function $\mu(-p)$ gives the oversupply level when the real-time price is $p < 0$. According to (3.8), the real-time price is the minimum price at which the supply minus oversupply meets the demand.

With the average price $\overline{P}(q^I, S^F)$ computed in (3.7), the aggregate (constant) output rate IGs are willing to set over $[0, T]$ is $S^I\big(\overline{P}(q^I, S^F)\big)$. The system operator needs to ensure consistency between what IGs are asked to produce and what they are willing to produce. Thus, q^I must satisfy the constraint $q^I = S^I\big(\overline{P}(q^I, S^F)\big)$. Imposing this constraint, however, may prevent the system from achieving the optimal q^{I*} that minimizes the total system cost. A significant result in Theorem II.5 from Chapter II is that the optimal q^{I*} actually satisfies this constraint. Indeed, by using (3.5) and the revised average price formula (3.7), the IG production level q^{I*} from Equation (2.16) becomes the solution to

$$q^{I*} = S^I\left(\overline{P}\left(q^{I*}, S^F\right)\right). \tag{3.9}$$

Equation (3.9) confirms that imposing the constraint $q^I = S^I\big(\overline{P}(q^I, S^F)\big)$ does not sacrifice system optimality. Equation (3.9) can also be written as $(S^I)^{-1}(q^{I*}) = \overline{P}(q^{I*}, S^F)$, which means that q^{I*} is the intersection of the IGs' inverse supply function $(S^I)^{-1}(q^I)$ and the IGs' inverse residual demand function $\overline{P}(q^I, S^F)$. These two functions are depicted as the solid curves in Figure 3.1.

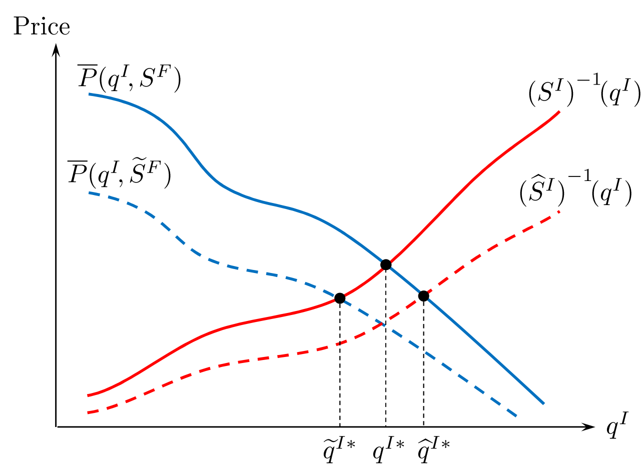


Figure 3.1: Optimal IG Production q^{I*}

How does the optimal IG production q^{I*} vary with the supply functions? When IGs bid more competitively by increasing their supply function to $\widehat{S}^I(p)$ or decreasing their inverse supply function to $(\widehat{S}^I)^{-1}(q^I)$ shown as the dashed curve in Figure 3.1, q^{I*} rises to $\widehat{q}^{I*}$, i.e., IGs' market share increases. When FGs bid more competitively by increasing their supply function to $\widetilde{S}^F(p)$, (3.8) implies that the real-time price decreases, and the average price decreases to $\overline{P}(q^I, \widetilde{S}^F)$, as shown in Figure 3.1. Consequently, q^{I*} reduces to $\widetilde{q}^{I*}$. In both cases, more competitive supply offers lead to a lower average market price. These results are in line with our intuition.

3.3.3 The Market Mechanism

Theorems II.4 and II.5 in Chapter II solve the system operator's problem of deciding the optimal production for all generators to minimize the expected total cost (implied by the generators' supply offers). We now formally define the market mechanism based on the these results.

1) Prior to $t = 0$, IGs and FGs submit supply functions $\{S_i(p) : i \in G^I\}$ and $\{S_j(p) : j \in G^F\}$, and VGs offer price $-r$ (assumed in §3.3.1).

2) The system operator clears the market prior to $t = 0$ according to the following steps:

   (i) Find the aggregate IG and FG supply functions:

$$S^I(p) = \sum_{i \in G^I} S_i(p) \qquad \text{and} \qquad S^F(p) = \sum_{j \in G^F} S_j(p).$$

   (ii) Compute the real-time price as a function of the IG output q^I, the load L, and the VG potential output W:

$$P(q^I, L, W, S^F) \;=\; \inf\left\{p : S^F(p) + W\mathbf{1}_{\{p \geq -r\}} - \mu(-p) \geq L - q^I\right\}, \qquad \text{or}$$

$$\begin{aligned} P(q^I, L, W, S^F) \;&=\; (S^F)^{-1}(L - W - q^I)\mathbf{1}_{A_4} - h'(q^I + W - L)\mathbf{1}_{A_3} \\ &\quad - r\mathbf{1}_{A_2} - h'(q^I - L)\mathbf{1}_{A_1}. \end{aligned}$$

(iii) Determine the IG output rate q^{I*} by

$$\left(S^I\right)^{-1}(q^{I*}) \;=\; \overline{P}(q^{I*}, S^F) \equiv \frac{1}{T}\int_0^T \mathsf{E}\left[P(q^{I*}, L_t, W_t, S^F)\right]\, dt. \qquad (3.10)$$

3) Production and market settlement:

(i) IG $i \in G^I$ produces $S_i(\overline{P}(q^{I*}, S^F))$ for all $t \in [0, T]$.

(ii) At any $t \in [0, T]$, the real-time price is $p_t \equiv P(q^{I*}, L_t, W_t)$, and FG $j \in G^F$ produces $S_j(p_t)$.

(iii) VGs produce W_t if $p_t > -r$, produce $L_t - q^I + \mu(r)$ if $p_t = -r$, and do not produce if $p_t < -r$.

(iv) All generators are paid p_t per unit of output at time t.

The above mechanism is common knowledge to all generators. In the next section, we will analyze the equilibrium behavior in a supply function competition.

3.4 Supply Function Competition

In the classical supply function equilibria (SFE) literature, each firm submits a supply function such that for each demand realization, the firm behaves as a monopolist with respect to its residual demand. Hence, the mark-up percentage is inversely proportional to the elasticity of the residual demand. Because this elasticity consists of derivatives of competitors' supply functions, a SFE satisfies a system of differential equations (*Klemperer and Meyer*, 1989).

Unlike the supply function competitions analyzed in the literature where generators essentially are all FGs, in our model, generators have different levels of flexibility. FGs in our model not only compete among themselves, but also compete with all IGs for market share, manifested in the IG production q^{I*} in Theorem II.5 and the discussion at the end of §3.3.2. Therefore, the residual demand facing an individual FG or IG depends on q^{I*}, which in turn depends on its own supply function. This is the key difference between our model and the classical SFE model.

For our model, it is possible to derive a system of differential equations for the equilibria; see Chapter IV for the detailed derivation and special solution to the system of differential equations when the real-time price is always positive. Appendix C also shows the derivation for the general problem. For the purpose of this chapter, we are interested in all possible price scenarios, especially the situations involving VG curtailment (see the four events illustrated in Figure 2.2). However, solving for the general SFE for our model raises several challenges. First, the first-order conditions are only necessary; the optimality of a generator's response depends on other conditions such as the shape of the supply functions and the probability distributions of uncertainties. These conditions are much more difficult to analyze than in the classical model (where the SFE does not depend on the probability distributions of the uncertainties). Second, there is usually a continuum of equilibria for the classic SFE model, thus a common approach in the literature is to focus on some special class of supply function equilibria. For our model with different classes of generators, the general theoretical analysis is even more difficult.

SFE with linear supply functions are considered in the classical works by *Klemperer and Meyer* (1989); *Green* (1996); *Rudkevich* (1999), among others. One of the goals of this chapter is to examine how firms' (in)flexibility affects SFE (see the first two questions raised in §3.1) and compare the results with the insights obtained from those in the literature. To compare our model with the classical model, we focus on SFE with linear supply functions. In fact, using the historical generator offer data from MISO[3], we find that more than 70% of the generators (excluding those who submit only one price-quantity pair) actually submit affine supply functions.

[3]Available at https://www.misoenergy.org/Library/MarketReports

3.4.1 Linear Supply Function Equilibrium

From this point onward, we consider the case where each generator's production cost rate is quadratic in its output rate (see, e.g., *Green* 1996):

$$C_k(q) = \frac{1}{2}c_k q^2, \qquad k \in G^I \cup C^F, \quad c_k > 0, \quad q \geq 0, \tag{3.11}$$

which implies a linear marginal cost $C_k'(q) = c_k q$. Hence, in a perfectly competitive market, generator k would submit the inverse marginal cost as its supply function, i.e., $S_k(p) = c_k^{-1}p^+$. In an imperfect competition, we assume generators submit linear supply functions:

$$S_k(p) = \beta_k p^+, \qquad k \in G^I \cup C^F, \quad \beta_k > 0, \quad p \in \Re. \tag{3.12}$$

That is, when the real-time price (or average price over $[0, T]$) is positive, the output rate that generator $j \in G^F$ (or $i \in G^I$) is willing to produce is linear in price. Our market equilibrium analysis is confined within the linear supply function strategies. The pure strategy set will be defined in §3.4.1 after we discuss the optimal IG production in §3.4.1.

We assume the oversupply penalty function is

$$h(e) = a_h e + \frac{1}{2}c_h e^2, \qquad a_h \geq 0, \quad c_h > 0, \quad e \geq 0. \tag{3.13}$$

Optimal Dispatch under Given Supply Functions

For given $q^I > 0$, Theorem II.4 gives the optimal $q_t^{F*} = (L_t - q^I - W_t)^+$ and $q_t^{V*} = \min\left\{W_t, \, (L_t - q^I + \mu(r))^+\right\}$, where $\mu(r) = (h')^{-1}(r) = (r - a_h)^+/c_h$ based on the penalty in (3.13).

The results on the optimal q^{I*} in Theorem II.5 are specialized below. The aggregate IG and FG supply functions are

$$S^I(p) = \beta^I p^+, \qquad \text{and} \qquad S^F(p) = \beta^F p^+,$$

where $\beta^I \stackrel{\text{def}}{=} \sum_{i \in G^I} \beta_i$ and $\beta^F \stackrel{\text{def}}{=} \sum_{j \in G^F} \beta_j$. Hereafter, we write the price $P(q^I, L_t, W_t, S^F)$ in (2.13) as $P(q^I, L_t, W_t, \beta^F)$, and the average price $\overline{P}(q^I, S^F)$ in (4.2) as $\overline{P}(q^I, \beta^F)$, expressed as follows:

$$\begin{aligned} P(q^I, L_t, W_t, \beta^F) = \frac{1}{\beta^F}(L_t - W_t - q^I)\mathbf{1}_{A_4} - \left[a_h + c_h(q^I - L_t + W_t)\right]\mathbf{1}_{A_3} \\ - r\mathbf{1}_{A_2} - \left[a_h + c_h(q^I - L_t)\right]\mathbf{1}_{A_1}, \end{aligned} \tag{3.14}$$

$$\overline{P}(q^I, \beta^F) = \frac{1}{T}\int_0^T \mathsf{E}\big[P(q^I, L_t, W_t, \beta^F)\big]dt. \tag{3.15}$$

In most of the practical situations, the system operator instructs IGs to produce a positive output and the average market price is also positive. Thus, we assume the optimal $q^{I*} > 0$. Equation (4.1) that determines q^{I*} can be written as

$$q^I = \beta^I \overline{P}(q^I, \beta^F). \tag{3.16}$$

There is a unique q^{I*} satisfying (3.16). We denote this unique q^{I*} as a function of β^I and β^F:

$$q^{I*} \equiv Q^I(\beta^I, \beta^F) \stackrel{\text{def}}{=} \big\{q^I : q^I = \beta^I \overline{P}(q^I, \beta^F)\big\}. \tag{3.17}$$

Lemma III.4. *The total IG output rate $Q^I(\beta^I, \beta^F)$ strictly increases in β^I and strictly decreases in β^F.*

The monotonicity of $Q^I(\beta^I, \beta^F)$ is intuitively illustrated in Figure 3.1 and formally proved in Lemma III.4 in Appendex C.

Pure Strategy Set

In the linear supply function competition, the supply function slopes, β_k, $k \in G^I \cup C^F$, are strategic variables. This section establishes the bounds on β_k. These bounds form a compact and convex pure strategy set, which is needed to establish the existence of the equilibrium in §3.4.2.

For generator k's supply function $S_k(p) = \beta_k p^+$, a larger β_k implies a more competitive supply offer. The discussion preceding (3.12) reveals that an upper bound for β_k is $c_k^{-1} < \infty$, which is what generator k would offer in face of perfect competition.

For FG $j \in G^F$, a lower bound on β_j can be found by solving a less competitive game in which IGs do not exist; for IG $i \in G^I$, a lower bound on β_i can be obtained by considering a less competitive game in which FGs do not exist and the demand is constant over $[0, T]$, but its level is uncertain prior to $t = 0$. These games are the same as the standard supply function game considered by *Klemperer and Meyer* (1989). *Rudkevich* (1999) studies the linear SFE for such games and shows that the slopes of the equilibrium supply functions are strictly positive and independent of the demand distribution. Hence, β_k is bounded from below by a strictly positive number, denoted as $\beta_k^{\min} > 0$, which is independent of the distribution of the uncertainties.

We define the pure strategy set of generator k as $[\beta_k^{\min}, c_k^{-1}]$. The slopes of the aggregate IG and FG supply functions are also bounded: $\beta^I \in [\beta^{I\min}, \beta^{I\max}]$ and $\beta^F \in [\beta^{F\min}, \beta^{F\max}]$, where

$$\beta^{I\min} = \sum_{i \in G^I} \beta_i^{\min}, \quad \beta^{I\max} = \sum_{i \in G^I} c_i^{-1}, \quad \beta^{F\min} = \sum_{j \in G^F} \beta_j^{\min}, \quad \beta^{F\max} = \sum_{j \in G^F} c_j^{-1}.$$

Using Lemma III.4, we can establish bounds on q^I as

$$q^{I\min} = Q^I(\beta^{I\min}, \beta^{F\max}) \qquad \text{and} \qquad q^{I\max} = Q^I(\beta^{I\max}, \beta^{F\min}). \tag{3.18}$$

Furthermore, we assume $\overline{P}(0, \beta^{F\max}) > 0$, i.e., when IGs do not exist and FGs reveal their true marginal cost, the average price is positive, which is a mild assumption. This assumption implies $q^{I\min} > 0$, because $q^{I\min}$ is the unique solution to $q^I = \beta^{I\min}\,\overline{P}(q^I, \beta^{F\max})$.

Individual IG's Problem

We now formulate how an individual IG $i \in G^I$ chooses β_i in response to all other generators' supply functions. Given an average price $\overline{P} > 0$, generator i will produce at rate $S_i(\overline{P}) = \beta_i \overline{P}$ throughout $[0, T]$, and incurs a cost rate of $\frac{1}{2}c_i(\beta_i \overline{P})^2$. Thus, the profit rate is $\beta_i \overline{P}^2 - \frac{1}{2}c_i(\beta_i\overline{P})^2 = \beta_i\big(1 - \frac{1}{2}c_i\beta_i\big)\overline{P}^2$. Note that $\overline{P}$ depends on q^I and β^F through (3.15), and q^I is affected by β_i through (3.16). Hence, IG i's optimization problem can be written as

$$\begin{aligned} \max_{\beta_i} \quad & \beta_i\Big(1 - \frac{1}{2}c_i\beta_i\Big)\overline{P}(q^I, \beta^F)^2 \\ \text{s.t.} \quad & (3.16) \text{ and } \beta_i \in [\beta_i^{\min}, c_i^{-1}]. \end{aligned} \tag{3.19}$$

The system-level optimization yields (3.16), and the firm-level objective is given by (3.19). Thus, this formulation is similar to the bi-level optimization procedure described in *Hobbs et al.* (2000).

Using (3.16) and (3.17), we can write the price function as

$$\overline{P}(q^I, \beta^F) = \frac{q^I}{\beta^I} = \frac{Q^I(\beta^I, \beta^F)}{\beta^I}.$$

We define $\beta_{-i} \stackrel{\text{def}}{=} \beta^I - \beta_i$ and rewrite the objective in (3.19) as a function of the strategic variables:

$$\pi_i(\beta_i; \beta_{-i}, \beta^F) \stackrel{\text{def}}{=} \frac{\beta_i\big(1 - \frac{1}{2}c_i\beta_i\big)}{(\beta_i + \beta_{-i})^2}\,\big(Q^I(\beta_i + \beta_{-i}, \beta^F)\big)^2. \tag{3.20}$$

The best response of IG i to β_{-i} and β^F is determined by optimizing $\max_{\beta_i \in [\beta_i^{\min}, c_i^{-1}]} \pi_i(\beta_i; \beta_{-i}, \beta^F)$.

Individual FG's Problem

An individual FG $j \in G^F$ chooses β_j in response to all other generator's supply functions. Observing price p_t at time t, generator j produces at rate $\beta_j p_t^+$ and incurs a cost rate of $\frac{1}{2} c_j (\beta_j p_t^+)^2$. Thus, the profit rate is $\beta_j (p_t^+)^2 - \frac{1}{2} c_j \big(\beta_j p_t^+\big)^2 = \beta_j \big(1 - \frac{1}{2} c_j \beta_j\big)(p_t^+)^2$. Note that $p_t = P(q^I, L_t, W_t, \beta^F)$ as defined in (3.14). Thus, generator j's problem is

$$\max_{\beta_j} \quad \int_0^T \mathsf{E}\Big[\beta_j\Big(1 - \frac{1}{2} c_j \beta_j\Big)\big(P(q^I, L_t, W_t, \beta^F)^+\big)^2\Big]\, dt \tag{3.21}$$

$$\text{s.t.} \quad (3.16) \ \text{ and } \ \beta_j \in [\beta_j^{\min}, c_j^{-1}].$$

Equations (3.14) and (3.17) lead to

$$P(q^I, L_t, W_t; \beta^F)^+ \ = \ \frac{(L_t - W_t - q^I)^+}{\beta^F} \ = \ \frac{\big(L_t - W_t - Q^I(\beta^I, \beta^F)\big)^+}{\beta^F}.$$

We define $\beta_{-j} \stackrel{\text{def}}{=} \beta^F - \beta_j$ and rewrite the objective in (3.21) as a function of the strategic variables:

$$\pi_j(\beta_j; \beta_{-j}, \beta^I) \stackrel{\text{def}}{=} \frac{\beta_j\big(1 - \frac{1}{2} c_j \beta_j\big)}{(\beta_j + \beta_{-j})^2} \int_0^T \mathsf{E}\Big[\Big(\big(L_t - W_t - Q^I(\beta^I, \beta_j + \beta_{-j})\big)^+\Big)^2\Big]\, dt. \tag{3.22}$$

Then, FG j's best response to β_{-j} and β^I is determined by optimizing

$$\max_{\beta_j \in [\beta_j^{\min}, c_j^{-1}]} \pi_j(\beta_j; \beta_{-j}, \beta^I).$$

Interactions Between IGs and FGs

The total IG production function $Q^I(\beta^I, \beta^F)$ is the only component in the profit functions (3.20) and (3.22) through which IGs and FGs interact. This implies that the competition between the two types of generators is over the average market share.

The variabilities in load and VG potential output play no (direct) role in lGs' profit function (3.20). Hence, IGs do not directly complete with FGs in meeting the variable demand. The variabilities directly affect FGs' profit in (3.22), and FGs compete among themselves to serve the variable demand.

The interactions between IGs and FGs also render the equilibrium dependent on the distributions of the uncertainties. If $Q^I(\beta^I, \beta^F)$ were constant in (3.22), then the distributions of the load L_t and VG potential output W_t would not affect the strategic choice of β_j. In fact, without IG-FG interactions, the FGs' problem reduces to that in the classical SFE model. In our model, however, the distributions of uncertainties affect the choice of β_j in (3.22), which in turn affect the strategic decisions of all other generators. This feature is in contrast with the classical SFE model. For example, *Klemperer and Meyer* (1989), *Green* (1996), *Holmberg* (2007), and *Anderson and Hu* (2008) demonstrate that supply function equilibria are independent of the demand distribution.

3.4.2 Existence of SFE under Normally Distributed Load and VG Output

Because the SFE depends on the distribution of uncertainties, the existence of an SFE is difficult to establish for general distribution of uncertainties. In this section, we show the existence of an SFE in a special case when the load and VG potential output are jointly normally distributed and the net demand variance is not too large. As common with models using normal distributions as an approximation, we assume that the probability of L_t or W_t being negative is sufficiently small such that it has a negligible effect on the equilibrium.

The fact that generators do not modify their decisions within the time horizon and that all generators are risk neutral implies that we can collapse all sample paths across time into a single probability distribution.

Denote the joint probability density function of L_t and W_t as $f_{L_t,W_t}(x,y)$, $(x,y) \in \Re^2$. Define

$$f_{L,W}(x,y) \overset{\text{def}}{=} \frac{1}{T}\int_0^T f_{L_t,W_t}(x,y)\,dt. \tag{3.23}$$

It can be verified that $f_{L,W}(x,y)$ is also a probability density function. Let L and W be the random variables that follow the distribution $f_{L,W}(x,y)$. Then, for any real-valued function $g(x,y)$, we have

$$\begin{aligned}\frac{1}{T}\int_0^T \mathsf{E}\big[g(L_t,W_t)\big]\,dt &= \frac{1}{T}\int_0^T\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{L_t,W_t}(x,y)\,dx\,dy\,dt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\left[\frac{1}{T}\int_0^T f_{L_t,W_t}(x,y)\,dt\right]dx\,dy \;=\; \mathsf{E}\big[g(L,W)\big].\end{aligned}$$

That is, the time-average of the expected value of $g(L_t,W_t)$ equals the expected value of $g(L,W)$ under the time-invariant probability distribution $f_{L,W}(x,y)$.

In this section, we assume L and W follow a bivariate normal distribution: $L \sim \mathcal{N}(\mu_L,\sigma_L^2)$ and $W \sim \mathcal{N}(\mu_W,\sigma_W^2)$ with a correlation coefficient ρ. We define the net demand random variable $D \overset{\text{def}}{=} L - W \sim \mathcal{N}(\mu_D,\sigma_D^2)$ where $\mu_D = \mu_L - \mu_W$, and $\sigma_D^2 = \sigma_L^2 + \sigma_W^2 - 2\sigma_L\sigma_W\rho$.

According to *Debreu* (1952), a sufficient condition for the existence of a Nash equilibrium of this game is that the IG and FG profit functions shown in (3.20) and (3.22) are quasi-concave in their own strategic variables. Proving the quasi-concavity under general conditions is difficult due to the complicated structure of the price function in (2.13), which gives an average price that is neither convex nor concave (see illustration in Figure 3.1). However, having $q^I < D$ (in region A_4) with a sufficiently high probability makes the average price curve close to linear, which bounds the average price's second derivative and bestows some structure to the problem. The following result gives a sufficient condition to establish such a bound on q^I.

Lemma III.5. *If* $\sigma_D \leq \sigma_D^* \equiv \sqrt{2\pi}\, \beta^{F\min}\left[\frac{\mu_D}{\beta^{I\max}} + \frac{\min\{r, a_h\}}{2}\right]$, *then* $q^{I\max} < \mu_D$.

Lemma III.5 shows that for a sufficiently small σ_D, the IG production is bounded above by μ_D. Indeed, the IGs' aggregate output does not exceed the average net demand for most situations in practice. The condition given in Lemma III.5 is not very stringent. For example, if $\min\{r, a_h\} = 0$ and $\beta^{F\min}$ is one tenth of $\beta^{I\max}$ (which according to our numerical tests is on the conservative end), then $\sigma_D^* = \sqrt{2\pi}\mu_D\, \beta^{F\min}/\beta^{I\max} \approx 0.25\mu_D$. Thus, the Lemma's condition holds if the standard deviation of the net demand is within 25% of its mean, which is a mild assumption in most practical situations. The importance of Lemma III.5 is that it provides sufficient conditions for bounding q^I, which ultimately leads to the following equilibrium existence theorem.

Theorem III.6. *When generators compete using linear supply functions and the standard deviation of the net demand* σ_D *is sufficiently small, there exists a (pure strategy) supply function equilibrium.*

The result above establishes the existence of a linear supply function equilibrium under certain sufficient conditions. Our numerical experiments, however, show that the equilibrium exists for a wider range of demand variances and for other load and VG output distributions. Indeed, a linear SFE is obtained in all our numerical examples. The following section presents the results and insights from some of these numerical experiments.

3.5 Numerical Study

In this section, we compute the SFE based on the model analyzed in §§3.3-3.4. We compare our results with the classical SFE model by *Klemperer and Meyer* (1989) and *Green* (1996). We also analyze the effect of increasing VG penetration and its curtailment on the SFE. Our analysis does not aim to predict the magnitude of these effects in reality, but to derive qualitative insights and provide policy recommendations.

3.5.1 Setups and Computational Procedure

We focus on the linear SFE described in §3.4.1. We consider a system consisting of four IGs indexed by $i \in G^I = \{1, 2, 3, 4\}$ and four FGs indexed by $j \in G^F = \{5, 6, 7, 8\}$. The production cost rates (in \$/hour) of these generators are quadratic in their power generation rates, as assumed in (3.11). That is, $C_k(q) = \frac{1}{2} c_k q^2$, where q is in MW, $c_k = 4$ \$/MWh/MW for $k \in G^I$, and $c_k = 12$ \$/MWh/MW for $k \in G^F$. The cost functions are kept identical within each generator group to facilitate comparison between IGs and FGs; our computational procedure allows for different cost functions. We assume the system's oversupply penalty for e MW of oversupply is $h(e) = 2e^2$ \$/hour. We set the operating time horizon $T = 1$ day, and assume that each generator submits a single supply function for the entire operating day; see discussion in §3.3.1.

We next specify the time-invariant probability distribution $f_{L,W}(x, y)$ defined in (3.23). We assume the load L and VG potential output W follow independent normal distributions, with $\mu_L = 100$, $\sigma_L = 15$, $\mu_W = 5$, and $\sigma_W = 1.75$ in the base case. These parameters are in MWh per 5 minutes. (Many electricity systems measure load and VG output at 5-minute intervals, which can be used to estimate these parameters.)

The VG penetration level is $\mu_W/\mu_L = 5\%$ (close to the current VG penetration in the U.S.). In addition to this base case, we also consider various VG penetration levels. Following *Wu and Kapuscinski* (2013), when VG penetration increases by m times (μ_W increases by m times), the standard deviation σ_W increases by m times if the existing and added VG outputs are perfectly correlated, or $\sqrt{m}$ times if they are independent. The realistic case is likely in between and we assume that σ_W increases by $m^{0.75}$ times. Specifically, we consider four additional VG penetration levels: 0%, 15%, 30%, and 50%. That is, $m = 0, 3, 6, 10$.

We consider the following VG curtailment policies: priority dispatch for VG (no curtailment), economic curtailment for VG (when subsidy $r = 0$), partial economic curtailment (when subsidy $r = 20$ or 40 \$/MWh).

The generators submit linear supply functions $S_k(p) = \beta_k p^+$, described in (3.12). The following procedure is used to compute the generators' equilibrium supply functions:

Step 1. Start with iteration $n = 0$ and set the initial slopes $\beta_k^n \leq c_k^{-1}, \forall\, k \in G^I \cup G^F$.

Step 2. Let $n := n + 1$. For every generator $k \in G^I \cup G^F$, find the optimal slope β_k^n that maximizes generator k's profit, assuming that none of the other generators modify their slopes (i.e. use the β_l^{n-1} slopes from the previous iteration for generators $l \neq k$). The market equilibrium condition (3.16) is used in this step to find q^I and the spot price function.

Step 3. If $\max\limits_{k \in G^I \cup G^F} \left\{ \left|\beta_k^n - \beta_k^{n-1}\right| / \beta_k^{n-1} \right\} < \varepsilon$, then terminate the procedure and the equilibrium slopes are $\{\beta_k^n\}$, otherwise go to Step 2.

We use $\varepsilon = 0.1\%$ in the convergence criterion. The program typically takes 4 to 5 iterations to converge. We select multiple different starting points for our iteration, and all of them lead to the same equilibrium in the numerical examples tested.

For the purpose of presenting the insights from our numerical experiments, it is more intuitive to describe generators' strategies using the slopes of their inverse supply functions:

$$\gamma_k \stackrel{\text{def}}{=} 1/\beta_k.$$

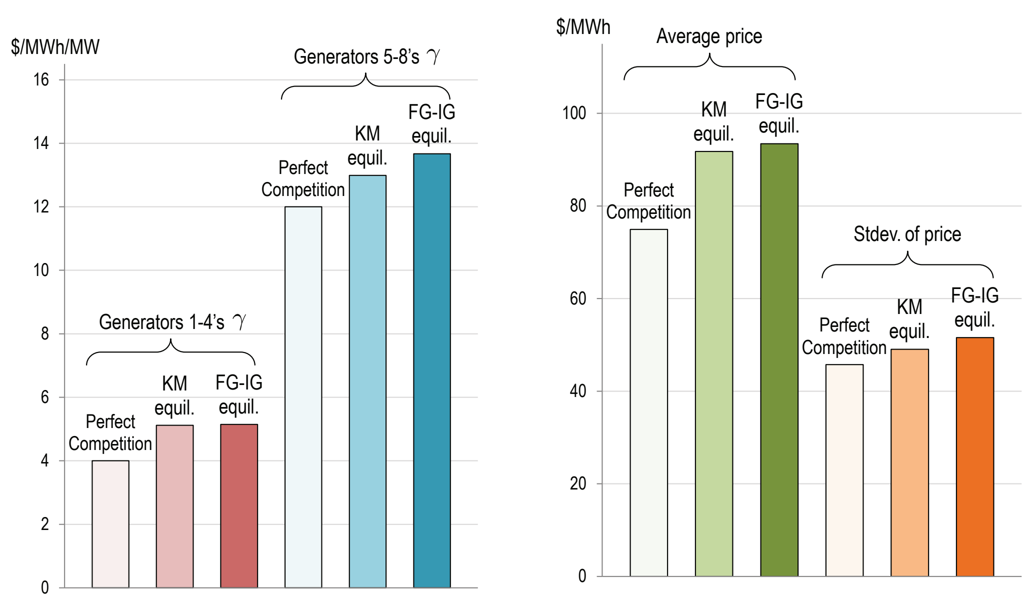
We refer to γ_k as *price offer slope*. In a perfect competition, the price for generator

k to produce q is equal to its marginal cost $c_k q$. In an oligopolistic competition, generator k offers price $\gamma_k q$, $\gamma_k \geq c_k$. The closer γ_k is to c_k, the lower the markup and the more competitive the price offer is.

3.5.2 FG-IG Equilibrium vs. Klemperer-Meyer Equilibrium

Our model extends the classical SFE model by *Klemperer and Meyer* (1989) to include the asymmetries in both cost and flexibility. We first compare the equilibrium in our model (referred to as FG-IG equilibrium) and the Klemperer-Meyer (KM) equilibrium, focusing on the linear SFE. The KM equilibrium with linear supply functions and asymmetric cost is solved by *Green* (1996) and *Rudkevich* (1999).

We compute the KM equilibrium under the assumption that generators 1-4 are also flexible and have the same cost functions as described in §3.5.1. In the KM equilibrium, generators 1-4 each offer price $5.0\,q$ and generators 5-8 each offer price $10.9\,q$. Figure 3.2(a) shows these price offer slopes, as well as the price offer slopes in the perfect competition ($\gamma_k = c_k$) and the FG-IG equilibrium.



(a) Price offer slope γ.

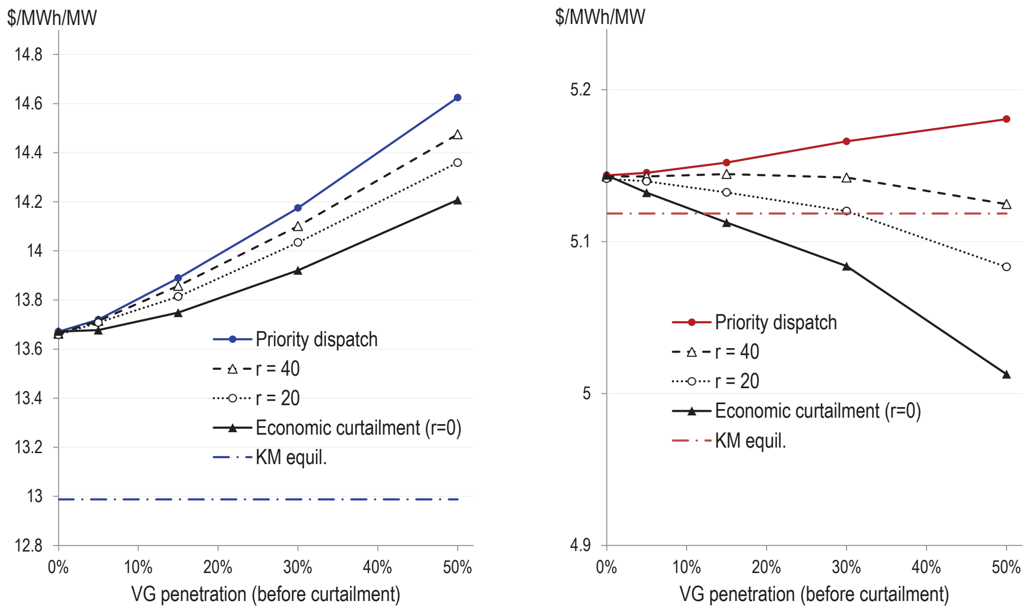
(b) Average and standard deviation of price.

Figure 3.2: Klemperer-Meyer Equilibrium vs. FG-IG Equilibrium without VG

Using the procedure in §3.5.1, we compute the FG-IG equilibrium when there is no VG in the system. The only difference in setup between our model and the KM model is that generators 1-4 are IGs. Thus, one might expect IGs 1-4 to behave more differently than FGs 5-8 compared to the KM equilibrium. However, Figure 3.2(a) shows that IGs' equilibrium price offer is slightly higher than in the KM model, whereas FGs' equilibrium price offer is significantly higher than in the KM model. The reasons stem from the asymmetry in flexibility. First, IGs do not compete with FGs in matching production with the uncertain load. Hence, the competition facing an FG in our model is less intense than that in the KM model, allowing FGs to raise their price offers. Second, FGs still compete with IGs for market share (i.e., IGs produce q^{I*} and FGs produce the rest, which are allocated by the system operator according to (3.17)). Hence, the competition facing an IG in our model is similar to that in the KM model. It is slightly less intense because FGs raise their price offers as explained in the first reason. As a result, IGs slightly raise their price offers above the KM equilibrium. The above finding suggests that the KM model tends to underestimate generators' price offers, more significantly so for FGs.

Figure 3.2(b) shows that the average and standard deviation of the real-time price in the FG-IG equilibrium are higher than those in the KM equilibrium. Because both IGs and FGs offer higher prices in the FG-IG equilibrium than in the KM equilibrium, the average price is also higher. The price volatility increases for two reasons. First, in the KM equilibrium, all generators adjust their outputs in response to load fluctuations, whereas in the FG-IG equilibrium, only four FGs respond to load fluctuations. Consequently, the price is more sensitive to load fluctuations in the FG-IG equilibrium than in the KM equilibrium. The higher FG price offers seen in Figure 3.2(a) further increase this sensitivity. Second, when the load drops below the IGs' production level, an oversupply situation occurs and the price becomes negative, whereas the price in the KM model is always positive when there is no VG.

In short, in view of both the generators' price offers and the equilibrium price, the KM model tends to overestimate the intensity of the competition in a market with inflexible generators.



(a) FGs' price offer slope γ.

(b) IGs' price offer slope γ.

Figure 3.3: Effects of VG on Equilibrium Price Offers.

3.5.3 Impact of Variable Generation (under Priority Dispatch) on SFE

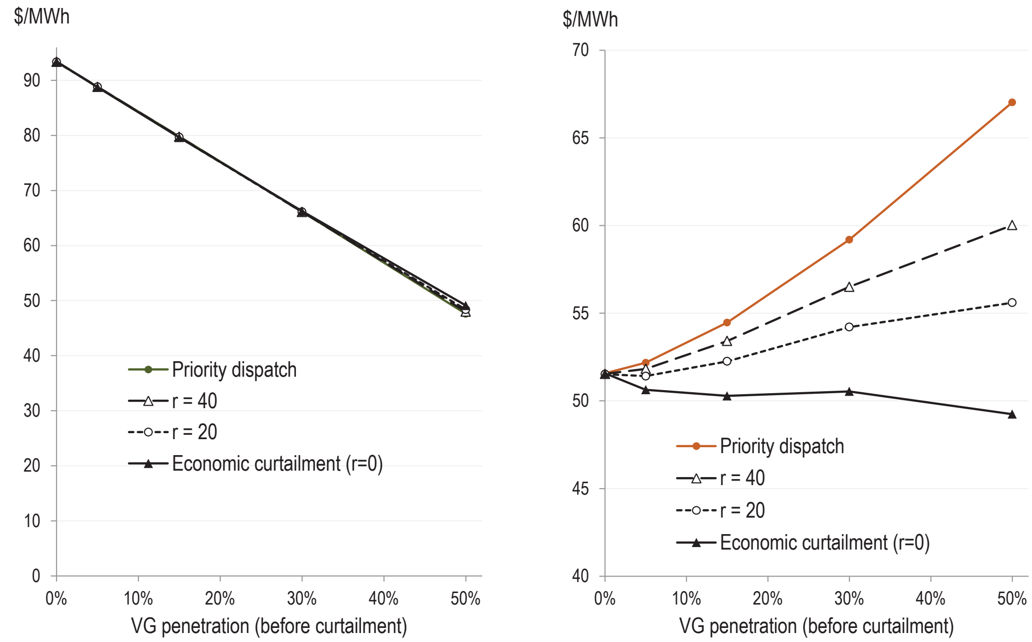
The impact of VG penetration on the FG-IG equilibrium depends on the dispatch policy for VG. In this section, we focus on the priority dispatch policy, i.e., the entire VG potential output is absorbed into the system.

The KM equilibrium is known to be independent of the distribution of the uncertainties (*Klemperer and Meyer*, 1989). Thus, the KM equilibrium price offers are invariant to the VG penetration levels and shown as the flat dashed lines in Figure 3.3.

In the FG-IG equilibrium, FGs and IGs together serve the net demand (load minus VG output). As VG penetration increases, if FGs and IGs keep their price offers unchanged, the system operator will allocate a larger market share to FGs and a smaller share to IGs to avoid significant oversupply penalty. To profit from this advantage, FGs raise their price offers as VG penetration increases, which is confirmed in Figure 3.3(a); the top curve is for priority dispatch.

On the IGs' side, as VG penetration increases, IGs face a price-quantity tradeoff: They can either increase price offers to raise the equilibrium price but get a smaller market share, or lower their price offers to gain more market share. Because under VG priority dispatch the system operator tends to keep a low IG output to mitigate the oversupply penalty, IGs' strategy of lowering price offers may not lead to an output increase sufficient to raise IGs' profit. The strategy of raising price offers turns out to be more profitable for IGs, partly because FGs also raise their price offers, seen in Figure 3.3(a). The higher IG price offers are confirmed in Figure 3.3(b); the top curve is for priority dispatch. Hence, under priority dispatch, as VG penetration increases, both IGs and FGs raise their price offers in the FG-IG equilibrium.

Although increased price offers tend to raise the market price, increased VG penetration also reduces the average net demand and exerts downward pressure on the price. The equilibrium price is a result of the combination of these two effects. The second effect dominates in determining the average equilibrium price, as shown in Figure 3.4(a), where the average equilibrium price declines as VG penetration increases. However, the first effect is important in affecting the price volatility. At a high VG penetration level, the VG output can still occasionally drop to a low level, requiring the system to ramp up FGs production to meet the demand. In such situations, FGs' increased price offers at high VG penetration levels lead to high equilibrium prices. On the other hand, when VG output surges, the system has to take all VG output because of the priority dispatch policy, resulting in an oversupply penalty and negative prices. Therefore, under priority dispatch, increasing VG penetration makes the price more volatile, as revealed by the top curve in Figure 3.4(b).



(a) Average price.

(b) Standard deviation of price.

Figure 3.4: Effects of VG on Equilibrium Price.

3.5.4 Impact of the Economic Curtailment Policy on SFE

The analysis in §3.5.3 assumes priority dispatch for VG, which is used in some electricity systems. Some other electricity systems allow curtailment of VG for economic reasons, but as discussed in §3.1, the production-based subsidy reduces the amount of VG curtailment, effectively increasing the priority for VG. In this section, we consider three subsidy levels: $r = 0$, 20, and 40 \$/MWh. We focus on discussing the economic curtailment case with zero subsidy, and refer to the cases of $r > 0$ as the partial economic curtailment cases.

The economic curtailment policy increases the competition among IGs and FGs in two ways. First, economic curtailment provides the system operator with an additional lever to manage uncertainty, and thus, the system operator allocates less production to FGs than under the priority dispatch policy. As a result, FGs offer more competitive prices to compete for market share. Second, economic curtailment significantly reduces the oversupply penalty, thereby altering the price-quantity tradeoff facing IGs (this tradeoff is described in §3.5.3). Consequently, IGs' strategy of lowering price offers can yield a market share increase that is sufficient to increase IGs' profit. These two effects of economic curtailment reinforce each other in equilibrium, because IGs reduce their price offers in response to FGs' reduced price offers and vice versa.

The combination of these effects yields the equilibrium price offers shown in Figure 3.3. The economic curtailment policy encourages both IGs and FGs to offer more competitive prices. In Figure 3.3(b), IGs' price offers may drop below the level predicted by the KM model.

Figure 3.3 also shows the effect of subsidies. Because subsidies effectively grant priority to VGs to some extent, a higher subsidy leads to less competitive price offers. Thus, the curves for the partial economic curtailment cases lie in between the curves for priority dispatch and economic curtailment.

Economic curtailment has little effect on the equilibrium average price, but the impact on price volatility is significant. Figure 3.4(a) shows that when VG penetration level is below 30%, average prices under various VG policies are indistinguishable. At higher VG penetration levels, the average price under economic curtailment is slightly higher because the curtailment reduces the severity of the negative prices. In contrast, the price standard deviation drops considerably under economic curtailment, as shown in Figure 3.4(b), because economic curtailment reduces extreme prices by making the market more competitive when prices are high and reducing the oversupply penalty when prices are negative.

3.5.5 Effects of Curtailment on Costs and Emissions

The economic curtailment policy also impacts the system operating cost, including the actual (not revealed) production cost and the oversupply penalty. Table 3.1 shows that, on average, one MWh of economic curtailment reduces the system operating cost by about \$30. This cost reduction effect is consistent across all VG penetration levels. This finding is also in line with the economic benefit of curtailment found in *Wu and Kapuscinski* (2013).

Table 3.1 also shows the effect of production-based subsidies on curtailment and the system operating cost. A higher subsidy reduces the amount of curtailment, but increases the system operating cost. In theory, when the subsidy approaches to infinity, no curtailment will occur, and the system operating cost equals that under the priority dispatch.

Interestingly, a higher subsidy also increases the value of per-MWh curtailment. For example, at 5% VG penetration with r = 20 \$/MWh, one MWh of economic curtailment reduces the system operating cost by \$49; with r = 40 \$/MWh, this value increases to \$67. This result is again consistent across all VG penetration levels. The implication is that the benefit of economic curtailment may be very high in countries and regions where VGs are heavily subsidized based on production.

An environmental benefit from increasing VG penetration is the reduced CO_2 emissions due to the replacement of the conventional production by the clean VG production. Table 3.1 confirms that the total CO_2 emission significantly decreases as VG penetration increases.

The impact of economic curtailment on CO_2 emission, however, is not as obvious and depends on the generators' fuel types. Because economic curtailment allows for more IG production and less FG production, if IGs have a higher/lower CO_2 emission rate than FGs, economic curtailment may increase/decrease total CO_2 emission. Table 3.1 demonstrates that when IGs are coal-fired/nuclear-power generators and FGs are natural gas combustion turbines, economic curtailment increases/decreases emissions. Furthermore, when subsidies exist, the emissions lie in between economic curtailment and priority dispatch cases.

3.6 Conclusion

Electricity markets have been gradually evolving toward deregulated structures that are meant to encourage competition and improve efficiency. The research in deregulated electricity markets, especially the supply function competition, has provided considerable insights into generators' bidding behavior and market power. This chapter provides new results that address how the competition is affected by generators' flexibility and variable generation. The two most important messages from this chapter are that inflexibility contributes to the market power and that the economic curtailment of variable generation increases the market competition and system efficiency.

Inflexibility contributes to the market power in the following way. Inflexible generators do not compete with flexible generators in matching production with uncertain demand, leading to increased market power for flexible generators, which in turn results in higher average price and price volatility than predicted by the classical SFE model.

Variable generation, when given priority in dispatch, exacerbates the effect of inflexibility on market competition, but the economic curtailment policy can intensify the market competition because economic curtailment serves as a partial substitute for flexible generators to balance against variability. Furthermore, economic curtailment improves system efficiency by reducing the oversupply penalty and using more inflexible generation which is less costly than flexible generation.

The insights from this chapter also provide several recommendations for the regulators and policy makers. First, in assessing the competitiveness of the electricity market, it is important to incorporate generators' flexibility/inflexibility. Flexible generators compete in balancing against variability and often set the market price. Encouraging the development of more flexible generators (e.g., fueled by natural gas) enhances the overall competitiveness of the electricity market. Second, in assessing the benefit of the economic curtailment policy, it is important to recognize that economic curtailment helps increase market competition and reduce price volatility. Policy makers need to revisit the policy of giving priority to variable generation from renewable sources, and consider a full range of benefits of economic curtailment. Other benefits of economic curtailment include reduced cycling cost and peaking cost (*Wu and Kapuscinski*, 2013), and improved production allocation in a network (*Ela*, 2009). Third, policy makers need to reconsider the design of incentives aimed to maximize the benefits of renewable energy. The design of subsidies should facilitate economic curtailment and avoid unintended consequences. Investment in research and development can push technology advancement that makes renewable energy generation more competitive in the near future even without subsidies.

It remains an area of future research to study the competition among generators with various ramping capabilities and the costs associated with ramping. A model with ramping would more accurately reflects the composition of generators in practice and would provide a more accurate estimate for the magnitude of the effects of inflexibility and economic curtailment. In this chapter, variable generators are assumed to be price-takers. A more comprehensive market competition model would include strategic behavior of variable generators as well. Furthermore, in some regions, variable generators are also allowed to submit supply functions, and certain market rules specify the actions to take when the realized output differs from what they offer to supply. We leave the analysis of such behavior and resulting supply function competition to future research.

Table 3.1: Effects of Curtailment on Costs and Emissions

Metrics	VG dispatch policy (or subsidies)	VG penetration			
		5%	15%	30%	50%
VG penetration (after curtailment)	$r = 40$	4.97%	14.91%	29.73%	49.10%
	$r = 20$	4.93%	14.79%	29.42%	48.34%
	Economic Curtailment	4.84%	14.53%	28.85%	46.99%
System operating cost (thousand $/day)	Priority Dispatch	1076.4	887.1	648.9	415.4
	$r = 40$	1075.8	885.4	643.7	397.6
	$r = 20$	1075.4	884.2	640.9	390.9
	Economic Curtailment	1075.0	883.1	638.7	386.7
System cost savings per MWh of curtailment ($/MWh)	$r = 40$	67.1	65.0	66.8	69.3
	$r = 20$	48.9	47.0	48.8	51.3
	Economic Curtailment	30.1	29.3	30.8	33.1
Total CO_2 emissions with coal-fired IGs (thousand tons/day)	Priority Dispatch	30.91	25.07	17.52	9.67
	$r = 40$	30.92	25.10	17.60	9.87
	$r = 20$	30.94	25.14	17.70	10.04
	Economic Curtailment	30.97	25.24	17.89	10.38
Total CO_2 emissions with nuclear IGs (thousand tons/day)	Priority Dispatch	4.12	3.51	2.79	2.09
	$r = 40$	4.11	3.50	2.75	1.99
	$r = 20$	4.10	3.48	2.71	1.90
	Economic Curtailment	4.09	3.44	2.63	1.76

Emission rates: 215 lb. of CO_2 per MBtu of coal, 117 lb. of CO_2 per MBtu of natural gas, no emission for nuclear power generators. Fuel price: $2.5 per MBtu of coal and $5 per MBtu of natural gas.

CHAPTER IV

Daytime Supply Function Equilibrium Model with Generation Inflexibility

4.1 Introduction

The electricity system model we have considered thus far permits the total electricity supply to exceed the demand. We have found that such events produce non-positive electricity prices. Excess electricity occurs when the electricity demand falls below its nominal level in a system with either of the two following characteristics:

1. The system has substantial IG production that cannot promptly react to the demand dips, or

2. The system has a large VG penetration with government subsidies that make it profitable for VGs to continue to generate electricity even when the prices become negative.

In practice, the electricity demand plummets during the night, while the daytime demand, especially during peak time, requires a large fleet of generators, with FGs supplying the marginal demand. This is precisely the A_4 scenario shown in Figure 2.2 with FGs as price setters.

In this chapter, we will consider the electricity system's problem during the daytime when flexible, inflexible, and variable generators all have positive production rates throughout the time horizon. In such circumstances, we assume that the load is large enough such that all the VG production is accepted (see Equation (2.11)). Therefore, we will regard VGs as negative demand and only consider flexible and inflexible generation in this chapter.

The main contribution in this chapter is that we derive the Oridinary Differential Equations (ODE) system for the SFE when there system has a mix of flexible and inflexible generators in §4.3. We then find an affine solution to the ODE system when generators have quadratic costs in §4.4. We then conclude this chapter in §4.5 by presenting some possible future research direction in this area.

4.2 The System Operator's Problem

Consider an operating time horizon $[0, T]$ during the daytime (peak hours). A system operator schedules production to satisfy a price inelastic demand D_t in time $t \in [0, T]$ using two types of generators:

- Inflexible generators (IG), indexed by $i \in G^I$, cannot adjust their output rates during $[0, T]$. The output rate of generators $i \in G^I$, denoted as $q_i \geq 0$, is determined by the system operator prior to time $t = 0$ and stays constant over $[0, T]$. We assume that i incurs an operating cost rate $C_i(q)$ for producing q MW.

- Flexible generators (FG), indexed by $j \in G^F$, can adjust their output rates instantaneously. We denote the output rate of generator $j \in G^F$ at time $t \in [0, T]$ by q_{jt}, and use a cost rate function $C_j(q)$.

We assume that every firm owns a single generator and we do not consider generator capacity limits in this problem.

Before time 0, every generator $k \in G^I \cup G^F$ submits a strictly increasing supply function $S_k(p)$. After receiving these offers, the system operator computes the revealed cost functions define as

$$\widehat{C}_k(q) \stackrel{\text{def}}{=} \int_0^q S_k^{-1}(x)dx, \qquad \forall\, k \in G^I \cup G^F,$$

where the inverse supply function is $S_k^{-1}(q) \stackrel{\text{def}}{=} \inf\{p : S_k(p) > q\}$. The objective of the system operator is to minimize the total expected revealed cost of serving the demand over $[0,T]$.

The system operator's problem can be formulated as first deciding the aggregate output rate for each type of generators and then allocating the aggregate output to individual generators. Let $q^I = \sum_{i \in G^I} q_i$ and $q_t^F = \sum_{j \in G^F} q_{jt}$ be the aggregate IG and FG output rates, respectively. The allocations of q^I and q_t^F to individual IGs and FGs are determined by minimizing the total revealed cost for each generator type:

$$C^I(q^I) \stackrel{\text{def}}{=} \min\Big\{ \sum_{i \in G^I} \widehat{C}_i(q_i) \;\; : q_i \geq 0, \; \sum_{i \in G^I} q_i = q^I \Big\},$$
$$C^F(q_t^F) \stackrel{\text{def}}{=} \min\Big\{ \sum_{j \in G^F} \widehat{C}_j(q_j) : q_j \geq 0, \; \sum_{j \in G^F} q_j = q_t^F \Big\}.$$

We can show that C^I and C^F are continuously differentiable, convex, and strictly increasing in q, and that $(C^I)'(q) = (S^I)^{-1}(q)$ and $(C^F)'(q) = (S^F)^{-1}(q)$, where S^I and S^F are aggregate supply functions defined as

$$S^I(p) \stackrel{\text{def}}{=} \sum_{i \in G^I} S_i(p) \qquad \text{and} \qquad S^F(p) \stackrel{\text{def}}{=} \sum_{j \in G^F} S_j(p).$$

Let $D_{\min} > 0$ be the minimum possible demand over $[0,T]$. Assume that no oversupply or shortage is allowed. Hence, the system operator must set $q^I \leq D_{\min}$ and $q_t^F = D_t - q^I$ at time t. Then, the FGs' revealed cost rate at time t is $C^F(D_t - q^I)$, and IGs' revealed cost rate is $C^I(q^I)$ at all times. The system operator determines q^I prior to time 0 by solving the following problem:

$$\min_{q^I \in [0, D_{\min}]} TC^I(q^I) + \mathsf{E}\Big[\int_0^T C^F(D_t - q^I)\,dt\Big].$$

By the convexity of the cost functions, we can use the first order condition to solve for the optimal q^I. Assume that the optimal solution, q^{I*}, is an interior solution: $q^{I*} \in (0, D_{\min})$, which is a typical case for the day time. The first order condition for q^I gives the following optimality condition

$$q^{I*} = S^I\big(\overline{P}(q^{I*}, S^F)\big), \tag{4.1}$$

where

$$\overline{P}(q^I, S^F) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T \mathsf{E}\big[(S^F)^{-1}(D_t - q^I)\big]\,dt, \tag{4.2}$$

$\overline{P}(q^I, S^F)$ is the average market price when the total IG production is q^I and the aggregate FG supply function is S^F. Note that $\overline{P}$ is decreasing in q^I and increasing in S^F (in the sense of $S^F(p)$ increasing uniformly).

4.3 Generators' Best Response Problem

4.3.1 IG Problem

In its best response problem, an individual IG $i \in G^I$ knows all other generator supply functions $S_{-i}(p) \stackrel{\text{def}}{=} \sum_{k \in G^I \setminus i} S_k(p)$ and $S^F(p)$. The function $\overline{P}(q^I, S^F)$ is essentially an inverse demand function facing IGs, and generator i can use the average price p as a decision variable given the residual demand function $Q^I(p) - S_{-i}(p)$, where

$$Q^I(p) \stackrel{\text{def}}{=} \inf\{q : \overline{P}(q, S^F) \geq p\}. \tag{4.3}$$

Notice that the total IG supply must match Q^I in equilibrium, which gives the relation $Q^I(p) = S^I(p)$. This equation is an alternative expression to (4.1). Thus, generator i's problem becomes:

$$\max_p \; \big(Q^I(p) - S_{-i}(p)\big)p - C_i\big(Q^I(p) - S_{-i}(p)\big).$$

The first order condition for this problem is

$$Q^I(p) - S_{-i}(p) + \Big(Q^{I\prime}(p) - S'_{-i}(p)\Big)\Big[p - C'_i\big(Q^I(p) - S_{-i}(p)\big)\Big] = 0.$$

By substituting $S_i(p)$ for $Q^I(p) - S_{-i}(p)$ the above condition can be written as

$$S'_{-i}(p) = \frac{S_i(p)}{p - C'_i(S_i(p))} + Q^{I\prime}(p), \qquad \forall\, i \in G^I. \tag{4.4}$$

This resembles the classical equilibrium condition derived by Klemperer and Meyer (1989). The difference is that the demand function in our setting shown in (4.3) is implied by FGs' supply functions. Hence, the last term in (C.12) captures how FGs' decisions affect IGs' supply functions.

The second derivative of the IG's objective is

$$\begin{aligned} &2\left(Q^{I\prime}(p) - S'_{-i}(p)\right) - \left(Q^{I\prime}(p) - S'_{-i}(p)\right)^2 C''_i(S_i(p)) \\ &+ \left(Q^{I\prime\prime}(p) - S''_{-i}(p)\right)\left[p - C'_i(S_i(p))\right]. \end{aligned} \tag{4.5}$$

From (4.3), we can express the first and second derivatives of Q^I as

$$Q^{I\prime}(p) = \frac{1}{\frac{\partial}{\partial q^I}\overline{P}(q^I, S^F)|_{q^I=S^I(p)}} \qquad \text{and} \qquad Q^{I\prime\prime}(p) = -\frac{\frac{\partial^2}{\partial {q^I}^2}\overline{P}(q^I, S^F)|_{q^I=S^I(p)}}{\left[\frac{\partial}{\partial q^I}\overline{P}(q^I, S^F)|_{q^I=S^I(p)}\right]^3}.$$

Notice that $\overline{P}$ shown in (4.2) is decreasing in q^I, and since all supply functions are increasing then Q^I is decreasing in p and $Q^{I\prime}(p) \leq 0$. Furthermore, if $\overline{P}$ is concave with respect to q^I then Q^I would be concave in p and $Q^{I\prime\prime}(p) \leq 0$. Therefore, the first term in (4.5) is negative because Q^I is decreasing and S_{-i} is increasing in p and the second term is negative because C_i is convex. If Q^I were concave and $S_i(p)$ were convex then the last term would become negative since $p - C_i'(S_i(p)) \geq 0$. Hence, the second order condition holds and (C.12) becomes sufficient if the following two conditions hold:

- The IG supply functions $S_i(p)$ are convex $\forall\ i \in G^I$.
- The average price function $\overline{P}(q^I, S^F)$ is concave with respect to q^I.

4.3.2 FG Problem

An individual FG $j \in G^F$ offers a supply function $S_j(p)$, knowing that $S_j(p)$ has two effects on the outcome of the game. The first effect is that $S_j(p)$ directly influences real-time prices $(S^F)^{-1}(D_t - q^I)$. This effect captures the fact that FG j directly competes with other FGs in satisfying the balance of the demand. This first effect also changes the average price in (4.2), which affects the market-clearing IG output rate q^{I*} in (4.1). This second effect captures the fact that FGs and IGs are also competing with each other. These two effects are intertwined because the IG output q^{I*} influences real-time price as well. Recognizing these effects, FG j decides its supply function $S_j(p)$. We first consider FG j choosing among all feasible supply functions that support a given q^I as the market-clearing IG output. Once a desired supply function is found to support each q^I, FG j will then optimize over q^I.

When the demand is D_t, generator j produces $D_t - q^I - S_{-j}(p_t)$ at a price p_t, which corresponds to a point on its supply function. Hence, we can equivalently use p_t as j's decision variable, under the constraint that the price should support q^I as the market-clearing IG output. From (4.1), this condition can be written as

$$(S^I)^{-1}(q^I) = \frac{1}{T}\int_0^T \mathsf{E}[p_t]\, dt.$$

Taking one step further, FG j chooses q^I and p_t jointly as long as they satisfy this constraint. Hence FG j's best response problem is

$$\begin{aligned} \max_{p_t,\, q^I} \quad & \int_0^T \mathsf{E}\big[\big(D_t - q^I - S_{-j}(p_t)\big)p_t - C_j\big((D_t - q^I - S_{-j}(p_t)\big)\big]\, dt \\ \text{s.t.} \quad & \int_0^T \mathsf{E}[p_t]\, dt = T(S^I)^{-1}(q^I) \qquad\qquad \perp \eta_j. \end{aligned}$$

If we denote by η_j the Lagrange multiplier of the constraint, then the Lagrangian of this problem becomes:

$$\int_0^T \mathsf{E}\big[\big(D_t - q^I - S_{-j}(p_t) - \eta_j\big)p_t - C_j\big(D_t - q^I - S_{-j}(p_t)\big)\big]\, dt + \eta_j\, T(S^I)^{-1}(q^I).$$

The first-order condition for p_t is

$$\big(D_t - q^I - S_{-j}(p_t) - \eta_j\big) - S'_{-j}(p_t)\big[p_t - C'_j(D_t - q^I - S_{-j}(p_t))\big] = 0.$$

Substituting $D_t - q^I - S_{-j}(p_t)$ by $S_i(p_t)$, the above condition can be written as

$$S'_{-j}(p) = \frac{S_j(p) - \eta_j}{p - C'_j(S_j(p))} \qquad \forall\, j \in G^F. \tag{4.6}$$

The second derivative of the Lagrangian with respect to p_t is

$$-2S'_{-j}(p_t) - (S'_{-j}(p_t))^2 C''_j(S_j(p_t)) - S''_{-j}(p_t)\big[p_t - C'_j(S_j(p_t))\big].$$

The first term is negative since S_{-j} is increasing, the second term is negative because C_j is convex, and the third term is negative if S_{-j} is convex since $p_t \geq C'_j(S_j(p_t))$. Therefore, the second-order condition holds for p_t and (C.17) becomes sufficient if all the FG supply functions are convex.

The first-order condition with respect to q^I is

$$-\int_0^T \mathsf{E}[p_t - C'_j(S_j(p_t))]\,dt + T\frac{\eta_j}{S^{I\prime}((S^I)^{-1}(q^I))} = 0.$$

By substituting $(S^I)^{-1}(q^I) = \frac{1}{T}\int_0^T \mathsf{E}[p_t]\,dt = \overline{P}(q^I, S^F)$, the dual variable η_j becomes

$$\eta_j = S^{I\prime}(\overline{P}(q^I, S^F))\frac{1}{T}\int_0^T \mathsf{E}[p_t - C'_j(S_j(p_t))]\,dt \qquad \forall\ j \in G^F. \tag{4.7}$$

Since the spot price p_t depends on q^I and S^F, then η_j depends on all generator bids. The second derivative of the Lagrangian with respect to q^I is

$$-\int_0^T C''_j(S_j(p_t))\,dt - T\eta_j\frac{S^{I\prime\prime}((S^I)^{-1}(q^I))}{S^{I\prime}((S^I)^{-1}(q^I))^3}.$$

Note that the second-order condition holds and (4.7) becomes sufficient if S^I is a linear function.

4.3.3 ODE Summary

The following is a summary of the ODE system that characterizes the first order conditions for the SFE.

$$
\begin{aligned}
S'_{-i}(p) &= \frac{S_i(p)}{p - C'_i(S_i(p))} + Q^{I\prime}(p), \qquad \forall\, i \in G^I, \\
S'_{-j}(p) &= \frac{S_j(p) - \eta_j}{p - C'_j(S_j(p))} \qquad \forall\, j \in G^F, \\
\eta_j &= S^{I\prime}(\overline{P}(q^I, S^F))\frac{1}{T}\int_0^T \mathsf{E}[p_t - C'_j(S_j(p_t))]\,dt \qquad \forall\, j \in G^F,
\end{aligned}
$$

where

$$
\begin{aligned}
&\overline{P}(q^I, S^F) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T \mathsf{E}\big[(S^F)^{-1}(D_t - q^I)\big]\,dt, \\
&Q^I(p) \stackrel{\text{def}}{=} \inf\{q : \overline{P}(q, S^F) \geq p\}, \\
&p_t = (S^F)^{-1}(D_t - q^I), \\
&q^I = S^I\big(\overline{P}(q^I, S^F)\big).
\end{aligned}
$$

It turns out that this system admits an affine supply function solution when the costs are quadratic as we will show next.

4.4 Affine SFE

We assume in this section that the production cost for generator $k \in G^I \cup G^F$ is

$$C_k(q) = a_k q + \frac{1}{2}c_k q^2,$$

and the supply functions take the form

$$S_k(p) = \beta_k(p - \alpha_k).$$

We will often refer to β_k as the slope bid and α_k as the intercept bid of generator k. For this linear supply function form to be valid we need to ensure that the average price exceeds α_i for $i \in G^I$ and the minimum price exceeds α_j for $j \in G^F$, otherwise generators may have negative production values. Furthermore, the average price must not fall below a_i for $i \in G^I$ and the minimum price must not fall below a_j for $j \in G^F$ because a rational generator would benefit from setting its supply bid to 0 for all prices below its initial marginal cost, rendering the linear supply function form invalid.

If we denote minimum demand by $D_{\min}$ and the average and minimum prices by $\overline{P}$ and $p_{\min}$, then the following list summarizes our assumptions about the problem:

(1) $0 < q^I < D_{\min}$. (2) $a_i < \overline{P}$ for $i \in G^I$. (3) $\alpha_i < \overline{P}$ for $i \in G^I$.

(4) $a_j < p_{\min}$ for $j \in G^F$. (5) $\alpha_j < p_{\min}$ for $j \in G^F$.

We will show in §4.4.5 conditions under which our assumption about these properties is valid. Note that the second order conditions for the SFE ODE system are satisfied for affine supply functions. Therefore, if we can find slopes β_k and intercepts α_k for all generators $k \in G^I \cup G^F$ that solve the ODE system (C.12), (C.17), and (4.7) then this solution would give a valid SFE.

4.4.1 The System Operator's Problem

Given the generator supply bids, the system operator sets the IG production level q^I such that (4.1) is satisfied. We will find in this subsection the equilibrium q^I, the average spot price, and the spot price as a function of the realized demand for given IG and FG slopes and intercepts β_k and α_k.

Define $\beta^I = \sum_i \beta_i$, $\gamma^I = \sum_i \beta_i\alpha_i$, $\beta^F = \sum_j \beta_j$, and $\gamma^F = \sum_j \beta_j\alpha_j$. The aggregate

IG and FG supply functions are

$$S^I(p) = \beta^I p - \gamma^I \qquad \text{and} \qquad S^F(p) = \beta^F p - \gamma^F,$$

and the aggregate inverse supply functions for IGs and FGs are

$$(S^I)^{-1}(q) = \frac{q + \gamma^I}{\beta^I} \qquad \text{and} \qquad (S^F)^{-1}(q) = \frac{q + \gamma^F}{\beta^F}.$$

The average price and its derivative with respect to its first argument are

$$\overline{P}(q^I, S^F) = \frac{1}{T}\int_0^T \mathsf{E}\left[(S^F)^{-1}(D_t - q^I)\right]\, dt = \mathsf{E}\left[\frac{\frac{1}{T}\int_0^T D_t\, dt - q^I + \gamma^F}{\beta^F}\right] = \frac{\overline{D} - q^I + \gamma^F}{\beta^F},$$

$$\frac{\partial}{\partial q^I}\overline{P}(q^I, S^F) = -\frac{1}{\beta^F}.$$

where $\overline{D}$ is the average demand in the operational time horizon. The IG production function $Q^I(p)$ shown in (4.3) and its derivative are

$$\begin{aligned} Q^I(p) =& \beta^F(\overline{D} + \gamma^F - p), \\ Q^{I\prime}(p) =& -\beta^F. \end{aligned}$$

From the equilibrium condition (4.1) we can get the following

$$q^{I*} = \frac{1}{\beta^I + \beta^F}\left[\beta^I(\overline{D} + \gamma^F) - \gamma^I\beta^F\right], \tag{4.8}$$

$$\overline{P}(q^{I*}, S^F) = (S^I)^{-1}(q^{I*}) = \frac{\overline{D} + \gamma^I + \gamma^F}{\beta^I + \beta^F}. \tag{4.9}$$

Furthermore, we can calculate the spot price when the demand is D_t as

$$\begin{aligned}P(D_t) =&(S^F)^{-1}(D_t - q^{I*}) = \frac{D_t - q^{I*} + \gamma^F}{\beta^F} \\ =&\frac{1}{\beta^I + \beta^F}\left[\gamma^I + \gamma^F - \frac{\beta^I}{\beta^F}\overline{D}\right] + \frac{D_t}{\beta^F} = \overline{P} + \frac{1}{\beta^F}\left(D_t - \overline{D}\right). \end{aligned} \tag{4.10}$$

4.4.2 The IG Problem

The IG ODE for generator i is given by

$$\begin{aligned} S'_{-i}(p) &= \frac{S_i(p)}{p - C'_i(S_i(p))} + Q^{I'}(p) \\ \Rightarrow \beta_{-i} &= \frac{\beta_i(p-\alpha_i)}{p - (a_i + c_i\beta_i(p-\alpha_i))} - \beta^F \\ \Rightarrow \beta_{-i} + \beta^F &= \frac{\beta_i p - \beta_i\alpha_i}{p(1-c_i\beta_i) - (a_i - c_i\beta_i\alpha_i)}. \end{aligned}$$

For the slope and intercept bids to be independent of p, the solution must satisfy

$$\begin{aligned} \beta_{-i} + \beta^F &= \frac{\beta_i}{1 - c_i\beta_i}, \qquad \text{and} \\ \frac{\beta_i}{1-c_i\beta_i} &= \frac{\beta_i\alpha_i}{a_i - c_i\beta_i\alpha_i}. \end{aligned} \tag{4.11}$$

If we solve for the α_i values in the last equation we get

$$\alpha_i = a_i, \tag{4.12}$$

which is the well known incentive compatibility result by Rudkevich (1999). This also shows that the β_i values are determined from the competition only among IGs, but with a price elastic demand with the FG slope bids setting the demand derivative. Therefore, we need to know β^F in order to determine the equilibrium IG slope bids.

4.4.3 The FG Problem

The FG ODE for generator j is given by

$$\begin{aligned}
S'_{-j}(p) &= \frac{S_j(p) - \eta_j}{p - C'_j(S_j(p))} \\
\Rightarrow \beta_{-j} &= \frac{\beta_j(p - \alpha_j) - \eta_j}{p - (a_j + c_j\beta_j(p - \alpha_j))} = \frac{\beta_j p - (\beta_j\alpha_j + \eta_j)}{p(1 - c_j\beta_j) - (a_j - c_j\beta_j\alpha_j)}.
\end{aligned}$$

To make β_j and α_j independent of p the solution needs to satisfy

$$\begin{aligned}
\beta_{-j} &= \frac{\beta_j}{1 - c_j\beta_j}, \qquad \text{and} \\
\frac{\beta_j}{1 - c_j\beta_j} &= \frac{\beta_j\alpha_j + \eta_j}{a_j - c_j\beta_j\alpha_j}.
\end{aligned} \tag{4.13}$$

This gives two results. First, FGs can determine their slopes independently from IG bids or their own intercept bids by solving a linear SFE competition with only FG participants and with inelastic demand. Second, Rudkevich's incentive compatibility result for the intercepts does not hold for FGs in general, and their intercept bids can be determined by

$$\alpha_j = a_j - \eta_j \left(\frac{1}{\beta_j} - c_j \right). \tag{4.14}$$

An FG's intercept bid equals its linear price coefficient if and only if its η_j value is zero. By substituting $(S^I)'(p) = \beta^I$ and (4.9) into (4.7), the dual variable η_j can be determined by

$$\eta_j = \beta^I \left[\frac{\overline{D} + \gamma^I + \gamma^F}{\beta^I + \beta^F}(1 - c_j\beta_j) - a_j + c_j\beta_j\alpha_j \right]. \tag{4.15}$$

This equation shows that η_j depends on the slope and intercept bids of all generators as well as the average demand, and hence even if β_j is known, Equation (4.14) alone is not sufficient to determine α_j. Instead, α_j and η_j can be found simultaneously by solving the system of linear equations (4.14) and (4.15) after determining $\beta_j \; \forall \; j \in G^F$ and β^I.

The dual variable η_j can be interpreted as generator j's reaction to q^I. A large η_j is indicative of j's resistant to IGs increasing their production quantity, while a small negative η_j suggests that j benefits from IGs raising their production. By substituting the average price $\overline{P}$ in (4.9) into (4.15), and rearranging the terms we get

$$\eta_j = \frac{1 - c_j\beta_j}{\frac{1}{\beta^I} + c_j(1 - c_j\beta_j)}(\overline{P} - a_j).$$

Since we only consider cases where the price exceeds the generator initial marginal costs, this formula shows that η_j is always positive for our problem, implying that FGs benefit from lowering q^I. This observation has an important implication: *FGs submit lower bid intercepts than their cost function intercepts.* This can be easily verified by substituting a positive η_j into (4.14) to get $\alpha_j < a_j$. This, however, does not imply that FGs produce at a loss because the price is assumed to never fall below a_j during the daytime.

4.4.4 Calculation Method

Given the problem parameters c_k and a_k for $k \in G^I \cup G^F$ and $\overline{D}$, we will develop in this section an algorithm to find the β_k and α_k bids for $k \in G^I \cup G^F$.

From §4.4.2 and §4.4.3, $\alpha_i = a_i$ and β_j can be determined independently from other bids, and hence the first step of the algorithm is to solve for β_j using (4.13). Given β_j, the IG slope bids can be found from (4.11), which is the second step of our algorithm. Finally, given all other bids, we can simultaneously find all the α_j and η_j values by solving the system of linear equations given by (4.14) and (4.15). An outline of this algorithm is shown below.

Algorithm IV.1.

Step 0: *Set* $\alpha_i = a_i$ *for* $i \in G^I$.

Step 1: *Find* β_j *for* $j \in G^F$ *by solving* $\beta_{-j} = \frac{\beta_j}{1-c_j\beta_j}$. *Set* $\beta^F = \sum_j \beta_j$.

Step 2: *Find* β_i *for* $i \in G^I$ *by solving* $\beta_{-i} + \beta^F = \frac{\beta_i}{1-c_i\beta_i}$. *Set* $\beta^I = \sum_i \beta_i$ *and* $\gamma^I = \sum_i \beta_i \alpha_i$.

Step 3: *Find* α_j *and* η_j *for* $j \in G^F$ *by solving the linear system*

$$\alpha_j + \left(\frac{1}{\beta_j} - c_j\right)\eta_j = a_j$$
$$-\frac{1-c_j\beta_j}{\beta^I+\beta^F}\sum_{k\in G^F}\beta_k\alpha_k - c_j\beta_j\alpha_j + \frac{1}{\beta^I}\eta_j = \frac{\overline{D}+\gamma^I}{\beta^I+\beta^F}(1-c_j\beta_j) - a_j.$$

Steps 1 and 2 of the algorithm can be solved using the method by Rudkevich (1999) shown in Appendix D.1, which gives $\beta_k \in (0, \frac{1}{c_k})$.

4.4.5 Parameter Ranges

To ensure that the linear SFE solution is valid we have to ensure that the 5 conditions listed in the beginning of this section hold. Note that conditions (2) and (3) are the same since IGs submit the same intercept bids as their initial marginal costs, and condition (4) implies (5) because FGs bid smaller intercepts than their initial marginal costs.

According to (4.10), the minimum price occurs at the minimum demand $D_{\min}$, and hence the fourth condition is given by

$$a^F_{\max} < P(D_{\min}) = \frac{1}{\beta^I+\beta^F}\left[\gamma^I + \gamma^F - \frac{\beta^I}{\beta^F}\overline{D}\right] + \frac{D_{\min}}{\beta^F}, \tag{4.16}$$

where $a^F_{\max} \equiv \max\{a_j : j \in G^F\}$. Note that $a^F_{\max} \geq \frac{\sum_j \beta_j a_j}{\sum_j \beta_j} = \frac{\gamma^F}{\beta^F}$. Accordingly, if

(4.16) holds then

$$\frac{\gamma^F}{\beta^F} < \frac{1}{\beta^I + \beta^F}\left[\gamma^I + \gamma^F - \frac{\beta^I}{\beta^F}\overline{D}\right] + \frac{D_{\min}}{\beta^F}$$
$$\Rightarrow D_{\min} > \gamma^F + \frac{1}{\beta^I + \beta^F}\left[\beta^I\overline{D} - \beta^F(\gamma^I + \gamma^F)\right] = \frac{1}{\beta^I + \beta^F}\left[\beta^I(\overline{D} + \gamma^F) - \gamma^I\beta^F\right] = q^I.$$

Therefore, if the minimum spot price exceeds all FG linear cost coefficients then the total IG production would never reach the minimum demand. The second condition is given by

$$a^I_{\max} < \overline{P} = \frac{\overline{D} + \gamma^I + \gamma^F}{\beta^I + \beta^F}$$
$$\Rightarrow \overline{D} > a^I_{\max}(\beta^F + \beta^I) - (\gamma^I + \gamma^F), \tag{4.17}$$

where $a^I_{\max} \equiv \max\{a_i : i \in G^I\}$. Similarly, we could write $a^I_{\max} \geq \frac{\sum_i \beta_i a_i}{\beta_i} = \frac{\gamma^I}{\beta^I}$ and combine this condition with (4.17) to get

$$\overline{D} > \frac{\beta^F}{\beta^I}\gamma^I - \gamma^F,$$

and from (4.8) we get

$$q^I > \frac{1}{\beta^I + \beta^F}\left[\beta^I\left(\left(\frac{\beta^F}{\beta^I}\gamma^I - \gamma^F\right) + \gamma^F\right) - \gamma^I\beta^F\right] = 0.$$

This implies that the first assumption is guaranteed whenever the second and fourth assumptions are satisfied. Ultimately, the five necessary assumptions about our model are valid for any set of parameters a_k, c_k, $\overline{D}$, and $D_{\min}$ for which the two inequalities (4.16) and (4.17) hold.

4.4.6 Analysis

In this subsection, we will analyze the SFE solution (from sections 4.4.2 and 4.4.3) and obtain some insights on the problem. We will then show the effect of disregarding the generation flexibility by comparing the SFE model studied in this chapter with the conventional SFE models in which all generators are assumed to be flexible.

Solution Interpretation

We have shown in §4.4.3 that IGs have no effect on the FG slope bids, but can influence the FG intercept bids. The more competitive IG bids are the higher their aggregate slope β^I becomes, which raises η_j and causes α_j to fall. This implies that increasing the IG competition causes FGs to lower their intercept bids, and hence makes FGs more competitive[1]. The IG intercept bids, on the other hand, are not affected by any bid, but their slope bids depend on the FG competition as shown in 4.4.2. When the FGs bid more competitively the sum of their slope bids β^F increases, causing the demand derivative for the IG competition to rise and intensifying the IG competition.

This outcome can be explained by two types of competition:

- Market share competition in which all generators participate to increase their average production rates, and
- Competition to balance the variable demand, in which only FGs participate[2].

IGs chose their slope bids based on the market share competition, while FGs base their slope bids on the variable demand competition and their intercept bids on the market share competition. The IG market share is given by (4.8) and the average FG

[1]FGs become more competitive because they would offer more production quantity at the same price when they have lower α_j bids.

[2]The absence of IG dependent terms in Equation (4.13) signifies their exclusion from competing over the variable demand.

market share is

$$q^F = \overline{D} - q^I = \frac{1}{\beta^I + \beta^F}\left[\beta^F\left(\overline{D} + \gamma^I\right) - \beta^I\gamma^F\right]. \tag{4.18}$$

An intensive market share competition is characterized by high IG slope bids and low FG intercept bids, while an intensive variable demand competition causes FGs to raise their slope bids.

The demand distribution plays no role in determining the supply function bids in the conventional KM model. Although the slope bids can be determined independently from demand for all generators, the FG intercept bids depend on the demand. According to (4.15), the α_j values decrease with the average demand, signifying aggressive participation in the market share competition. When the average demand increases the average market price also increases at a rate $\frac{1}{\beta^F}$. At the higher average market prices, FGs find it beneficial to lower their price intercepts in order to increase their average market share, even if the average market price slightly drops.

Comparison With the All FG Model

We will consider two problems in this subsection:

- Problem AF (*All Flexible*): All generators in G^F and G^I are flexible in this problem, and
- Problem FI (*Flexible/Inflexible*): Generators in G^F are flexible while generators in G^I are inflexible in this problem.

We will use the same variables for these two problems but with an AF and FI superscripts. For a set of generators G and a scalar $b \geq 0$ define the bid solution to a linear SFE as

$$B(G,b) = \{\beta_k, k \in G : \sum_{l \in G \setminus k} \beta_l + b = \frac{\beta_k}{1 - c_k \beta_k} \forall k \in G\}. \tag{4.19}$$

B gives the linear supply function solution to the set of generators G with demand derivative b. Using this notation, the slope bids for the AF problem is $B(G^F \cup G^I, 0)$ and the solution to the FI problem is $B(G^F, 0)$ for generators in G^F and $B(G^I, \beta^F)$ for generators in G^I. In order to compare the bid solutions from the two problems, we show in the following Lemma that the generator slope bids are increasing in b.

Lemma IV.2. *The bids β_k for $k \in G$ given by $B(G,b)$ are increasing in b.*

Proof of this lemma and other results in this chapter can be found in Appendix D.

If we think of b as the demand derivative, then this Lemma can be restated as follows: The generator competition in a linear SFE intensifies when the demand curve becomes steeper. Using Lemma IV.2, we can get the following result.

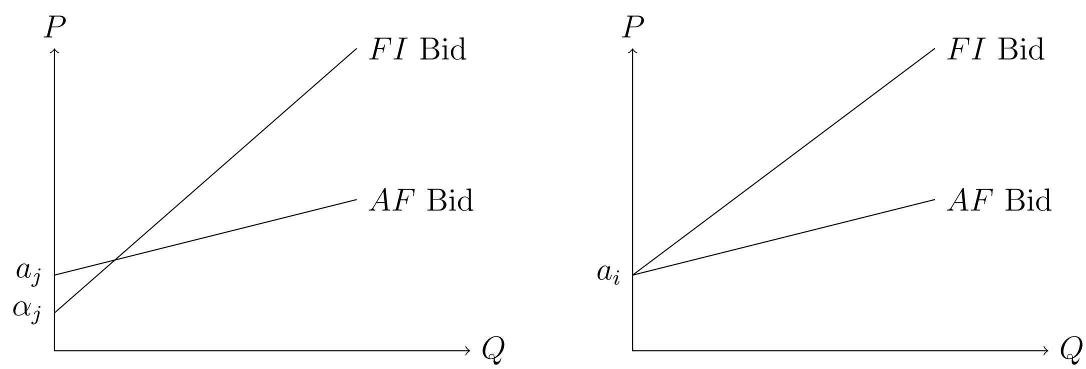
Theorem IV.3. $\beta_k^{AF} > \beta_k^{FI}\ \forall\ k \in G^F \cup G^I$.

Theorem IV.3 shows that generators submit more competitive price bids in the AF problem compared to the same generator bids under the FI problem. The intuition behind this result is that the exclusion of IGs from the demand balancing competition gives FGs the opportunity to markdown their β_j values in the FI problem. Consequently, IGs lower their slope bids in the market share competition in response to the aggressive FG bids. One thing to note is that the set of generators in G^I are only slightly less competitive in the FI problem (compared to the generators in G^F) because their β_i markdowns are not caused by the exclusion of competitors, but are rather in response to the lower bids of the generators in G^F.

The significance of Theorem IV.3 is that it confirms that failing to account for generator inflexibility overestimates the slope bids for all generators. If an all flexible generator model were used to study the competition in a system with flexible and inflexible generators then only FGs may come out as irrational because IGs would appear to submit rational best response bids to the seemingly irrational FGs. Not accounting for generator inflexibility can also have a profound impact on the electricity spot price as the following result shows.

Theorem IV.4. $\overline{P}^{FI} > \overline{P}^{AF}$ *and* $Var(P^{FI}(D)) > Var(P^{AF}(D))$.

In addition to overestimating the competition intensity, Theorem IV.4 shows that not accounting for generation inflexibility underestimates the average spot price and the price variance. Figure 4.1 illustrates the difference in bid between the AF and FI problems for a FG (Figure 4.1a) and an IG (Figure 4.1b) and Figure 4.2 shows the aggregate price curves for the two problems.



(a) FG bids (b) IG bids

Figure 4.1: Generator bidding under the the AF and FI problems.

4.5 Future Work

We developed in this chapter an ODE system for finding the SFE when there are flexible and inflexible generators. We then found the unique affine SFE when all generators have quadratic cost functions. There are several potential extensions to this work. First, studying the general non-linear equilibrium solution can give a theoretical contribution to the SFE literature. In the original supply function

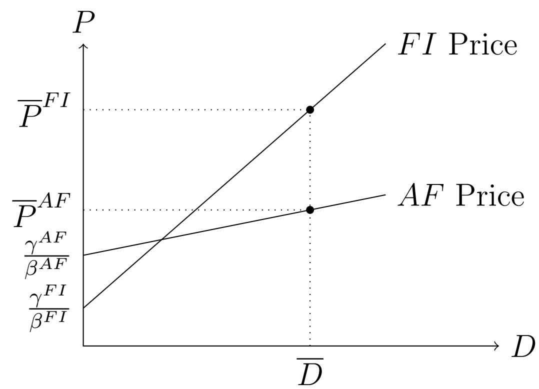


Figure 4.2: Aggregate supply curves for the AF and FI problems.

literature, ownership of multiple generators by a single firm does not complicate the problem since all generator's can be lumped into a single generator. However, in our model, if the firm's generation portfolio consists of flexible and inflexible generators, then the problem becomes significantly more complicated. Studying the impact of multiple firm ownership is another modification to the problem that we intend to consider.

Electricity contracts have been studied for the traditional SFE with all flexible generators. Flexible generators may have additional incentives to enter into long term contracts to protect themselves against low prices. Incorporating contract positions for generators could be an interesting third research direction. Lastly, integrating variable generation into the model and studying its impact is also a relevant research direction given the recent interest in renewable energy.

CHAPTER V

Price Bidding in Electricity Markets

5.1 Introduction

We study in this chapter two electricity auction models where generators submit price bids for their generation capacity to an ISO, who dispatches generators to satisfy demand in a uniform price auction. In the first model, we study the effect of transmission capacity constraints on the competition when there are only two generators. We find that the addition of these constraints does not change the bidding policy of generators, instead it changes the critical demand levels at which generators revise their bids. We then consider a problem with multiple symmetric generators and a random demand. We consider a symmetric mixed strategy Nash equilibrium and give the ODE condition for the equilibrium, then solve for closed form solutions for the duopoly case and the multiple generator case with uniform demand.

5.2 Background

The literature on price bidding in electricity auctions was first introduced by *von der Fehr and Harbor* (1993) to study the electricity market in the United Kingdom. In their model, firms that owned generators with constant marginal costs submitted price bids to the system operator that indicate the prices at which they would offer the entire capacities of their generators. Although their model considered the competition between multiple firms, where each firm can own several generation units with different production costs and capacities, their analysis only focused on Pure Strategy Nash Equilibrium (PSNE) solutions for two firms, each with a single generator.

Supatgiat et al. (2001) consider PSNE solutions for a problem with multiple firms, where each firm owns a single generator, and with discrete bid increments. They characterize properties of the equilibria and develop an algorithm that finds the most aggressive PSNE. *Brunekreeft* (2001) studied a multiple firm problem in which every firm owns multiple units. The paper uses a very specific generation unit order and demand pattern and does not fully characterize the equilibrium, but rather provides bidding bounds. *Garcıa-Dıaz and Marın* (2003) studied a more general multi-unit multiple firm auction for a fixed deterministic demand. *Crawford et al.* (2007) focused on asymmetric equilibria for multi-unit auction duopoly with complete information.

5.2.1 Model Description

A set of generators $i = 1 \cdots, n$ bid selling prices $\mathbf{b} = (b_i)_{i=1}^n$ (in \$ per power unit) for their electricity production to satisfy a demand $D \in [\underline{D}, \overline{D}]$. In this chapter, we assume that the demand is not affected by the electricity prices. The cost of producing q power units from generator i is $c_i q$ (\$ per time unit), where c_i is generator i's marginal production cost (in \$ per power unit). We will assume that generators are indexed based on non-decreasing marginal costs, so $c_i \leq c_j$ if $i < j$. Generators also have production capacity limits k_i (energy per unit time). We will denote the aggregate capacity for generators 1 through i by $K_i = \sum_{j=1}^{i} k_j$, and we will assume that $K_n \geq \overline{D}$ to ensure sufficient capacity for all demand realizations. We will also assume that the system has a maximum price bid $\overline{p}$ (\$ per power unit). This implies that a rational generator i bids in the range $b_i \in [c_i, \overline{p}]$. We will assume that all generators are rational and have perfect information about all the generation costs, production capacities, and the customer demand.

An ISO that minimizes the production cost to consumers can use any portion of any generator's capacity as long as generators are paid at least their bids for every energy unit. We consider a uniform price auction in which all generators are paid the same price for every unit sold. The uniform market clearing price is the lowest price at which the demand can be satisfied, defined as

$$p(\mathbf{b}, D) \stackrel{\text{def}}{=} \min \left\{ x \in \mathbb{R} : \sum_{i=1}^{n} k_i \mathbf{1}_{\{b_i \leq x\}} \geq D \right\}. \tag{5.1}$$

Recall that $\mathbf{1}$ is the indicator function where $\mathbf{1}_{\{A\}} = 1$ if event A is true and 0 if A is false. All generators with bids strictly below p are completely utilized while generators with bids above p are not used. Generators with bids exactly equal to p may be partially loaded. If multiple generators bid p then the generators with the least index i will be given priority in satisfying the marginal demand $D - \sum_{i=1}^{n} \mathbf{1}_{\{b_i < p\}} k_i$[1]. Therefore, generator i gets the following production allocation,

$$q_i(\mathbf{b}, D) = \mathbf{1}_{\{b_i < p(D,\mathbf{b})\}} k_i + \mathbf{1}_{\{b_i = p(D,\mathbf{b})\}} \left(D - \sum_{j=1}^{n} \mathbf{1}_{\{b_j < p(D,\mathbf{f})\}} k_j - \sum_{j=1}^{i-1} \mathbf{1}_{\{b_j = p(D,\mathbf{b})\}} k_j \right)^{+}, \tag{5.2}$$

and makes a profit of

$$\pi_i(\mathbf{b}, D) = q_i(\mathbf{b}, D)(p(\mathbf{b}, D) - c_i). \tag{5.3}$$

Finding an equilibrium for this problem is not trivial, even for the duopoly case. *von der Fehr and Harbor* (1993) showed that this problem may have multiple PSNE

[1] Others, including *von der Fehr and Harbor* (1993) and *Garcıa-Dıaz and Marın* (2003), have made a similar assumption. Ideally, more sophisticated tie breakers that gives equal priority to all generators should be used, but such allocation rules would complicate the model without adding insights.

when the demand is deterministic, and there may not exist a PSNE when the demand is stochastic.

5.2.2 Model Applications

The United Kingdom's switch from uniform price to discriminatory price auctions in the early 2000's triggered an abundance of studies comparing the two price mechanisms. California considered switching to discriminatory pricing shortly after the UK experience, which resulted in numerous studies both for and against the switch. For example, *Kahn et al.* (2001) argue agains the shift while *Rassenti et al.* (2003) conclude that discriminatory pricing can improve the market performance by substantially lowering the price volatility. The explicit price decision in the generator's objective makes the price auction a popular model for comparing the two market mechanisms. For example, *Federico and Rahman* (2003) use a price auction model to characterize the different trade-offs between these mechanisms. *Fabra et al.* (2006) study the uniform and discriminatory auction cases for a duopoly problem with a known demand, using both PSNE and MSNE. They also consider several extensions to the problem such as price-elastic demand and stochastic demand.

Unlike the SFE models, the price bidding models are commonly used to construct increasing step function supply offers, which is required by most system operators. To construct such offers, generator price curves are first linearized. Each linear segment of the cost is then treated as an independent generator with its own marginal cost and capacity, and a supply offer curve is constructed by optimizing the price bid over the demand range.

Price auction models with deterministic demand are also applicable in markets with short-lived auctions in which firms regularly update their bids. Their applicability under such settings is due to two factors: (1) When the duration between bids is short the demand variability becomes small, and hence the demand can be approximated by a constant value between bid durations, and (2) generators operate within a tight production range in the short duration between bids, and hence their marginal costs can be approximated by a constant value. Moreover, under the short-lived auction setting the price bidding model can be used to predict renewable plant bids since their available capacity can be accurately predicted over short durations.

5.3 Duopoly in a Transmission Constrained Network

In a duopoly model with $c_1 < c_2$, the cheap generator can undercut the expensive generator by bidding $b_1 = c_2$ when the demand is low, while both generators must produce to satisfy a demand that exceeds k_1. To maximize their profits, one of the generators would have an incentive to bid $\overline{p}$ in high demand scenarios. *von der Fehr and Harbor* (1993) showed that only one generator bids $\overline{p}$ in a PSNE. However, they showed that there may be multiple pure strategy equilibria with different price setters. *Fabra et al.* (2006) showed that the cheap generator may break the monopoly and bid $\overline{p}$ even if the demand is below its capacity. This occurs when the cheap generator's capacity is large enough compared to the expensive generator's capacity, and the price ceiling is high enough such that the cheap generator would be willing to loose market share in order to increase its selling price. They show that the critical demand at which the clearing price jumps from c_2 to $\overline{p}$ and generators change their bidding behavior is $D^* = k_1 \wedge \left(\frac{\overline{p}-c_1}{\overline{p}-c_2} k_2\right)$. Notice that we use the notation $\wedge$ and $\vee$ for the min and max operators, respectively, throughout this chapter.

In this section, we extend this problem to include the transmission network. The introduction of transmission network constraints change the ISO's problem. Instead of setting a clearing price according to (5.1) and allocating production according to (5.2), the ISO solves a transmission constrained optimal dispatch problem after getting the generator bids. Consequently, the two generators anticipate the auction's new structure and bid accordingly. We find that the solution structure of the auction with network constraints is similar to the original problem, but with a lower critical demand level D^* that depends on the network topology and line capacities. Appendix A shows some background and derivation of the DC optimal dispatch problem that is used in this section.

5.3.1 Production Allocation

We consider a general network topology with a set of nodes $\mathcal{N}$ and a set of transmission lines $\mathcal{L}$, where each transmission line $l \in \mathcal{L}$ has capacity $\bar{f}_l$. The transmission lines are assumed to be directed arcs in this model, with the the same capacity in either direction. We will assume that every transmission line is represented twice in $\mathcal{L}$ to model flows in both directions. That is, if line $(i, j) \in \mathcal{L}$ the we assume that line $(j, i) \in \mathcal{L}$ with $\bar{f}_{(i,j)} = \bar{f}_{(j,i)}$. Each node $i \in \mathcal{N}$ has a fixed demand $d_i \geq 0$. Generators 1 and 2 are located in two arbitrary nodes in the network, which we denote by nodes 1 and 2. We will assume without loss of generality that these two nodes are different[2]. We will denote the total demand by $D \stackrel{\text{def}}{=} \sum_{i \in \mathcal{N}} d_i$ and the demand vector by $\mathbf{d} = (d_i)_{i \in \mathcal{N}}$. Also, let $\mathbb{D} \subset \mathbb{R}_+^{|\mathcal{N}|}$ be the set of feasible demand vectors (i.e. the combinations of demands in the network that can be satisfied by the two generators), which we will determine later in this section. We will denote by $\bar{d}_i$ the maximum demand in node i.

Upon receiving the generator bids b_1 and b_2, the ISO uses these bids to construct the generator revealed cost functions $C_i(q) = b_i q$ for $i \in \mathcal{N}$. Because the demand is fixed in this problem, the decision values $q_i = -d_i$ are known and the revealed costs $C_i(q_i) = 0$ for all nodes $i \in \mathcal{N} \setminus \{1, 2\}$ in the problem (A.5)-(A.8). Nodes 1 and 2 have nodal injections $q_1 - d_1$ and $q_2 - d_2$ and revealed costs $b_1 q_1$ and $b_2 q_2$, respecitvely, where q_1 and q_2 are the only unknown variables in the ISO's problem. According to (A.5), the ISO's objective in this problem is to minimize $b_1 q_1 + b_2 q_2$.

[2]If the two generators happen to connect to the same node we can *split* this node into two nodes with an infinite capacity line in between.

By selecting node 2 as the slack node we can set $\beta_l^2 = 0$ for all lines and (A.7) can be expressed without q_2[3]. We can also write $\sum_{i\in\mathcal{N}} \beta_l^i q_i$ as $\beta_l^1 q_1 - \beta_l^\top \mathbf{d}$. By substituting $q_2 = D - q_1$ the objective becomes $\min_{q_1}(b_1 - b_2)q_1 + b_2 D$, where $D = \sum_{i\in\mathcal{N}} d_i$ as before. Notice that $b_2 D$ can be dropped from the objective since it is constant in the ISO's problem. The constraint that $q_2 \in [0, k_2]$ can be written as $q_1 \in [D - k_2, D]$, and when combined with generator 1's constraint $q_1 \in [0, k_1]$ we get the condition $(D - k_2)^+ \le q_1 \le D \wedge k_1$. We can express the problem as

$$\begin{aligned} \max_{q_1} \quad & (b_2 - b_1)q_1 \\ \text{s.t.} \quad & \beta_l^1 q_1 \le \overline{f}_l + \beta_l^\top \mathbf{d} \;\; \forall \;\; l \in \mathcal{L} \\ & (D - k_2)^+ \le q_1 \le D \wedge k_1. \end{aligned}$$

To ensure the existence of a feasible solution, we assume the following:

(1) *No generation capacity limitation:* $\sum_{i\in\mathcal{N}} \bar{d}_i \in (0, k_1 + k_2]$.

(2) *No transmission line limitation:* $\bar{f}_l \ge 0$ is large enough such that $\exists \; q_1 \in [(D - k_2)^+, D \wedge k_1]$ that satisfies $\bar{f}_l \ge \beta_l^1 q_1 - \beta_l^\top \mathbf{d} \; \forall \; \mathbf{d} \in \mathbb{D}$ and $l \in \mathcal{L}$.

Condition (2) ensures the existence of a pair of generator injections $q_1 \ge 0$ and $q_2 \ge 0$ that can satisfy demand without violating any line capacity. If $\beta_l^1 = 0$ for some line l, then condition (2) cannot be binding and q_1 is not restricted by line l's capacity. Let

$$\mathcal{L}_1 = \{l \in \mathcal{L} : \beta_l^1 > 0\}, \qquad \mathcal{L}_2 = \{l \in \mathcal{L} : \beta_l^1 < 0\}.$$

[3]Refer to Appendix A for the definition of the slack node and the power transfer distribution factors (PTDF) β_l^k, which can be interpreted as the portion of the flow from node k's power injection passing through line l on its way to the slack node. We also use β_l as the column vector of PTDFs for line l and $\beta_l^\top$ for its transpose (row) vector.

To satisfy the line capacity constraints, q_1 must be chosen such that

$$\frac{\bar{f}_{l_2} + \beta_{l_2}^\top \mathbf{d}}{\beta_{l_2}^1} \leq q_1 \leq \frac{\bar{f}_{l_1} + \beta_{l_1}^\top \mathbf{d}}{\beta_{l_1}^1} \quad \forall \ l_1 \in \mathcal{L}_1 \text{ and } l_2 \in \mathcal{L}_2.$$

Note that the left most term in the inequality decreases to $-\infty$ as $\bar{f}_{l_2} \to \infty$ and the right most term increases to ∞ as $\bar{f}_{l_1} \to \infty$. In other words, condition (2) is always feasible for sufficiently large $\bar{f}_l$ values. Define the *critical ratio* for a line as

$$r_l(\mathbf{d}) = \frac{\bar{f}_l + \beta_l^\top \mathbf{d}}{\beta_l^1} \qquad \forall \ l \in \mathcal{L}_1 \cup \mathcal{L}_2.$$

$r_l(\mathbf{d})$ is the maximum (minimum) injection q_1 for which the flow in line $l \in \mathcal{L}_1$ ($\mathcal{L}_2$) is feasible when the demand vector is $\mathbf{d}$. We will also define $l^*(\mathbf{d}) \in \text{argmin}_{l \in \mathcal{L}_1} \{r_l(\mathbf{d})\}$, which is the set of first lines in the network that get congested when increasing a feasible q_1, and $l_*(\mathbf{d}) \in \text{argmax}_{l \in \mathcal{L}_2} \{r_l(\mathbf{d})\}$ as the set of first lines that become congested when decreasing q_1 from a feasible level. Denote these lines' critical ratios by

$$r^*(\mathbf{d}) = \min_{l \in \mathcal{L}_1} \{r_l(\mathbf{d})\}, \qquad r_*(\mathbf{d}) = \max_{l \in \mathcal{L}_2} \{r_l(\mathbf{d})\}.$$

$r^*(\mathbf{d})$ and $r_*(\mathbf{d})$ are the maximum and minimum levels for which q_1 is feasible, and thus $q_1 \in [r_*(\mathbf{d}), r^*(\mathbf{d})]$. By convention, $r^*(\mathbf{d}) = \infty$ if $\mathcal{L}_1 = \phi$ and $r_*(\mathbf{d}) = -\infty$ if $\mathcal{L}_2 = \phi$. If we now combine all constraints we get

$$\max\{D - k_2, r_*(\mathbf{d}), 0\} \leq q_1 \leq \min\{D, k_1, r^*(\mathbf{d})\}. \tag{5.4}$$

By the same argument as before (when $\bar{f}_l \to \infty$), we can find $\bar{f}_l$ values for which $\max\{(D - k_2)^+, r_*(\mathbf{d})\} \leq \min\{D, k_1, r^*(\mathbf{d})\}$, which is a necessary condition for the existence of a feasible solution[4]. Using (5.4), the quantity allocation problem becomes

[4] $\max\{D - k_2, 0\} \leq \min\{D, k_1\}$ follows from the range of $D \in (0, k_1 + k_2]$.

trivial; set q_1 to $\min\{D, k_1, r^*(\mathbf{d})\}$ if $b_1 \le b_2$ and to $\max\{(D-k_2)^+, r_*(\mathbf{d})\}$ if $b_1 > b_2$. Therefore, we can write a closed form solution for the quantity allocations.

$$q_1(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1 \le b_2\}} \min\{D, k_1, r^*(\mathbf{d})\} + \mathbf{1}_{\{b_1 > b_2\}} \max\{(D-k_2)^+, r_*(\mathbf{d})\}, \tag{5.5}$$

$$q_2(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1 \le b_2\}} \max\{(D-k_1)^+, D - r^*(\mathbf{d})\} + \mathbf{1}_{\{b_1 > b_2\}} \min\{D, k_2, D - r_*(\mathbf{d})\}. \tag{5.6}$$

Notice that Equation (5.6) is attained by subtracting $q_1(\mathbf{d},\mathbf{b})$ from D.

We can solve a similar problem for generator 2, where generator 1 is used as the slack node and q_1 is replaced by $D - q_2$. To distinguish between the two problems, we will use $\tilde{\beta}$, $\tilde{l}^*$, $\tilde{l}_*$, $\tilde{r}^*$, and $\tilde{r}_*$ in place of β, l^*, l_*, r^*, and r_*. By solving the same problem when q_2 is the decision variable we get

$$q_2(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1 > b_2\}} \min\{D, k_2, \tilde{r}^*(\mathbf{d})\} + \mathbf{1}_{\{b_1 \le b_2\}} \max\{(D-k_1)^+, \tilde{r}_*(\mathbf{d})\}, \tag{5.7}$$

$$q_1(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1 > b_2\}} \max\{(D-k_2)^+, D - \tilde{r}^*(\mathbf{d})\} + \mathbf{1}_{\{b_1 \le b_2\}} \min\{D, k_1, D - \tilde{r}_*(\mathbf{d})\}. \tag{5.8}$$

Define the *effective capacities* for generators 1 and 2 as $\tilde{k}_1(\mathbf{d}) = k_1 \wedge r^*(\mathbf{d})$ and $\tilde{k}_2(\mathbf{d}) = k_2 \wedge \tilde{r}^*(\mathbf{d})$. We interpret the effective capacity as the limit of a generator's production, which is determined by the generator's capacity or the transmission line bottleneck, whichever is lower. We can therefore write the first term in (5.5) as $\mathbf{1}_{\{b_1 \le b_2\}}(D \wedge \tilde{k}_1(\mathbf{d}))$ and the first term in (5.7) as $\mathbf{1}_{\{b_1 > b_2\}}(D \wedge \tilde{k}_2(\mathbf{d}))$. The first term in (5.6) can be written as

$$\begin{aligned}\mathbf{1}_{\{b_1 \le b_2\}} \max\{D - k_1, D - r^*(\mathbf{d}), 0\} &= \mathbf{1}_{\{b_1 \le b_2\}} \left(\max\{D - k_1, D - r^*(\mathbf{d})\}\right)^+ \\ &= \mathbf{1}_{\{b_1 \le b_2\}} \left(D - \tilde{k}_1(\mathbf{d})\right)^+.\end{aligned}$$

Likewise, the first term in (5.8) can be simplified to

$$\mathbf{1}_{\{b_1>b_2\}}\max\{D-k_2, D-\tilde{r}^*(\mathbf{d}), 0\} = \mathbf{1}_{\{b_1>b_2\}}\left(D-\tilde{d}_2(\mathbf{d})\right)^+.$$

Using these results, we can write q_1 and q_2 as

$$q_1(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1\le b_2\}}\left(D\wedge\tilde{k}_1(\mathbf{d})\right)+\mathbf{1}_{\{b_1>b_2\}}\left(D-\tilde{k}_2(\mathbf{d})\right)^+, \tag{5.9}$$

$$q_2(\mathbf{b},\mathbf{d}) = \mathbf{1}_{\{b_1>b_2\}}\left(D\wedge\tilde{k}_2(\mathbf{d})\right)+\mathbf{1}_{\{b_1\le b_2\}}\left(D-\tilde{k}_1(\mathbf{d})\right)^+. \tag{5.10}$$

Equations (5.9) and (5.10) have an intuitive interpretation: a generator with the lower bid satisfies the entire demand if possible, otherwise it produces at the level where it reaches its effective capacity, and if it has a higher bid then it generates the portion of the demand the other generator could not satisfy.

5.3.2 Clearing Prices

The basic rule in a uniform price auction is that all generators are paid the same clearing price for their production. When network constraints are introduced, generators can exercise locational convenience to raise their prices. This causes price discrimination between nodes whenever a transmission line becomes congested. In such a scenario, no single uniform price can be used in all locations to clear the market, instead each node must have its own uniform price that is used for all agents connected to the node, which can be calculated from (A.9). The node-dependent price is known as the locational marginal price (LMP).

When a generator is marginal (i.e. supplies the additional unit of demand in the node) and produces at partial capacity then, according to (A.9)[5], the generator is paid its own bid. From (5.9), this scenario occurs for generator 1 when it has a larger

[5]The capacity constraint is not binding for a generator operating at partial capacity, and hence the dual variable $\gamma_i = 0$. This gives a price $p_i = C_i'(q_i) = b_i$.

bid and the total demand exceeds the second generator's effective capacity ($b_1 > b_2$ and $D > \tilde{k}_2(\mathbf{d})$), or if generator 1's bid does not exceed the other generator's bid but the total demand is less than its effective capacity ($b_1 \leq b_2$ and $D < \tilde{k}_1(\mathbf{d})$). From (5.10), the same scenario occurs for generator 2 if it submits at least as large a bid as b_1 and there is sufficient demand ($b1 \leq b_2$ and $D > \tilde{k}_1(\mathbf{d})$), or if generator 2 submits a smaller bid and the total demand is smaller than its effective capacity ($b_1 > b_2$ and $D < \tilde{k}_2(\mathbf{d})$).

A generator's price may depend on the other generator's bid when the generator is non-marginal, which can either happen because (1) the generator's bid is high and the demand is low, or (2) the generator's bid is low and the demand is high. We need not consider the first case because a generator does not produce when its bid is high and the other generator can satisfy the entire demand on its own. In the second case, the low bidding generator i produces $\tilde{k}_i(\mathbf{d})$. If the generator is operating at partial capacity ($\tilde{k}_i(\mathbf{d}) < k_i$) then using the first part of (A.9) its LMP must equal its own bid. If, on the other hand, generator i is operating at full capacity ($\tilde{k}_i(\mathbf{d}) = k_i$) then $\gamma_i > 0$ and using the second part of (A.9) the generator is paid $\lambda - \sum_l \beta_l^i \mu_l$ where λ is the other generator's bid and μ_l are the dual variables for the line capacity constraints. However, the fact that generator i reaches its full capacity implies that the network is not congested, and hence $\mu_l = 0$ for all lines. Therefore, generator i is paid the other generator's bid in this case.

By combining the scenarios from the previous discussion, we can write the nodal prices for generators 1 and 2 as

$$\begin{aligned} p_1(\mathbf{b}, \mathbf{d}) = &\left[\mathbf{1}_{\{b_1 > b_2\}} \mathbf{1}_{\{D > \tilde{k}_2(\mathbf{d})\}} + \mathbf{1}_{\{b_1 \leq b_2\}} \mathbf{1}_{\{D \leq \tilde{k}_1(\mathbf{d})\}}\right] b_1 \\ &+ \mathbf{1}_{\{b_1 \leq b_2\}} \mathbf{1}_{\{D > \tilde{k}_1(\mathbf{d})\}} (\mathbf{1}_{\{\tilde{k}_1(\mathbf{d}) < k_1\}} b_1 + \mathbf{1}_{\{\tilde{k}_1(\mathbf{d}) = k_1\}} b_2), \end{aligned} \tag{5.11}$$

$$\begin{aligned} p_2(\mathbf{b}, \mathbf{d}) = &\left[\mathbf{1}_{\{b_1 \leq b_2\}} \mathbf{1}_{\{D \geq \tilde{k}_1(\mathbf{d})\}} + \mathbf{1}_{\{b_1 > b_2\}} \mathbf{1}_{\{D \leq \tilde{k}_2(\mathbf{d})\}}\right] b_2 \\ &+ \mathbf{1}_{\{b_1 > b_2\}} \mathbf{1}_{\{D > \tilde{k}_2(\mathbf{d})\}} (\mathbf{1}_{\{\tilde{k}_2(\mathbf{d}) < k_2\}} b_2 + \mathbf{1}_{\{\tilde{k}_2(\mathbf{d}) = k_2\}} b_1). \end{aligned} \tag{5.12}$$

Notice that the case when $b_1 > b_2$ and $D \le \tilde{k}_2(\mathbf{d})$ was not considered for p_1 because generator 1 is not allocated any production in this case. Similarly, the scenario $b_1 \le b_2$ and $D < \tilde{k}_1(\mathbf{d})$ does not appear in p_2 because generator 2 is not active in this case. Now that we have characterized the ISO generation allocation decision and clearing prices as functions of generator bids and demand realizations, we will consider the generator bidding problem next.

5.3.3 The Generator Problem

A generator's profit function in this problem has the form shown in Equation (5.3) but with a location dependent nodal price instead of the uniform price. By substituting the quantities and prices rom Equations (5.9), (5.10), (5.11), and (5.12), we can write the two generator profit functions as

$$\begin{aligned}\pi_1(\mathbf{b},\mathbf{d}) =& \mathbf{1}_{\{b_1 \le b_2\}}\Big[\mathbf{1}_{\{D \le \tilde{k}_1(\mathbf{d})\}}(b_1 - c_1)D \\ &+ \mathbf{1}_{\{D > \tilde{k}_1(\mathbf{d})\}}(\mathbf{1}_{\{\tilde{k}_1(\mathbf{d}) = k_1\}}b_2 + \mathbf{1}_{\{\tilde{k}_1(\mathbf{d}) < k_1\}}b_1 - c_1)\tilde{k}_1(\mathbf{d})\Big] \\ &+ \mathbf{1}_{\{b_1 > b_2\}}(b_1 - c_1)(D - \tilde{k}_2(\mathbf{d}))^+,\end{aligned} \tag{5.13}$$

$$\begin{aligned}\pi_2(\mathbf{b},\mathbf{d}) =& \mathbf{1}_{\{b_1 > b_2\}}\Big[\mathbf{1}_{\{D \le \tilde{k}_2(\mathbf{d})\}}(b_2 - c_2)D \\ &+ \mathbf{1}_{\{D > \tilde{k}_2(\mathbf{d})\}}(\mathbf{1}_{\{\tilde{k}_2(\mathbf{d}) = k_2\}}b_1 + \mathbf{1}_{\{\tilde{k}_2(\mathbf{d}) < k_2\}}b_2 - c_2)\tilde{k}_2(\mathbf{d})\Big] \\ &+ \mathbf{1}_{\{b_1 \le b_2\}}(b_2 - c_2)(D - \tilde{k}_1(\mathbf{d}))^+.\end{aligned} \tag{5.14}$$

In a best response problem, a generator determines the bid that maximizes its own profit in response to the other generator's (known) bid. Our goal in this subsection is to characterize the best response bids to find bidding equilibria.

In generator 1's best response problem, we assume that it knows the other generator's bid b_2, the demand in every node $\mathbf{d}$, and finds b_1 that maximizes its own profit. Generator 1's profit function depends on whether it bids below or above b_2. If generator 1 bids $b_1 \le b_2$ then according to (5.13), it is a weakly dominant strategy to bid b_2, and if it bids above b_2 then bidding $\overline{p}$ is a dominant strategy. Generator 2 has a similar strategy, but with bidding $b_1 - \varepsilon$[6] as a weakly dominant strategy when $b_2 < b_1$ and $\overline{p}$ as a dominant strategy when $b_2 \geq b_1$. Because of their rationality, generator i bids $b_i \in [c_i, \overline{p}]$, $i = 1, 2$, otherwise a generator could make a negative profit when $b_i < c_i$. Since generator 2 never bids below c_2 then it is weakly dominant for generator 1 to never bid below c_2. We will consider the weekly dominant strategy in which both generators bid in $[c_2, \overline{p}]$.

If it is more profitable for generator 1 to bid higher than generator 2 then, from (5.13), it is always optimal to bid $\overline{p}$. Therefore, generator 1 only bids below $\overline{p}$ to undercut the second generator. Consequently, a bid $b_1 \in (c_2, \overline{p})$ can only be optimal if $b_1 \leq b_2$. This solution can be an equilibrium if it is profitable for the second generator to bid higher, in which case it bids $\overline{p}$ according to (5.14). However, if generator 2 finds it profitable to bid low then it can undercut b_1. This results in a Bertrand competition in which generators undercut one another, which has an equilibrium with both generators bidding c_2. As a result, there can only be two possible equilibrium nodal prices in pure strategy; c_2 and $\overline{p}$. In fact, this result has the same structure of the original auction problem with no transmission constraints (see for example *Fabra et al.* (2006)). Nevertheless, it turns out that transmission constraints change the generator preferences on whether to bid high or low as we will see next.

5.3.4 Comparison with the Original Problem

The transmission line capacities affect the solution only when they constrain a generator's production. For large enough line capacities the critical ratios $r^*(\mathbf{d}) > k_1$ and $\tilde{r}^*(\mathbf{d}) > k_2$, which makes $\tilde{k}_i(\mathbf{d}) = k_i$ for $i = 1, 2$. When this happens, the

[6] ε is a small positive number used to break the tie with generator 1.

production allocation in Equations (5.9) and (5.10) become

$$
\begin{aligned}
q_1(\mathbf{b}, D) =& \mathbf{1}_{\{b_1 \leq b_2\}} (D \wedge k_1) + \mathbf{1}_{\{b_1 > b_2\}} (D - k_2)^+ ,\\
q_2(\mathbf{b}, D) =& \mathbf{1}_{\{b_1 > b_2\}} (D \wedge k_2) + \mathbf{1}_{\{b_1 \leq b_2\}} (D - k_1))^+ ,
\end{aligned}
$$

the prices in (5.11) and (5.12) become

$$
\begin{aligned}
p_1(\mathbf{b}, D) &= \left[\mathbf{1}_{\{b_1 > b_2\}}\mathbf{1}_{\{D > k_2\}} + \mathbf{1}_{\{b_1 \leq b_2\}}\mathbf{1}_{\{D \leq k_1\}}\right] b_1 + \mathbf{1}_{\{b_1 \leq b_2\}}\mathbf{1}_{\{D > k_1\}} b_2,\\
p_2(\mathbf{b}, D) &= \left[\mathbf{1}_{\{b_1 \leq b_2\}}\mathbf{1}_{\{D \geq k_1\}} + \mathbf{1}_{\{b_1 > b_2\}}\mathbf{1}_{\{D \leq k_2\}}\right] b_2 + \mathbf{1}_{\{b_1 > b_2\}}\mathbf{1}_{\{D > k_2\}} b_1,
\end{aligned}
$$

and the profits in (5.13) and (5.14) become

$$
\begin{aligned}
\pi_1(\mathbf{b}, D) =& \mathbf{1}_{\{b_1 \leq b_2\}} \Big[\mathbf{1}_{\{D \leq k_1\}}(b_1 - c_1)D + \mathbf{1}_{\{D > k_1\}}(b_2 - c_1)k_1\Big]\\
&+ \mathbf{1}_{\{b_1 > b_2\}}(b_1 - c_1)(D - k_2)^+,\\
\pi_2(\mathbf{b}, D) =& \mathbf{1}_{\{b_1 > b_2\}} \Big[\mathbf{1}_{\{D \leq k_2\}}(b_2 - c_2)D + \mathbf{1}_{\{D > k_2\}}(b_1 - c_2)k_2\Big]\\
&+ \mathbf{1}_{\{b_1 \leq b_2\}}(b_2 - c_2)(D - k_1)^+.
\end{aligned}
$$

In this problem, the first generator bids c_2 when the demand is low, but when it rises above k_1 then the equilibrium price is $\overline{p}$ (either generator may be marginal). However, when $k_1 > k_2$ there may be a demand below k_1 but above k_2 at which generator 1 prefers to change its policy and raise its bid to $\overline{p}$ at some production loss to the second generator, which would be compensated by the high clearing price. This critical demand value is the point at which generator 1 is indifferent between bidding high and bidding low, which satisfies the equation

$$
(c_2 - c_1)D = (\overline{p} - c_1)(D - k_2).
$$

This gives the indifference demand level $D = \frac{\overline{p} - c_1}{\overline{p} - c_2} k_2$. Therefore, the price rises from

c_2 to $\overline{p}$ when the demand either exceeds this level or exceeds k_1, whichever occurs first. Hence, as D increases, the price starts at c_2 for low demands, jumps to $\overline{p}$ at $D^* = k_1 \wedge \left(\frac{\overline{p}-c_1}{\overline{p}-c_2} k_2\right)$, and remains at $\overline{p}$ for high demand values.

To compare the transmission constrained problem with the original problem, we will assume that the demand increases uniformly in every node as D increases. That is, if $D^0 > 0$ with a demand vector $\mathbf{d}^0$, then $\mathbf{d}^1 = \frac{D^1}{D^0}\mathbf{d}^0$ for the demand level D^1 with the demand vector $\mathbf{d}^1$. For a sufficiently small demand, the transmission network is not capacitated and the cheaper generator can satisfy the entire demand by bidding c_2, which gives a price of c_2 in all nodes. As the demand is increased the price in node 1 would jump to $\overline{p}$ at some point. If $\tilde{k}_1(\mathbf{d}) = k_1$ for some $\mathbf{d}$ then the transmission constraint does not limit generator 1's production and the same analysis for the unconstrained case holds. However, if $\tilde{k}_1(\mathbf{d}) < k_1$ for demand vector $\mathbf{d}$, then generator 1 is constrained not by its own capacity, but by the network. Under this scenario, node 1's price is high if the demand exceeds either k_1 or $\frac{\overline{p}-c_1}{\overline{p}-c_2}\tilde{k}_2(\mathbf{d})$ as before, but we also need to considers the case when $\tilde{k}_1(\mathbf{d}) < D < k_1$. In such a scenario, according to (5.13), generator 1 always gets its price when it bids low, but also gets to sell its maximum potential $\tilde{k}_1(\mathbf{d})$. The indifference demand level between bidding c_2 and $\overline{p}$ when $\tilde{k}_1(\mathbf{d}) < D < k_1$ satisfies

$$\begin{aligned}(c_2 - c_1)\tilde{k}_1(\mathbf{d}) &= (\overline{p} - c_1)(D - \tilde{k}_2(\mathbf{d})) \\ \Rightarrow D &= \frac{c_2 - c_1}{\overline{p} - c_1}\tilde{k}_1(\mathbf{d}) + \tilde{k}_2(\mathbf{d}).\end{aligned}$$

Consequently, node 1's price will be high if D exceeds both this level and $\tilde{k}_1(\mathbf{d})$. Therefore, the critical demand level at which prices jump from c_2 to $\overline{p}$ in node 1 is

$$D^* = \min\left\{k_1, \frac{\overline{p} - c_1}{\overline{p} - c_2}\tilde{k}_2(\mathbf{d}^*), \max\left\{\tilde{k}_1(\mathbf{d}^*), \frac{c_2 - c_1}{\overline{p} - c_1}\tilde{k}_1(\mathbf{d}^*) + \tilde{k}_2(\mathbf{d}^*)\right\}\right\}. \tag{5.15}$$

Indeed, this formula is a generalization of the duopoly problem when network constraints are considered. To derive the non-constrained network solution with a critical demand $D^* = k_1 \wedge \left(\frac{\overline{p}-c_1}{\overline{p}-c_2} k_2\right)$ from (5.15) we can take all line capacities $\overline{f}_l$, $l \in \mathcal{L}$, to be sufficiently large, which makes $\tilde{k}_i(\mathbf{d}) = k_i$, $i = 1, 2$, for every feasible $\mathbf{d}$. The max term in (5.15) becomes $\max\{k_1, \frac{c_2-c_1}{\overline{p}-c_1} k_1 + k_2\}$. If $k_1 \geq \frac{c_2-c_1}{\overline{p}-c_1} k_1 + k_2$ then this max term becomes k_1, which makes $D^* = k_1 \wedge \left(\frac{\overline{p}-c_1}{\overline{p}-c_2} k_2\right)$. On the other hand, if $k_1 < \frac{c_2-c_1}{\overline{p}-c_1} k_1 + k_2$ then this max term becomes $\frac{c_2-c_1}{\overline{p}-c_1} k_1 + k_2$, which can be disregarded because it is obviously larger than the first term in the min operator in (5.15). Therefore, in either case the price jumps from c_2 to $\overline{p}$ at the critical demand level $D^* = k_1 \wedge \left(\frac{\overline{p}-c_1}{\overline{p}-c_2} k_2\right)$ when the network is not congested, which is identical to the solution when there are no network constraints. By taking $\tilde{k}_i(\mathbf{d}) = k_i$, $i = 1, 2$, it can also be verified that the system has the same uniform price and all generators would behave as in the non-network constrained problem.

We have shown in this subsection that including network congestion in our two stage market mechanism starting with an auction then moving to a network economic dispatch problem requires some work in calculating power transfer distribution factors and effective capacities under different demand realizations, but the equilibrium structure remains unchanged. The main insight in this section is that accounting for transmission capacity constraints in the Bertrand duopoly model does not change the solution structure nor does it significantly complicate the calculations. This result holds because setting a slack generator reduces the number of decision variables in the ISO problem (A.5)-(A.8) to one, and the allocation to this generator can be expressed in terms of the relative values of the two generator bids. Unfortunately, we could not use a a similar approach when the network has more than two generators because (A.5)-(A.8) becomes a multivariable linear program.

5.4 Symmetric Mixed Strategy Nash Equilibrium for Multiple Generators with Stochastic Demand

The price bidding equilibrium model becomes more difficult when the demand is modeled as a stochastic process. PSNE may not even exists in this stochastic demand case[7]. *Fabra et al.* (2006) derive the MSNE for the two generator game. When there are multiple generators all the different bid orders have to be accounted for when solving for a mixed strategy equilibrium. This problem becomes increasingly more complicated as the number of generator's increase because accounting for all scenarios becomes a combinatorial problem. Even if brute force enumeration is used to attain all possible scenarios, getting the mixed strategies requires solving an ODE system, one ODE for every generator, with all these combinatorial terms. Solving this system could be numerically achievable for small problems, but may not provide useful insights.

To study the competitive behaviour of multiple firms we assume in this section that all firms have identical costs and capacities, and we focus on the class of symmetric equilibria. These assumptions make the model much more tractable and can lead to closed form solutions for many special cases. We will go over the problem's model in §5.4.1 and derive an ODE that can be used to attain the mixed strategy equilibria. We will then present a duopoly example in §5.4.2 and an oligopoly example with a uniform demand in §5.4.3. Our main goal from this section is to find the generator equilibrium strategies in order to understand how the number of generators, their costs, the market price ceiling, and the demand distribution affect the competition and the market prices.

[7] See for example Proposition 4 in *von der Fehr and Harbor* (1993).

5.4.1 The Model

Our problem has n generators each with a capacity k and a fixed marginal cost c. Let D be the random variable representing the stochastic consumer demand. Let $\mathbf{b} = (b_1, \cdots, b_n)$ be the generator bid vector. Let $b_{(i)}$ be the ith smallest bid, so $b_{(1)} \leq b_{(2)} \leq \cdots \leq b_{(n)}$. We will partition the demand range into n left continuous equilength intervals, where the jth interval is $I_j = ((j-1)k, jk]$. Generator i's payoff from bidding b_i where the remaining bids are b_{-i} is given by

$$\pi_i(b_i, b_{-i}) = \sum_{j=1}^{n} \mathbf{1}_{\{D \in I_j\}} \left[\mathbf{1}_{\{b_i = b_{(j)}\}} (D - (j-1)k)(b_i - c) + \mathbf{1}_{\{b_{(j)} > b_i\}} k(b_{(j)} - c) \right].$$

Now consider the case when generator i chooses its bid when the other generators randomize their bids. Let B be the random variable of the other generator bids with distribution F_B, density f_B, $\overline{F}_B(x) = 1 - F_B(x)$, and support $[\underline{b}, \overline{p}]$. We will denote the ith order statistic (ith smallest) random variable among the $n-1$ generators by $B_{(i)}$. We will denote the probability of the demand being in interval I_j by $PD_j \stackrel{\text{def}}{=} \mathbb{P}\{D \in I_j\}$ and the expected excess demand in interval I_j by $ED_j \stackrel{\text{def}}{=} \mathsf{E}[D | D \in I_j] - (j-1)k$. If generator i bids b_i and the remaining $n-1$ generators bid according to the mixed strategy then generator i's expected payoff is

$$\begin{aligned} \mathsf{E}[\pi_i(b_i)] = \sum_{j=1}^{n} PD_j \Big[& \mathbb{P}\{B_{(j-1)} \leq b_i < B_{(j)}\} ED_j (b_i - c) \\ & + \mathbb{P}\{B_{(j-1)} > b_i\} k (\mathsf{E}[B_{(j-1)} | B_{(j-1)} > b_i] - c) \Big]. \end{aligned} \tag{5.16}$$

Notice that this expectation is taken with respect to the demand D and the $n-1$ generator bids that are drawn from F_B, while b_i in this formula is an input variable. We use the convention $B_{(0)} < c$ and $B_{(n)} > \overline{p}$ in (5.16). The following theorem gives the ODE that characterizes the symmetric MSNE for this game.

Theorem V.1. *The symmetric mixed strategy bid distribution is the solution to the*

ODE

$$\sum_{j=1}^{n} PED_j \frac{(n-1)!}{(j-1)!(n-j)!} F_B(x)^{j-1} \overline{F}_B(x)^{n-j}$$
$$+ (x-c) f_B(x) \sum_{j=1}^{n-1} (\Delta PED_j - kPD_{j+1}) \frac{(n-1)!}{(j-1)!(n-j-1)!} F_B(x)^{j-1} \overline{F}_B(x)^{n-j-1} = 0. \tag{5.17}$$

The proof of this theorem can be found in Appendix E.

We will go over two examples next where the solution to (5.17) can be expressed in closed form.

5.4.2 Duopoly Model Example

In a duopoly model with two symmetric generators, the demand has support $(0, 2k]$. If D never exceeds k then a trivial pure strategy equilibrium exists where at least one generator bids c and satisfies the entire demand. We will therefore assume that D has a a continuous distribution F_D with $\overline{F}_D(k) = 1 - F_D(k) > 0$[8]. We can write the problem parameters as $PD_1 = F_D(k)$, $PD_2 = \overline{F}_D(k)$, $PED_1 = F_D(k)\mathsf{E}[D|D \le k]$, $PED_2 = \overline{F}_D(k)(\mathsf{E}[D|D > k] - k)$, and $\Delta PED_1 = PED_2 - PED_1 = \mathsf{E}\left[D\left(\mathbf{1}_{\{D>k\}} - \mathbf{1}_{\{D\le k\}}\right)\right] - k\overline{F}_D(k)$. We can substitute $n = 2$ in (5.17) to get the following differential equation for the symmetric duopoly model.

$$PED_1 \overline{F}_B(x) + PED_2 F_B(x) + (\Delta PED_1 - k\overline{F}_D(k))(x-c) f_B(x) = 0$$
$$\Rightarrow PED_1 + \Delta PED_1 F_B(x) + (\Delta PED_1 - kPD_2)(x-c) f_B(x) = 0.$$

[8] If $F_D(k) = 1$ then the problem becomes trivial and both generators bid c with probability 1.

This differential equation has the solution

$$F_B(x) = \begin{cases} K_1(x-c)^\alpha - \frac{PED_1}{\Delta PED_1}; & \text{if } \Delta PED_1 \neq 0, \\ \frac{PDE_1}{k\overline{F}_D(k)} \ln(x-c) + K_2; & \text{if } \Delta PED_1 = 0, \end{cases}$$

where K_1 and K_2 are constants determined by the boundary conditions and

$$\alpha = -\frac{\Delta PED_1}{\Delta PED_1 - k\overline{F}_D(k)}.$$

Notice that $\Delta PED_1 < k\overline{F}_D(k)$ because otherwise

$$\begin{aligned} &\mathsf{E}\left[D\left(\mathbf{1}_{\{D>k\}} - \mathbf{1}_{\{D\leq k\}}\right)\right] \geq 2k\overline{F}_D(k) \\ \Rightarrow &\overline{F}_D(k)\left(2k - \mathsf{E}[D|D>k]\right) \leq -F_D(k)\mathsf{E}[D|D\leq k]. \end{aligned}$$

If $\overline{F}_D(k) > 0$ and F_D is continuous on $(0, 2k]$, then the LHS of this inequality would be positive and the RHS would be non-positive, which is a contradiction. Therefore, $\alpha = 0$ if $\Delta PED_1 = 0$ and $\alpha > 0$ when $\Delta PED_1 \neq 0$.

Using the the boundary condition $F_B(\overline{p}) = 1$ we get the following integration constants

$$\begin{aligned} K_1 &= \left(1 + \frac{PED_1}{\Delta PED_1}\right)(\overline{p} - c)^{-\alpha}, \\ K_2 &= 1 - \frac{PDE_1}{k\overline{F}_D(k)} \ln(\overline{p} - c). \end{aligned}$$

By substituting these constants the distribution becomes

$$F_B(x) = \begin{cases} \frac{PED_2}{\Delta PED_1}\left(\frac{x-c}{\overline{p}-c}\right)^\alpha - \frac{PED_1}{\Delta PED_1}; & \text{if } \Delta PED_1 \neq 0, \\ 1 - \frac{PDE_1}{k\overline{F}_D(k)} \ln\left(\frac{\overline{p}-c}{x-c}\right); & \text{if } \Delta PED_1 = 0. \end{cases} \tag{5.18}$$

The lower bid bound $\underline{b}$ can be calculated by setting $F_B(\overline{p}) = 0$, which gives

$$\underline{b} = \begin{cases} c + (\overline{p} - c)\left(\frac{PED_1}{PED_2}\right)^{1/\alpha}; & \text{if } \Delta PED_1 \neq 0, \\ c + (\overline{p} - c)e^{-\frac{k\overline{F}_D(k)}{PDE_1}}; & \text{if } \Delta PED_1 = 0. \end{cases} \tag{5.19}$$

The lower bidding bound is greater than c, with $\underline{b} \downarrow c$ when $F_D(k) \downarrow 0$. This confirms that (5.17) gives the symmetric version of the stochastic model by *Fabra et al.* (2006) when $n = 2$.

5.4.3 Uniform Demand Model Example

We consider in this subsection the case when the demand is uniformly distributed in $(0, nk]$. The problem parameters become $PD_j = \frac{1}{n}$, $ED_j = \frac{k}{2}$, $PED_j = \frac{k}{2n}$, and $\Delta PED_j = 0 \; \forall \; j$, and the differential equation (5.17) simplifies to

$$\frac{k}{2n} - \frac{k}{n} x f(x)(n-1) = 0.$$

Therefore,

$$f_B(x) = \frac{1}{2(n-1)(x-c)}. \tag{5.20}$$

The cdf of the mixed strategy has the form

$$F_B(x) = \frac{1}{2(n-1)} \ln(x - c) + K,$$

where K is an integration constant. By using the boundary condition $F_B(\overline{p}) = 1$ then

$$K = 1 - \frac{1}{2(n-1)} \ln(\overline{p} - c).$$

Therefore, the mixed strategy cdf is

$$F_B(x) = 1 - \frac{1}{2(n-1)} \ln\left(\frac{\overline{p}-c}{x-c}\right). \tag{5.21}$$

The minimum bid that satisfies $F_B(\underline{b}) = 0$ is

$$\underline{b} = c + (\overline{p}-c)e^{-2(n-1)}. \tag{5.22}$$

The expected generator bid is

$$\begin{aligned}
\mathsf{E}[B] = \int_{\underline{b}}^{\overline{p}} x f_B(x)\, dx = \frac{1}{2(n-1)} \left[c \ln(x-c) + x\right]_{\underline{b}}^{\overline{p}} \\
= \frac{1}{2(n-1)} \left[c\ln(\overline{p}-c) + \overline{p} - c\ln\left((\overline{p}-c)e^{-2(n-1)}\right) - c - (\overline{p}-c)e^{-2(n-1)}\right] \\
= \frac{1}{2(n-1)} \left[\overline{p} + 2(n-1)c - c - (\overline{p}-c)e^{-2(n-1)}\right] = c + \frac{\overline{p}-c}{2(n-1)}\left[1 - e^{-2(n-1)}\right].
\end{aligned}$$

The expected value of the jth largest bid is

$$\mathsf{E}[B_{(j)}] = \int_{\underline{b}}^{\overline{p}} x f_{B_{(j)}}(x)\, dx = \frac{n!}{(j-1)!(n-j)!} \int_{\underline{b}}^{\overline{p}} x f(x) F_B(x)^{j-1} \overline{F}_B(x)^{n-j}\, dx.$$

Notice that the distribution of this order statistic is different from the one used in the derivation of Equation (5.17) because all n generators are accounted for in this ordering, while only $n-1$ generators were used in Equation (5.17)'s derivation. We can use this expectation to calculate the expected spot price

$$\begin{aligned}
\mathsf{E}[P] =& \sum_{j=1}^{n} PD_j \mathsf{E}[B_{(j)}] = \frac{1}{n} \sum_{j=1}^{n} \mathsf{E}[B_{(j)}] \\
=& \int_{\underline{b}}^{\overline{p}} x f_B(x) \left[\sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!(n-j)!} F_B(x)^{j-1} \overline{F}_B(x)^{n-j} \right] dx = \int_{\underline{b}}^{\overline{p}} x f_B(x)\, dx = \mathsf{E}[B] \\
=& c + \frac{\overline{p} - c}{2(n-1)} \left[1 - e^{-2(n-1)} \right].
\end{aligned}$$

This result shows that the average price increases linearly with the price ceiling $\overline{p}$ and the generator marginal cost c, but decreases nonlinearly with the number of generators n.

CHAPTER VI

Incentive Design for Optimal Electricity Network Transmission Expansion

6.1 Introduction

The sustainable development of electric power networks is a pressing issue in many regions, and according to *Hogan* (2008) there is currently very little incentives for investors. The electricity network is recognized as a natural monopoly (*Hogan* (1992)), and most of the transmission expansions in electricity networks are traditionally carried out by a central planner, such as the government or a regulated transmission company. The central planner chooses the transmission line paths and determines their capacities to maximizes the collective social welfare. This problem can be formulated as a mixed integer programming problem (*Binato et al.* (2001), *Alguacil et al.* (2003), and *de la Torre et al.* (2008)). Congestion-driven methods, such as *Shrestha and Fonseka* (2004), are also commonly used to make investment decisions. There is an abundance of transmission expansion planning models in the literature. *Latorre et al.* (2003) classifies some of these publications.

One distinct feature of electric networks is the *loop flow* phenomenon, which entails that an electricity injection in a node flows in every transmission line in the network according to Kirchhoff's laws. This causes free riding and public goods sharing issues and can deter network investments. Nevertheless, several mechanisms have been proposed to attract investors. Among the most popular mechanisms are the coalition investment and the merchant investment models. We will consider these two models in this chapter.

In the coalition investment model, generation companies and electricity retailers invest in transmission lines to access distant markets. This model became popular after its successful implementation in Argentina in the mid 1990's as depicted in *Littlechild and Skerk* (2008). Cooperative game theory is commonly used to study the investor coalition formation. For example, *Contreras and Wu* (1999) studies the coalition formation and the investment cost allocation using Shapely values, while *Contreras and Wu* (2000) considers kernel-stable coalitions. The benefits from a transmission expansion are unevenly realized by the network agents. In fact, some generators and customers may even be harmed by transmission expansions. Subsequently, coalitions are formed among agents that benefit from the investment, and an agent's investment share often depends on the agent's benefit from the transmission expansion. The physical network usage can also form a basis for allocating the investment costs as proposed by *Conejo et al.* (2007). Although some non-investors may also benefit from the expansion, the investors usually share most of the benefits.

In the merchant investment model, an independent investor seeks monetary gains from the transmission rent. This gains come in the form of financial transmission rights (FTR) that specify nodal injections in the network that the investor is entitled to. The investor ultimately benefits from collecting the monetary value of these nodal injections. The FTR concept has been adopted in many networks, and their benefits can also be realized by coalition investors. Several variants and hybrids of the merchant investment model have been considered, such as the merchant-regulatory mechanism proposed by *Hogan et al.* (2010). New transmission capacity often changes the topology of the network and affects the profits of generators, consumers, as well as transmission line owners. In fact, many of the ongoing transmission market design efforts aim at encouraging investments and protecting market participants from future network modifications.

Building new transmission lines relieves congestion in the network, which can affect the revenue of existing transmission line owners. *Hogan* (2011) uses a simple two node example to illustrate this phenomenon. To limit this effect, some have proposed mechanisms that reward existing FTR owners whenever a transmission expansion is made. For example, *Kristiansen and Rosellón* (2013) propose a method for awarding incremental FTRs that maintains revenue adequacy for existing FTR holders. We will not address this effect in this chapter and leave it as a possible future research direction. Another approach proposed by *Contreras et al.* (2009) is to use Shapley values to allocate the total rewards from the investments to the different investors based on the value added by each transmission line.

Transmission capacity expansions can also influence the generation competition. *Borenstein et al.* (2000) study the capacity of a transmission line connecting two nodes with symmetric generating firms and customers. They show that increasing the transmission line capacity can substantially improve competition. *Sauma and Oren* (2009) reach the same conclusion from studying a similar two node problem but with asymmetric generators and consumers between the two nodes. They also conclude that the cheaper generator has the correct capacity investment incentive.

In this chapter, we analyze the incentives of the coalition and merchant investment models for a general network topology. Although certain transmission expansions may reduce the operating efficiency and/or network capacity, we only consider expansions that are overall beneficial. We assume that the expansion paths are determined and study the capacity increment under each investment model. We first present the socially optimal expansion increments in §6.3.1, then consider the merchant investment model in §6.3.2 and the coalition investment model in §6.3.3. We show that neither of these models achieves the socially optimal capacity increment. We then consider in §6.3.4 and §6.3.5 two investment setups that are based on the merchant investor mechanism. We show that these two models yield near optimal investment levels. We then compare the different mechanisms using the IEEE 14 bus test system in 6.4 and conclude this chapter with some future research considerations in 6.5.

6.2 Background

We will use in this chapter the DC power network model shown in Appendix A. Specifically, we will use the optimal dispatch problem but without generator capacities, which is given by the subset of the problem (A.5)-(A.7) rather than the complete problem (A.5)-(A.8).

6.2.1 Problem Description

Consider an electricity network with a set of nodes N and directed transmission line arcs L. Without loss of generality, assume that each node $i \in N$ has at most a single generator and a single customer. A generator / load located in node i is indexed by i. Generator i incurs a cost rate $C_i(q)$ (in \$ per unit time) from generating q electric power units, and load i has a benefit rate $B_i(q)$ (in \$ per unit time) from consuming q power units. The functions C_i are increasing convex functions and B_i are increasing concave functions $\forall\ i \in N$. Starting with a line capacity $\mathbf{K}_0$, the network operator wants to make capacity increments to each line according to the increment vector $\mathbf{\Delta}$. Such an investment relieves congestion and improves the social welfare, but costs $I(\mathbf{K}_0, \mathbf{\Delta})$, where I is a jointly increasing and concave function in $\mathbf{\Delta}$.

6.2.2 Optimal Dispatch Problem

Consider a network with the set of line capacities $\mathbf{K}$. The ISO allocates production and consumption levels, $\mathbf{q^s}$ and $\mathbf{q^d}$, to maximize the social welfare defined as

$$SW(\mathbf{K}) \stackrel{\text{def}}{=} \sum_i \left(B_i(q_i^d(\mathbf{K}) - C_i(q_i^s(\mathbf{K}))\right).$$

The ISO must ensure the feasibility of nodal injections by constraining them to the set

$$\mathcal{Q}(\mathbf{K}) = \left\{(\mathbf{q^s}, \mathbf{q^d}) \in \mathbb{R}_+^N \times \mathbb{R}_+^N : -\mathbf{K} \leq H(\mathbf{q^s} - \mathbf{q^d}) \leq \mathbf{K}\right\}.$$

This ISO's allocation problem, known as the optimal dispatch problem, is given by

$$OD(\mathbf{K}) = \max_{(\mathbf{q^s},\mathbf{q^d})\in\mathcal{Q}(\mathbf{K})} \left\{\sum_i [B_i(q_i^d) - C_i(q_i^s)]\right\}.$$

We can consider this formulation as special case of problem (A.5)-(A.7) when the injection into node i is $q_i = q_i^s - q_i^d$ and the generation cost in node i is $C_i(q_i^s) - B_i(q_i^d)$. We will denote the solution to the optimal dispatch problem when the line capacities are $\mathbf{K}$ as $\mathbf{q^s}(\mathbf{K})$ and $\mathbf{q^d}(\mathbf{K})$ with individual nodal injection into node i as $q_i^s(\mathbf{K})$ and extraction (i.e. negative injection) from node i as $q_i^d(\mathbf{K})$. To satisfy the first order conditions, a solution to the optimal dispatch problem must satisfy[1]

$$B_i'(q_i^d(\mathbf{K})) = C_i'(q_i^s(\mathbf{K})) \qquad \forall\, i.$$

In other words, in an optimal solution the marginal benefit from consumption in a node equals the marginal cost of producing electricity from the node. This marginal cost / benefit equals the electricity LMP $p_i(\mathbf{K})$, which is the uniform price customers

[1] This is equivalent to the first order condition in Equation (A.9), but since the generation capacity constraint is not accounted for we can assume that γ_i is zero for all nodes.

in node i pay for their entire consumption and suppliers in node i receive for their entire production. Therefore, the surpluses of suppliers and customers in node i are

$$\pi_i^s(\mathbf{K}) = p_i(\mathbf{K})q_i^s(\mathbf{K}) - C_i(q_i^s(\mathbf{K})), \qquad \pi_i^d(\mathbf{K}) = B_i(q_i^d(\mathbf{K})) - p_i(\mathbf{K})q_i^d(\mathbf{K}).$$

Because different nodes may have different prices and supply may not equal the demand at a node, the consumer payments to the system may exceed the amount collected by the generators. This difference given by

$$TR(\mathbf{K}) \stackrel{\text{def}}{=} \sum_i p_i(\mathbf{K})(q_i^d(\mathbf{K}) - q_i^s(\mathbf{K}))$$

is known as the *transmission rent*, and is paid to the transmission line owners. This formula can be written in vector format as $\mathbf{p}(\mathbf{K})^\top T(\mathbf{K})$ where

$$T(\mathbf{K}) \stackrel{\text{def}}{=} \mathbf{q^d}(\mathbf{K}) - \mathbf{q^s}(\mathbf{K}).$$

The total producer and consumer surplus is

$$\begin{aligned}\Pi(\mathbf{K}) &\stackrel{\text{def}}{=} \sum_i \left(\pi_i^s(q_i^s(\mathbf{K})) + \pi_i^d(q_i^d(\mathbf{K}))\right) \\ &= \sum_i \left[p_i(\mathbf{K})\left(q_i^s(\mathbf{K}) - q_i^d(\mathbf{K})\right) + B_i(q_i^d(\mathbf{K})) - C_i(q_i^s(\mathbf{K}))\right],\end{aligned}$$

and the social welfare can be expressed as

$$SW(\mathbf{K}) = \Pi(\mathbf{K}) + TR(\mathbf{K}).$$

In other words, the social welfare of the system is given by the total benefits of the producers, consumers, as well as the transmission line owners. The difference in nodal prices in our model is caused by transmission line congestion[2], which implies

[2]Transmission losses may also play a role in the difference in nodal prices, but we do not consider

that owners of congested transmission lines get larger transmission rent allocations. The most common allocation method is based on point to point FTRs.

6.2.3 Financial Transmission Rights

The owners of a transmission network are collectively paid the transmission rent $TR(\mathbf{K}) = \mathbf{p}(\mathbf{K})^\top T(\mathbf{K})$. Correspondingly, the owners of the transmission network have the right to collect the profits on the injection vector $T(\mathbf{K})$ given the price vector $\mathbf{p}(\mathbf{K})$. The vector $T(\mathbf{K})$ is known as the aggregate node to node FTR. If there are multiple transmission line owners, each owner is assigned a FTR vetor τ, and all FTRs sum to $T(\mathbf{K})$.

Definition VI.1 (Financial Transmission Rights (FTR))**.** A point to point *financial transmission right* is a non-zero vector $\tau \in \mathbb{R}^N$, where a positive entry signifies power injection at a node and a negative entry signifies power generation at the node.

A transmission owner k with FTR τ^k is *obligated* to collect or make payments that match the generated and consumed generation values according to τ^k at the spot prices. The value of a transmission right τ^k is $\sum_i p_i \tau_i^k$, where p_i is the spot price at node i. The total FTRs in a network $T = \sum_k \tau^k$ is *simultaneously feasible* if $-T \in \mathcal{Q}(\mathbf{K})$. A necessary condition for financial feasibility is *revenue adequacy*, which requires that the surplus the ISO collects be at least as large as the required FTR paymens. It can be shown that a simultaneously feasible FTR satisfies the revenue adequacy condition[3]:

$$\sum_i p_i(\mathbf{K})(q_i^d(\mathbf{K}) - q_i^s(\mathbf{K})) \geq \sum_i p_i(\mathbf{K})(q_i^d - q_i^s) \;\; \forall \;\; (\mathbf{q^s}, \mathbf{q^d}) \in \mathcal{Q}(\mathbf{K}).$$

This result obviously applies to an FTR $T(\mathbf{K})$ since $-T(\mathbf{K}) \in \mathcal{Q}(\mathbf{K})$. If the sum of FTRs in a network with line capacities $\mathbf{K}_0$ is $T(\mathbf{K}_0)$, then an investor that raises the

their effect in this chapter.

[3]See for example *Hogan* (2002).

line capacity to $\mathbf{K}_1$ is assigned the incremental FTR

$$\tau(\mathbf{K}_1, T(\mathbf{K}_0)) = T(\mathbf{K}_1) - T(\mathbf{K}_0) = \left(\mathbf{q^d}(\mathbf{K}_1) - \mathbf{q^d}(\mathbf{K}_0)\right) - \left(\mathbf{q^s}(\mathbf{K}_1) - \mathbf{q^s}(\mathbf{K}_0)\right). \tag{6.1}$$

Naturally, no rational agent would make the capacity investment unless

$$\mathbf{p}(\mathbf{K}_1)^\top \tau(\mathbf{K}_1, T(\mathbf{K}_0)) \geq 0.$$

6.3 Transmission Investment Models

6.3.1 Centralized Transmission Investment Problem

In this problem a social welfare maximizing entity, such as the government or a regulated transmission company, expands the transmission line capacities to relieve congestion and improve the aggregate profit of producers and consumers. Starting with the transmission line capacity vector $\mathbf{K}_0$, expanding the transmission lines by the capacity vector $\mathbf{\Delta}$ gives a social welfare improvement

$$\begin{aligned} SW(\mathbf{K}_0, \mathbf{\Delta}) &\stackrel{\text{def}}{=} OD(\mathbf{K}_0 + \mathbf{\Delta}) - \sum_l I_l(K_l, \Delta_l) \\ &= \sum_i \left[B_i(q_i^d(\mathbf{K}_0 + \mathbf{\Delta})) - C_i(q_i^s(\mathbf{K}_0 + \mathbf{\Delta}))\right] - \sum_l I_l(K_l, \Delta_l). \end{aligned}$$

To get the optimal investment, the central planner chooses $\mathbf{\Delta}$ that maximizes

$$SW^*(\mathbf{K}_0) \stackrel{\text{def}}{=} \max_{\mathbf{\Delta} \geq \mathbf{0}} \left\{SW(\mathbf{K}_0, \mathbf{\Delta})\right\}.$$

We will focus on the investment in a single line l throughout this chapter. Let the incremental capacity vector $\mathbf{\Delta}_l$ be the vector of zeros for all entries $k \neq l$ and Δ_l for line l's entry. The first order optimality condition for Δ_l is

$$\begin{aligned}\frac{\partial I}{\partial \Delta_l}(\mathbf{K}_0, \boldsymbol{\Delta}_l) &= \sum_i \left[B_i'(q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}_l)) \frac{\partial q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}_l)}{\partial \Delta_l} - C_i'(q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}_l)) \frac{\partial q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}_l)}{\partial \Delta_l} \right] \\ &= \sum_i p_i(\mathbf{K}_0 + \boldsymbol{\Delta}_l) \left(\frac{\partial q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}_l)}{\partial \Delta_l} - \frac{\partial q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}_l)}{\partial \Delta_l} \right). \qquad (6.2)\end{aligned}$$

Therefore, the central planner should invest in increasing the capacity of line l so that the marginal benefits from the capacity expansion equals the marginal cost of the investment.

6.3.2 Merchant Investment Model

A merchant investor is a transmission line investor that neither supplies or consumes power nor owns existing FTRs in the network. If the transmission capacity vector is initially $\mathbf{K}_0$ with total FTRs $T(\mathbf{K}_0)$ allocated to existing network owners, the merchant would gain the FTR

$$\tau(\mathbf{K}_0 + \boldsymbol{\Delta}, T(\mathbf{K}_0)) = \mathbf{q^d}(\mathbf{K}_0 + \Delta) - \mathbf{q^s}(\mathbf{K}_0 + \Delta) - T(\mathbf{K}_0)$$

from an investment in capacity $\boldsymbol{\Delta}$, but would also incur the investment costs $I(\mathbf{K}_0, \boldsymbol{\Delta})$. If we consider the problem of investing in a single line l, then the merchant's surplus from adding capacity Δ_l to line l would be

$$\pi(\mathbf{K}_0, \boldsymbol{\Delta}_l) = \mathbf{p}(\mathbf{K}_0 + \boldsymbol{\Delta}_l)^T \tau(\mathbf{K}_0 + \boldsymbol{\Delta}_l, T(\mathbf{K}_0)) - I(\mathbf{K}_0, \boldsymbol{\Delta}_l). \qquad (6.3)$$

The optimal merchant investment Δ_l makes the marginal benefits from incremental FTRs equal to the marginal investment as the following first order condition indicates.

$$\frac{\partial I}{\partial \Delta}(\mathbf{K}_0, \mathbf{\Delta}_l) = \sum_i p_i(\mathbf{K}_0 + \mathbf{\Delta}_l)\left(\frac{\partial q_i^d(\mathbf{K}_0 + \mathbf{\Delta}_l)}{\partial \Delta_l} - \frac{\partial q_i^s(\mathbf{K}_0 + \mathbf{\Delta}_l)}{\partial \Delta_l}\right)$$
$$+ \sum_i \frac{\partial p_i}{\partial \Delta_l}(\mathbf{K}_0 + \mathbf{\Delta}_l)(q_i^d(\mathbf{K}_0 + \mathbf{\Delta}_l) - q_i^s(\mathbf{K}_0 + \mathbf{\Delta}_l) - T_i(\mathbf{K}_0)). \tag{6.4}$$

The last term in this formula is generally negative, which gives merchants an incentive to underinvest when compared to the socially optimal investment in (6.2). The intuition for this underinvestment is that a merchant benefits from congestion because it causes large price discrepancies between nodes, and hence the merchant would be willing to lower the magnitude of the FTR injections to raise electricity prices. If the expansion does not cause significant price changes in the network, then the second line in (6.4) becomes small and the transmission expansion would be close to the social optimal. Although the merchant investor model may seem as an attractive alternative to the centralized transmission approach, this model also has its fair share of criticism. *Joskow and Tirole* (2005), for example, point out that many challenges of the electricity markets that are not accounted for in the merchant investor model can cause significant inefficiencies.

6.3.3 Coalition Investment Model

In this model, a coalition of agents P make a transmission investment of capacity $\mathbf{\Delta}$. A coalition may include both suppliers and consumers. We denote the set of nodes with suppliers and consumers in P as P^s and P^d respectively. The coalition's gain has 2 parts: (1) the incremental FTRs from the transmission expansion, and (2) the increased surplus due to the nodal price and quantity changes. If the coalition invests in increasing the line capacities from $\mathbf{K}_0$ by $\mathbf{\Delta}$ with existing aggregate FTR

$T(\mathbf{K}_0)$, then the coalition's FTR gain is

$$\sum_i p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) \left[q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}) - q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}) - T_i(\mathbf{K}_0) \right],$$

and its surplus gain is

$$\sum_{i \in P^s} \left[p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}) - p_i(\mathbf{K}_0) q_i^s(\mathbf{K}_0) - C_i(q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta})) + C_i(q_i^s(\mathbf{K}_0)) \right]$$
$$+ \sum_{i \in P^d} \left[B_i(q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta})) - B_i(q_i^d(\mathbf{K}_0)) - p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}) + p_i(\mathbf{K}_0) q_i^d(\mathbf{K}_0) \right].$$

By combining the different cost and revenue terms, we can write coalition P's profit function as

$$\begin{aligned} \pi_P(\mathbf{K}_0, \boldsymbol{\Delta}) = & \sum_{i \in P^s} \left(q_i^s(\mathbf{K}_0) \left(p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) - p(\mathbf{K}_0) \right) + C_i(q_i^s(\mathbf{K}_0)) - C_i(q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta})) \right) \\ & + \sum_{i \in P^d} \left(B_i(q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}) - B_i(q_i^d(\mathbf{K}_0)) + q_i^d(\mathbf{K}_0) \left(p_i(\mathbf{K}_0) - p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) \right) \right) \\ & + \sum_{i \notin P^s} p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) \left(q_i^s(\mathbf{K}_0) - q_i^s(\mathbf{K}_0 + \boldsymbol{\Delta}) \right) \\ & + \sum_{i \notin P^d} p_i(\mathbf{K}_0 + \boldsymbol{\Delta}) \left(q_i^d(\mathbf{K}_0 + \boldsymbol{\Delta}) - q_i^d(\mathbf{K}_0) \right) - I(\mathbf{K}_0, \boldsymbol{\Delta}). \end{aligned}$$

This model, in general, does not give the socially optimal investment that satisfies (6.2). Coalitions with expensive generators and customers with access to cheap generation tend to have an incentive to underinvest in order to limit competition, while coalitions with cheap generators and expensive customers tend have an incentive to over-invest in order to reach other customers or access to cheap generation. Clearly, the central planner's problem and the merchant model are special cases of this model with $P = N$ and $P = \phi$ respectively.

6.3.4 Competitive Merchant Investment Model

We consider in this model the problem where M identical merchant investors expand the network simultaneously. A merchant investor j makes a decision to expand the line l by Δ_l^j (expansion vector is $\mathbf{\Delta}_l^j$ with Δ_l^j as line l's entry and 0 everywhere else) according to equation (6.4), but with an initial transmission line capacity $\mathbf{K}_0^j = \mathbf{K}_0 + \sum_{k \neq j} \mathbf{\Delta}_l^k$ and an initial FTR allocation $T^j = T(\mathbf{K}_0) + \sum_{k \neq j} \tau(\mathbf{K}_0^k, T^k)$. Hence, merchant j's expansion satisfies

$$\begin{aligned} \frac{\partial I}{\partial \Delta_l}(\mathbf{K}_0^j, \mathbf{\Delta}_l^j) = & \sum_i p_i(\mathbf{K}_0^j + \mathbf{\Delta}_l^j) \left(\frac{\partial q_i^d}{\partial \Delta_l} - \frac{\partial q_i^s}{\partial \Delta_l} \right) (\mathbf{K}_0^j + \mathbf{\Delta}_l^j) \\ & + \sum_i \frac{\partial p_i}{\partial \Delta_l}(\mathbf{K}_0^j + \mathbf{\Delta}_l^j)(q_i^d(\mathbf{K}_0^j + \mathbf{\Delta}_l^j) - q_i^s(\mathbf{K}_0^j + \mathbf{\Delta}_l^j) - T_i^j). \end{aligned} \tag{6.5}$$

By symmetry, $\Delta_l^j = \Delta_l^k = \Delta_l$ and $T^j = T^k = \hat{T}\ \forall\ j$ and k. Therefore, $\mathbf{K}_0^j = \mathbf{K}_0 + (M-1)\mathbf{\Delta}_j$ and $\hat{T} = T(\mathbf{K}_0) + (M-1)\tau(\mathbf{K}_0 + (M-1)\mathbf{\Delta}_l, \hat{T})$. We can further expand $\hat{T}$ as

$$\begin{aligned} \hat{T} = & T + (M-1)\left(\mathbf{q^d}(\mathbf{K}_0 + (M-1)\mathbf{\Delta}_l) - \mathbf{q^s}(\mathbf{K}_0 + (M-1)\mathbf{\Delta}_l) - \hat{T}\right) \\ = & \frac{T}{M} + \frac{M-1}{M}\left(\mathbf{q^d}(\mathbf{K}_0 + (M-1)\mathbf{\Delta}_l) - \mathbf{q^s}(\mathbf{K}_0 + (M-1)\mathbf{\Delta}_l)\right). \end{aligned}$$

This result can be used to calculate a single merchant's FTR as

$$\begin{aligned}
\tau(\mathbf{K}_0 + M\boldsymbol{\Delta}, \hat{T}) =& \mathbf{q^d}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \mathbf{q^s}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \hat{T} \\
=& \left[\mathbf{q^d}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \frac{M-1}{M}\mathbf{q^d}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] \\
& - \left[\mathbf{q^s}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \frac{M-1}{M}\mathbf{q^s}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] - \frac{T}{M} \\
=& \left[\mathbf{q^d}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \mathbf{q^d}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] \\
& - \left[\mathbf{q^s}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - \mathbf{q^s}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] \\
& + \frac{1}{M}\left[\mathbf{q^d}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l) - \mathbf{q^s}(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l) - T\right].
\end{aligned}$$

Under the assumptions that q_i^s and q_i^d are smooth, bounded, and monotone with respect to Δ_l, then $\tau(\mathbf{K}_0 + M\boldsymbol{\Delta}_l, \hat{T})$ goes to $\mathbf{0}$ as $M \to \infty$. We can substitute this result back into (6.5) to get

$$\begin{aligned}
\frac{\partial I}{\partial \Delta_l}(\mathbf{K}_0^j, \boldsymbol{\Delta}_l^j) =& \sum_i p_i(\mathbf{K}_0 + M\boldsymbol{\Delta}_l)\left(\frac{\partial q_i^d}{\partial \Delta_l} - \frac{\partial q_i^s}{\partial \Delta_l}\right)(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) \\
& + \sum_i \frac{\partial p_i}{\partial \Delta_l}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l)\left[q_i^d(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - q_i^d(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] \\
& - \sum_i \frac{\partial p_i}{\partial \Delta_l}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l)\left[q_i^s(\mathbf{K}_0 + M\boldsymbol{\Delta}_l) - q_i^s(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l)\right] \\
& + \frac{1}{M}\sum_i \frac{\partial p_i}{\partial \Delta_l}(\mathbf{K}_0 + M\boldsymbol{\Delta}_l)\left[q_i^d(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l) - q_i^s(\mathbf{K}_0 + (M-1)\boldsymbol{\Delta}_l) - T\right].
\end{aligned}$$

We will let $M \to \infty$ to reflect a perfectly competitive market. If there exists a solution $\tilde{\Delta}_l = \lim_{M\to\infty} M\Delta_l < \infty$ then all the terms except for the first sum cancel as $M \to \infty$ [4]. The increment in this case would be $\tilde{\boldsymbol{\Delta}}_l$[5] and the competitive equilibrium would

[4] Under the assumptions that q_i^s and q_i^d are smooth, bounded, and monotone and that $\frac{\partial p_i}{\partial \Delta_{l'}}$ is bounded.

[5] $\tilde{\boldsymbol{\Delta}}_l$ is a vector of zeros for lines $k \neq l$ and $\tilde{\Delta}_l$ for line l.

satisfy

$$\frac{\partial I}{\partial \Delta_l}(\mathbf{K}_0, \tilde{\mathbf{\Delta}}_l) = \sum_i p_i(\mathbf{K}_0 + \tilde{\mathbf{\Delta}}_l) \left(\frac{\partial q_i^d}{\partial \tilde{\Delta}_l} - \frac{\partial q_i^s}{\partial \tilde{\Delta}_l} \right) (\mathbf{K}_0 + \tilde{\mathbf{\Delta}}_l). \tag{6.6}$$

If we further assume that investment costs are not path dependent[6], then (6.6) and (6.2) become equivalent, and the competitive equilibrium coincides with the socially optimal solution.

6.3.5 Capacity Bidding Model

In this model, a public auction is conducted in which transmission investment companies compete for building a transmission line l in a network with initial transmission capacity $\mathbf{K}_0$. Investment companies bid for the capacity they are willing to add to line l, and **the bidder with the largest capacity increment bid wins the auction**, builds the line with the proposed increment, and collects a FTR according to (6.1). In case of a tie, the winner is chosen randomly from the maximum capacity bidders with equal winning probability for each player. We assume in this model that all bidders are merchant investors with profit functions given by (6.3) for the winning bidder. We will denote the optimal merchant investment that solves (6.4) by Δ_l^*, and the capacity at which an agent is indifferent between investing and not investing by $\overline{\Delta}_l$. This indifference capacity is the solution to the following equation.

$$I(\mathbf{K}_0, \overline{\Delta}_l) = \sum_i p_i(\mathbf{K}_0 + \overline{\mathbf{\Delta}}_l) \left(q_i^d(\mathbf{K}_0 + \overline{\mathbf{\Delta}}_l) - q_i^s(\mathbf{K}_0 + \overline{\mathbf{\Delta}}_l) - T_i \right).$$

Consider agent k's best response problem to a set of generators $-k$ with a maximum capacity bid Δ_l^{-k}. Agent k's preferred scenario is to win the auction with a bid

[6] An investment in a transmission line is not path dependent if expanding a line's capacity from A to $A+B$ then from $A+B$ to $A+B+C$ is the same as expanding it from A to $A+C$ and then from $A+C$ to $A+B+C$. In terms of I this can be expressed as $I(A+B, A+B+C) = I(A+C, A+B+C)$. This assumption in transmission investment is also made by *Vogelsang* (2001). An investment function with a constant marginal cost falls under this category.

Δ_l^*. Therefore, agent k always bids Δ_l^* when $\Delta_l^{-k} < \Delta_l^*$. If the highest bid from other generators equals or exceeds Δ_l^* then agent k would want to win if his winning bid does not exceed $\overline{\Delta}_l$. Therefore, as long as $\Delta_l^{-k} < \overline{\Delta}_l$ then k would rather win the bid. Since agent k's profit is decreasing in Δ_l^k in the region $[\Delta_l^*, \overline{\Delta}_l]$, then k minimizes his bid in this region. Bidding exactly Δ_l^{-k} would create a tie, in which case agent k's expected profit would be divided by the number of agents bidding Δ_l^{-k}, and hence agent k bids slightly over Δ_l^{-k} to secure a win. If $\Delta_l^{-k} = \overline{\Delta}_l$ then agent k makes a negative profit if he bids over $\overline{\Delta}_l$ and makes a zero profit if he ties the bid or looses the auction. Therefore, agent k is indifferent between tying the largest bid and loosing the auction and bids $\Delta_l^k \leq \overline{\Delta}_l$. Finally, if Δ_l^{-k} exceeds $\overline{\Delta}_l$ agent k cannot profitably win the auction, and would bid $\Delta_l^k < \Delta_l^{-k}$. Agent k's bid Δ_l^k is summarized in the following best response correspondence.

$$BR_k(\Delta_l^{-k}) = \begin{cases} \Delta_l^*, & \text{if } \Delta_l^{-k} < \Delta_l^*; \\ \Delta_l^{-k} + \varepsilon, & \text{if } \Delta_l^{-k} \in [\Delta_l^*, \overline{\Delta}_l); \\ \leq \overline{\Delta}_l, & \text{if } \Delta_l^{-k} = \overline{\Delta}_l; \\ < \Delta_l^{-k}, & \text{if } \Delta_l^{-k} > \overline{\Delta}_l. \end{cases}$$

ε in this expression is a small positive increment that agent k makes to win the auction. Figure 6.1 illustrates the best response bids for a two agent problem. Agent 1's bid in this Figure when $\Delta_l^2 \in [\Delta_l^*, \overline{D}_l)$ is slightly below the dashed line, while agent 2's bid when $\Delta_l^1 \in [\Delta_l^*, \overline{D}_l)$ is slightly above the dashed line. The two best response correspondences only intersect in $(\overline{\Delta}_l, \overline{\Delta}_l)$. Therefore, this problem has a unique PSNE in which every agent bids $\overline{\Delta}_l$.

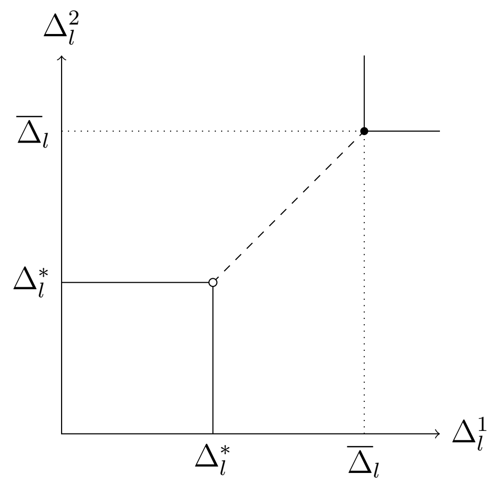


Figure 6.1: Best responses plot for 2 investors in the capacity bidding model.

Linear Investment Model

We consider in this subsection a linear investment cost model

$$I(\mathbf{K}_0, \Delta_l) = I_l \Delta_l.$$

We assume that I_l is a constant marginal expansion cost for all agents. Since the only equilibrium winning bid in this problem is $\overline{\Delta}_l$, the investment expansion would satisfy

$$\begin{aligned} I_l\overline{\Delta}_l &= \sum_i p_i(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l)\left(q_i^d(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l) - q_i^s(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l) - T_i\right) \\ \Rightarrow I_l &= \sum_i p_i(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l)\left(\frac{q_i^d(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l)}{\overline{\Delta}_l} - \frac{q_i^s(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l)}{\overline{\Delta}_l} - \frac{T_i}{\overline{\Delta}_l}\right). \end{aligned}$$

By substituting $T = \mathbf{q}^d(\mathbf{K}_0) - \mathbf{q}^s(\mathbf{K}_0)$, then

$$I_l = \sum_i p_i(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l)\left(\frac{q_i^d(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l) - q_i^d(\mathbf{K}_0)}{\overline{\Delta}_l} - \frac{q_i^s(\mathbf{K}_0 + \overline{\boldsymbol{\Delta}}_l) - q_i^s(\mathbf{K}_0)}{\overline{\Delta}_l}\right).$$

If $\bar{\Delta}$ is small compared to the existing transmission line capacity in the network and if the change in supply and demand quantities due to this change is relatively small, then we can approximate

$$\frac{q_n^d(\mathbf{K}_0+\overline{\mathbf{\Delta}}_l)-q_n^d(\mathbf{K}_0)}{\overline{\Delta}_l}\approx\frac{\partial q_n^d}{\partial\overline{\Delta}_l}(\mathbf{K_0}+\overline{\mathbf{\Delta}}_l)\text{ and }\frac{q_n^s(\mathbf{K}_0+\overline{\mathbf{\Delta}}_l)-q_n^s(\mathbf{K}_0)}{\overline{\Delta}_l}\approx\frac{\partial q_n^s}{\partial\Delta_l}(\mathbf{K_0}+\overline{\mathbf{\Delta}}_l).$$

Hence,

$$I_{l'}\approx\sum_n p_n(\mathbf{K}_0+\bar{\mathbf{\Delta}})\left(\frac{\partial q_n^d}{\partial\Delta_{l'}}(\mathbf{K_0}+\bar{\mathbf{\Delta}})-\frac{\partial q_n^s}{\partial\Delta_{l'}}(\mathbf{K_0}+\bar{\mathbf{\Delta}})\right).$$

By comparing this to (6.2) we can conclude that this mechanism gives an approximate socially optimal solution when the investment cost is linear and the increment is not very large compared to the existing network. We will compare the solutions from the different mechanisms next.

6.4 Numerical Example

6.4.1 Problem Setup

We conduct in this section a numerical assessment of five transmission investment scenarios on the IEEE 14 bus transmission network used in *Freris and Sasson* (1968). The network has 5 generators, which we assume to be identical with production cost rate $q+\frac{1}{2}q^2$ for producing q power units. The network has 10 customer loads, which we also assume to be identical with a benefit function $5q-6q^2$. Each generator and load is located in one of 14 nodes in a network that consists of 20 transmission lines. Figure 6.2 shows a schematic of this network and Table 6.1 shows the line resistance R and conductance X information. Note that the per-unit (pu) system is used for all quantities throughout this example[7].

Initially, we assume that all generators and transmission lines have large enough capacities and never reach their limits in an optimal dispatch. We call this case the

[7]The per-unit system is a standard normalization convention in power systems that is used to compared actual quantities to base values. For example, refer to *Glover et al.* (2011) for more details.

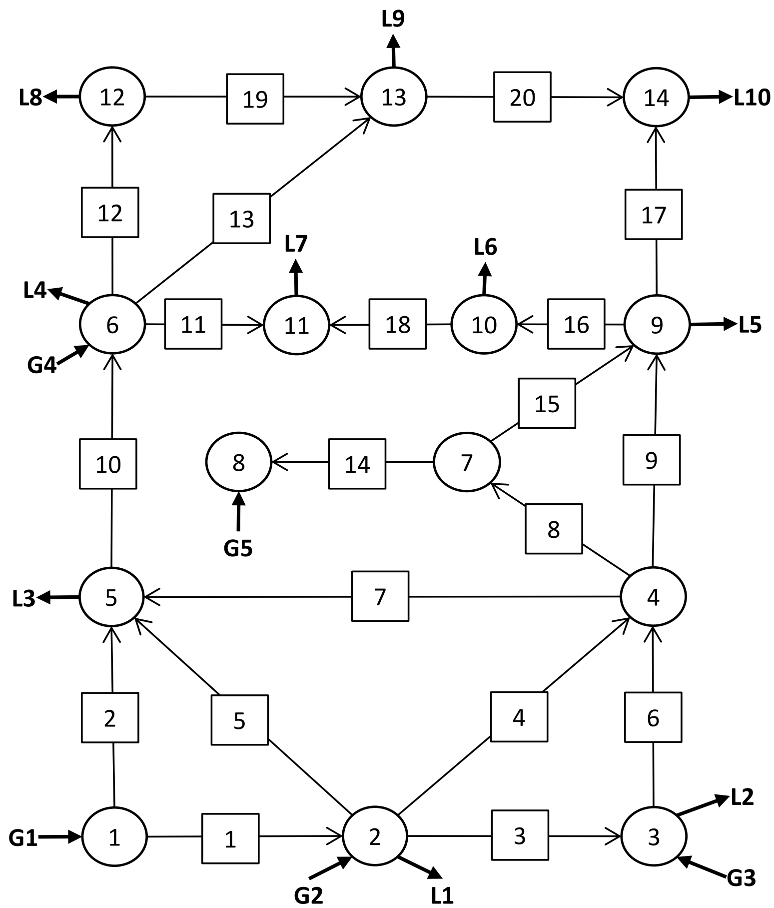


Figure 6.2: IEEE 14 bus transmission network. The busses are shown as the nodes indexed by 1-14 inside the circles and the transmission lines are shown as the directed arcs between the nodes with indices given in boxes. Generators (Loads) are indexed by G (L) and are connected to their respective nodes in this drawing.

infinite capacity case. We then downgrade line 15 alone and assume it has a zero capacity, but without removing the line from the network. This causes line 15 to bind the network operation and limit the rest of the line power flows and the nodal injections in the network. We refer to this second case as the *zero capacity* case.

After finding the optimal dispatch solution under line 15's two extreme cases, we allow investors to expand this line and collect the incremental FTRs from this expansions and incur a constant marginal cost of 1 $ per capacity unit for their investment. We consider 5 investment scenarios: the social welfare solution (SW) shown in §6.3.1, the merchant investment model (MI) from §6.3.2, the capacity auction model (CA) from §6.3.5, and two coalition examples (P1) and (P2) from the coalition investment model presented in §6.3.3. We let the P1 coalition consist of the generators and customers that benefit from the line congestion and the P2 coalition be the set of generators and customers that are hurt by the congestion.

6.4.2 Limiting Cases

In the infinite capacity case, the generators and loads are free to exchange electricity as if all agents were at the same location. As a result, the LMPs are identical ($ 2 per power unit) for all nodes as shown in Table 6.2. Because of the generator and load cost symmetry, all generators produce 0.5 power units and make a profit of $0.25 per time unit as shown in Table 6.3, while every load consumes 0.25 power units and makes a profit of $0.875 per time unit as shown in Table 6.4. Line 15 has a flow magnitude of 0.6404 power units, which is the largest among all transmission lines as shown in 6.1. Note that negative flows in this Table indicate power flow opposite to the predefined flow direction.

Reducing the capacity of line 15 to 0 affects the entire network because it changes the feasible set $\mathcal{Q}$, which can have a profound impact on the optimal dispatch solution. Indeed, as Table 6.1 shows, all line flows change when line 15's capacity drops

Table 6.1: Transmission line information for the IEEE 14 bus system.

Line	Bus		R	X	**Infinite Capacity**		**Zero Capacity**	
	From	To			Power Flow	μ	Power Flow	μ
			pu	pu	pu	\$/pu	pu	\$/pu
1	1	2	0.01938	0.05917	0.2239	0	0.1727	0
2	1	5	0.05403	0.22304	0.2761	0	0.0604	0
3	2	3	0.04699	0.19797	-0.0132	0	0.0624	0
4	2	4	0.05811	0.17632	0.2248	0	0.0209	0
5	2	5	0.05695	0.17388	0.2624	0	0.0153	0
6	3	4	0.06701	0.17103	0.2368	0	-0.0455	0
7	4	5	0.01335	0.04211	0.1417	0	-0.0245	0
8	4	7	0	0.20912	0.1404	0	0	0
9	4	9	0	0.55618	0.1794	0	0	0
10	5	6	0	0.25202	0.4302	0	-0.2367	0
11	6	11	0.09498	0.1989	0.1582	0	0.1743	0
12	6	12	0.12291	0.25581	0.213	0	0.1453	0
13	6	13	0.06615	0.13027	0.309	0	0.23	0
14	7	8	0	0.17615	-0.5	0	0	0
15	7	9	0	0.11001	0.6404	0	0	-7.2238
16	9	10	0.03181	0.0845	0.3418	0	-0.0267	0
17	9	14	0.12711	0.27038	0.228	0	0.0113	0
18	10	11	0.08205	0.19207	0.0918	0	-0.0689	0
19	12	13	0.22092	0.19988	-0.037	0	-0.0182	0
20	13	14	0.17093	0.34802	0.022	0	0.0622	0

to zero, with some flows increasing in magnitude, some decreasing in magnitude, and some flows changing direction. The μ values in this table are the dual variables of the flow capacity constraint. μ is zero when the line capacity constraint is not biding and non-zero when it is. Because line 15 is congested in this problem, incremental capacity to this line would have a marginal benefit of \$7.22 per power unit, which is attributed to the increase in both agent surpluses and transmission rent. Although the gain in transmission rent may be initially attractive, the capacity increments have reducing marginal gains that give independent merchants an incentive to under-invest. Existing agents and transmission line owners investing in the network may have different capacity increment incentives when accounting for the additional surpluses to their existing generator, load, and FTR portfolios. Naturally, no investor would exceed a capacity increment of 0.6404 power units since it relieves line 15 of its congestion, and exceeding this level would incur additional investment costs without impacting the optimal dispatch solution.

The drop in line 15's capacity restricts the trade between generators and loads that would otherwise have access to the entire market. As a result, generators and loads at different loctions in the network have varying injections and marginal costs and benefits, causing locational price discrimination. Although the average nodal price increases by about %12, half of the nodes see a reduction in prices as shown in Table 6.2. The prices at node 7 and 8 even become negative, indicating that additional consumption in these busses can relieve congestion in the network and allow more efficient trades to take place in another part of the network. Consequently, generator 5 located in node 8 refrains from production when line 15 has a zero capacity.

This limited market access reduces the electricity trade from 2.5 to 1.608 power units and causes a social welfare drop of about 36%. Nevertheless, some agents can benefit from the reduced competition. For example, generator 4 can reach customers that would have been served by other generators had there been sufficient network capapcity. As a result, generator 4's production increases by almost %92 and its profits increase by over 3.5 fold. Similarly, the isolation of other customers gives loads 1, 2, and 3 greater access to cheap generation. This causes these loads to consume %18 more electricity and to raise their gains by %39 when the network is congested.

Table 6.2: Node information for the IEEE 14 bus system.

Node	**Infinite Capacity**		**Zero Capacity**	
	LMP $/pu	Angle, θ degrees	LMP $/pu	Angle, θ degrees
1	2	0	1.4663	0
2	2	-0.0147	1.4445	-0.0113
3	2	-0.0119	1.3865	-0.0244
4	2	-0.0586	1.3317	-0.0154
5	2	-0.0652	1.5449	-0.0143
6	2	-0.1736	2.9194	0.0454
7	2	-0.088	-1.1188	-0.0154
8	2	0.0001	-1.1188	-0.0154
9	2	-0.1584	4.8159	-0.0154
10	2	-0.1914	4.4937	-0.0128
11	2	-0.2122	3.7352	0.0028
12	2	-0.2407	3.0377	-0.0004
13	2	-0.2242	3.2044	0.0077
14	2	-0.2337	4.1178	-0.0192

Table 6.3: Generator information for the IEEE 14 bus system.

Generator	Node	**Infinite Capacity**				**Zero Capacity**			
		Prod pu	Rev $/pu	Cost $/pu	Profit $/pu	Prod pu	Rev $/pu	Cost $/pu	Profit $/pu
1	1	0.5	1	0.75	0.25	0.2331	0.3419	0.2875	0.0544
2	2	0.5	1	0.75	0.25	0.2222	0.321	0.2716	0.0494
3	3	0.5	1	0.75	0.25	0.1932	0.2679	0.2306	0.0373
4	6	0.5	1	0.75	0.25	0.9597	2.8018	1.8808	0.9211
5	8	0.5	1	0.75	0.25	0	0	0	0
Total		2.5	5	3.75	1.25	1.6082	3.7326	2.6705	1.0622

Table 6.4: Load information for the IEEE 14 bus system.

Load	Node	Infinite Capacity				Zero Capacity			
		Cons $/pu	Benefit $/pu	Cost $/pu	Surplus $/pu	Cons $/pu	Benefit $/pu	Cost $/pu	Surplus $/pu
1	2	0.25	0.875	0.5	0.375	0.2963	0.9547	0.428	0.5267
2	3	0.25	0.875	0.5	0.375	0.3011	0.9616	0.4175	0.5441
3	5	0.25	0.875	0.5	0.375	0.2879	0.9422	0.4448	0.4974
4	6	0.25	0.875	0.5	0.375	0.1734	0.6865	0.5062	0.1804
5	9	0.25	0.875	0.5	0.375	0.0153	0.0753	0.0739	0.0014
6	10	0.25	0.875	0.5	0.375	0.0422	0.2003	0.1896	0.0107
7	11	0.25	0.875	0.5	0.375	0.1054	0.4604	0.3937	0.0667
8	12	0.25	0.875	0.5	0.375	0.1635	0.6572	0.4967	0.1604
9	13	0.25	0.875	0.5	0.375	0.1496	0.6138	0.4795	0.1343
10	14	0.25	0.875	0.5	0.375	0.0735	0.3352	0.3027	0.0324
Total		2.5	8.75	5	3.75	1.6082	5.8872	3.7326	2.1545

6.4.3 Investment Scenarios

§6.4.2 shows the optimal dispatch outcomes for the two extreme cases with line 15 having zero and infinite capacity. In this section, we consider the problem in which an investor builds a line, incurs the investment cost, and gains the incremental transmission rent. As we have seen throughout this chapter, an investor's optimal capacity expansion decision depends on the model's incentives. The socially optimal investment from the centralized model is used as a benchmark in this section. We refer to the socially optimal solution as SW. Using the coalition investment model in §6.3.3, we consider two coalition options: P1 is the set of agents that benefit form the congestion of line 15, and P2 are the agents that are harmed by the congestion. We also use two merchant investment models: MI is the model with a single merchant that chooses the optimal solution as shown in 6.3.2, and CA is the capacity auction model discussed in §6.3.5. When considering the optimal investment we limit our search to capacities no greater than 0.6404 power units for all models as discussed in §6.4.2.

The outcomes of the five investment scenarios are shown in Figure 6.3. The horizontal axis in this Figure is the capacity investment level and the vertical axis is the investor surplus. Each of the 4 plots shows a utility curve for a scenario with the best investment marked on the curve. The merchant investor curve has two investment outcomes; the utility maximizing solution for the single merchant investor Δ_l^{MI} and the zero surplus solution for the capacity auction model Δ_l^{CA}. The centralized optimal investment level for which the marginal welfare improvement equals the marginal investment cost is 0.495 pu. Figure 6.3 shows that agents in coalition P1 would under invest by about %82 if given the opportunity, while agents in coalition P2 would over-invest by about %23. These deviations from the socially optimal levels are caused by the non-FTR related surplus losses (gains) agents in P1 (P2) realize. On the other hand, the optimal investment capacity for a merchant is %64 lower than the socially optimal value, but if multiple merchants are asked to bid for the line's capacity as in §6.3.5 then the winning bidder would build the socially optimal capacity.

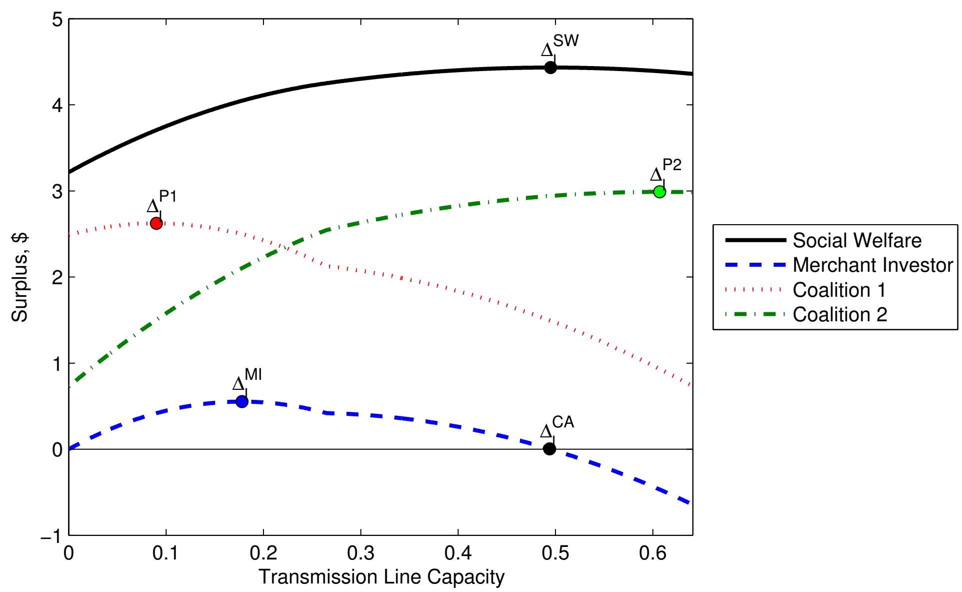


Figure 6.3: Capacity expansions under different investors for the IEEE 14 bus transmission network for the linear investment cost function $I(K) = K$.

The same model is tested for two other investment cost functions. Figures 6.4 and 6.5 show the solutions under the four investment options for a concave and a convex investment cost function, respectively. Although the capacity auction solution is not socially optimal when the investment cost is either concave or convex, the result is very close to the social optimal and gives a substantial improvement over all other investment models.

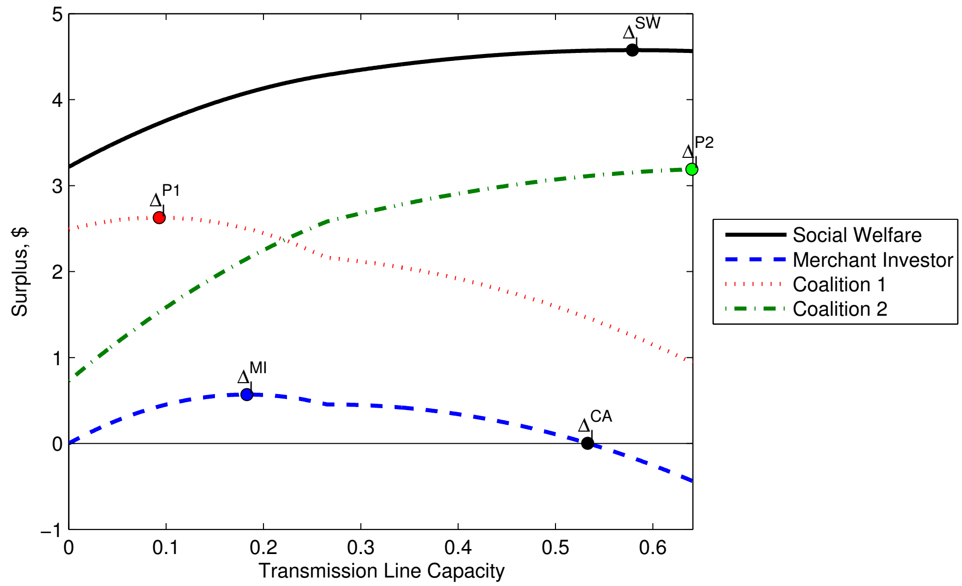


Figure 6.4: Capacity expansions under different investors for the IEEE 14 bus transmission network for the concave investment cost function $I(K) = K - 0.5K^2$.

6.5 Conclusion and Future Work

We have studied in this chapter several transmission network expansion mechanisms and used a network example to compare their performance to the central planner's solution. The analysis in §6.3.5 shows that the transmission expansion from the merchant investor capacity auction mechanism can approximate the social optimal, and the numerical experiment confirms it. Our analysis assumes that merchant investors do not already own network assets and do not account for risk or future ex-

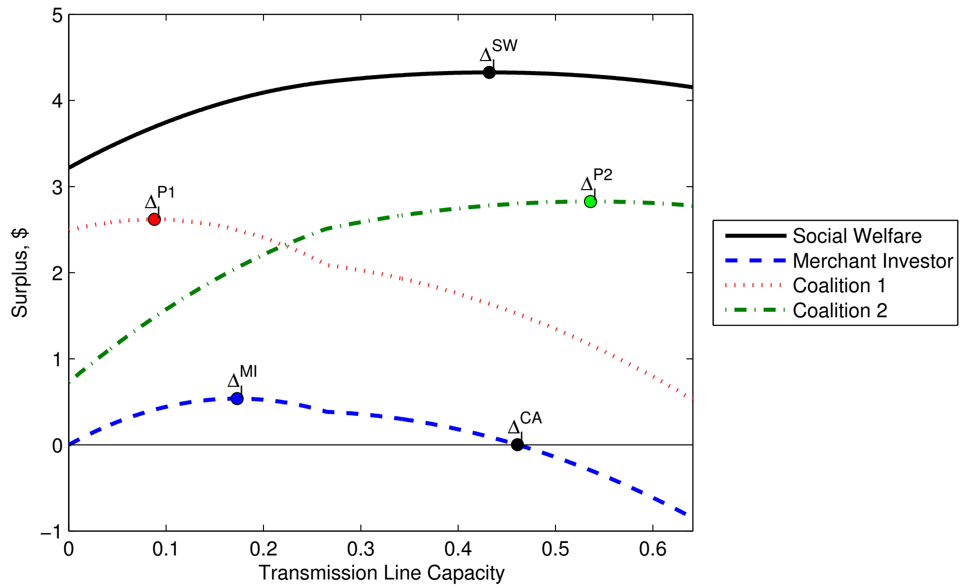


Figure 6.5: Capacity expansions under different investors for the IEEE 14 bus transmission network for the convex investment cost function $I(K) = K + 0.5K^2$.

pansions. One possible area of future work is to include investors with existing FTRs. In such a model, FTRs of existing transmission line owners would lose value due to the reduction in price discrepancy, which can cause them to underinvest. Although most of the analysis conducted in this section pertains to single line investments, many of these results can be extended to the multiple line investment problem. Nevertheless, having investors choose both the lines and the capacity expansions could be an interesting research direction.

The transmission expansion benefit usually discussed in the literature are based on operational efficiency improvements. However, as discussed by *Hogan et al.* (2010), a lot of the transmission expansions are prompted by the system's reliability standards rather than congestion relief. They argue that the current network reliability standards cannot be justified economically and express the importance of resolving this issue. Finding mechanisms that give investors the correct incentives for reliability investments could be an important step in closing the reliability standards and economic benefits gap.

After a network expansion is made, strategic generators can change their bidding behavior to capture some of the transmission rent as shown in *Oren* (1997). In fact, *Joskow and Tirole* (2000) show that more ownership of transmission rights by generating firms can influence their market power. Such a research direction would be in the merit of this dissertation because it integrates the competitive behavior of agents studied in the previous chapters with the investment incentives in this chapter.

CHAPTER VII

Conclusion

The evolution of electricity systems and the emergence of new technologies can have profound consequences on the market structure and competition. In fact, this thesis was partly inspired by the recent interest in renewable energy technologies, and especially by their impact on the electricity market. As it turns out, the penetration and subsidies to renewables could indeed change the strategic behavior of generating firms as shown in Chapter III. The widespread of renewables could also require substantial transmission investments if they are located in remote areas, as is common with many wind plants, which can be regarded as another form of government subsidy. Developing incentive mechanisms for connecting these remote plants to the grid could be integral to the promotion of renewable technologies and an interesting future research topic.

With the adoption of the smart grid technologies, the market competition issues can become very different. For example, the widespread of cheap energy storage technologies would reduce the negative impact of intermittency, reduce the reliance on flexible generation, and downgrade the market power caused by fast ramping capabilities. Competition can also be intensified by increased consumer market participation and the popularity of smart appliances that can adjust their operating levels in response to real-time spot prices.

Some of the new technologies can alter the physical network constraints. Transmission switching technologies, as in *Hedman et al.* (2011), would regularly alter the transmission network topology, which may require the development of new FTR calculation methods and new reward systems for transmission line owners and network investors. Moreover, the strategic response of generating firms would influence the optimal transmission switching scheme. The special structure of electricity markets and the ongoing technology changes continuously invoke research interest and the need for market restructuring.

APPENDICES

APPENDIX A

Electrical Power Network Background

Electric power flow can be modeled with varying degrees of complexity. The more detailed models are usually used for operational purposes, while simpler models are used for market design and economic analysis. We use the DC representation of power networks in this study, which is consistent with the market design problems in the literature.

Consider a connected network with a set of nodes $\mathcal{N}$ and a set of directed arcs $\mathcal{L} \subseteq \{(i, j) : i, j \in \mathcal{N}\}$. A node can represent a generator or a consumer, while an arc represents a transmission or a distribution line. We use directed arcs rather than non-directed edges to represent power lines because the flow in one direction of a line does not necessarily have the same magnitude as the flow in the opposite direction due to transmission losses. Note that if $(i, j) \in \mathcal{L}$ then we required that $(j, i) \in \mathcal{L}$.

A line (i, j) has four attributes: *susceptance* $B_{ij} \geq 0$, *conductance* $G_{ij} \geq 0$, *flow* f_{ji}, and *capacity* $\bar{f}_{ij} > 0$. Flow through a line is a function while the other attributes are transmission line properties, where $B_{ij} = B_{ji}$, $G_{ij} = G_{ji}$, and $\bar{f}_{ij} = \bar{f}_{ji}$. f_{ij} represents the flow leaving node i towards node j, which may be different from the flow entering node j coming from node i on the other end of the line. Generally $f_{ij} \neq -f_{ji}$ unless we ignore the line losses. A node i has three attributes: *voltage* V_i, *phase angle* θ_i, and *injection* q_i. We wil use the notation $\theta_{ij} = \theta_i - \theta_j$ and $V_{ij} = V_i - V_j$ for the phase angle and voltage difference between two nodes. V_i and θ_i are variables that depend on the state of the network while q_i is a decision variable that represents the exogenous power generation or consumption. In contrast to some other network problems, power flow through lines is not a decision variable, it is determined by the state of the network. We will derive the power flow equations next.

A.1 The AC Power Flow Equations

The susceptance and conductance of a line (i, k) can be calculated from the lines resistance R_{ik} and reactance X_{ik}, both measured in Ohms (Ω). The impedance of a line is defined as $Z_{ik} \stackrel{\text{def}}{=} R_{ik} + jX_{ik}$, where $j = \sqrt{-1}$, and its admittance is given by $Y_{ik} \stackrel{\text{def}}{=} 1/Z_{ik} = \frac{1}{R_{ik}+jX_{ik}} = \frac{R_{ik}-jX_{ik}}{R_{ik}^2+X_{ik}^2}$. The line's conductance is defined as $G_{ik} \stackrel{\text{def}}{=} \frac{R_{ik}}{R_{ik}^2+X_{ik}^2}$ and its susceptance as $B_{ik} \stackrel{\text{def}}{=} \frac{X_{ik}}{R_{ik}^2+X_{ik}^2} \Rightarrow Y_{ik} = G_{ik} - jB_{ik}$. If the current flow through the line is I_{ik}, then the voltage drop across the line is $V_{ik} = V_i - V_k = I_{ik}Z_{ik}$ and $I_{ik} = V_{ik}Y_{ik}$. The power loss in the line is $S_{ik} = V_{ik}I_{ik}^*$, where a superscript $*$ denotes the complex conjugate. S, measured in volt-ampere (VA), is called the complex power and consists of a real or active power flow component P measures in watts (W), and a reactive power flow component Q measured in volt-ampere reactive (var). The relationship between these power components is given by $S = P + jQ$. Note that we use S_i, P_i, and Q_i for power injections into node i and S_{ik}, P_{ik}, and Q_{ik} for the power flow through line (i, k). Due to line losses, the flow in one end of a line may be different from the flow at the other end of the line. As a convention, we define the flows S_{ik}, P_{ik}, and Q_{ik} at node i's end of the line.

If S_i is injected into node i then using $P_i = \text{Re}(S_i)$ and $Q_i = \text{Im}(S_i)$ we can calculate the real and a reactive power components as

$$P_i = \sum_{k \in \mathcal{N}} V_i V_k \left(G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik} \right),$$
$$Q_i = \sum_{k \in \mathcal{N}} V_i V_k \left(G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik} \right).$$

For two different busses i and k ($i \neq k$), the power flow equations are:

$$\begin{aligned}
\frac{\partial P_i}{\partial \theta_k} &= V_i V_k \left(G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik} \right), \\
\frac{\partial P_i}{\partial V_k} &= V_i \left(G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik} \right), \\
\frac{\partial Q_i}{\partial \theta_k} &= - V_i V_k \left(G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik} \right), \\
\frac{\partial Q_i}{\partial V_k} &= V_i \left(G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik} \right).
\end{aligned}$$

If $i = k$ then

$$\begin{aligned}
\frac{\partial P_i}{\partial \theta_i} &= -B_{ii} V_i^2 - Q_i, & \frac{\partial P_i}{\partial V_i} &= G_{ii} V_i + \frac{P_i}{V_i}, \\
\frac{\partial Q_i}{\partial \theta_i} &= -G_{ii} V_i^2 + P_i, & \frac{\partial Q_i}{\partial V_i} &= -B_{ii} V_i + \frac{Q_i}{V_i}.
\end{aligned}$$

We can therefore write

$$\begin{aligned}
\partial P_i &= V_i \sum_{k:(i,k) \in \mathcal{L}} \left[V_k \left(G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik} \right) \partial\theta_k + \left(G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik} \right) \partial V_k \right], \\
\partial Q_i &= V_i \sum_{k:(i,k) \in \mathcal{L}} \left[-V_k \left(G_{ik} \cos\theta_{ik} + B_{ik} \sin\theta_{ik} \right) \partial\theta_k + \left(G_{ik} \sin\theta_{ik} - B_{ik} \cos\theta_{ik} \right) \partial V_k \right].
\end{aligned}$$

Shunt capacitors are often connected in parallel to transmission lines to reduce current losses due to electromagnetizm. If line (i, k) has a shunt capacitance C_{ik}^s (in Ω), then the real and reactive power flows along line (i, k) can be calculated as

$$P_{ik} = G_{ik}\left[V_i^2 - V_i V_j \cos(\theta_{ik})\right] + B_{ik} V_i V_j \sin\theta_{ik}$$
$$Q_{ik} = B_{ik}\left[V_i^2 - V_i V_j \cos(\theta_{ik})\right] - G_{ik} V_i V_j \sin\theta_{ik} - V_i^2 C_{ik}^s.$$

A.2 The DC Power Flow Model

The AC power flow model gives a non-linear set of equations with a non-convex solution space. Solving large power flow problems using the AC power flow model can be challenging, which is why the simpler DC linear model is often used to approximate the solution. The DC power flow model is commonly used in problems that do not directly affect the network's physical operation. In particular, the DC models is widely used in modeling electricity markets and in resolving financial transactions. Starting from the general AC model, we will derive the DC power flow model using the necessary approximations next.

The focus in the DC power flow is on the real power P_{ij}. Therefore, we will set $f_{ij} \stackrel{\text{def}}{=} P_{ij}$ in the DC power flow model. From the power flow equations, the flow on line (i, j) can be calculated as $f_{ij}(\theta_{ij}, V_i, V_j) = G_{ij} V_i^2 - G_{ij} V_i V_j \cos(\theta_{ij}) + B_{ij} V_i V_j \sin(\theta_{ij}) \ \forall \ (i, j) \in \mathcal{L}$. Power network analysis is usually simplified by adopting a "per unit" system in which voltages are normalized to 1 without loss of generality (i.e. $V_i = 1 \ \forall \ i \in \mathcal{N}$). Consequently, we can simplify the line flow expression to $f_{ij}(\theta_{ij}) = G_{ij} - G_{ij}\cos(\theta_{ij}) + B_{ij}\sin(\theta_{ij})$. Furthermore, the phase angles are usually small under normal operating conditions, and it is typical to adopt the second order Taylor approximations for the sin and cos functions ($\sin(\theta) \approx \theta$ and $\cos(\theta) \approx 1 - \frac{\theta^2}{2}$). Thus, the power flow equations can be approximated by $f_{ij}(\theta_{ij}) = B_{ij}\theta_{ij} + \frac{G_{ij}}{2}\theta_{ij}^2$ for $(i, j) \in \mathcal{L}$. Although transmission capacities are not hard constraints, meaning that they could be violated for short durations as proposed by *Hedman et al.* (2011), we will restrict power flow to rated line capacities in our model, so the ISO must ensure that $f_{ij}(\theta_{ij}) \leq \bar{f}_{ij}\ \forall\ (i,j) \in \mathcal{L}$. Note that there is no need to specify a lower bound on the flow through line (i,j) since it is guaranteed by the constraint on the flow through line (j,i).

Nodes have two sets of constraints in the DC model; nodal balance and injection capacities. To get a balanced model, nodal injections must satisfy $q_i = \sum_{j\in\mathcal{N}:(i,j)\in\mathcal{L}} f_{ij}(\theta_{ij})\ \forall\ i \in \mathcal{N}$. Because of the network structure of the problem, one nodal balance constraint becomes redundant. Therefore, we will disregard the nodal balance equation for one of the nodes, which we refer to as the *reference* or *slack node* and denote it node 0. Let $\mathcal{N}' = \mathcal{N} \setminus \{0\}$ and set the reference phase angle $\theta_0 = 0$. Injection capacity constraints represent generation and consumption limits at nodes. If we denote $k_i^{\min}$ and $k_i^{\max}$ as the capacity limits, then the capacity constraints at node i become $q_i \in [k_i^{\min}, k_i^{\max}]$. Denote $C_i(q)$ as the a convex differentiable cost function for injections at node i. $C_i(q)$ represents an increasing convex generation cost rate for producing $q > 0$ in node i and $-C_i(q)$ is an increasing concave benefit rate for consuming $-q > 0$ in node i. Now that we have defined the electrical power network, we will next consider a market setting for this problem starting with a centralized market and then move to a decentralized market.

An ISO seeks to find feasible nodal injections that minimizes the total system cost. The ISO problem can be formulated as follows:

$$\min_{q_i,\theta_j; i\in\mathcal{N}, j\in\mathcal{N}'} \quad \sum_{i\in\mathcal{N}} C(q_i) \tag{A.1}$$

$$\text{Subject to} \quad q_i = \sum_{j\in\mathcal{N}:(i,j)\in\mathcal{L}} f_{ij}(\theta_{ij})\ \ \forall\ \ i \in \mathcal{N}' \tag{A.2}$$

$$f_{ij}(\theta_{ij}) \leq \bar{f}_{ij}\ \ \forall\ \ (i,j) \in \mathcal{L} \tag{A.3}$$

$$k_i^{\min} \leq q_i \leq k_i^{\max}\ \ \forall\ \ i \in \mathcal{N}. \tag{A.4}$$

This problem is also called the network's *optimal dispatch* problem. The nonlinearity of the constraints makes problem (A.1)-(A.4) difficult to solve because the solution space is in general non-convex. *Chao and Peck* (1996) show that for small θ_{ij} values the flows remain in a locally convex region under normal operating conditions. However, even with convexity assumptions, the nonlinearity of the constraints complicates the problem. The source of the nonlinearity are the G_{ij} variables, which are the loss coefficients. In fact, the total loss in the line connecting nodes i and j is $f_{ij} + f_{ji} = G_{ij}\theta_{ij}^2$. To resolve this issue, losses are usually not directly account for. Instead, it is typical to initially assume that $G_{ij} = 0 \ \forall \ (i, j) \in \mathcal{L}$ so that the system becomes lossless. The losses are then made up by some generator(s); the slack generator is usually used if it is not at its capacity. The discrepancy in the model is then calculated and included in the linear system, which is solved again. This process is repeated until an ε-solution is obtained, where the total discrepancy is no more than a small ε value. For example, *Chao and Peck* (1996) and *Wu and Varaiya* (1999) assume linear systems then extend their market mechanisms for non-linear systems using this approach. In the DC model, we assume that electric networks are lossless and set $G_{ij} = 0 \ \forall \ (i, j) \in \mathcal{L}$.

The zero conductance assumption simplifies the line flow formula to $f_{ij}(\theta_{ij}) = -f_{ji}(\theta_{ji}) = B_{ij}\theta_{ij}$. By substituting this into (A.2) we get

$$q_i = \sum_{j \in \mathcal{N}:(i,j) \in \mathcal{L}} B_{ij}(\theta_i - \theta_j) \ \ \forall \ \ i \in \mathcal{N}'.$$

This equation gives a the linear system $\mathbf{q} = A\theta$, where $\mathbf{q}$ is a column vector of the nodal injections without the reference injection, θ is a column vector of phase angles without the reference angle, and A is an $N \times N$ vector with entries a_{ij} given by

$$a_{ij} = \begin{cases} -B_{ij}; & \text{if } i \neq j, \\ \sum_{k=1}^{n} B_{ik}; & \text{if } i = j. \end{cases}$$

Alternatively, we can calculate $\theta = A^{-1}\mathbf{q}$ and write θ_i as the linear combination $\theta_i = \sum_{j\in\mathcal{N}'} \alpha_{ij} q_j$ where α_{ij} are the entries of the A^{-1} matrix. The flow through line f_{ij} is thus given by

$$f_{ij}(\mathbf{q}) = B_{ij}(\theta_i(\mathbf{q}) - \theta_j(\mathbf{q})) = B_{ij} \sum_{k\in\mathcal{N}'} (\alpha_{ik} - \alpha_{jk}) q_k = \sum_{k\in\mathcal{N}'} \beta_{ij}^k q_k \qquad \forall\ (i,j) \in \mathcal{N},$$

where $\beta_{ij}^k = B_{ij}(\alpha_{ik} - \alpha_{jk})$ are the PTDF. β_{ij}^k can be interpreted as the portion of flow from node k's injection ending up in the reference node that flows through line (i, j). We will use the convention $\beta_{ij}^0 = 0$ for the reference node. We can therefore drop the θ variables from (A.1)-(A.4) for lossless systems and reformulate the problem as

$$\min_{q_i, i\in\mathcal{N}} \quad \sum_{i\in\mathcal{N}} C_i(q_i) \tag{A.5}$$

$$\text{subject to} \quad \sum_{i\in\mathcal{N}} q_i = 0 \qquad \perp \lambda \tag{A.6}$$

$$\sum_{i\in\mathcal{N}} \beta_l^i q_i \leq \bar{f}_l \ \ \forall\ l \in \mathcal{L} \qquad \perp \mu_l \tag{A.7}$$

$$k_i^{\min} \leq q_i \leq k_i^{\max} \ \ \forall\ i \in \mathcal{N} \qquad \perp \gamma_i^-, \gamma_i^+. \tag{A.8}$$

Problem (A.5)-(A.8) is much simpler to solve than (A.1)-(A.4) because it has a convex objective function and linear constraints.

Let λ be the dual variable to the constraint (A.6), μ_l as the dual variable to (A.7), γ_i^- as the dual variable to the lower bound inequality for q_i and γ_i^+ as the dual variable for the upper bound inequality in (A.8). The Lagrangian of this problem is

$$\begin{aligned}
\mathcal{L}(\mathbf{q}, \lambda, \mu, \gamma^-, \gamma^+) = & \sum_{i\in\mathcal{N}} \left[C_i(q_i) - \lambda q_i - \gamma_i^-(q_i - k_i^{\min}) - \gamma_i^+(k_i^{\max} - q_i)\right] \\
& - \sum_{l\in\mathcal{L}} \mu_l \left(\overline{f}_l - \sum_{i\in\mathcal{N}} \beta_l^i q_i\right) \\
= & \sum_{i\in\mathcal{N}} \left[C_i(q_i) - q_i\left(\lambda + \gamma_i^- - \gamma_i^+ - \sum_{l\in\mathcal{L}} \beta_l^i \mu_l\right) + k_i^{\min}\gamma_i^- - k_i^{\max}\gamma_i^+\right] \\
& - \sum_{l\in\mathcal{L}} \mu_l \overline{f}_l.
\end{aligned}$$

The Karush-Kuhn-Tucker (KKT) condition for q_i gives the LMP at every node.

$$p_i = C_i'(q_i) + \gamma_i = \lambda - \sum_{l\in\mathcal{L}} \beta_l^i \mu_l, \tag{A.9}$$

where $\gamma_i = \gamma_i^+ - \gamma_i^-$. A positive μ_l implies that line l is operating at its capacity (in the predefined direction), a $\gamma_i^- > 0$ implies that the nodal injection is at its minimum value[1], and a $\gamma_i^+ > 0$ occurs when a nodal injection reaches its maximum capacity[2]. Notice that the LMP at the slack node is λ.

[1]This occurs when the maximum demand level is reached and additional electricity usage does not improve the consumer surplus.

[2]This happens when generators in node i reach their production rate capacity limits.

APPENDIX B

Chapter II Appendix

B.1 Chapter II Proofs

Proof of Lemma II.3.

Proof. Consider the optimal allocation problem in (2.2), rewritten below

$$C^F(q) = \min\Big\{ \sum_{j\in G^F} C_j(q_j) : q_j \geq 0, \sum_{j\in G^F} q_j = q \Big\}. \tag{B.1}$$

For any $\overline{q} > 0$, the objective in (B.1) is convex on a closed convex set $\big\{(q,\, q_j,\, j \in G^F) : q \in [0, \overline{q}], q_j \in [0, \overline{q}], \sum_{j\in G^F} q_j = q\big\}$. Hence, the theorem on convexity preservation under minimization (*Heyman and Sobel*, 1984, p. 525) implies that $C^F(q)$ is convex in q.

For a given $q > 0$, let $\{q_j^*\}$ be the minimizer for (B.1). We show $\{q_j^*\}$ has two properties:

1) If $q_j^*, q_k^* > 0$, then $C_j'(q_j^*) = C_k'(q_k^*)$. To see this, note that if $C_j'(q_j^*) < C_k'(q_k^*)$, we can strictly reduce the objective by increasing q_j^* by ε and reducing q_k^* by ε, where $\varepsilon > 0$ is small.

2) If $q_j^* = 0$ and $q_k^* > 0$, then $C_j'(0) \geq C_k'(q_k^*)$. To see this, if $C_j'(0) < C_k'(q_k^*)$, we can strictly reduce the objective by setting $q_j^* = \varepsilon$ and reducing q_k^* by ε, where $\varepsilon > 0$ is small.

Denote $p \equiv C_j'(q_j^*)$ for $q_j^* > 0$. Note that $p > 0$ because $C_j(q_j)$ is convex and strictly increasing in q_j for $q_j \geq 0$. Define $G_+^F = \{j \in G^F : C_j'(q_j^*) = p\}$. Then, $C^F(q) = \sum_{j \in G_+^F} C_j(q_j^*)$. For $j \notin G_+^F$, we have $q_j^* = 0$ and $\widehat{C}_j'(0) > p$. Then, for sufficiently small $\varepsilon > 0$, we have

$$C^F(q + \varepsilon) = \sum_{j \in G_+^F} C_j(q_j^* + \varepsilon_j), \tag{B.2}$$

for some $\varepsilon_j \geq 0$ and $\sum_{j \in G_+^F} \varepsilon_j = \varepsilon$. Using Taylor series, (B.2) can be written as

$$C^F(q + \varepsilon) = \sum_{j \in G_+^F} \left[C_j(q_j^*) + \varepsilon_j C_j'(q_j^*) + \mathcal{O}(\varepsilon_j^2)\right] = C^F(q) + \varepsilon p + \mathcal{O}(\varepsilon^2).$$

Similarly, we can show that $C^F(q) - C^F(q - \varepsilon) = \varepsilon p + \mathcal{O}(\varepsilon^2)$. Hence, $C^F(q)$ is differentiable with derivative $(C^F)'(q) = p > 0$.

Similar results can be shown for IGs' problem in (2.1), which completes the proof.

□

Proof of Theorem II.4

Proof. We first prove that (2.11) is optimal in the case of $L_t - q^I - W_t \geq 0$. In this case, constraints (2.8)-(2.9) imply that $q_t^F \geq L_t - q^I - W_t \geq 0$. If we set q_t^F at the lower bound $L_t - q^I - W_t$, then $q_t^V = W_t$ and $e_t = 0$, which clearly minimize the objective in (2.7).

When $L_t - q^I - W_t < 0$, we have $q_t^{F*} = 0$ because: (i) if $q_t^F > 0$ and $e_t > 0$, then a lower q_t^F reduces the objective in (2.7); (ii) if $q_t^F > 0$ and $e_t = 0$, then $q_t^V = L_t - q^I - q_t^F < W_t$, and we can reduce q_t^F and increase q_t^V to lower the objective in (2.7). Hence, $q_t^{F*} = 0$. We determine q_t^{V*} by

$$\min\left\{ -r\, q_t^V + h(q^I + q_t^V - L_t) \;:\; 0 \le q_t^V \le W_t \right\},$$

where we set $h(e) = 0$ for $e < 0$. An interior optimal solution satisfies $h'(q^I + q_t^{V*} - L_t) = r$, or $q_t^{V*} = L_t - q^I + \mu(r)$, which is indeed optimal if $0 < L_t - q^I + \mu(r) < W_t$. If $L_t - q^I + \mu(r) \ge W_t$, then $q_t^{V*} = W_t$. If $L_t - q^I + \mu(r) < 0$, then $q_t^{V*} = 0$. This proves that (2.11) is optimal.

For any $\overline{q} > \overline{L}$, the objective function in (2.7) is convex on a closed convex set $\{(q^I, L_t, W_t, q_t^F, q_t^V) : q^I \in [0, \overline{q}],\ L_t \in [\underline{L}, \overline{L}],\ W_t \in [0, K],\ q_t^F \in [0, \overline{q}],\ (2.8),\ \text{and}\ (2.9)\}$. By the theorem on convexity preservation under minimization (*Heyman and Sobel*, 1984, p. 525), we conclude that $\widetilde{C}(q^I, L_t, W_t)$ is jointly convex in (q^I, L_t, W_t). □

Proof of Theorem II.5

Proof. We first derive an expression for $\mathsf{E}\big[\widetilde{C}(q^I, L_t, W_t)\big]$, which is useful for deriving the first-order condition for (2.15). Using Theorem II.4, we can write

$$\widetilde{C}(q^I, L_t, W_t) = C^F(q_t^{F*}) - r\, q_t^{V*} + h(q^I + q_t^{F*} + q_t^{V*} - L_t),$$

where q_t^{F*} and q_t^{V*} are given in Figure 2.2 under the four events. The indicators of these events can be written as

$$\begin{aligned}
\mathbf{1}_{A_1} &= \mathbf{1}_{\{L_t \le q^I - \mu(r)\}}, \\
\mathbf{1}_{A_2} &= -\mathbf{1}_{\{L_t \le q^I - \mu(r)\}} + \mathbf{1}_{\{L_t < q^I + W_t - \mu(r)\}}, \\
\mathbf{1}_{A_3} &= -\mathbf{1}_{\{L_t < q^I + W_t - \mu(r)\}} + \mathbf{1}_{\{L_t \le q^I + W_t\}}, \\
\mathbf{1}_{A_4} &= \mathbf{1}_{\{L_t > q^I + W_t\}}.
\end{aligned} \tag{B.3}$$

We denote $\widetilde{C}_{A_i}(\cdot,\cdot,\cdot) = \widetilde{C}(\cdot,\cdot,\cdot)$ when A_i occurs. Then, using the optimal policy in Figure 2.2, we have

$$\begin{aligned}
\widetilde{C}_{A_1}(q^I, L_t, W_t) &= h(q^I - L_t), \\
\widetilde{C}_{A_2}(q^I, L_t, W_t) &= -r\,(L_t - q^I + \mu(r)) + h(\mu(r)), \\
\widetilde{C}_{A_3}(q^I, L_t, W_t) &= -r\,W_t + h(q^I + W_t - L_t), \\
\widetilde{C}_{A_4}(q^I, L_t, W_t) &= C^F(L_t - q^I - W_t) - r\,W_t.
\end{aligned} \tag{B.4}$$

Using the equations in (B.3), we can write the expected real-time cost as

$$\begin{aligned}
\mathsf{E}\big[\widetilde{C}(q^I, L_t, W_t)\big] &= \sum_{i=1}^{4} \mathsf{E}\big[\widetilde{C}_{A_i}(q^I, L_t, W_t)\mathbf{1}_{A_i}\big] \\
&= \mathsf{E}\Big[\big(\widetilde{C}_{A_1}(q^I, L_t, W_t) - \widetilde{C}_{A_2}(q^I, L_t, W_t)\big)\mathbf{1}_{\{L_t \le q^I - \mu(r)\}}\Big] \\
&\quad + \mathsf{E}\Big[\big(\widetilde{C}_{A_2}(q^I, L_t, W_t) - \widetilde{C}_{A_3}(q^I, L_t, W_t)\big)\mathbf{1}_{\{L_t < q^I + W_t - \mu(r)\}}\Big] \\
&\quad + \mathsf{E}\Big[\widetilde{C}_{A_3}(q^I, L_t, W_t)\mathbf{1}_{\{L_t \le q^I + W_t\}}\Big] - \mathsf{E}\Big[\widetilde{C}_{A_4}(q^I, L_t, W_t)\mathbf{1}_{\{L_t > q^I + W_t\}}\Big].
\end{aligned} \tag{B.5}$$

It can be verified that $\widetilde{C}(q^I, L_t, W_t)$ is differentiable in q^I except at $q^I = L_t - W_t$, where the left derivative is $-{C^F}'(0)$ and the right derivative is $h'(0)$. Because L_t and W_t have continuous distributions, $\mathsf{E}\big[\widetilde{C}(q^I, L_t, W_t)\big]$ is differentiable in q^I everywhere. Next, we compute its derivative.

The first three expectations in (B.5) all have the form $\mathsf{E}\big[g(q^I, L_t, W_t)\mathbf{1}_{\{L_t \le b(q^I, W_t)\}}\big]$, for some functions $g(q^I, L_t, W_t)$ and $b(q^I, W_t)$. Let the joint probability density function of L_t and W_t be $f_t(l, w)$, $l \in [\underline{L}, \overline{L}], w \in [0, K]$, and let "$\vee$" and "$\wedge$" denote the max and min operations. We have

$$\frac{d}{dq^I}\mathsf{E}\Big[g(q^I,L_t,W_t)\mathbf{1}_{\{L_t\le b(q^I,W_t)\}}\Big] \;=\; \frac{d}{dq^I}\Bigg[\int\limits_0^K \int\limits_{\underline{L}}^{\underline{L}\vee b(q^I,w)\wedge\overline{L}} g(q^I,l,w)f_t(l,w)\,dl\,dw\Bigg]$$

$$\begin{aligned}
&=\mathsf{E}\bigg[\frac{\partial g(q^I,L_t,W_t)}{\partial q^I}\mathbf{1}_{\{L_t\le b(q^I,W_t)\}}\bigg]\\
&\quad+\int\limits_0^K g\big(q^I,b(q^I,w),w\big)\; f_t\big(b(q^I,w),w\big)\frac{\partial b(q^I,w)}{\partial q^I}\mathbf{1}_{\{b(q^I,w)\in[\underline{L},\overline{L}]\}}\,dw.
\end{aligned}\tag{B.6}$$

The last expectation in (B.5) is in the form of $\mathsf{E}\big[g(q^I,L_t,W_t)\mathbf{1}_{\{L_t>b(q^I,W_t)\}}\big]$ and its derivative is

$$\frac{d}{dq^I}\mathsf{E}\Big[g(q^I,L_t,W_t)\mathbf{1}_{\{L_t> b(q^I,W_t)\}}\Big] \;=\; \frac{d}{dq^I}\Bigg[\int\limits_0^K \int\limits_{\underline{L}\vee b(q^I,w)\wedge\overline{L}}^{\overline{L}} g(q^I,l,w)f_t(l,w)\,dl\,dw\Bigg]$$

$$\begin{aligned}
&=\mathsf{E}\bigg[\frac{\partial g(q^I,L_t,W_t)}{\partial q^I}\mathbf{1}_{\{L_t> b(q^I,W_t)\}}\bigg]\\
&\quad-\int\limits_0^K g\big(q^I,b(q^I,w),w\big)\; f_t\big(b(q^I,w),w\big)\frac{\partial b(q^I,w)}{\partial q^I}\mathbf{1}_{\{b(q^I,w)\in[\underline{L},\overline{L}]\}}\,dw.
\end{aligned}\tag{B.7}$$

Now, applying (B.6)-(B.7) to the derivatives of the expectations in (B.5), we find that the integral term $g\big(q^I,b(q^I,w),w\big)$ in (B.6)-(B.7) becomes

$$\begin{aligned}
\big(\widetilde{C}_{A_1}(q^I,l,w)-\widetilde{C}_{A_2}(q^I,l,w)\big)\big|_{l=q^I-\mu(r)} &= 0,\\
\big(\widetilde{C}_{A_2}(q^I,l,w)-\widetilde{C}_{A_3}(q^I,l,w)\big)\big|_{l=q^I+w-\mu(r)} &= 0,\\
\widetilde{C}_{A_3}(q^I,l,w)\big|_{l=q^I+w} &= -r\,w,\\
\widetilde{C}_{A_4}(q^I,l,w)\big|_{l=q^I+w} &= -r\,w,
\end{aligned}$$

where we used the value of $\widetilde{C}_{A_i}(\cdot,\cdot,\cdot)$ given by (B.4). This leads to

$$\begin{aligned}\frac{d}{dq^I}\mathsf{E}\big[\widetilde{C}(q^I,L_t,W_t)\big] &= \sum_{i=1}^{4}\mathsf{E}\Big[\frac{\partial\widetilde{C}_{A_i}(q^I,L_t,W_t)}{\partial q^I}\mathbf{1}_{A_i}\Big]\\ &= \mathsf{E}\Big[-(C^F)'(L_t-W_t-q^I)\mathbf{1}_{A_4}+h'(q^I+W_t-L_t)\mathbf{1}_{A_3}\\ &+r\mathbf{1}_{A_2}+h'(q^I-L_t)\mathbf{1}_{A_1}\Big]\\ &= \mathsf{E}\big[-P(q^I,L_t,W_t,C^F)\big],\end{aligned}$$

where the last equality follows from the definition in (2.13).

Now, consider the optimization problem in (2.15). If $q^{I*}>0$, then it must satisfy

$$\begin{aligned}T(C^I)'(q^{I*})+\int_0^T\frac{d}{dq^I}\mathsf{E}\big[\widetilde{C}(q^{I*},L_t,W_t)\big]dt &= T(C^I)'(q^{I*})-\int_0^T\mathsf{E}\big[P(q^{I*},L_t,W_t,C^F)\big]dt\\ &=0,\end{aligned}$$

which is equivalent to $(C^I)'(q^{I*})=\overline{P}(q^{I*},C^F)$, completing the proof. □

Proof of Theorem II.6

Proof. We can rewrite (4.1) as

$$\begin{aligned}C^{I\prime}(q^I)=&\frac{1}{T}\int_0^T\mathsf{E}\Big[C^{F\prime}(L_t-K\rho_t-q^I)\mathbf{1}_{\{q^I<L_t-K\rho_t\}}-r\mathbf{1}_{\{L_t+\mu(r)-K\rho_t\le q^I<L_t+\mu(r)\}}\\ &-h'(q^I+K\rho_t-L_t)\mathbf{1}_{\{L_t-K\rho_t\le q^I<L_t+\mu(r)-K\rho_t\}}-h'(q^I-L_t)\mathbf{1}_{\{q^I\ge L_t+\mu(r)\}}\Big]\,dt.\end{aligned}\tag{B.8}$$

By differentiating this equation we get

$$
\begin{aligned}
dq^I \Big[& TC^{I''}(q^I) + \int_0^T \mathsf{E}\big[C^{F''}(L_t - K\rho_t - q^I)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
& + h''(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}} + h''(q^I - L_t)\mathbf{1}_{\{q^I \ge L_t + \mu(r)\}}\big]\, dt\Big] \\
+ dW & \int_0^T \mathsf{E}\big[\rho_t\big(C^{F''}(L_t - K\rho_t - q^I)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
& + h''(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}}\big)\big]\, dt \\
+ dr & \int_0^T \mathbb{P}\{L_t + \mu(r) - K\rho_t \le q^I < L_t + \mu(r)\}\, dt = 0.
\end{aligned}
$$

The partial derivative of q^I with respect to r is $\frac{\partial q^I}{\partial r} = -\frac{A_r}{B}$, and with respect to K is $\frac{\partial q^I}{\partial K} = -\frac{A_K}{B}$, where

$$
\begin{aligned}
A_r &= \int_0^T \mathbb{P}\{L_t + \mu(r) - K\rho_t \le q^I < L_t + \mu(r)\}\, dt, \\
A_K &= \int_0^T \mathsf{E}\big[\rho_t\big(C^{F''}(L_t - K\rho_t - q^I)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
&\quad + h''(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}}, \\
B &= TC^{I''}(q^I) + \int_0^T \mathsf{E}\big[C^{F''}(L_t - K\rho_t - q^I)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
&\quad + h''(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}} + h''(q^I - L_t)\mathbf{1}_{\{q^I \ge L_t + \mu(r)\}}\big]\, dt.
\end{aligned}
$$

The fact that both partial derivatives of q^I are non-positive implies that q^I is decreasing in both r and K.

To determine the impact of r and K on C^a we will consider the two cases when

$\mu(r) = 0$ and when $\mu(r) > 0$. The average production cost when $\mu(r) = 0$ is

$$C^a(r,K) = TC^I(q^I) + \int_0^T \mathsf{E}\big[C^F((L_t - q^I - K\rho_t)^+) + h(q^I - L_t)\mathbf{1}_{\{q^I \geq L_t\}} + h(0)\mathbf{1}_{\{L_t - K\rho_t \leq q^I < L_t\}}\big]\,dt.$$

The derivative of C^a with respect to r is

$$\frac{\partial C^a}{\partial r} = \frac{\partial q^I}{\partial r}\Bigg[TC^{I\prime}(q^I) + \int_0^T \mathsf{E}\left[-C^{F\prime}(L_t - q^I - K\rho_t)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} + h'(q^I - L_t)\mathbf{1}_{\{q^I \geq L_t\}}\right]\,dt\Bigg].$$

We can use (B.8) when $\mu(r) = 0$ to get

$$\frac{\partial C^a}{\partial r} = -r\frac{\partial q^I}{\partial r}\int_0^T \mathbb{P}\{q^I < L_t \leq q^I + K\rho_t\}\,dt.$$

This expression is clearly nonnegative, which makes C^a increasing in r. The derivative of C^a with respect to K is

$$\frac{\partial C^a}{\partial K} = -r\frac{\partial q^I}{\partial K}\int_0^T \mathbb{P}\{q^I < L_t \leq q^I + K\rho_t\}\,dt - \int_0^T \mathsf{E}\left[\rho_t C^{F\prime}(L_t - q^I - K\rho_t)\mathbf{1}_{\{q^I < L_t - K\rho_t\}}\right]\,dt.$$

The first term is nonnegative and the second term is non-positive. If $r = 0$ the first term is zero, making C^a a decreasing function of K. This result does not hold for general r as the first term can outweigh the second term for $r > 0$.

The average production cost when $\mu(r) > 0$ is

$$C^a(r,K) = TC^I(q^I) + \int_0^T \mathsf{E}\big[C^F((L_t - q^I - K\rho_t)^+) + h(q^I - L_t)\mathbf{1}_{\{q^I \ge L_t+\mu(r)\}}$$

$$h(\mu(r))\mathbf{1}_{\{L_t - K\rho_t + \mu(r) \le q^I < L_t + \mu(r)\}}$$

$$+ h(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}}\big]\, dt.$$

The derivative of C^a with respect to r is

$$\frac{\partial C^a}{\partial r} = \frac{\partial q^I}{\partial r}\Bigg[TC^{I\prime}(q^I) + \int_0^T \mathsf{E}\big[- C^{F\prime}(L_t - q^I - K\rho_t)\mathbf{1}_{\{q^I < L_t - K\rho_t\}}$$

$$+ h'(q^I - L_t)\mathbf{1}_{\{q^I \ge L_t + \mu(r)\}} + h'(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \le q^I < L_t + \mu(r) - K\rho_t\}}\big]\, dt\Bigg]$$

$$+ \mu'(r)h'(\mu(r)) \int_0^T \mathbb{P}\{q^I - \mu(r) < L_t \le q^I + K\rho_t - \mu(r)\}\, dt.$$

Notice that $\mu'(r) = 1/h''(\mu(r)) > 0$ and $h'(\mu(r)) = r$ for $\mu(r) > 0$. We can also use (B.8) to get

$$\frac{\partial C^a}{\partial r} = \left(\mu'(r)h'(\mu(r)) - r\frac{\partial q^I}{\partial r}\right) \int_0^T \mathbb{P}\{q^I - \mu(r) < L_t \le q^I + K\rho_t - \mu(r)\}\, dt > 0.$$

Therefore, C^a is increasing in r.

Finally, the derivative of C^a with respect to K when $\mu(r) > 0$ is

$$\begin{aligned}
\frac{\partial C^a}{\partial K} = \frac{\partial q^I}{\partial K}\Big[& T{C^I}'(q^I) + \int_0^T \mathsf{E}\big[- {C^F}'(L_t - q^I - K\rho_t)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
& + h'(q^I - L_t)\mathbf{1}_{\{q^I \geq L_t + \mu(r)\}} + h'(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \leq q^I < L_t + \mu(r) - K\rho_t\}}\big]\, dt\Big] \\
& + \int_0^T \mathsf{E}\Big[\rho_t\Big(- {C^F}'(L_t - q^I - K\rho_t)\mathbf{1}_{\{q^I < L_t - K\rho_t\}} \\
& + h'(q^I + K\rho_t - L_t)\mathbf{1}_{\{L_t - K\rho_t \leq q^I < L_t + \mu(r) - K\rho_t\}}\Big)\Big]\, dt \\
= & - r\frac{\partial q^I}{\partial K}\int_0^T \mathbb{P}\{q^I - \mu(r) < L_t \leq q^I + K\rho_t - \mu(r)\}\, dt \\
& + \int_0^T \mathsf{E}\Big[\rho_t\Big(- {C^F}'(L_t - q^I - K\rho_t)\mathbf{1}_{\{L_t > q^I + K\rho_t\}} \\
& + h'(q^I + K\rho_t - L_t)\mathbf{1}_{\{q^I + K\rho_t - \mu(r) < L_t \leq q^I + K\rho_t\}}\Big)\Big]\, dt.
\end{aligned}$$

The first and last terms in this formula are positive while the second term is negative, which implies that C^a in general is not monotone with respect to K. However, if $r = 0$ then the first and last terms becomes 0 and C^a becomes negative. Hence, C^a decreases with K when $r = 0$. □

APPENDIX C

Chapter III Appendix

C.1 Chapter III Proofs

Proof of Lemma III.2

Proof. For a given $q > 0$, let $p \equiv (C^F)'(q)$ and $\{q_j^*\}$ be the optimal FG allocation, and define $G_+^F = \{j \in G^F : \widehat{C}_j'(q_j^*) = p\}$. We want to show that $p = (S^F)^{-1}(q)$. From Lemma II.3, $p = \widehat{C}_j'(q_j^*) = S_j^{-1}(q_j^*)$ or $S_j(p) = q_j^*$ for $j \in G_+^F$. For $j \notin G_+^F$, then from Lemma II.3 we have $\widehat{C}_j'(0) > p$, which implies that $S_j^{-1}(0) = p_j^{\min} > p$ and in turn leads to $S_j(p) = 0 = q_j^*$ due to Assumption III.1(i). Hence, $S^F(p) = \sum_{j \in G^F} S_j(p) = \sum_{j \in G^F} q_j^* = q$, which leads to $p = \left(S^F\right)^{-1}(q)$. Because $S^F(p)$ also satisfies Assumption III.1, $\left(S^F\right)^{-1}(q)$ is continuous in q. Therefore, $(C^F)'(q) = \left(S^F\right)^{-1}(q)$. □

Proof of Corollary III.3

Proof. We will verify that $P(q^I, L_t, W_t, S^F)$ has the alternative expression in (3.8) by considering the disjoint regions A_1-A_4 in Figure 2.2. The inequality in (3.8) is

$$S^F(p) + W_t \mathbf{1}_{\{p \geq -r\}} - \mu(-p) \ \geq \ L_t - q^I. \tag{C.1}$$

In region A_1, $L_t - q^I \le -\mu(r)$, and (C.1) clearly holds if $p = p^o \equiv -h'(q^I - L_t)$. Note that $p^o \le -r$. Thus, for any other price $p_1 < p^o$, the left side of (C.1) becomes $-\mu(-p_1)$, which is strictly less than $L_t - q^I$. Hence, p^o is the minimum price for (C.1) to hold.

In region A_2, $L_t - q^I \in (-\mu(r), W_t - \mu(r))$. If $p = -r$, then (C.1) holds because $W_t - \mu(r) > L_t - q^I$. For any other $p_1 < -r$, (C.1) does not hold because $-\mu(-p_1) < -\mu(r) < L_t - q^I$.

In region A_3, $L_t - q^I \in [W_t - \mu(r), W_t]$. If $p = -h'(q^I + W_t - L_t) \in [-r, 0]$, then (C.1) holds with equality: $W_t - (q^I + W_t - L_t) = L_t - q^I$.

Lastly, in region A_4, $L_t - q^I > W_t$. If $p = (C^F)'(L_t - W_t - q^I) > (C^F)'(0)$, then (C.1) also holds with equality: $(L_t - W_t - q^I) + W_t = L_t - q^I$.

Hence, the minimum price p for (C.1) to hold is exactly $P(q^I, L_t, W_t, S^F)$. $\square$

Proof of Lemma III.4

Proof. Recall that the average price $\overline{P}(q^I, \beta^F)$ decreases in q^I, as we discussed after the definition in (3.7). Furthermore, because L_t and W_t have continuous distributions, $\overline{P}(q^I, \beta^F)$ is differentiable in q^I everywhere. Denote $\overline{P}_1 \equiv \partial\overline{P}/\partial q^I$. We have $\overline{P}_1 \le 0$.

Equation (3.16), $q^I - \beta^I\overline{P}(q^I, \beta^F) = 0$, implicitly determines q^I as a function of β_k, $k \in G^I \cup G^F$. Thus, the lemma's results can be seen from the following partial derivatives:

$$\begin{aligned} \frac{\partial q^I}{\partial \beta_i} &= \frac{\overline{P}}{1 - \beta^I\overline{P}_1} = \frac{q^I}{\beta^I(1 - \beta^I\overline{P}_1)} > 0, \qquad i \in G^I, \\ \frac{\partial q^I}{\partial \beta_j} &= \frac{\beta^I\overline{P}_2}{1 - \beta^I\overline{P}_1} \ < \ 0, \qquad j \in G^F, \end{aligned} \tag{C.2}$$

where $\overline{P}_2 \equiv \partial\overline{P}/\partial\beta^F < 0$ is established below.

We will express $\overline{P}(q^I, \beta^F)$ and derive $\overline{P}_2$. To simplify notations, let random variables L and W follow the probability distribution $f_{L,W}(x, y)$ defined in (3.23). Let

$D = L - W$ denote the net demand.

For a continuous random variable X, we use $f_X(x)$ and $F_X(x)$ to denote the probability density and cumulative distribution functions, and we let $\overline{F}_X(x) = 1 - F_X(x)$.

Then, we can write the average price function in (3.15) as

$$\overline{P}(q^I, \beta^F) = \frac{1}{\beta^F}\mathsf{E}\left[(D - q^I)^+\right] - c_h \int_{q^I-\mu(r)}^{q^I} (q^I - x) f_D(x)\, dx - c_h \int_{-\infty}^{q^I-\mu(r)} (q^I - x) f_L(x)\, dx$$
$$- a_h F_D(q^I) + (a_h - r) F_D(q^I - \mu(r)) + (r - a_h) F_L(q^I - \mu(r)). \qquad \text{(C.3)}$$

Thus,

$$\overline{P}_2 \equiv \frac{\partial \overline{P}}{\partial \beta^F} = -\frac{1}{(\beta^F)^2}\mathsf{E}\left[(D - q^I)^+\right] \;<\; 0,$$
$$\frac{\partial q^I}{\partial \beta_j} = \frac{\beta^I \overline{P}_2}{1 - \beta^I \overline{P}_1} = -\frac{\beta^I \mathsf{E}\left[(D - q^I)^+\right]}{(\beta^F)^2 (1 - \beta^I \overline{P}_1)} \;<\; 0, \qquad j \in G^F. \qquad \text{(C.4)}$$

This completes the proof. □

Proof of Lemma III.5

Proof. We first bound the average price in (C.3). Note that

$$\int_{q^I-\mu(r)}^{q^I} (q^I - x) f_D(x)\, dx \geq 0, \qquad \text{and}$$
$$\int_{-\infty}^{q^I-\mu(r)} (q^I - x) f_L(x)\, dx > \mu(r) F_L(q^I - \mu(r))$$
$$= \frac{(r - a_h)^+}{c_h} F_L(q^I - \mu(r)) \;\geq\; \frac{(r - a_h)}{c_h} F_L(q^I - \mu(r)).$$

Using these inequalities, the average price in (C.3) is bounded above by

$$\overline{P}(q^I, \beta^F) < \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - a_h F_D(q^I) + (a_h - r)F_D(q^I - \mu(r)). \tag{C.5}$$

If $a_h \geq r$, then $\mu(r) = 0$ and (C.5) becomes $\overline{P}(q^I, \beta^F) < \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - rF_D(q^I)$. If $a_h < r$, then (C.5) implies $\overline{P}(q^I, \beta^F) < \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - a_h F_D(q^I)$. Combining these two cases, we obtain

$$\overline{P}(q^I, \beta^F) \; < \; \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - \min\{r, a_h\}F_D(q^I). \tag{C.6}$$

We can express and bound $\mathsf{E}\left[(D-q^I)^+\right]$ as follows:

$$\begin{aligned}
\mathsf{E}\left[(D-q^I)^+\right] &= \int_{q^I}^{\infty} (x - \mu_D + \mu_D - q^I) f_D(x)\, dx \\
&= \int_{q^I}^{\infty} \frac{x-\mu_D}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{(x-\mu_D)^2}{2\sigma_D^2}\right) dx + \int_{q^I}^{\infty} (\mu_D - q^I) f_D(x)\, dx \\
&= \frac{\sigma_D}{\sqrt{2\pi}} \int_{\frac{q^I-\mu_D}{\sigma_D}}^{\infty} y \exp\left(-y^2/2\right) dy + (\mu_D - q^I)\overline{F}_D(q^I) \\
&\leq \frac{\sigma_D}{\sqrt{2\pi}} + (\mu_D - q^I)\overline{F}_D(q^I).
\end{aligned} \tag{C.7}$$

The inequalities (C.6) and (C.7) lead to

$$\overline{P}(q^I, \beta^F) \; < \; \frac{\sigma_D}{\sqrt{2\pi}\beta^F} + \frac{(\mu_D - q^I)\overline{F}_D(q^I)}{\beta^F} - \min\{r, a_h\}F_D(q^I). \tag{C.8}$$

Using (3.16), (3.18), and (C.8), we have

$$\begin{aligned} q^{I\max} &= \beta^{I\max}\overline{P}(q^{I\max}, \beta^{F\min}) \\ &< \beta^{I\max}\left[\frac{\sigma_D}{\sqrt{2\pi}\beta^{F\min}} + \frac{(\mu_D - q^{I\max})\overline{F}_D(q^{I\max})}{\beta^{F\min}} - \min\{r, a_h\}F_D(q^{I\max})\right]. \end{aligned} \tag{C.9}$$

We now prove $q^{I\max} < \mu_D$. If the opposite is true, $q^{I\max} \geq \mu_D$, then $F_D(q^{I\max}) \geq \frac{1}{2}$ and (C.9) implies

$$q^{I\max} < \beta^{I\max}\left[\frac{\sigma_D}{\sqrt{2\pi}\beta^{F\min}} - \frac{\min\{r, a_h\}}{2}\right] \leq \beta^{I\max}\left[\frac{\sigma_D^*}{\sqrt{2\pi}\beta^{F\min}} - \frac{\min\{r, a_h\}}{2}\right] = \mu_D,$$

where $\sigma_D^* \equiv \sqrt{2\pi}\,\beta^{F\min}\left[\frac{\mu_D}{\beta^{I\max}} + \frac{\min\{r, a_h\}}{2}\right]$. This contradicts $q^{I\max} \geq \mu_D$. Therefore, we conclude that $q^{I\max} < \mu_D$ when $\sigma_D \leq \sigma_D^*$. □

Proof of Theorem III.6. Because generator k's pure strategy set is a finite interval $[\beta_k^{\min}, c_k^{-1}]$, it suffices to show that, $\forall\, k \in G^I \cup G^F$, generator k's profit function is quasi-concave with respect to β_k to prove the existence of a pure strategy Nash equilibrium (*Debreu*, 1952).

The proof of the quasi-concavity will use the derivatives of $\overline{P}(q^I, \beta^F)$. Differentiating $\overline{P}(q^I, \beta^F)$ in (C.3) with respect to q^I and using $\mu(r) = (r - a_h)^+/c_h$, we obtain

$$\begin{aligned}\overline{P}_1(q^I,\beta^F) \equiv \frac{\partial \overline{P}}{\partial q^I} &= -\frac{1}{\beta^F}\overline{F}_D(q^I) - c_h\left[F_D(q^I) - F_D(q^I-\mu(r)) + F_L(q^I-\mu(r))\right] \\ &\quad - a_h f_D(q^I) + \left[c_h\mu(r) + (a_h - r)\right]\left[f_D(q^I-\mu(r)) - f_L(q^I-\mu(r))\right] \\ &= -\frac{1}{\beta^F}\overline{F}_D(q^I) - c_h\left[F_D(q^I) - F_D(q^I-\mu(r)) + F_L(q^I-\mu(r))\right] \\ &\quad - a_h f_D(q^I) + (a_h - r)^+\left[f_D(q^I-\mu(r)) - f_L(q^I-\mu(r))\right],\end{aligned} \tag{C.10}$$

$$\begin{aligned}\overline{P}_{11}(q^I,\beta^F) \equiv \frac{\partial^2 \overline{P}}{\partial {q^I}^2} &= \frac{1}{\beta^F} f_D(q^I) - c_h\left[f_D(q^I) - f_D(q^I-\mu(r)) + f_L(q^I-\mu(r))\right] \\ &\quad - a_h f'_D(q^I) + (a_h - r)^+\left[f'_D(q^I-\mu(r)) - f'_L(q^I-\mu(r))\right].\end{aligned} \tag{C.11}$$

By Lemma III.5, if $\sigma_D \leq \sigma_D^*$, we have $q^{I\,\max} < \mu_D$. When $q^I < \mu_D$, and $\sigma_D \to 0$, all the distribution functions in (C.10)-(C.11) approach zero, except for $\overline{F}_D(q^I)$, which approaches one. Therefore, when σ_D is small, $\overline{P}_1$ is close to $-1/\beta^F$ and $\overline{P}_{11}$ is close to zero.

Quasi-concavity of IG's profit function. The profit function of IG $i \in G^I$ is expressed as $\pi_i(\beta_i;\beta_{-i},\beta^F)$ in (3.20). To prove its quasi-concavity in β_i, we will show that its derivative $\partial\pi_i/\partial\beta_i$ can cross the zero value from above at most once as β_i increases, while holding β_{-i} and β^F constant.

In (3.20), the function $Q^I(\beta^I,\beta^F)$ is used to emphasize the dependence of the aggregate IG output q^I on β^I and β^F. In what follows, we use q^I to denote $Q^I(\beta^I,\beta^F)$ when no confusion will rise. Note that $\partial q^I/\partial\beta_i \equiv \partial Q^I/\partial\beta_i$ is given by (C.2). Differentiating (3.20) with respect to β_i, we obtain

$$
\begin{aligned}
\frac{\partial \pi_i}{\partial \beta_i} &= \frac{\beta_i(2-c_i\beta_i)}{(\beta_i+\beta_{-i})^2}\, q^I \frac{\partial q^I}{\partial \beta_i} + \frac{\beta_{-i}(1-c_i\beta_i)-\beta_i}{(\beta_i+\beta_{-i})^3}(q^I)^2 \\
&= \frac{\beta_i(2-c_i\beta_i)}{(\beta^I)^2}\frac{(q^I)^2}{\beta^I(1-\beta^I\overline{P}_1)} + \frac{\beta_{-i}(1-c_i\beta_i)-\beta_i}{(\beta^I)^3}(q^I)^2 \\
&= \frac{(q^I)^2}{(\beta^I)^3(1-\beta^I\overline{P}_1)}\Big[\beta_i(2-c_i\beta_i) + \big(\beta_{-i}(1-c_i\beta_i)-\beta_i\big)(1-\beta^I\overline{P}_1)\Big] \\
&= \frac{(q^I)^2}{(\beta^I)^3(1-\beta^I\overline{P}_1)}\Big[\beta^I(1-c_i\beta_i) - \big(\beta_{-i}(1-c_i\beta_i)-\beta_i\big)\beta^I\overline{P}_1\Big] \\
&= \frac{(q^I)^2}{(\beta^I)^2(1-\beta^I\overline{P}_1)} X(\beta_i;\beta_{-i},\beta^F),
\end{aligned}
$$

where $X(\beta_i;\beta_{-i},\beta^F) \stackrel{\text{def}}{=} 1-c_i\beta_i + \big(\beta_i(1+c_i\beta_{-i})-\beta_{-i}\big)\overline{P}_1$. To show $\partial\pi_i/\partial\beta_i$ can cross zero value from above at most once, it suffices to show X decreases in β_i. Differentiating X with respect to β_i,

$$
\frac{\partial X}{\partial \beta_i} = -c_i + (1+c_i\beta_{-i})\overline{P}_1 + \big(\beta_i(1+c_i\beta_{-i})-\beta_{-i}\big)\overline{P}_{11}\,\frac{q^I}{\beta^I(1-\beta^I\overline{P}_1)},
$$

where $\overline{P}_{11}$ is derived in (C.11). Note that $-c_i+(1+c_i\beta_{-i})\overline{P}_1<0$. Thus, if $\overline{P}_{11}(q^I,\beta^F)$ is sufficiently small, we can establish $\partial X/\partial\beta_i \le 0$. Based on the discussion after (C.10) and (C.11), there exists $\widehat{\sigma}_D$, such that when $\sigma_D < \widehat{\sigma}_D$, we have $\partial X/\partial\beta_i \le 0$ and, therefore, π_i is quasi-concave in β_i.

Quasi-concavity of FG's profit function. Using the probability distribution in (3.23), and denote $D = L-W$ and $q^I = Q^I(\beta^I,\beta^F)$, we can write FG's profit function in (3.22) as

$$
\pi_j(\beta_j;\beta_{-j},\beta^I) = \frac{\beta_j\big(1-\frac{1}{2}c_j\beta_j\big)}{(\beta_j+\beta_{-j})^2}\mathsf{E}\left[\big((D-q^I)^+\big)^2\right].
$$

We will show that $\partial\pi_j/\partial\beta_j$ can cross the zero value at most once from above when β_j increases.

Differentiating π_j with respect to β_j and using $\partial q^I/\partial\beta_j$ from (C.4) and the following fact

$$\frac{\partial}{\partial q^I}\mathsf{E}\left[\left((D-q^I)^+\right)^2\right] = \frac{\partial}{\partial q^I}\int_{q^I}^{\infty}(x-q^I)^2 f_D(x)dx = \int_{q^I}^{\infty}-2(x-q^I)f_D(x)dx$$
$$= -2\mathsf{E}\left[(D-q^I)^+\right],$$

we obtain

$$\frac{\partial \pi_j}{\partial \beta_j} = \frac{\beta_{-j}(1-c_j\beta_j)-\beta_j}{(\beta^F)^3}\mathsf{E}\left[\left((D-q^I)^+\right)^2\right] + \frac{\beta_j(2-c_j\beta_j)}{(\beta^F)^2}\mathsf{E}\left[(D-q^I)^+\right]\frac{\beta^I\mathsf{E}\left[(D-q^I)^+\right]}{(\beta^F)^2(1-\beta^I\overline{P}_1)}$$
$$= \frac{\beta_j\mathsf{E}\left[((D-q^I)^+)^2\right]}{(\beta^F)^3(1-\beta^I\overline{P}_1)}\Big[Y(\beta_j,\beta_{-j},\beta^I)+\beta^I Z(\beta_j,\beta_{-j},\beta^I)\Big],$$

where

$$Y(\beta_j,\beta_{-j},\beta^I) \stackrel{\text{def}}{=} \Big(\frac{\beta_{-j}}{\beta_j}-(1+c_j\beta_{-j})\Big)(1-\beta^I\overline{P}_1),$$
$$Z(\beta_j,\beta_{-j},\beta^I) \stackrel{\text{def}}{=} \frac{2-c_j\beta_j}{\beta_j+\beta_{-j}}\psi(q^I),$$
$$\psi(q^I) \stackrel{\text{def}}{=} \frac{\left(\mathsf{E}\left[(D-q^I)^+\right]\right)^2}{\mathsf{E}\left[((D-q^I)^+)^2\right]}.$$

It suffices to show that Y and Z decrease in β_j. Differentiating Y with respect to β_j,

$$\frac{\partial Y}{\partial \beta_j} = -\frac{\beta_{-j}}{\beta_j^2}(1-\beta^I\overline{P}_1) + \Big(\frac{\beta_{-j}}{\beta_j}-(1+c_j\beta_{-j})\Big)\beta^I\overline{P}_{11}\frac{\beta^I\mathsf{E}\left[(D-q^I)^+\right]}{(\beta^F)^2(1-\beta^I\overline{P}_1)}.$$

By the same argument used for the quasi-concavity of π_i, we see that when σ_D is sufficiently small, $\overline{P}_1$ is close to $-1/\beta^F$ and $\overline{P}_{11}$ is close to zero. Thus, there exists $\widetilde{\sigma}_D$, such that when $\sigma_D < \widetilde{\sigma}_D$, we have $\partial Y/\partial\beta_j \le 0$.

Next, we show that Z decreases in β_j. Note that $\frac{\partial}{\partial q^I}\mathsf{E}\left[(D-q^I)^+\right] = -\overline{F}_D(q^I)$

and

$$\begin{aligned}\psi'(q^I) &= -\frac{2\mathsf{E}\big[(D-q^I)^+\big]\overline{F}_D(q^I)}{\mathsf{E}\big[\big((D-q^I)^+\big)^2\big]} + \frac{2\big(\mathsf{E}\big[(D-q^I)^+\big]\big)^3}{\big(\mathsf{E}\big[\big((D-q^I)^+\big)^2\big]\big)^2}\\ &= -\frac{2\psi(q^I)(\overline{F}_D(q^I)-\psi(q^I))}{\mathsf{E}[(D-q^I)^+]}.\end{aligned}$$

Using this derivative and $\partial q^I/\partial\beta_j$ in (C.4), we have

$$\begin{aligned}\frac{\partial Z}{\partial\beta_j} &= \frac{-c_j\beta^F-(2-c_j\beta_j)}{(\beta^F)^2}\,\psi(q^I) + \frac{2-c_j\beta_j}{\beta^F}\,\psi'(q^I)\frac{\partial q^I}{\partial\beta_j}\\ &= \frac{-c_j\beta^F-(2-c_j\beta_j)}{(\beta^F)^2}\,\psi(q^I) + \frac{2-c_j\beta_j}{\beta^F}\,\frac{2\psi(q^I)(\overline{F}_D(q^I)-\psi(q^I))\beta^I}{(\beta^F)^2\,(1-\beta^I\overline{P}_1)}\\ &= \frac{\psi(q^I)}{(\beta^F)^2}\left[-c_j\beta^F-(2-c_j\beta_j)\left(1-\frac{2(\overline{F}_D(q^I)-\psi(q^I))\beta^I}{\beta^F(1-\beta^I\overline{P}_1)}\right)\right].\end{aligned}$$

We will show that $\psi(q^I)$ is close to $\overline{F}_D(q^I)$ when σ_D is sufficiently small and $q^I<\mu_D$ to complete the proof.

For a normal random variable $X\sim\mathcal{N}(\mu,\sigma)$, we can show that $\mathsf{E}[X^+]=\mu\overline{F}_X(0)+\sigma^2 f_X(0)$ and $\mathsf{E}\big[(X^+)^2\big]=(\mu^2+\sigma^2)\overline{F}_X(0)+\mu\sigma^2 f_X(0)$. Then,

$$\begin{aligned}\mathsf{E}\left[(D-q^I)^+\right] &= (\mu_D-q^I)\overline{F}_D(q^I)+\sigma_D^2 f_D(q^I),\\ \mathsf{E}\left[((D-q^I)^+)^2\right] &= ((\mu_D-q^I)^2+\sigma_D^2)\overline{F}_D(q^I)+\sigma_D^2(\mu_D-q^I)f_D(q^I),\\ \psi(q^I) &= \frac{\big[(\mu_D-q^I)\overline{F}_D(q^I)+\sigma_D^2 f_D(q^I)\big]^2}{\big[(\mu_D-q^I)^2+\sigma_D^2\big]\overline{F}_D(q^I)+\sigma_D^2(\mu_D-q^I)f_D(q^I)}.\end{aligned}$$

If $\sigma_D\le\sigma_D^*$, we have $q^{I\max}<\mu_D$ (Lemma III.5). The above expression for $\psi(q^I)$ implies that as $\sigma_D\to 0$, we have $\overline{F}_D(q^I)\to 1$, $f_D(q^I)\to 0$, and $\psi(q^I)\to 1$. Hence, there exists $\sigma_D^\dagger$, such that when $\sigma_D<\sigma_D^\dagger$, we have $\partial Z/\partial\beta_j\le 0$.

To summarize, when $\sigma_D<\min\{\sigma_D^*,\widehat{\sigma}_D,\widetilde{\sigma}_D,\sigma_D^\dagger\}$, the profit function π_j is quasi-concave in β_j. This establishes the existence of a pure strategy equilibrium, i.e., the linear supply function equilibrium. □

C.2 SFE ODE Derivation for IG, FG, and VG model

We derive in this appendix the ODE that characterize the SFE for the problem with the three generator types IG, FG, and VG. We first present the problem IGs and FGs face and then find their equilibrium ODEs. The FG ODEs are found using two different methods. The first method uses optimal control in which the optimal supply function trajectory of every agent is chosen to satisfies the Euler-Lagrange condition. The second method is the price control approach, which is widely used in the SFE literature. Generators in the price approach method use the residual demand as their supply functions and choose a price trajectory that maximizes their profit. We show that these methods give the same SFE ODEs. We will assume that all IGs and FGs have initial marginal costs $C_i'(0) = a^I$ and $C_j'(0) = a^F$). We also make the generator rationality assume to ensure that the IG and FG supply functions produce nothing when the price is below their initial marginal cost (i.e. $S_i(p) = 0$ for $p < a^I$ and $S_j(p) = 0$ for $p < a^F$).

To study their bidding strategies, we find the best response of every generator, assuming all other generators' supply functions are given. As a convention, we define

$$S_{-i}(p) \stackrel{\text{def}}{=} S^I(p) - S_i(p), \qquad \text{and} \qquad S_{-j}(p) \stackrel{\text{def}}{=} S^F(p) - S_j(p).$$

C.2.1 IG Bidding Strategy

Knowing all FGs' supply functions, generator $i \in G^I$ can compute the real-time price $P(q^I, L_t, W_t, S^F)$ in (3.6) and derive the average market price as a function of aggregate IG production q^I:

$$\overline{P}(q^I) \stackrel{\text{def}}{=} \frac{1}{T} \int_0^T \mathsf{E}\left[P(q^I, L_t, W_t, S^F)\right] dt.$$

Note that $\overline{P}(q^I)$ decreases in q^I. That is, the more IGs produce, the lower the average real-time price, which also equals the price that IGs are paid because of the market-clearing condition in (3.10). Hence, $\overline{P}(q^I)$ is essentially an inverse demand function facing IGs, and we can use the classical approach in *Klemperer and Meyer* (1989) to find the equilibrium condition.

IG i can equivalently use the market-clearing price as a decision variable. If the market-clearing price is p, then the total IG output is $q^I = \overline{P}^{-1}(p)$, and other IGs supply $S_{-i}(p)$. Thus, IG i's problem becomes:

$$\max_p \ \big(\overline{P}^{-1}(p) - S_{-i}(p)\big)p - C_i\big(\overline{P}^{-1}(p) - S_{-i}(p)\big).$$

The first order condition of this problem is

$$\overline{P}^{-1}(p) - S_{-i}(p) + \left(\overline{P}^{-1\prime}(p) - S'_{-i}(p)\right)\left[p - C'_i\big(\overline{P}^{-1}(p) - S_{-i}(p)\big)\right] = 0.$$

Substituting $\overline{P}^{-1}(p) - S_{-i}(p)$ by $S_i(p)$, and noting that $\overline{P}^{-1\prime}(p) = \big(\overline{P}'(q^I)\big)^{-1}$, the above condition can be written as

$$S'_{-i}(p) = \frac{S_i(p)}{p - C'_i(S_i(p))} + \frac{1}{\overline{P}'(S^I(p))}, \qquad \forall\, i \in G^I. \tag{C.12}$$

This is the classical equilibrium condition derived by *Klemperer and Meyer* (1989). The difference is that the demand function in our setting is implied by FGs' supply functions. Hence, the last term in (C.12) captures how FGs' decisions affect IGs' supply functions.

C.2.2 FG Bidding Strategy

An individual FG $j \in G^F$ offers supply function $S_j(p)$, knowing that $S_j(p)$ has two effects on the outcomes of the game. The first effect is that $S_j(p)$ directly influences real-time prices $P(q, L_t, W_t, S^F)$ in (3.8). This effect captures the fact that FG j directly competes with other FGs in satisfying the balance of the demand. This first effect also changes the average price on the right hand side of (3.10), which affects the market-clearing IG output rate q^{I*} in (3.10). This second effect captures the fact that FGs and IGs are also competing with each other. These two effects are intertwined because the IG output q^{I*} influences real-time price through (3.8) as well.

Recognizing the above effects, FG j decides its supply function $S_j(p)$. We first consider FG j choosing among all feasible supply functions that support a given q^I as the market-clearing IG output. Once a desired supply function is found to support each q^I, FG j will then optimize over q^I.

For notational convenience, we define *net demand* as load minus VG capacity, $D_t \stackrel{\text{def}}{=} L_t - W_t$. Let the cumulative distribution function of D_t be F_{D_t}, and its complement be $\overline{F}_{D_t} = 1 - F_{D_t}$. Whenever the net demand $D_t > q^I$, FGs must produce exactly the remaining demand, $D_t - q^I > 0$, and the real-time price p_t must satisfy

$$S^F(p_t) = D_t - q^I > 0.$$

The above equation suggests that if FG j desires a market price $p_t > 0$ when the net demand is D_t, it must produce $D_t - q^I - S_{-j}(p_t)$ at price p_t, which corresponds to a point on its supply function. Hence, we can equivalently use p_t as FG j's decision variable. FG j decides price $p_t > 0$ for each level of net demand $D_t > q^I$, under the constraint that the prices should support q^I as the market-clearing IG output. That is, the time-average real-time price should match ${S^I}^{-1}(q^I)$ as per (3.10):

$$T{S^I}^{-1}(q^I) = \int_0^T \mathsf{E}\big[p_t\big]dt. \tag{C.13}$$

Note that FG j makes price decision only when $D_t > q^I$. When $D_t \leq q^I$, the real-time price is zero or negative, which is independent of any supply function, for given q^I. The expected negative price averaged over $[0,T]$ is

$$\begin{aligned} P^-(q^I) &\stackrel{\text{def}}{=} \int_0^T \mathsf{E}\big[\min\big\{0,\ P(q^I, L_t, W_t)\big\}\big]dt \\ &= \int_0^T \mathsf{E}\big[-h'(q^I + W_t - L_t)\mathbf{1}_{A_3} - r\mathbf{1}_{A_2} - h'(q^I - L_t)\mathbf{1}_{A_1}\big]dt. \end{aligned} \tag{C.14}$$

Then, the constraint in (C.13) can be written as

$$\int_0^T \mathsf{E}\big[p_t \big| D_t > q^I\big]\overline{F}_{D_t}(q^I)dt \;=\; TS^{I^{-1}}(q^I) - P^-(q^I) \;\equiv\; g(q^I). \tag{C.15}$$

Hence, FG j's problem is choose $p_t > 0$ for each level of $D_t > q^I$ so that the expected positive price equals $g(q^I)$. This equality ensures that q^I is indeed the market-clearing IG output rate. Taking one step further, FG j chooses q^I and p_t jointly as long as they satisfy the constraint (C.15). Hence, FG j's problem can be formulated as:

$$\max_{p_t,\, q^I} \int_0^T \mathsf{E}\Big[\big(D_t - q^I - S_{-j}(p_t)\big)p_t - C_j\big((D_t - q^I - S_{-j}(p_t)\big)\Big|D_t > q^I\Big]\overline{F}_{D_t}(q^I)\,dt \tag{C.16}$$

$$\text{s.t.} \int_0^T \mathsf{E}\big[p_t \big| D_t > q^I\big]\overline{F}_{D_t}(q^I)dt \;=\; g(q^I),$$

where p_t is implicitly a function of D_t. An FG can find its optimal decision by satisfying the first order conditions of (C.16), and the system's equilibrium conditions can be attained by solving the system of ODEs given by (C.12) and the optimality conditions for each FG from (C.16).

The Lagrangian of the problem in (C.16) is (η_j is the Lagrangian multiplier):

$$\int_0^T \mathsf{E}\big[\big(D_t - q^I - S_{-j}(p_t) - \eta_j\big)p_t - C_j\big(D_t - q^I - S_{-j}(p_t)\big)\big|D_t > q^I\big]\overline{F}_{D_t}(q^I)\,dt + \eta_j\, g(q^I).$$

The first-order condition for p_t is

$$\big(D_t - q^I - S_{-j}(p_t) - \eta_j\big) - S'_{-j}(p_t)\big[p_t - C'_j(D_t - q^I - S_{-j}(p_t))\big] = 0.$$

Substituting $D_t - q^I - S_{-j}(p_t)$ by $S_i(p_t)$, the above condition can be written as

$$S'_{-j}(p) = \frac{S_j(p) - \eta_j}{p - C'_j(S_j(p))}, \qquad \forall\ j \in G^F. \tag{C.17}$$

We can derive η_j from the first-order condition for q^I to get

$$\eta_j = \frac{\int_{a^F}^{\infty}(xS_j(x) - C_j(S_j(x)))S^{F\prime}(x)\psi(S^F(x), q^I)\,dx}{g'(q^I) + \int_{a^F}^{\infty} xS^{F\prime}(x)\psi(S^F(x), q^I)\,dx} \qquad \forall\ j \in G^F, \tag{C.18}$$

where

$$\psi(S^F(x), q^I) = \int_0^T \big[f_{D_t}(S^F(x) + q^I)f_{D_t}(q^I) - f'_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\big]\,dt.$$

Two derivation of (C.18) are shown next; using the optimal control approach and using the price control approach.

Optimal Control Approach

We can use the relation $S^F(p_t) + q^I = D_t$ to get p_t's distribution $F_{p_t}(x) = F_{D_t}(S^F(x) + q^I)$ when $p_t > a^F$. This makes $f_{p_t}(x) = S^{F\prime}(x)f_{D_t}(S^F(x) + q^I)$ for $p_t > a^F$ and $D_t > q^I$. Let b bet the price ceiling that corresponds to the maximum demand. The probability that $p_t > a^F$ is $\mathbb{P}\{D_t > q^I\} = \overline{F}_{D_t}(q^I)$. j's best response problem becomes

$$\begin{aligned}
&\max_{S_j, q^I} \int_{a^F}^{b} (xS_j(x) - C_j(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dx \\
&\text{s.to} \;\; g(q^I) = \int_{a^F}^{b} xS^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dx \qquad \perp \eta_j.
\end{aligned}$$

The Lagrangian of the optimal control problem is

$$\begin{aligned}
\mathcal{L} =& (xS_j(x) - C_j(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt \\
&+ \eta_j \left(\frac{g(q^I)}{b - a^F} - xS^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt \right) \\
=& (x(S_j(x) - \eta_j) - C_j(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt \\
&+ \frac{\eta_j}{b - a^F} g(q^I).
\end{aligned}$$

The partial derivative of the Lagrangian with respect to S_j is

$$\begin{aligned}
\mathcal{L}_{S_j} =& (x - C_j'(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt \\
&+ (x(S_j(x) - \eta_j) - C_j(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}'(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt,
\end{aligned}$$

and the partial derivative of the Lagrangian with respect to S_j' is

$$\mathcal{L}_{S_j'} = (x(S_j(x) - \eta_j) - C_j(S_j(x))) \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt.$$

The derivative of this expression with respect to x is

$$\frac{d}{dx}\mathcal{L}_{S_j'} = \left[S_j(x) - \eta_j + S_j'(x)(x - C_j'(S_j(x)))\right] \int_0^T f_{D_t}(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt$$
$$+ (x(S_j(x) - \eta_j) - C_j(S_j(x)))S^{F\prime}(x) \int_0^T f_{D_t}'(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\,dt.$$

The Euler-Lagrange optimality condition $\mathcal{L}_{S_j} = \frac{d}{dx}\mathcal{L}_{S_j'}$ gives the differential equation

$$S_{-j}'(x) = \frac{S_j(x) - \eta_j}{x - C_j'(S_j(x))}.$$

η_j can be calculated from the first order condition of q^I as

$$\eta_j = \frac{\int_{a^F}^{b}(xS_j(x) - C_j(S_j(x)))S^{F\prime}(x)\psi(S^F(x), q^I)\,dx}{g'(q^I) + \int_{a^F}^{b} xS^{F\prime}(x)\psi(S^F(x), q^I)\,dx},$$

where

$$\psi(S^F(x), q^I) = \int_0^T \left[f_{D_t}(S^F(x) + q^I)f_{D_t}(q^I) - f_{D_t}'(S^F(x) + q^I)\overline{F}_{D_t}(q^I)\right] dt.$$

Price Control Approach

In this approach, generator j sees a residual demand of $D_t - q^I - S_{-j}(p_t)$ in time t, provided that $D_t > q^I$, and chooses p_t to maximize his gain. If we denote by $\overline{D}$ the maximum demand, then generator j's best response problem becomes

$$\max_{p_j, q^I} \int_0^T \mathsf{E}\left[p_t(D_t - q^I - S_{-j}(p_t)) - C_j(D_t - q^I - S_{-j}(p_t))|D_t > q^I\right]\overline{F}_{D_t}(q^I)\,dt$$
$$\text{s.to} \quad g(q^I) = \int_0^T \mathsf{E}\left[p_t|D_t > q^I\right]\overline{F}_{D_t}(q^I)\,dt \qquad \perp \eta_j$$

where the expectation is taken over the stochastic process D_t and p_t is implicitly a function of D_t. The Lagrangian of this problem is

$$\int_0^T \mathsf{E}\big[p_t(D_t - q^I - S_{-j}(p_t) - \eta_j) - C_j(D_t - q^I - S_{-j}(p_t))|D_t > q^I\big]\overline{F}_{D_t}(q^I)\,dt$$
$$+\eta_j g(q^I).$$

The first order condition for p_t is

$$D_t - q^I - S_{-j}(p_t) - \eta_j - S'_{-j}(p_t)\big(p_t - C'_j(D_t - q^I - S_{-j}(p_t))\big) = 0$$

or

$$S'_{-j}(p) = \frac{S_j(p) - \eta_j}{p - C'_j(S_j(p))}.$$

We can derive η_j from the first order condition for q^I. The Lagrangian can be written in integral form as

$$\int_{q^I}^{\infty} \big[p(y - q^I - S_{-j}(p) - \eta_j) - C_j(y - q^I - S_{-j}(p))\big] \int_0^T f_{D_t}(y)\overline{F}_{D_t}(q^I)\,dt\,dy$$
$$+\eta_j g(q^I)$$

where p is an implicit function of y. The first order condition with respect to q^I is

$$0 = (a^F\eta_j + C_j(0)) \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt + \eta_j g'(q^I)$$
$$- \int_{q^I}^{\infty} \big(p - C_j'(y - q^I - S_{-j}(p))\big) \int_0^T f_{D_t}(y)\overline{F}_{D_t}(q^I)\,dt\,dy$$
$$- \int_{q^I}^{\infty} \big[p(y - q^I - S_{-j}(p) - \eta_j) - C_j(y - q^I - S_{-j}(p))\big] \int_0^T f_{D_t}(y) f_{D_t}(q^I)\,dt\,dy.$$

We used the relations $p = a^F$ when $y = q^I$ and $S_{-j}(a^F) = 0$ in deriving this equation. We can rearrange terms to get

$$\eta_j\Bigg(g'(q^I) + a^F \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt + \int_{q^I}^{\infty} p \int_0^T f_{D_t}(y) f_{D_t}(q^I)\,dt\,dy\Bigg)$$
$$= \int_{q^I}^{\infty} \big[p(y - q^I - S_{-j}(p)) - C_j(y - q^I - S_{-j}(p))\big] \int_0^T f_{D_t}(y) f_{D_t}(q^I)\,dt\,dy$$
$$+ \int_{q^I}^{\infty} \big(p - C_j'(y - q^I - S_{-j}(p))\big) \int_0^T f_{D_t}(y)\overline{F}_{D_t}(q^I)\,dt\,dy$$
$$- C_j(0) \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt.$$

We can now make a change of variables from y to p, where $y = q^I + S^F(p)$ which makes $dy = S^{F\prime}(p)\,dp$.

$$\eta_j\left(g'(q^I) + a^F \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt + \int_{a^F}^{\infty} pS^{F\prime}(p) \int_0^T f_{D_t}(S^F(p)+q^I)f_{D_t}(q^I)\,dt\,dp\right) \tag{C.19}$$

$$\begin{aligned}
&= \int_{a^F}^{\infty} \big[pS_j(p) - C_j(S_j(p))\big] S^{F\prime}(p) \int_0^T f_{D_t}(S^F(p)+q^I)f_{D_t}(q^I)\,dt\,dp \\
&+ \int_{a^F}^{\infty} \big(p - C_j'(S_j(p))\big) S^{F\prime}(p) \int_0^T f_{D_t}(S^F(p)+q^I)\overline{F}_{D_t}(q^I)\,dt\,dp \\
&- C_j(0) \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt.
\end{aligned}$$

Consider the expression

$$\int_{a^F}^{\infty} \big[pS_j(p) - C_j(S_j(p))\big] S^{F\prime}(p) \int_0^T f'_{D_t}(S^F(p)+q^I)\overline{F}_{D_t}(q^I)\,dt\,dp.$$

We can integrate it by parts to get

$$\begin{aligned}
& C_j(0) \int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt \\
& - \int_{a^F}^{\infty} \big(S_j(p) + S_j'(p)(p - C_j'(S_j(p)))\big) \int_0^T f_{D_t}(S^F(p)+q^I)\overline{F}_{D_t}(q^I)\,dt\,dp.
\end{aligned}$$

By substituting $S_j'(p) = S^{F\prime}(p) - S_{-j}'(p)$ we can write

$$S_j(p) + S_j'(p)(p - C_j'(S_j(p))) = S_j(p) - S_{-j}'(p)(p - C_j'(S_j(p))) + S^{F\prime}(p)(p - C_j'(S_j(p)))$$
$$= \eta_j + (p - C_j'(S_j(p)))S^{F\prime}(p).$$

By substituting this into the original formula we get

$$\int_{a^F}^{\infty} \big[pS_j(p) - C_j(S_j(p))\big] S^{F\prime}(p) \int_0^T f'_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp$$
$$= C_j(0)\int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt - \eta_j \int_{a^F}^{\infty}\int_0^T f_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp$$
$$- \int_{a^F}^{\infty} (p - C'_j(S_j(p)))S^{F\prime}(p)\int_0^T f_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp.$$

We can substitute this back into (C.19) to get

$$\eta_j\Bigg(g'(q^I) + a^F\int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt$$
$$+ \int_{a^F}^{\infty} pS^{F\prime}(p)\int_0^T f_{D_t}(S^F(p) + q^I)f_{D_t}(q^I)\,dt\,dp + \int_{a^F}^{\infty}\int_0^T f_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp\Bigg)$$
$$= \int_{a^F}^{\infty}\big[pS_j(p) - C_j(S_j(p))\big]S^{F\prime}(p)\psi(S^F(p), q^I)\,dt\,dp.$$

We can write

$$\int_{a^F}^{\infty}\int_0^T f_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp = -a^F\int_0^T f_{D_t}(q^I)\overline{F}_{D_t}(q^I)\,dt$$
$$- \int_{a^F}^{\infty} pS^{F\prime}(p)\int_0^T f'_{D_t}(S^F(p) + q^I)\overline{F}_{D_t}(q^I)\,dt\,dp,$$

which makes

$$\eta_j\left(g'(q^I) + \int\limits_{a^F}^{\infty} pS^{F\prime}(p)\psi(S^F(p), q^I)\,dt\,dp\right)$$
$$= \int\limits_{a^F}^{\infty} \big[pS_j(p) - C_j(S_j(p))\big]\, S^{F\prime}(p)\psi(S^F(p), q^I)\,dt\,dp.$$

APPENDIX D

Chapter IV Appendix

D.1 Solving Linear SFEs

We solve Green's (1996) affine SFE problem using Rudkevich (1999). In this problem, the demand is assumed to be price elastic with the form $D(p) = \theta - bp$, where θ is a random demand shock and b is the price elasticity of the demand. Generators have quadratic cost functions of the form $C_k(q) = a_k + \frac{1}{2}c_k q^2$ and submit linear supply functions of the form $S_k(p) = \beta_k(p - \alpha_k)$. Rudkevich gives a procedure to solves for the β_k coefficients in non-linear system

$$\sum_{i \neq k} \beta_i = \frac{\beta_k}{1 - c_k \beta_k} - \gamma.$$

In this procedure, we first find $U \in (0, 1)$ that satisfies

$$U = 1 + n\frac{U + \varepsilon}{2} - \sum_{k=1}^{n} \sqrt{s_k^2 + \frac{(U + \varepsilon)^2}{4}},$$

where U is a market power index (market is competitive when U is close to 1 and non-competitive when U is close to 0), n is the number of generators, $c_M = \frac{1}{\sum_{k=1}^{n} c_k}$,

$\varepsilon = bc_M$, and $s_k = \frac{c_M}{c_k}$. The β_k bid of generator k can then be calculated from

$$\beta_k = \frac{1}{c_M}\left(s_k + \frac{U+\varepsilon}{2} - \sqrt{s_k^2 + \frac{(U+\varepsilon)^2}{4}}\right). \tag{D.1}$$

The Matlab function `fsolve` with a starting point of 1 can be used to solve for U.

D.2 Chapter IV Proofs

Proof of Lemma IV.2

Proof. The β_k bids can be calculated using Equation (D.1) in Appendix D.1, where $\varepsilon = bc_M$. The derivative of β_k with respect to ε is

$$\frac{\partial \beta_k}{\partial \varepsilon} = \frac{1}{c_M}\left(\frac{U'+1}{2} - \frac{\frac{1}{2}(U+\varepsilon)(U'+1)}{2\sqrt{s_k^2 + \frac{(U+\varepsilon)^2}{4}}}\right),$$

where $U' = \frac{\partial U}{\partial \varepsilon}$. Since $s_k > 0 \;\forall\; k \in G$, the derivative can be bounded as follows

$$\begin{aligned}\frac{\partial \beta_k}{\partial \varepsilon} >& \frac{1}{c_M}\left(\frac{U'+1}{2} - \frac{\frac{1}{2}(U+\varepsilon)(U'+1)}{2\sqrt{\frac{(U+\varepsilon)^2}{4}}}\right)\\ =& \frac{1}{2c_M}(U'+1)\left(1 - \frac{U+\varepsilon}{U+\varepsilon}\right) = 0.\end{aligned}$$

Therefore, β_k is increasing in $b \;\forall\; k \in G$. □

Proof of Theorem IV.3

Proof. Using the bid solution function B defined in (4.19), we can find the slope bids for the AF problem β_k^{AF} for $k \in G$ by solving $B(G, 0)$. In other words, the set of generator bids β_k^{AF} solve

$$\sum_{l \in G \setminus k} \beta_l^{AF} = \frac{\beta_k^{AF}}{1 - c_k \beta_k^{AF}} \qquad \forall k \in G.$$

We can rewrite this system as

$$\sum_{l \in G^F \setminus j} \beta_l^{AF} + (\beta^I)^{AF} = \frac{\beta_j^{AF}}{1 - c_j \beta_j^{AF}} \qquad \forall\ j \in G^F,$$
$$\sum_{l \in G^I \setminus i} \beta_l^{AF} + (\beta^F)^{AF} = \frac{\beta_i^{AF}}{1 - c_i \beta_j^{AF}} \qquad \forall\ i \in G^I.$$

Therefore, bids β_j^{AF} for $j \in G^F$ are the solution to $B(G^F, (\beta^I)^{AF})$ and the bids β_i^{AF} for $i \in G^I$ are the solution to $B(G^I, (\beta^F)^{AF})$.

For the FI problem, we can get β_j^{FI} for $j \in G^F$ by solving $B(G^F, 0)$. Since $(\beta^I)^{AF} > 0$, we can use Lemma IV.2 to concluded that $\beta_j^{AF} > \beta_j^{FI}$ for $j \in G^F$. This also implies that $(\beta^F)^{AF} > (\beta^F)^{FI}$.

After finding $(\beta^I)^{FI}$, we can calculate β_i^{FI} for $i \in G^I$ by solving $B(G^I, (\beta^F)^{FI})$. Since $(\beta^F)^{AF} > (\beta^F)^{FI}$, we can use Lemma IV.2 again to show that $\beta_i^{AF} > \beta_i^{FI}$ for $i \in G^I$. This shows that $\beta_k^{AF} > \beta_k^{FI}$ for all generators $k \in G$.

□

APPENDIX E

Chapter V Appendix

E.1 Chapter V Proofs

Proof of Theorem V.1

We will use the following distributions to derive the different term in (5.16).

$$
\begin{aligned}
&\mathbb{P}\{B_{(j-1)} \le b_i < B_{(j)}\} = \frac{(n-1)!}{(j-1)!(n-j)!} F_B(b_i)^{j-1} \overline{F}_B(b_i)^{n-j}, \\
&\mathbb{P}\{B_{(j)} \le x\} = F_{B_{(j)}}(x) = \sum_{l=j}^{n-1} \frac{(n-1)!}{(l)!(n-l-1)!} F_B(x)^l \overline{F}_B(x)^{n-l-1}, \\
&\mathbb{P}\{B_{(j-1)} > b_i\} = \sum_{l=1}^{j-1} \frac{(n-1)!}{(l-1)!(n-l)!} F_B(b_i)^{l-1} \overline{F}_B(b_i)^{n-l}, \\
&f_{B_{(j)}}(x) = \frac{(n-1)!}{(j-1)!(n-j-1)!} f_B(x) F_B(x)^{j-1} \overline{F}_B(x)^{n-j-1}, \\
&\mathsf{E}\left[B_{(j-1)} \mathbf{1}_{\{B_{(j-1)} > b_i\}}\right] = \int_{b_i}^{\overline{p}} x f_{B_{(j-1)}}(x)\, dx \\
&\qquad = \frac{(n-1)!}{(j-2)!(n-j)!} \int_{b_i}^{\overline{p}} f_B(x) F_B(x)^{j-2} \overline{F}_B(x)^{n-j}\, dx.
\end{aligned}
$$

We can write (5.16) as $\mathsf{E}[\pi_i(b_i)] = A_1 + A_2 - A_3$, where

$$\begin{aligned} A_1 &= \sum_{j=1}^{n} PD_j \mathbb{P}\{B_{(j-1)} \leq b_i < B_{(j)}\} ED_j (b_i - c), \\ A_2 &= \sum_{j=1}^{n} PD_j k \mathsf{E}[B_{(j-1)} \mathbf{1}_{\{B_{(j-1)} > b_i\}}], \\ A_3 &= \sum_{j=1}^{n} PD_j \mathbb{P}\{B_{(j-1)} > b_i\} ck. \end{aligned}$$

By using the order statistic distribution we get

$$A_1 = \sum_{j=1}^{n} PD_j ED_j \frac{(n-1)!}{(j-1)!(n-j)!} F_B(b_i)^{j-1} \overline{F}_B(b_i)^{n-j} (b_i - c).$$

Notice that $\mathbb{P}\{B_{(0)} > b_i\} = 0$, and hence we could start the sum in A_2's formula from $j = 2$. We can also use the conditional expectation of the order statistic to get

$$\begin{aligned} A_2 =& k \sum_{j=2}^{n} PD_j \frac{(n-1)!}{(j-2)!(n-j)!} \int_{b_i}^{\overline{p}} x f_B(x) F_B(x)^{j-2} \overline{F}_B(x)^{n-j} \, dx \\ =& k \sum_{j=1}^{n-1} PD_{j+1} \frac{(n-1)!}{(j-1)!(n-j-1)!} \int_{b_i}^{\overline{p}} x f_B(x) F_B(x)^{j-1} \overline{F}_B(x)^{n-j-1} \, dx. \end{aligned}$$

A_3 can also be expanded by using the order statistic distribution.

$$\begin{aligned} A_3 =& ck \sum_{j=2}^{n} PD_j \sum_{l=1}^{j-1} F(b_i)^{l-1} \overline{F}(b_i)^{n-l} \frac{(n-1)!}{(l-1)!(n-l)!} \\ =& ck \sum_{l=1}^{n-1} \sum_{j=l+1}^{n} PD_j F(b_i)^{l-1} \overline{F}(b_i)^{n-l} \frac{(n-1)!}{(l-1)!(n-l)!} \\ =& ck \sum_{j=1}^{n-1} \overline{PD}_j \frac{(n-1)!}{(j-1)!(n-j)!} F(b_i)^{j-1} \overline{F}(b_i)^{n-j}, \end{aligned}$$

where $\overline{PD}_j = \sum_{l=j+1}^{n} PD_l$. In a MSNE, generators other than i mix their strategies such that i becomes indifferent in choosing b_i, which is the same as having $\frac{\partial \mathsf{E}[\pi_i(b_i)]}{\partial b_i} = 0$.

We will find the derivative of A_1, A_2, and A_3 with respect to b_i next and then combine these terms to get the indifference condition.

$$
\begin{aligned}
\frac{\partial A_1}{\partial b_i} =& \sum_{j=1}^{n} PD_j ED_j \frac{(n-1)!}{(j-1)!(n-j)!} F(b_i)^{j-1} \overline{F}(b_i)^{n-j} \\
&+ f(b_i)(b_i - c) \sum_{j=2}^{n} PD_j ED_j \frac{(n-1)!}{(j-1)!(n-j)!} (j-1) F(b_i)^{j-2} \overline{F}(b_i)^{n-j} \\
&- f(b_i)(b_i - c) \sum_{j=1}^{n-1} PD_j ED_j \frac{(n-1)!}{(j-1)!(n-j)!} (n-j) F(b_i)^{j-1} \overline{F}(b_i)^{n-j-1} \\
=& \sum_{j=1}^{n} PED_j \frac{(n-1)!}{(j-1)!(n-j)!} F(b_i)^{j-1} \overline{F}(b_i)^{n-j} \\
&+ f(b_i)(b_i - c) \sum_{j=1}^{n-1} \Delta PED_j \frac{(n-1)!}{(j-1)!(n-j-1)!} F(b_i)^{j-1} \overline{F}(b_i)^{n-j-1}
\end{aligned}
$$

where $PED_j = PD_j ED_j$ and $\Delta PED_j = PD_{j+1} ED_{j+1} - PD_j ED_j$.

$$
\frac{\partial A_2}{\partial b_i} = - k b_i f(b_i) \sum_{j=1}^{n-1} PD_{j+1} \frac{(n-1)!}{(j-1)!(n-j-1)!} F(b_i)^{j-1} \overline{F}(b_i)^{n-j-1}.
$$

$$
\begin{aligned}
\frac{\partial A_3}{\partial b_i} =& ckf(b_i)\sum_{j=2}^{n-1}\overline{PD}_j\frac{(n-1)!}{(j-1)!(n-j)!}(j-1)F(b_i)^{j-2}\overline{F}(b_i)^{n-j} \\
&- ckf(b_i)\sum_{j=1}^{n-1}\overline{PD}_j\frac{(n-1)!}{(j-1)!(n-j)!}(n-j)F(b_i)^{j-1}\overline{F}(b_i)^{n-j-1} \\
=& - ckf(b_i)(n-1)\overline{PD}_{n-1}F(b_i)^{n-2} \\
&+ ckf(b_i)\sum_{j=1}^{n-2}(\overline{PD}_{j+1}-\overline{PD}_j)\frac{(n-1)!}{(j-1)!(n-j-1)!}F(b_i)^{j-1}\overline{F}(b_i)^{n-j-1} \\
=& - ckf(b_i)(n-1)PD_nF(b_i)^{n-2} \\
&- ckf(b_i)\sum_{j=1}^{n-2}PD_{j+1}\frac{(n-1)!}{(j-1)!(n-j-1)!}F(b_i)^{j-1}\overline{F}(b_i)^{n-j-1} \\
=& - ckf(b_i)\sum_{j=1}^{n-1}PD_{j+1}\frac{(n-1)!}{(j-1)!(n-j-1)!}F(b_i)^{j-1}\overline{F}(b_i)^{n-j-1}.
\end{aligned}
$$

We can combine these results to get

$$
\begin{aligned}
\frac{\partial \mathsf{E}[\pi_i(x)]}{\partial x} =& \sum_{j=1}^{n}PED_j\frac{(n-1)!}{(j-1)!(n-j)!}F(x)^{j-1}\overline{F}(x)^{n-j} \\
&+ (x-c)f(x)\sum_{j=1}^{n-1}\Delta PED_j\frac{(n-1)!}{(j-1)!(n-j-1)!}F(x)^{j-1}\overline{F}(x)^{n-j-1} \\
&- kxf(x)\sum_{j=1}^{n-1}PD_{j+1}\frac{(n-1)!}{(j-1)!(n-j-1)!}F(x)^{j-1}\overline{F}(x)^{n-j-1} \\
&+ ckf(x)\sum_{j=1}^{n-1}PD_{j+1}\frac{(n-1)!}{(j-1)!(n-j-1)!}F(x)^{j-1}\overline{F}(x)^{n-j-1} = 0.
\end{aligned}
$$

We can further simplify this expression to get the following ODE.

$$
\begin{aligned}
&\sum_{j=1}^{n}PED_j\frac{(n-1)!}{(j-1)!(n-j)!}F(x)^{j-1}\overline{F}(x)^{n-j} \\
&+ (x-c)f_B(x)\sum_{j=1}^{n-1}(\Delta PED_j - kPD_{j+1})\frac{(n-1)!}{(j-1)!(n-j-1)!}F(x)^{j-1}\overline{F}(x)^{n-j-1} = 0.
\end{aligned}
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Albadi, M., and E. El-Saadany (2007), Demand response in electricity markets: An overview, in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–5, IEEE.

Alguacil, N., A. L. Motto, and A. J. Conejo (2003), Transmission expansion planning: a mixed-integer lp approach, *Power Systems, IEEE Transactions on*, *18*(3), 1070–1077.

Anderson, E. J. (2013), On the existence of supply function equilibria, *Mathematical Programming*, *140*(2), 323–349.

Anderson, E. J., and X. Hu (2008), Finding supply function equilibria with asymmetric firms, *Operations Research*, *56*(3), 697–711.

Anderson, E. J., and A. B. Philpott (2002), Using supply functions for offering generation into an electricity market, *Operations Research*, *50*(3), 477–489.

Baldick, R., R. Grant, and E. Kahn (2004), Theory and application of linear supply function equilibrium in electricity markets, *Journal of Regulatory Economics*, *25*(2), 143–167.

Bentek Energy (2010), How less became more: Wind, power and unintended consequences in the Colorado energy market, available at http://www.bentekenergy.com/WindCoalandGasStudy.aspx.

Berry, C. A., B. F. Hobbs, W. A. Meroney, R. P. O'Neill, and W. R. Stewart Jr (1999), Understanding how market power can arise in network competition: a game theoretic approach, *Utilities Policy*, *8*(3), 139–158.

Binato, S., M. V. F. Pereira, and S. Granville (2001), A new benders decomposition approach to solve power transmission network design problems, *Power Systems, IEEE Transactions on*, *16*(2), 235–240.

Bolle, F. (1992), Supply function equilibria and the danger of tacit collusion: The case of spot markets for electricity, *Energy Economics*, *14*(2), 94–102.

Borenstein, S., J. Bushnell, and S. Stoft (2000), The competitive effects of transmission capacity in a deregulated electricity industry, *RAND Journal of Economics*, *31*(2), 294–325.

Brunekreeft, G. (2001), A multiple-unit, multiple-period auction in the british electricity spot market, *Energy economics*, *23*(1), 99–118.

Buygi, M. O., H. Zareipour, and W. D. Rosehart (2012), Impacts of large-scale integration of intermittent resources on electricity markets: A supply function equilibrium approach, *IEEE Systems Journal*, *6*(2), 220–232.

Chao, H.-P., and S. Peck (1996), A market mechanism for electric power transmission, *Journal of Regulatory Economics*, *10*, 25–59.

Conejo, A. J., J. Contreras, D. A. Lima, and A. Padilha-Feltrin (2007), Zbus transmission network cost allocation, *IEEE Trans. Power Syst*, *22*(1), 342–349.

Contreras, J., and F. F. Wu (1999), Coalition formation in transmission expansion planning, *Power Systems, IEEE Transactions on*, *14*(3), 1144–1152.

Contreras, J., and F. F. Wu (2000), A kernel-oriented algorithm for transmission expansion planning, *Power Systems, IEEE Transactions on*, *15*(4), 1434–1440.

Contreras, J., G. Gross, J. M. Arroyo, and J. I. Muñoz (2009), An incentive-based mechanism for transmission asset investment, *Decision Support Systems*, *47*(1), 22–31.

Crawford, G. S., J. Crespo, and H. Tauchen (2007), Bidding asymmetries in multi-unit auctions: implications of bid function equilibria in the british spot market for electricity, *International Journal of Industrial Organization*, *25*(6), 1233–1268.

de la Torre, S., A. J. Conejo, and J. Contreras (2008), Transmission expansion planning in electricity markets, *Power Systems, IEEE Transactions on*, *23*(1), 238–248.

Debreu, G. (1952), A social equilibrium existence theorem, *Proceedings of the National Academy of Sciences*, *38*, 886–893.

Denholm, P., E. Ela, B. Kirby, and M. Milligan (2010), The role of energy storage with renewable electricity generation, *Tech. rep.*, NREL/TP-6A2-47187, National Renewable Energy Laboratory.

EIA (2010), Electric power annual, *U.S. Department of Energy.*

Ela, E. (2009), Using economics to determine the efficient curtailment of wind energy, national Renewable Energy Laboratory, NREL/TP-550-45071.

Ela, E., and D. Edelson (2012), Participation of wind power in LMP-based energy markets, *IEEE Transactions on Sustainable Energy*, *3*(4), 777–783.

Ela, E., M. Milligan, B. Parsons, D. Lew, and D. Corbus (2009), The evolution of wind power integration studies: Past, present, and future, *Proceedings of Power & Energy Society General Meeting, 2009.*

Ela, E., B. Kirby, N. Navid, and J. Smith (2011), Effective ancillary services market designs on high wind power penetration systems.

Electric Reliability Council of Texas (ERCOT) (2012), ERCOT Business Practice: Setting the shadow price caps and power balance penalties in security constrained economic dispatch, version 4.0, available at http://ercot.com/calendar/2012/06/20120620-TAC.

EnerNex (2011), Eastern wind integration and transmission study, national Renewable Energy Laboratory, NREL/SR-5500-47078. Available at http://www.nrel.gov/wind/publications.html.

Erlich, I., and M. Wilch (2010), Primary frequency control by wind turbines, in *Power and Energy Society General Meeting, 2010 IEEE*, pp. 1–8, IEEE.

Fabra, N., N.-H. von der Fehr, and D. Harbord (2006), Designing electricity auctions, *RAND Journal of Economics*, *37*(1), 23–46.

Federico, G., and D. Rahman (2003), Bidding in an electricity pay-as-bid auction, *Journal of Regulatory Economics*, *24*(2), 175–211.

Fink, S., C. Mudd, K. Porter, and B. Morgenstern (2009), Wind energy curtailment case studies, *NREL subcontract report, NREL/SR-550-46716*.

Fischer, C., and R. G. Newell (2008), Environmental and technology policies for climate mitigation, *Journal of environmental economics and management*, *55*(2), 142–162.

Freris, L., and A. Sasson (1968), Investigation of the load-flow problem, in *Proceedings of the Institution of Electrical Engineers*, vol. 115, pp. 1459–1470, IET.

Garcıa-Dıaz, A., and P. L. Marın (2003), Strategic bidding in electricity pools with short-lived bids: an application to the spanish market, *International Journal of Industrial Organization*, *21*(2), 201–222.

GE Energy (2010), Western wind and solar integration study, national Renewable Energy Laboratory, NREL/SR-550-47434. Available at http://www.nrel.gov/wind/publications.html.

Genc, T. S., and S. S. Reynolds (2011), Supply function equilibria with capacity constraints and pivotal suppliers, *International Journal of Industrial Organization*, *29*(4), 432–442.

Glover, J. D., M. S. Sarma, and T. J. Overbye (2011), *Power system analysis and design*, Cengage Learning.

Green, R. (1996), Increasing competition in the British electricity spot market, *Journal of Industrial Economics*, *44*(2), 205–216.

Green, R. J., and D. M. Newbery (1992), Competition in the British electricity spot market, *Journal of political economy*, *100*(5), 929–953.

Hand, M., S. Baldwin, E. DeMeo, J. Reilly, T. Mai, D. Arent, G. Porro, M. Meshek, and D. Sandor (2012), Renewable electricity futures study, *NREL Paper*.

Hart, E. K., E. D. Stoutenburg, and M. Z. Jacobson (2012), The potential of intermittent renewables to meet electric power demand: Current methods and emerging analytical techniques, *Proceedings of the IEEE*, *100*(2), 322–334.

Hedman, K. W., R. P. O'Neill, E. B. Fisher, and S. S. Oren (2011), Smart flexible just-in-time transmission and flowgate bidding, *IEEE Transactions on Power Systems*, *26*(1), 93–102.

Heyman, D. P., and M. J. Sobel (1984), *Stochastic Models in Operations Research*, vol. 2, McGraw-Hill, New York.

Hobbs, B. F., C. B. Metzler, and J.-S. Pang (2000), Strategic gaming analysis for electric power systems: An MPEC approach, *IEEE Transactions on Power Systems*, *15*(2), 638–645.

Hogan, W. (2002), Financial transmission right formulations, *Report, Center for Business and Government, John F. Kennedy School of Government, Harvard University, Cambridge, MA*.

Hogan, W. (2011), Transmission benefits and cost allocation.

Hogan, W., J. Rosellón, and I. Vogelsang (2010), Toward a combined merchant-regulatory mechanism for electricity transmission expansion, *Journal of Regulatory Economics*, *38*(2), 113–143.

Hogan, W. W. (1992), Contract networks for electric power transmission, *Journal of Regulatory Economics*, *4*(3), 211–242.

Hogan, W. W. (2008), Electricity market structure and infrastructure, *Acting in Time on Energy Policy, Harvard University*.

Holmberg, P. (2007), Supply function equilibrium with asymmetric capacities and constant marginal costs, *Energy Journal*, *28*(2), 55–82.

Holmberg, P. (2008), Unique supply function equilibrium with capacity constraints, *Energy Economics*, *30*(1), 148–172.

Holmberg, P., and D. Newbery (2010), The supply function equilibrium and its policy implications for wholesale electricity auctions, *Utilities Policy*, *18*(4), 209–226.

Holmberg, P., and A. B. Philpott (2012), Supply function equilibria in networks with transport constraints, working Paper, Research Institute of Industrial Economics.

Joskow, P., and J. Tirole (2005), Merchant transmission investment*, *The Journal of industrial economics*, *53*(2), 233–264.

Joskow, P. L., and J. Tirole (2000), Transmission rights and market power on electric power networks, *The Rand Journal of Economics*, pp. 450–487.

Kahn, A. E., P. C. Cramton, R. H. Porter, , A. E. Kahn, P. C. Cramton, R. H. Porter, and R. D. Tabors (2001), Uniform pricing or pay-as-bid pricing: A dilemma for california and beyond, *The Electricity Journal*, *14*(6), 70–79.

Katzenstein, W., and J. Apt (2009), Air emissions due to wind and solar power, *Environmental Science and Technology*, *43*(2), 253–258.

Kirby, B., M. Milligan, and E. Ela (2010), Providing minute-to-minute regulation from wind plants, *Proceedings of the 9th Annual Large-Scale Integration of Wind Power into Power Systems and Transmission Networks for Offshore Wind Power Plant*.

Klemperer, P. D., and M. A. Meyer (1989), Supply function equilibria in oligopoly under uncertainty, *Econometrica*, *57*(6), 1243–1277.

Kristiansen, T., and J. Rosellón (2013), A merchant mechanism for electricity transmission expansion, in *Financial Transmission Rights*, pp. 157–189, Springer.

Latorre, G., R. D. Cruz, J. M. Areiza, and A. Villegas (2003), Classification of publications and models on transmission expansion planning, *Power Systems, IEEE Transactions on*, *18*(2), 938–946.

Li, G., J. Shi, and X. Qu (2011), Modeling methods for GenCo bidding strategy optimization in the liberalized electricity spot market–A state-of-the-art review, *Energy*, *36*(8), 4686–4700.

Liang, J., S. Grijalva, and R. G. Harley (2011), Increased wind revenue and system security by trading wind power in energy and regulation reserve markets, *Sustainable Energy, IEEE Transactions*, *2.3*, 340–347.

Littlechild, S. C., and C. J. Skerk (2008), Transmission expansion in argentina 1: the origins of policy, *Energy Economics*, *30*(4), 1367–1384.

Menanteau, P., D. Finon, and M. Lamy (2003), Prices versus quantities: choosing policies for promoting the development of renewable energy, *Energy policy*, *31*(8), 799–812.

Midcontinent Independent System Operator (MISO) (2013), Business Practices Manual: Energy and Operating Reserve Markets, https://www.misoenergy.org/Library/BusinessPracticesManuals.

Miller, N., K. Clark, and M. Shao (2011), Frequency responsive wind plant controls: Impacts on grid performance, in *Power and Energy Society General Meeting, 2011 IEEE*, pp. 1–8, IEEE.

Neuhoff, K. (2005), Large-scale deployment of renewables for electricity generation, *Oxford Review of Economic Policy*, *21*(1), 88–110.

Oren, S. S. (1997), Economic inefficiency of passive transmission rights in congested electricity systems with competitive generation, *ENERGY JOURNAL-CAMBRIDGE MA THEN CLEVELAND OH-*, *18*, 63–84.

Palmer, K., and D. Burtraw (2005), Cost-effectiveness of renewable electricity policies, *Energy Economics*, *27*(6), 873–894.

Rassenti, S. J., V. L. Smith, and B. J. Wilson (2003), Discriminatory price auctions in electricity markets: Low volatility at the expense of high price levels, *Journal of Regulatory Economics*, *23*(2), 109–123.

REN21 (2013), *Renewables Global Futures Report*, paris, France.

Renewable Energy Policy Network for the 21st Century (REN21) (2013), *Renewables Global Status Report*, paris, France.

Rivier Abbad, J. (2010), Electricity market participation of wind farms: the success story of the spanish pragmatism, *Energy policy*, *38*(7), 3174–3179.

Rudkevich, A. (1999), Supply function equilibrium in power markets: learning all the way, tabors Caramanis & Associates technical paper 1299-1702.

Rudkevich, A., M. Duckworth, and R. Rosen (1998), Modeling electricity pricing in a deregulated generation industry: The potential for oligopoly pricing in a poolco, *Energy Journal*, *19*(3), 19–48.

Sauma, E. E., and S. S. Oren (2009), Do generation firms in restructured electricity markets have incentives to support social-welfare-improving transmission investments?, *Energy Economics*, *31*(5), 676–689.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988), *Spot Pricing of Electricity*, Kluwer Academic Publishers.

Shrestha, G., and P. Fonseka (2004), Congestion-driven transmission expansion in competitive power markets, *Power Systems, IEEE Transactions on*, *19*(3), 1658–1665.

Sioshansi, R. (2011), Increasing the value of wind with energy storage, *Energy Journal*, *32*(2), 1–29.

Sioshansi, R., and S. Oren (2007), How good are supply function equilibrium models: an empirical analysis of the ercot balancing market, *Journal of Regulatory Economics*, *31*(1), 1–35.

Smith, J. C., M. Milligan, E. A. DeMeo, and B. Parsons (2007), Utility wind integration and operating impact state of the art, *IEEE Transactions on Power Systems*, *22*(3), 900–908.

Supatgiat, C., R. Q. Zhang, and J. R. Birge (2001), Equilibrium values in a competitive power exchange market, *Computational Economics*, *17*(1), 93–121.

Ventosa, M., Á. Baíllo, A. Ramos, and M. Rivier (2005), Electricity market modeling trends, *Energy Policy*, *33*(7), 897–913.

Vogelsang, I. (2001), Price regulation for independent transmission companies, *Journal of Regulatory Economics*, *20*(2), 141–165.

von der Fehr, N.-H. M., and D. Harbor (1993), Spot market competition in the uk electricity inustry, *The Economic Journal*, *103*, 531–546.

Wilson, R. (2008), Supply function equilibrium in a constrained transmission system, *Operations Research*, *56*(2), 369–382.

Wu, F. F., and P. Varaiya (1999), Coordinated multilateral trades for electric power networks: theory and implementation, *Electrical Power and Energy Systems*, *21*, 75–102.

Wu, O. Q., and R. Kapuscinski (2013), Curtailing intermittent generation in electrical systems, *Manufacturing & Service Operations Management*, articles in Advance.