MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

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

Majid Al-Gwaiz
Department of Industrial and Operations Engineering
University of Michigan

Xiuli Chao
Department of Industrial and Operations Engineering
University of Michigan

Owen Q. Wu
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1218
April 2014

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

Majid Al-Gwaiz*, Xiuli Chao*, and Owen Q. Wu**

*Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan

**Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan

April 3, 2014

Abstract

In this paper 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 classic supply function equilibrium model overestimates 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, reduces price volatility, and improves the system's overall efficiency. Moreover, we show that these benefits are most significant in the absence of the production-based subsidies.

Key words: electricity market; supply function equilibrium; flexible/inflexible generators; variable generation; economic curtailment; production-based subsidies

1. Introduction

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 changing 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 seminal studies and the following stream of research provide important economic insights and policy recommendations, which we will review in §2.

Most SFE models for electricity markets assume that all firms have the flexibility to adjust their power output at different prices. This assumption can be justified in two situations. First, each firm owns a portfolio of power generators and offers the aggregate supply as a function of the market price. The portfolio consists of inflexible generators (e.g., nuclear and some coal-fired generators) as well as flexible generators (e.g., oil- and gas-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). The theoretical framework of SFE is applicable to both situations, as discussed in Anderson and Philpott (2002) and Holmberg and Newbery (2010).

The assumption of production-adjustment flexibility, however, may not always be appropriate. As industry deregulation continues, firms downsize their portfolios by selling off part of their generation assets and independent power producers emerge to participate in the power markets. As a result, in many current markets, firms that own mainly inflexible generators cannot change their power output in a short time, whereas firms owning mostly flexible generators can quickly adjust their output. 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.

The classic SFE model does not address competitions involving inflexible firms, but intuitively the level of flexibility directly affects a firm's production, revenue, and its competitive behavior.

Thus, the first two natural questions we ask are: *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 classic SFE model may over- or under-estimate the intensity of the market competition. The understanding of the effect of generation flexibility/inflexibility on competition is especially important given the rapid evolution of generation mix, with coal-fired generation shifting toward more flexible generation fueled by natural gas.

An important part of the evolution is the increasing use of renewable energy sources, notably solar and wind power, which are often referred to as variable generation. 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, which raises another question: *How does variable generation impact the competition between flexible and inflexible firms?*

The impact of variable generation depends on the system's priority dispatching rule. 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). Consequently, many system operators started to develop market mechanisms for economic curtailment. 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.

Our objective in this paper is to address the four questions raised above through theoretical and computational analysis of a stylized model that captures the most relevant tradeoffs. We assume that each firm owns either inflexible generators (IG) or fully flexible generators (FG) or variable generators (VG). An IG firm produces power at a constant rate over an operating horizon (e.g., several hours to one day); an FG firm can adjust its output to meet the demand fluctuations. IG and FG firms submit supply functions to a system operator. VG output is considered as negative demand when VG has priority in dispatch; when economic curtailment is allowed, VG firms are assumed to be price-takers and submit their marginal cost determined by the production-based subsidies. We formulate the system operator's optimal dispatch problem and derive the market clearing condition. We then characterize and compute the SFE between FG and IG firms with linear supply functions, commonly adopted both in practice and in the research literature. We further study the impact of the amount of variable generation (referred to as VG penetration), its dispatch policy, and the subsidies.

The main insights from this paper are summarized below. First, by assuming all firms are flexible, the classic SFE model overestimates the intensity of the supply function competition. In our equilibrium model, because IGs do not compete with FGs in matching production with uncertain demand, FGs face less competition and offer significantly lower output than predicted by the classic SFE model. FGs' less competitive behavior induces IGs (who still compete with all other generators for market share) to offer slightly lower output than in the classic model. Consequently, our model leads to a higher average price and a higher price volatility than predicted by the classic SFE model.

Second, when the rising VG penetration increases the overall variability facing the system, if VGs have priority, the system operator has to balance the increased variability by using FGs rather than IGs, which allows FGs to have an advantage in the market-share competition with IGs. To profit from this advantage, FGs reduce their supply functions, i.e., offer lower output at each price. Thus, as VG penetration increases, the market becomes less competitive.

Third, economic curtailment of VG provides the system operator with an additional lever to balance against variability and serves as a partial substitute for FGs. Thus, economic curtailment intensifies the market competition: IGs and FGs offer more competitive supply functions than if VGs have priority; IGs' supply functions may be even more competitive than in the classic 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 economic curtailment, but emissions may increase or decrease depending on the generators' fuel types.

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

2. Literature Review

In their original work, Klemperer and Meyer (1989) show the existence of a family of SFE for competing firms with identical cost functions and without capacity constraints. They characterize 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 our paper.

Many studies focus on the case of symmetric equilibria, in which firms offer identical supply functions. Green and Newbery (1992) calibrate the SFE model for the British electricity industry and 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.

When firms differ in costs, the general asymmetric equilibria are difficult to find and, thus, linear supply functions are often used to simplify the analysis. Green (1996) solves the asymmetric equilibrium with linear supply functions 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. In this paper, we also analyze SFE with linear supply functions and extend the above studies by considering asymmetries in both cost and flexibility.

Physical constraints such as capacities and network transmission constraints are important areas in the SFE literature. SFE models with capacity constraints are considered by Green and Newbery (1992), Baldick et al. (2004), Holmberg (2007), Anderson and Hu (2008), Genc and Reynolds (2011) and Anderson (2013). SFE models with network transmission constraints are studied by Berry et al. (1999), Wilson (2008), and Holmberg and Philpott (2012). This paper complements these studies by focusing on firms' production-adjustment constraints. A key feature of our model is that the firms' (in)flexibility is incorporated in the system operator's optimal dispatch problem, and the resulting optimality condition serves as a constraint in the firm-level profit-maximization problem. Our approach shares similar features with the MPEC (mathematical program with equilibrium constraints) approach introduced by Hobbs, Metzler, and Pang (2000).

Constraints may also rise from market rules. Supatgiat, Zhang, and Birge (2001) study the Nash equilibria when the price bids are restricted to a discrete set and each firm offers a single price-quantity pair. They characterize the firms' equilibrium behavior and market clearing price.

Several empirical studies have been conducted to compare the SFE prediction with actual market data. Sioshansi and Oren (2007) find evidence that generators in the Texas electricity market bid less competitively than predicted by the SFE model. Willems et al. (2009) find similar evidence in the German electricity market and include constant correction terms in their model. The first insight obtained in this paper, mentioned in the introduction, is consistent with these empirical findings.

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). 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) considers a Stackelberg game with wind-power generators deciding output followed by a supply function competition among conventional generators. Assuming wind-power generators are price-takers and have priority, Buygi, Zareipour, and Rosehart (2012) analyze a 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 a lower market price. In this paper 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. 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 paper complements these works by studying the impact of economic curtailment on market competition. We find an additional benefit of economic curtailment—economic curtailment intensifies market competition.

3. The System Model

This section lays the foundation for our subsequent analysis. We first describe the generators' problem in the electricity system and then formulate the system operator's problem. Our model is a combination of Nash game and Stackelberg game: The generators bid simultaneously in a Nash game, and the system operator follows with an optimal dispatch decision. The generators' competition is based on the response of the system operator, and hence they are also the Stackelberg leaders.

3.1 Generator Types and Costs

The length of the operating horizon is denoted as T, which in practice can be several hours to one day. 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 paper.

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 in the system operator's problem prior to $t = 0$ and stays constant over $[0, T]$. Let $\widehat{C}_i(q_i)$ 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 $\widehat{C}_j(q_{jt})$ denote its operating cost rate at time t. For the purpose of the analysis in this paper, 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 $t \in [0, T]$. VGs may adjust their actual output below W_t, known as curtailment, for the reasons described below in §3.3. Curtailment can be achieved 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 1 *(i) For any generator $k \in G^I \cup G^F$, the cost rate function $\widehat{C}_k(q)$ is convex, strictly increasing, and continuously differentiable in q, and $\widehat{C}_k(0) = 0$; (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) approximates the reality well. Part (ii) states that VGs receive production-based subsidy and implies that the marginal cost of VGs is $-r$. Part (iii) means that FGs are fully flexible in adjusting their output levels and VG curtailment involves negligible operating cost.

Let $L_t \in [\underline{L}, \overline{L}]$ denote the price-insensitive load at time $t \in [0, T]$, where $0 \leq \underline{L} < \overline{L}$. 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 two sources of uncertainties in our model and they can be correlated.

3.2 Supply Offers and Stated Costs

Prior to $t = 0$, IGs and FGs simultaneously submit supply functions to the system operator. Generator $k \in G^I \cup G^F$ submits a supply function $S_k(p)$, which specifies the output rate it is willing to produce when the price is p ($p \in \Re$, the set of real numbers). The supply functions are used by the system operator to calculate the generators' stated cost functions, which will be defined in (1) below.

The supply functions are valid for the entire operating horizon $[0, T]$. In the PJM market, for example, each generator submits one supply function for each operating day (see generator offer data at http://www.pjm.com/markets-and-operations/energy/real-time/historical-bid-data.aspx). In the MISO market, although generators are allowed to submit hourly offers, most generators submit the same supply functions for the entire day. (Using MISO's historical generator offer data available at https://www.misoenergy.org/Library/MarketReports, 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 2 *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 2(i) implies that no generator is willing to produce when the price is too low. Part (ii) is consistent with practice, e.g., MISO's Business Practice Manual states that the price-quantity pairs that form a supply function must be (weakly) increasing for price and strictly increasing for quantity (MISO, 2013, p. 92). 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 $\beta_k > 0$, satisfies Assumption 2. (Throughout the paper, $x^+ = \max\{x, 0\}$ for any $x \in \Re$.)

Based on the submitted supply function $S_k(p)$, the system operator computes the stated cost function of generator k as follows (we use "$\stackrel{\text{def}}{=}$" for definitions):

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

where $S_k^{-1}(q) \stackrel{\text{def}}{=} \inf\{p : S_k(p) > q\}$ is the inverse supply function. The definition in (1) is commonly used in practice by the system operators.

Unlike IGs and FGs, VGs are unable to guarantee an output rate because of their inherent intermittency. Thus, 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 an individual VG's output does not influence the market price. Therefore, VGs will offer a price equal to their marginal cost $-r$ (see Assumption 1(ii)). This means that VGs' stated cost at output rate q is $-rq$, and that VGs will produce the potential output W_t when the price exceeds $-r$, completely curtail output when the price drops below $-r$, and are willing to produce at any rate in $[0, W_t]$ when the price is $-r$.

3.3 System Operator's Problem

The objective of the system operator is to minimize the total expected stated cost of serving the load over $[0, T]$. This objective is consistent with the practice (see, e.g., MISO, 2013, Attachment B, §4.1.5) and the literature (see, e.g., Anderson and Philpott 2002). The system operator determines IGs' output rates prior to $t = 0$, which will be fixed during $[0, T]$. In real time when the load and VG potential output are realized, the system operator dispatches FGs and VGs and computes the market price.

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, q_t^F, and q_t^V denote the aggregate output rate at time t for IGs, FGs, and VGs, respectively. The allocation of q_t^V to VGs has no effect on the stated cost $-rq_t^V$. The allocations of q^I and q_t^F are determined by minimizing the total stated cost for each type of generators:

$$C^I(q^I) \stackrel{\text{def}}{=} \min\Big\{ \sum_{i\in G^I} C_i(q_i) \quad : q_i \geq 0, \ \sum_{i\in G^I} q_i = q^I \Big\}, \tag{2}$$

$$C^F(q_t^F) \stackrel{\text{def}}{=} \min\Big\{ \sum_{j\in G^F} C_j(q_{jt}) : q_{jt} \geq 0, \ \sum_{j\in G^F} q_{jt} = q_t^F \Big\}. \tag{3}$$

The following lemma summarizes the properties of $C^I(q)$ and $C^F(q)$ and their relationship with the the aggregate supply functions, defined as:

$$S^I(p) \stackrel{\text{def}}{=} \sum_{i\in G^I} S_i(p), \qquad \text{and} \qquad S^F(p) \stackrel{\text{def}}{=} \sum_{j\in G^F} S_j(p). \tag{4}$$

The proofs for the lemma and all other technical results are included in the appendix.

Lemma 1 *The total stated cost functions $C^I(q)$ and $C^F(q)$ are continuously differentiable, convex, and strictly increasing in q. Furthermore, $(C^I)'(q) = (S^I)^{-1}(q)$ and $(C^F)'(q) = (S^F)^{-1}(q)$.*

Lemma 1 confirms that the aggregate supply functions in (4) are consistent with the inverse marginal stated cost functions.

In an electricity system, the total generation and the load should be balanced at any time. Imbalance leads to extra operating cost. For example, in the case of oversupply, the system operator must take mitigating actions, such as providing monetary incentives for some consumers to increase the load, reducing generation to an emergency minimum level, or even shutting down some IGs at significant wear-and-tear costs. We model the extra costs for handling oversupply situations using a penalty function $h(e)$, which represents the extra cost rate when the total output exceeds the load by $e \geq 0$. 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 3 *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 that all demand is met. Using the aggregate outputs q^I, q_t^F, and q_t^V as decision variables, the system operator's problem of minimizing the total expected stated cost can be written as

$$\begin{aligned}
\min \quad & TC^I(q^I) + \mathsf{E}\left[\int_0^T \Big(C^F(q_t^F) - r\, q_t^V + h(e_t)\Big)\, dt\right] && (5)\\
\text{s.t.} \quad & e_t \equiv q^I + q_t^F + q_t^V - L_t \geq 0, && \forall\, t \in [0,T], \quad (6)\\
& q_t^V \leq W_t, && \forall\, t \in [0,T], \quad (7)\\
& q^I,\, q_t^F,\, q_t^V \geq 0, && \forall\, t \in [0,T]. \quad (8)
\end{aligned}$$

Note that this optimization problem contains two stochastic processes: the load process $\{L_t; 0 \leq t \leq T\}$ and the VG potential output $\{W_t; 0 \leq t \leq T\}$. The two processes can be correlated and we assume that their (joint) probability distribution is known to all firms (L_t is usually much more predictable than W_t). The expectation is taken with respect to these two processes. The decision q^I is made before time 0, while q_t^F and q_t^V are determined at time t when L_t and W_t are realized. The inequality in (6) ensures sufficient supply to meet the load, whereas excess supply (if $e_t > 0$) is penalized in the objective (5).

4. Optimal Dispatch and Market Mechanism

The system operator acts as a Stackelberg game follower, who solves the problem in (5)-(8) after the generators submit their supply functions. To solve (5)-(8), 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 §4.1 and §4.2, respectively.

4.1 Optimal Flexible and Variable Generation for Given q^I

Suppose the 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 outputs by the following convex program:

$$\widetilde{C}(q^I, L_t, W_t) \stackrel{\text{def}}{=} \min_{\{q_t^F,\, q_t^V\}} \; C^F(q_t^F) - r\, q_t^V + h(e_t) \tag{9}$$

$$\text{s.t.} \quad e_t \equiv q^I + q_t^F + q_t^V - L_t \geq 0, \tag{10}$$

$$q_t^V \leq W_t, \qquad q_t^F,\, q_t^V \geq 0. \tag{11}$$

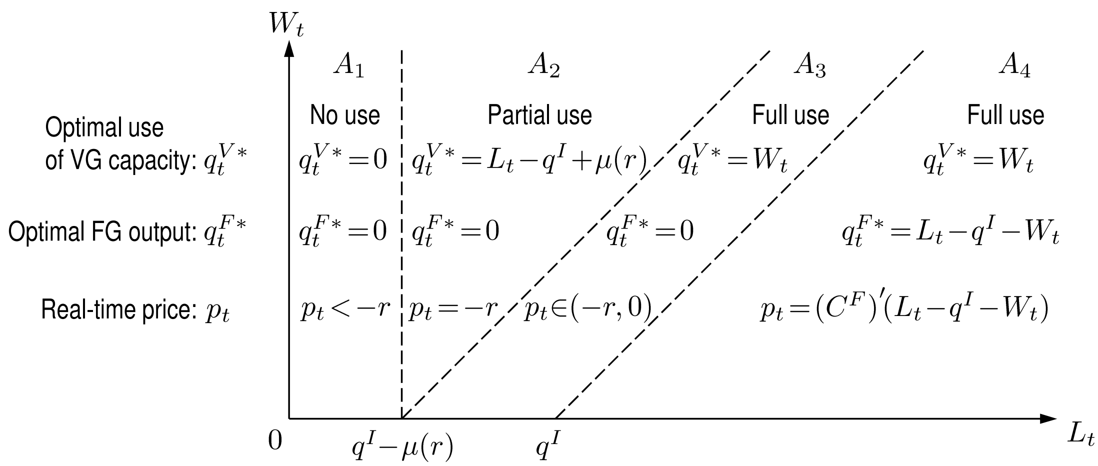
The following theorem provides an explicit solution to the problem in (9)-(11).

Theorem 1 *For a given IG output rate $q^I \geq 0$, under the realized VG potential output W_t and the load L_t, the optimal FG and VG production rates at time t are*

$$q_t^{F*} = (L_t - q^I - W_t)^+ \qquad \text{and} \qquad q_t^{V*} = \min\left\{W_t,\, (L_t - q^I + \mu(r))^+\right\}, \tag{12}$$

where $\mu(r) \stackrel{\text{def}}{=} (h')^{-1}(r) = \inf\{q \geq 0 : h'(q) > r\}$. Furthermore, the induced real-time cost rate $\widetilde{C}(q^I, L_t, W_t)$ in (9) is jointly convex in (q^I, L_t, W_t).

Figure 1: Real-Time Operating Policy and Price



$A_1 : L_t - q^I \leq -\mu(r)$ $\qquad$ $A_3 : L_t - q^I \in [W_t - \mu(r),\, W_t]$

$A_2 : L_t - q^I \in (-\mu(r),\, W_t - \mu(r))$ $\qquad$ $A_4 : L_t - q^I > W_t$

The optimal solutions in (12) under various realized values of L_t and W_t are illustrated in Figure 1. If the load L_t drops below the IG output q^I to such an extent that the marginal oversupply penalty exceeds the per-unit subsidy, $h'(q^I - L_t) \geq r$ (i.e., $q^I - L_t \geq \mu(r)$ or event A_1), then the VG output does not bring net benefit to the system and is completely curtailed. When $r > h'(q^I - L_t)$, some or all of the VG potential output is used, corresponding to the next three cases.

In event A_2, VG output is partially curtailed 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 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 imply that FGs produce if and only if the load cannot be satisfied by IGs and

VGs. The above discussion also leads to a complementary property of the optimal operating policy:

$$q_t^{F*}\left(q_t^{V*} - W_t\right) = 0. \tag{13}$$

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

Figure 1 also shows the price, p_t, which equals the system's marginal cost, i.e., the cost of serving an additional unit of load at time t. When the load exceeds the combined output of IGs and VGs (event A_4), the additional load is served by FGs and thus the price is FGs' marginal cost: $p_t = (C^F)'(L_t - W_t - q^I) > 0$. Using Lemma 1, we can also write $p_t = (S^F)^{-1}(L_t - W_t - q^I)$.

When the load can be met by IGs and VGs, the price becomes zero or negative:

a) The price is zero when VG output is partially curtailed (event A_2 occurs) and no subsidy is provided ($r = 0$). An additional unit of load can be served by VGs at zero cost.

b) The price is negative when the additional load lowers the total stated 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 price p_t as a function of q^I, L_t, and W_t by (14) below. In this expression, the dependence on supply function $S^F(\cdot)$ is also emphasized.

$$P(q^I, L_t, W_t, S^F) \stackrel{\text{def}}{=} (S^F)^{-1}(L_t - W_t - q^I)\mathbf{1}_{A_4} - h'(q^I + W_t - L_t)\mathbf{1}_{A_3} - r\mathbf{1}_{A_2} - h'(q^I - L_t)\mathbf{1}_{A_1}, \tag{14}$$

where $\mathbf{1}_{A_i}$ is the indicator function for event A_i. Clearly, for given q^I, the price does not depend on IGs' supply function $S^I(\cdot)$.

Using (14), the time-average of the expected price can be written as

$$\overline{P}(q^I, S^F) \stackrel{\text{def}}{=} \frac{1}{T}\int_0^T \mathsf{E}\left[P(q^I, L_t, W_t, S^F)\right] dt, \tag{15}$$

where the expectation is taken prior to $t = 0$. The function $\overline{P}(q^I, S^F)$ relates the average price to the IG output q^I for given FG supply function, thus $\overline{P}(q^I, S^F)$ can be interpreted as IGs' inverse residual demand function. Note that $\overline{P}(q^I, S^F)$ decreases in q^I, because $P(q^I, L_t, W_t, S^F)$ in (14) decreases in q^I due to the monotonicity of $S^F(\cdot)$ and $h'(\cdot)$.

4.2 Optimal Inflexible Generation

With the average price $\overline{P}(q^I, S^F)$ computed in (15), the aggregate (constant) output rate IGs are willing to set over $[0, T]$ is $S^I\left(\overline{P}(q^I, S^F)\right)$. The system operator needs to ensure consistency between what IGs are asked to produce and what they are willing to produce: $q^I = S^I\left(\overline{P}(q^I, S^F)\right)$. 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 2 below shows that the optimal q^{I*} satisfies this constraint.

With the real-time minimum cost $\widetilde{C}(q^I, L_t, W_t)$ given by (9), the problem in (5)-(8) can be reformulated as

$$\min_{q^I \geq 0} \; TC^I(q^I) + \mathsf{E}\left[\int_0^T \widetilde{C}(q^I, L_t, W_t)\, dt\right]. \tag{16}$$

From Lemma 1 and Theorem 1, $C^I(q^I)$ and $\widetilde{C}(q^I, L_t, W_t)$ are convex in q^I, which implies that the objective function in (16) is convex in q^I.

Theorem 2 *The optimal IG output q^{I*} satisfies*

$$q^{I*} = S^I\big(\overline{P}(q^{I*}, S^F)\big). \tag{17}$$

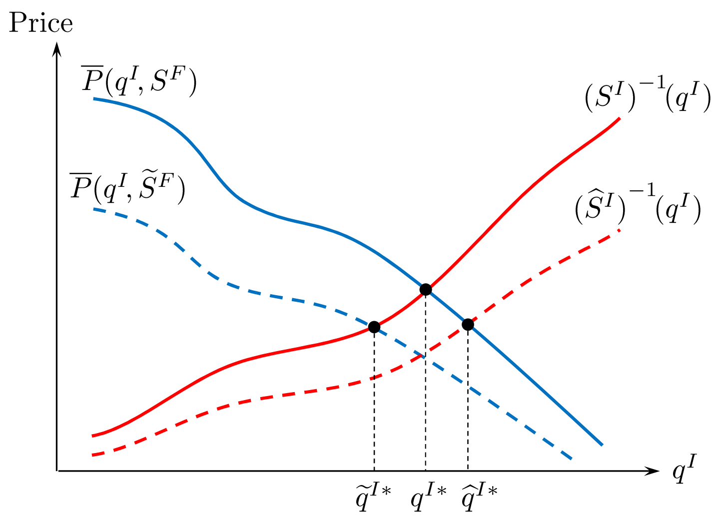
Furthermore, the price function in (14) *can be expressed as*

$$P(q^I, L_t, W_t; S^F) = \inf\big\{p : S^F(p) + W_t\mathbf{1}_{\{p \geq -r\}} - \mu(-p) \geq L_t - q^I\big\}. \tag{18}$$

Equation (18) provides a supply-function based method for calculating the price originally defined in (14). In (18), $S^F(p)$ 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 price is $p < 0$.

Theorem 2 confirms that the relation $q^I = S^I\big(\overline{P}(q^I, S^F)\big)$ must hold at optimality. Equation (17) 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

Figure 2: Optimal IG Production q^{I*}



two functions are depicted as the solid curves in Figure 2.

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 a dashed curve in Figure 2, q^{I*} rises to $\widehat{q}^{I*}$, i.e., IGs' market share increases. When FGs bid more competitively by increasing their supply function to $\widetilde{S}^F(p)$, the price decreases according to (18), and the average price decreases to $\overline{P}(q^I, \widetilde{S}^F)$, as shown in Figure 2. Consequently, q^{I*} decreases to $\widetilde{q}^{I*}$. In both cases, more competitive supply offers lead to a lower average price.

4.3 The Market Mechanism

Theorems 1 and 2 solve the system operator's problem of deciding the optimal production to minimize the expected total stated cost. We now summarize the market mechanism based on the above results.

1) Prior to $t = 0$, IGs and FGs simultaneously 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 §3.2).
2) Prior to $t = 0$, the system operator determines the IG output rate using 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 price as a function of the IG output q^I, load L, and VG potential output W:
   $$P(q^I, L, W, S^F) = \inf\{p : S^F(p) + W\mathbf{1}_{\{p\geq -r\}} - \mu(-p) \geq L - q^I\}, \qquad \text{or}$$
   $$P(q^I, L, W, S^F) = (S^F)^{-1}(L - W - q^I)\mathbf{1}_{A_4} - h'(q^I + W - L)\mathbf{1}_{A_3} - r\mathbf{1}_{A_2} - h'(q^I - L)\mathbf{1}_{A_1}.$$
   (iii) Determine the IG output rate q^{I*} by solving
   $$(S^I)^{-1}(q^{I*}) \ = \ \overline{P}(q^{I*}, S^F) \equiv \frac{1}{T}\int_0^T \mathsf{E}\left[P(q^{I*}, L_t, W_t, S^F)\right] dt. \tag{19}$$
3) Production and payment:
   (i) IG $i \in G^I$ produces $S_i(\overline{P}(q^{I*}, S^F))$ for all $t \in [0, T]$.
   (ii) At time $t \in [0, T]$, the price is $p_t \equiv P(q^{I*}, L_t, W_t, S^F)$, 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. With the knowledge of the system operator's dispatch and market mechanism, the IGs and FGs compete in a Nash game through supply functions, analyzed in the next section.

5. Linear Supply Function Competition

In the classic supply function equilibria (SFE) model, in a firm's best response problem, the firm optimizes its profit with respect to its residual demand function. Because the optimality condition involves the derivatives of competitors' supply functions, a SFE is characterized by a system of differential equations (Klemperer and Meyer 1989). The differential equation approach is analytically challenging, especially when firms are asymmetric. Thus, SFE with linear supply functions are considered in the classic works by Klemperer and Meyer (1989), Rudkevich (1999), among others; see Baldick et al. (2004) for a summary of the advantages of linear SFE. Green (1996, 1999) and Baldick et al. (2004) use linear SFE to derive economic insights and policy implications.

One of our goals in this paper is to examine how firms' (in)flexibility affects SFE and compare our results and insights with the classic works. This comparison is made possible by focusing on solving for linear SFE and comparing the slopes of the equilibrium supply functions in our model with those in the classic models. Solving for the general SFE with asymmetric firms is difficult in the classic setting, and even more difficulty in our setting where generators not only have asymmetric cost but also different flexibility levels.

From this point onward, we assume each generator's production cost rate is quadratic in its output rate (also assumed in the classic works):

$$C_k(q) = \frac{1}{2} c_k q^2, \qquad k \in G^I \cup C^F, \quad c_k > 0, \quad q \geq 0, \tag{20}$$

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:

$$S_k(p) = \beta_k p^+, \qquad k \in G^I \cup C^F, \quad \beta_k > 0, \quad p \in \Re. \tag{21}$$

That is, when the price is positive, the output rate that a generator is willing to produce is linear in price. The generator's pure strategy set will be defined in §5.2.

We also assume quadratic oversupply penalty function:

$$h(e) = a_h e + \frac{1}{2} c_h e^2, \qquad a_h \geq 0, \quad c_h > 0, \quad e \geq 0, \tag{22}$$

which gives $\mu(r) = (h')^{-1}(r) = (r - a_h)^+ / c_h$.

5.1 Optimal Dispatch under Given Supply Functions

For given $q^I > 0$, Theorem 1 gives the optimal q_t^{F*} and q_t^{V*} as functions of q^I. The results on the optimal q^{I*} in Theorem 2 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 $\beta^F \stackrel{\text{def}}{=} \sum_{j \in G^F} \beta_j$. With linear supply functions, we can express the price functions $P(q^I, L_t, W_t, S^F)$ and $\overline{P}(q^I, S^F)$ in (14)-(15) as functions of β^F, written as follows:

$$\begin{aligned} P(q^I, L_t, W_t, \beta^F) &= \frac{1}{\beta^F}(L_t - W_t - q^I)\mathbf{1}_{A_4} - \left[a_h + c_h(q^I - L_t + W_t)\right] \mathbf{1}_{A_3} \\ &\qquad\qquad - r\mathbf{1}_{A_2} - \left[a_h + c_h(q^I - L_t)\right] \mathbf{1}_{A_1}, \end{aligned} \tag{23}$$

$$\overline{P}(q^I, \beta^F) = \frac{1}{T}\int_0^T \mathsf{E}\big[P(q^I, L_t, W_t, \beta^F)\big]\,dt. \tag{24}$$

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 $q^{I*} > 0$. Then, equation (17) that determines q^{I*} can be written as

$$q^I = \beta^I \overline{P}(q^I, \beta^F). \tag{25}$$

There is a unique q^{I*} satisfying (25). We denote this unique q^{I*} as a function of β^I and β^F:

$$q^{I*} \equiv Q^I(\beta^I, \beta^F) \stackrel{\text{def}}{=} \big\{q^I : q^I = \beta^I \overline{P}(q^I, \beta^F)\big\}. \tag{26}$$

Lemma 2 *The optimal 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 2 and formally stated in Lemma 2.

5.2 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 used to establish the existence of the equilibrium in §5.5.

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 (21) 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 essentially the standard supply function games considered by Klemperer and Meyer (1989). Rudkevich (1999) studies the linear SFE for these 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.

Using the upper and lower bounds, 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

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

Using Lemma 2, we can establish bounds on q^I as

$$q^{I\min} = Q^I(\beta^{I\min}, \beta^{F\max}) \qquad \text{and} \qquad q^{I\max} = Q^I(\beta^{I\max}, \beta^{F\min}). \tag{27}$$

In deriving (25), we assumed $q^{I*} > 0$. A sufficient condition for $q^{I*} > 0$ is $\overline{P}(0, \beta^{F\max}) > 0$, i.e., the average price is positive when IGs do not produce and FGs bid perfectly competitively, which is a mild condition. To see the sufficiency, note that $q^{I\min}$ in (27) is the unique solution to $q^I = \beta^{I\min}\overline{P}(q^I, \beta^{F\max})$. Thus, $\overline{P}(0, \beta^{F\max}) > 0$ implies $q^{I\min} > 0$, which ensures $q^{I*} > 0$.

5.3 Individual IG and FG'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 a rate $S_i(\overline{P}) = \beta_i\overline{P}$ throughout $[0,T]$, incurring 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 (24), and q^I is affected by β_i through (25). Hence, IG i's optimization problem can be written as

$$\begin{aligned} \max_{\beta_i} \quad & \beta_i\Big(1-\frac{1}{2}c_i\beta_i\Big)\overline{P}(q^I, \beta^F)^2 \\ \text{s.t.} \quad & (25) \;\; \text{and} \;\; \beta_i \in [\beta_i^{\min}, c_i^{-1}]. \end{aligned} \tag{28}$$

This formulation is similar to the bi-level optimization procedure by Hobbs et al. (2000). The system-level optimization yields (25) and the firm-level objective is given by (28).

Using (25) and (26), 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 (28) as a function of the strategic variables:

$$\pi_i(\beta_i; \beta_{-i}, \beta^F) \stackrel{\text{def}}{=} \frac{\beta_i\big(1 - \frac{1}{2}c_i\beta_i\big)}{(\beta_i + \beta_{-i})^2}\big(Q^I(\beta_i + \beta_{-i}, \beta^F)\big)^2. \tag{29}$$

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

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 a 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 (23). Thus, generator j's problem is

$$\begin{aligned} \max_{\beta_j} \quad & \int_0^T \mathsf{E}\Big[\beta_j\Big(1 - \frac{1}{2}c_j\beta_j\Big)\big(P(q^I, L_t, W_t, \beta^F)^+\big)^2\Big]\,dt \\ \text{s.t.} \quad & (25) \ \text{ and } \ \beta_j \in [\beta_j^{\min}, c_j^{-1}]. \end{aligned} \tag{30}$$

Equations (23) and (26) lead to

$$P(q^I, L_t, W_t; \beta^F)^+ \;=\; \frac{1}{\beta^F}(L_t - W_t - q^I)^+ \;=\; \frac{1}{\beta^F}\big(L_t - W_t - Q^I(\beta^I, \beta^F)\big)^+.$$

We define $\beta_{-j} \stackrel{\text{def}}{=} \beta^F - \beta_j$ and rewrite the objective in (30) as a function of the strategic variables:

$$\pi_j(\beta_j; \beta_{-j}, \beta^I) \stackrel{\text{def}}{=} \frac{\beta_j\big(1 - \frac{1}{2}c_j\beta_j\big)}{(\beta_j + \beta_{-j})^2}\int_0^T \mathsf{E}\Big[\Big(\big(L_t - W_t - Q^I(\beta^I, \beta_j + \beta_{-j})\big)^+\Big)^2\Big]\,dt. \tag{31}$$

Then, FG j's best response to β_{-j} and β^I is determined by optimizing $\max\limits_{\beta_j \in [\beta_j^{\min}, c_j^{-1}]} \pi_j(\beta_j; \beta_{-j}, \beta^I)$.

5.4 Interactions Between IGs and FGs

The profit functions in (29) and (31) provide important insights on how IGs and FGs compete:

- IGs and FGs interact only through the function $Q^I(\beta^I, \beta^F)$, which is IGs' market share. This interaction implies that the competition between IGs and FGs is over the market share.
- The variabilities in load and VG potential output play no (direct) role in IGs' profit function (29). Thus, IGs do not directly compete with FGs in meeting the variable demand. On the other hand, the variabilities in L_t and W_t directly affect FGs' profit function in (31). Hence, FGs compete among themselves to serve the variable demand.

We will use these insights to explain some of the equilibrium behaviors observed in the numerical analysis in §6.

The strategic interaction between IGs and FGs also renders the best responses dependent on the distribution of the uncertainties. Without this strategic interaction, $Q^I(\beta^I, \beta^F)$ would be constant and, consequently, the distributions of L_t and W_t would not affect FG j's best response determined by (31). With the FG-IG interaction through $Q^I(\beta^I, \beta^F)$, the distributions of L_t and W_t affect the optimal choice of β_j in (31), which in turn affects the strategic decisions of all other generators. This feature is in contrast with the classic SFE model, in which supply function equilibria are found to be independent of the demand distribution; see, e.g., Klemperer and Meyer (1989), Green (1996), Holmberg (2007), and Anderson and Hu (2008).

5.5 Existence of Equilibrium under Normally Distributed Uncertainties

Proving the existence and uniqueness of the equilibrium for the game specified in §§5.2-5.3 presents analytical challenges. In general, existence of a pure-strategy Nash equilibrium can established by proving one of the following: a) the best response functions constitute a contraction mapping and the decision space is compact, b) each player's payoff function is quasi-concave in its own decision and the decision space is compact, convex, and independent of other players' decisions, and c) each player's payoff function is supermodular with respect to its own decision and other players' decisions, and that the decision space is a lattice. Proving uniqueness usually has to rely on approach a).

As discussed earlier, our model represents a combination of Nash game and Stackelberg game. Both IGs' and FGs' payoff functions (29) and (31) depend on the distribution of uncertainties and the optimal dispatch in §5.1. Thus the firm-level optimization problem belongs to the class of MPEC (mathematical program with equilibrium constraints) problems that are highly complex. However, we are able to use approach b) to prove the existence under the assumption that the load and VG potential output (when collapsed across time) are jointly normally distributed. We are unable to establish the uniqueness of the equilibrium, but our extensive numerical results show that the equilibrium is unique, which we will discuss in §6.

For the process $\{(L_t, W_t) : t \in [0, T]\}$, we let $f_{L_t,W_t}(x, y)$ be the joint probability density function of L_t and W_t. Note that load L_t and VG potential output W_t can be cross-sectionally and serially correlated. We collapse the distribution across time and define

$$f_{L,W}(x, y) \stackrel{\text{def}}{=} \frac{1}{T} \int_0^T f_{L_t,W_t}(x, y)\, dt. \tag{32}$$

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 (32).

From this point onward, we assume $f_{L,W}(x, y)$ is a bivariate normal density function, and $L \sim$

$\mathcal{N}(\mu_L, \sigma_L^2)$ and $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$ with a correlation coefficient ρ. Define the net demand random variable $D \stackrel{\text{def}}{=} L - W \sim \mathcal{N}(\mu_D, \sigma_D^2)$ where $\mu_D = \mu_L - \mu_W$, and $\sigma_D^2 = \sigma_L^2 + \sigma_W^2 + 2\rho\sigma_L\sigma_W$. As common with models using normal distributions to approximate nonnegative random variables, the results in this section are proven when the variance of the normal distribution is not too large.

Proving quasi-concavity of the profit functions (29) and (31) under general conditions is difficult due to the complicated structure of the price function in (14), which renders the average price in (15) neither convex nor concave in q^I. However, if the probability of $q^I < L - W$ (event A_4 in Figure 1) is sufficiently high, the average price function is approximately linear, which bounds its second-order derivative with respect to q^I and leads to the quasi-concavity of the profit functions. The formal proof requires a lemma stated below.

Lemma 3 *If* $\sigma_D \leq \sigma_D^* \equiv \sqrt{2\pi}\,\beta^{F\,\min}\left[\dfrac{\mu_D}{\beta^{I\,\max}} + \dfrac{\min\{r, a_h\}}{2}\right]$, *then* $q^{I\,\max} < \mu_D$.

Lemma 3 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 μ_D for most situations in practice. The condition given in Lemma 3 is not stringent. For example, if $a_h = 0$ 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 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. Lemma 3 leads to the following equilibrium existence theorem.

Theorem 3 *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.*

In Theorem 3, the upper-bound on σ_D that ensures the existence of a linear supply function equilibrium is provided in the proof in the online appendix. Our numerical experiments, however, show that the equilibrium exists for a wider range of σ_D and for other forms of load and VG output distributions.

6. Numerical Study

In this section, we compute SFE based on the model analyzed in §§4-5 and compare our results with the classic SFE model by Klemperer and Meyer (1989) and Green (1996). We also analyze the effect of increasing VG penetration and its dispatch policy (priority dispatch vs. economic curtailment) on SFE. Our analysis aims to derive qualitative insights and provide policy recommendations.

6.1 Setups and Computational Procedure

We consider a market consisting of four IGs indexed by $i \in G^I = \{1, 2, 3, 4\}$ and four FGs indexed by $j \in G^F = \{5, 6, 7, 8\}$. Their production cost rates (in \$/hour) are given in (20): $C_k(q) = \frac{1}{2} c_k q^2$, where q is in MW and c_k is measured in \$/MWh/MW. To facilitate comparison between IGs' and FGs' equilibrium behavior, we keep generators identical within each generator type. We assume the cost coefficients are

$$c_i = \frac{1}{3}, \text{ for } i \in G^I, \qquad c_j = 1, \text{ for } j \in G^F.$$

The system's oversupply penalty for e MW of oversupply is assumed to be $h(e) = \frac{1}{2} c_h e^2$ with $c_h = 1/3$. We report the equilibrium results under the above cost parameters, but we have also examined other cost parameters with $c_i < c_j$, where $c_i, c_j \in \{1/6, 1/4, 1/3, 1/2, 2/3, 1, 2, 3\}$, and $c_h \in \{1/4, 1/3, 1/2, 1, 2\}$. We find that the qualitative results described in this section are robust across all problem instances we examined. We remark on the robustness of the results along with the discussions in the rest of this section.

We next specify the time-invariant probability distribution, $f_{L,W}(x, y)$, defined in (32). Within a given operating horizon $[0, T]$, we assume the load and VG potential output follow independent normal distributions, with $\mu_L = 100$, $\sigma_L = 15$, $\mu_W = 5$, and $\sigma_W = 1.75$, measured in MWh per 5 minutes. (Many electricity systems measure load and VG output at 5-minute intervals.) The length of the horizon T is typically several hours to one day, but because we will report costs and emissions in hourly rates, the specific value of T does not affect our results.

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 to $m\mu_W$), the standard deviation σ_W increases to $m\sigma_W$ if the existing and added VG outputs are perfectly correlated, or to $\sqrt{m}\,\sigma_W$ if they are independent. The realistic case is likely in between and we assume that σ_W increases to $m^{0.75}\sigma_W$. Specifically, we consider five VG penetration levels: 0%, 5%, 15%, 30%, and 50%. That is, $m = 0, 1, 3, 6, 10$.

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

The generators submit linear supply functions $S_k(p) = \beta_k p^+$, as in (21). The following iterative procedure is used to compute the generators' equilibrium supply function slopes:

Step 1. Set $n = 0$ and choose an initial slope $\beta_k^0 \in [\beta_k^{\min}, c_k^{-1}]$ for every generator $k \in G^I \cup G^F$.

Step 2. Increase n by 1. For every generator $k \in G^I \cup G^F$, find $\beta_k^n \in [\beta_k^{\min}, c_k^{-1}]$ that maximizes generator k's profit, assuming that none of the other generators modify their slopes (i.e., use β_l^{n-1} for generator $l \neq k$).

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

In Step 2, we numerically find that the objective function is strictly quasi-concave, which ensures that any local maximum is the unique global maximum. In Step 3, we use $\varepsilon = 0.1\%$ for the convergence criterion. The procedure typically takes only 4 to 5 iterations to converge.

To numerically examine the uniqueness of the equilibrium, we initiate the procedure with many different starting points in Step 1 and find that the procedure always converges to the same equilibrium. Furthermore, when all generators are assumed to be flexible, the procedure produces exactly the same results as the classic SFE model with linear supply functions.

6.2 FG-IG Equilibrium vs. Klemperer-Meyer Equilibrium

The supply function slope β_k is useful for the theoretical analysis in §5, but for the purpose of describing the insights from our numerical experiments, it is more intuitive to use *price offer slope* $\gamma_k \stackrel{\text{def}}{=} 1/\beta_k \in [c_k, 1/\beta_k^{\min}]$. A lower γ_k means a more competitive price offer. In a perfect competition, generator k's price offer is equal to its marginal cost $c_k q$.

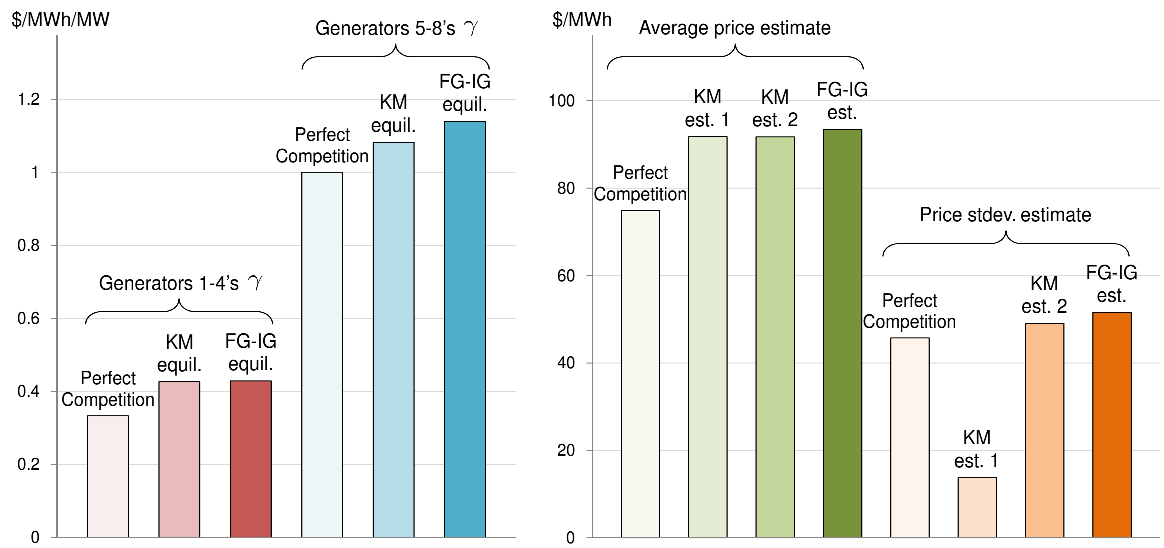
As our model extends the classic SFE model by Klemperer and Meyer (1989) to include asymmetries in both cost and flexibility, we first compare the SFE in our model (referred to as FG-IG equilibrium) with the Klemperer-Meyer model (referred to as KM equilibrium), focusing on the SFE with linear supply functions.

The KM model ignores the generators' inflexibility and treats generators 1-4 as if they are flexible to find the equilibrium supply functions. Following Green (1996) and Rudkevich (1999), we solve the KM equilibrium with linear supply functions under the same cost functions described in §6.1. We find that, in the KM model, generators 1-4 each offer price $0.427\,q$ and generators 5-8 each offer price $1.082\,q$, shown in Figure 3(a).

Using the procedure in §6.1, we compute the FG-IG equilibrium without VG in the system. Because our model recognizes the inflexibility of generators 1-4, one might expect IGs 1-4 to behave more differently than FGs 5-8 compared to the KM equilibrium. However, in the FG-IG equilibrium, IGs offer price $0.429\,q$, which is only 0.5% higher than in the KM model, whereas FGs offer price $1.139\,q$, 5.3% higher than in the KM model. In terms of the markup $(\gamma_k - c_k)q$, FGs' markup is $0.139\,q$, 69% higher than their markup $0.082\,q$ in the KM model; IGs' markup is only 2.4% higher

Figure 3: Klemperer-Meyer Equilibrium vs. FG-IG Equilibrium without VG

(a) Price offer slope γ (b) Average and standard deviation of price



than in the KM model. The price offer slopes under different models are shown in Figure 3(a).

The reasons underlying the difference between the FG-IG and KM equilibria stem from the reduced competition due to inflexibility. All eight generators are treated as flexible in the KM model, but only four of them are actually FGs. IGs do not compete with FGs in matching production with the variable demand; FGs compete among themselves to serve the variable demand (see the discussion in §5.4). Thus, the competition facing an FG in our model is less intense than that in the KM model, allowing FGs to raise their price offers significantly above what the KM model predicts.

On the other hand, IGs' price offers in our equilibrium is similar to that in the KM model, because an IG faces direct competition from all other generators. In particular, IGs still compete with FGs for market share (IGs produce q^{I*} and FGs serve the rest of the demand). This competition is only slightly less intense than in the KM model because FGs raise their price offers as explained above. As a result, IGs slightly raise their price offers above the KM equilibrium. The above finding suggests that the KM model underestimates generators' price offers, more significantly so for FGs.

Next, we compare how the two models differ in market price estimation. Estimating the market price involves two steps: First, estimate the equilibrium supply functions using a SFE model; second, compute the price statistics under the estimated supply functions. A forecaster who uses the classic SFE model in the first step may or may not consider inflexibility in the second step. We refer to the estimates without recognizing inflexibility in either step as "KM est. 1" and the estimates with inflexibility consideration in only the second step as "KM est. 2". The FG-IG model recognizes inflexibility in both steps. We also compute the price under perfect competition with inflexibility consideration as a benchmark.

Figure 3(b) shows that the mean and standard deviation of the price estimated by the KM model (both KM est. 1 and est. 2) are lower than those estimated by the FG-IG model. The KM est. 1 for the price standard deviation is considerably lower, because ignoring inflexibility in the second step leads to an incorrect assumption that all generators can mitigate load variability.

Although KM est. 2 considers inflexibility in the second step, it underestimates the average price and price standard deviation. KM est. 2 for the average price is lower because the KM model underestimates the equilibrium price offers for both IGs and FGs. KM est. 2 underestimates the price volatility for two reasons. First, the KM model underestimates FGs' price offer slopes and, thus, it underestimates the magnitude of price fluctuations when the price is positive. Second, because the KM model underestimates the price offers more for FGs, it underestimates IGs' market share and, thus, it underestimates the magnitude of the negative prices.

In view of both the generators' price offers and the equilibrium price, the KM model overestimates the intensity of the competition in a market with inflexible generators. This result is robust across all cost parameters we have examined.

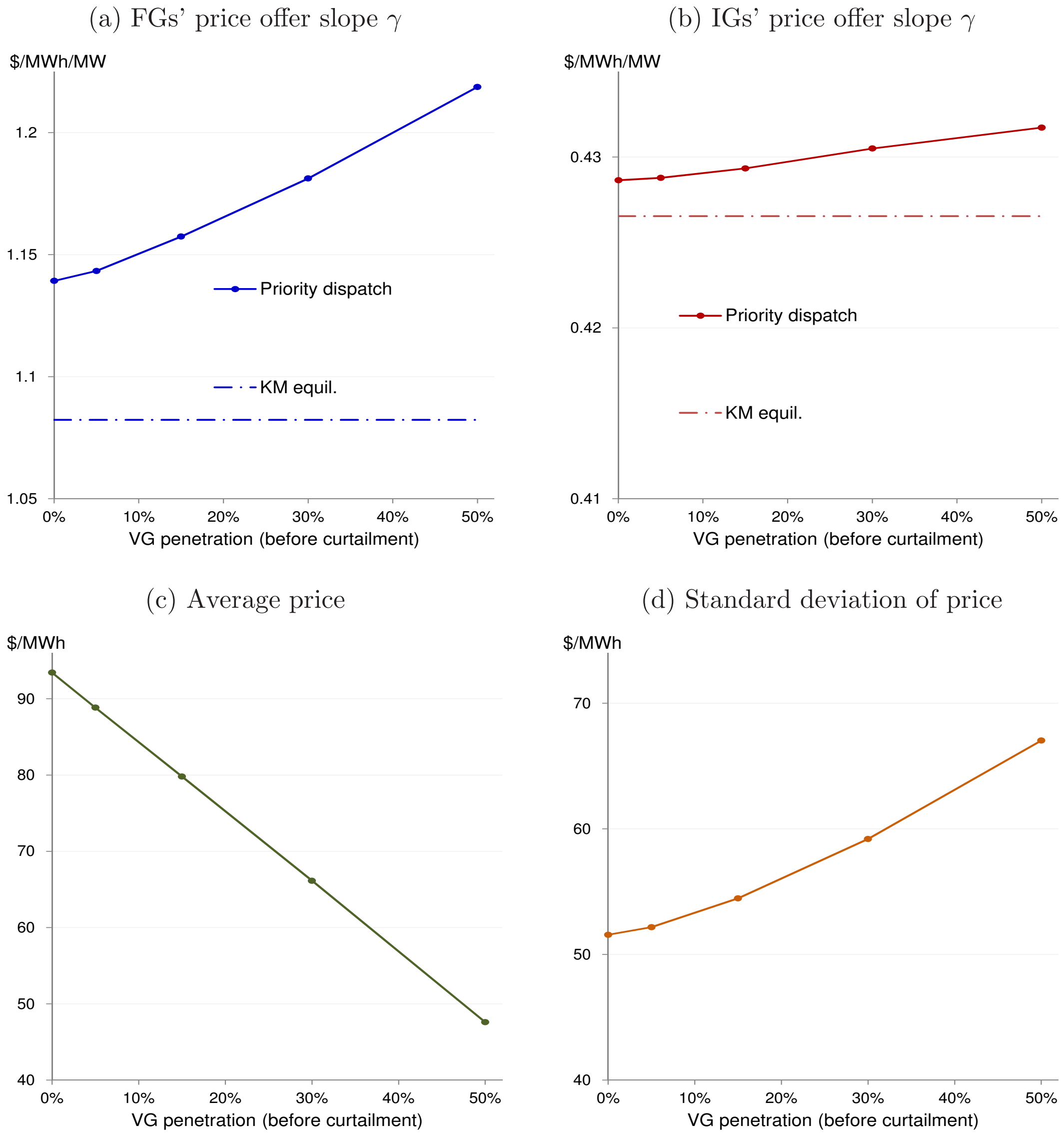
6.3 Impact of Variable Generation (under Priority Dispatch) on SFE

Because the KM equilibrium is known to be independent of the distribution of the uncertainties, variable generation has no impact on the price offers in the KM equilibrium. In Figure 4(a)-(b), the KM equilibrium price offers are invariant to the VG penetration levels.

In the FG-IG equilibrium, when the VG penetration increases, the overall variability (including demand and VG output variabilities) increases. If VGs have priority in dispatch, the only lever for balancing against the increased variability is adjusting the FGs' output. Increased variability also increases the chance of oversupply, which makes a lower IG output more desirable for the system. Both of these reasons give FGs an advantage in the market-share competition with IGs. To profit from this advantage, FGs raise their price offers as the VG penetration increases, which is confirmed in Figure 4(a); the top curve is for priority dispatch.

On the other hand, as the 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 a low IG output is desirable for the system to mitigate the oversupply penalty when VGs have priority, IGs' strategy of lowering price offers may not lead to an output increase that is sufficient to raise IGs' profits. Thus, raising price offers is preferred by IGs, as confirmed by the top curve in Figure 4(b). Hence, under the priority dispatch

Figure 4: Effects of VG on the Equilibrium under Priority Dispatch



policy for VGs, both IGs and FGs raise their price offers as the VG penetration increases, and the KM model increasingly underestimates generators' price offers and overestimates the intensity of the market competition.

Although inflated price offers tend to raise the market price, the increased VG penetration reduces the average net demand and tends to reduce the price. The equilibrium price is a result of the combination of these two effects. The second effect dominates in determining the average price, as shown in Figure 4(c): the average price declines as the VG penetration increases.

Under the priority dispatch policy, increasing VG penetration makes the price more volatile, as revealed in Figure 4(d). The reasons are twofold. At a high VG penetration level, the VG output can still occasionally drop to a low level, requiring FGs to ramp up production, which escalates the market price due to FGs' increased price offers. When the VG output surges, however, the system has to take all the VG output due to its priority, resulting in possibly very negative prices at high VG penetration levels.

We have examined the above results under various cost parameters. The qualitative trends discussed in this section are robust.

6.4 Impact of the Economic Curtailment Policy on SFE

The analysis in §6.3 assumes VGs have priority; in this section, we consider the policy that allows economic curtailment of VGs. We focus on the economic curtailment case under zero subsidy ($r = 0$) and compare it with the cases under subsidies $r = 20$ and 40 \$/MWh.

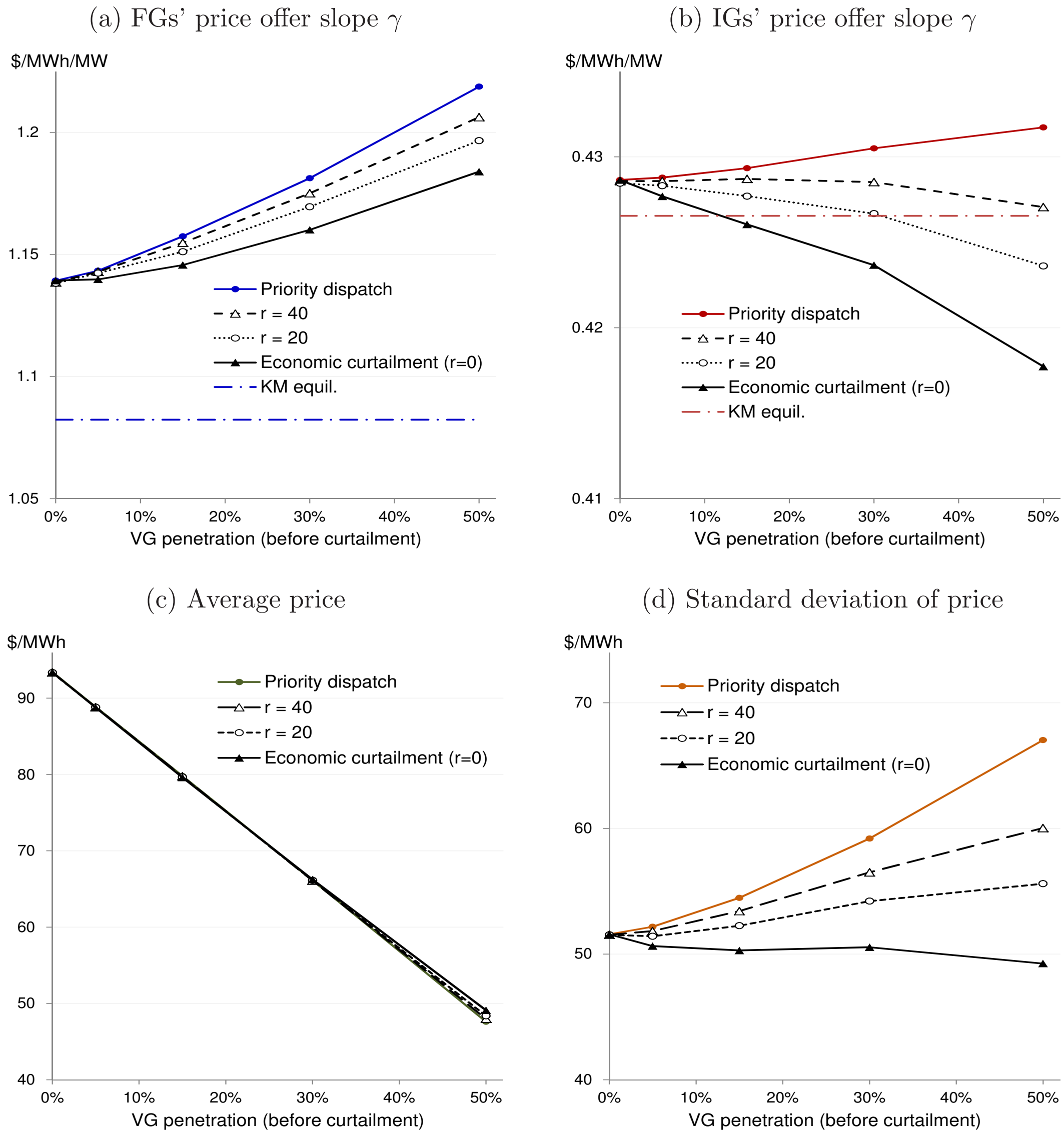
The FG-IG equilibrium price offers are shown in Figure 5(a)-(b). The economic curtailment policy encourages both IGs and FGs to offer lower (more competitive) prices compared to the priority dispatch policy. As the VG penetration increases, Figure 5(a) shows that FGs' price offer slope increases slower than that under the priority dispatch policy, while Figure 5(b) shows that IGs' price offer slope declines and may even drop below the price offer slope predicted by the KM model.

The economic curtailment policy increases the market competition 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 §6.3). Consequently, IGs' strategy of lowering price offers can yield an output increase that is sufficient to increase IGs' profits. These two effects of the economic curtailment policy reinforce each other in equilibrium, because IGs reduce their price offers in response to FGs' reduced price offers and vice versa.

Figure 5 also shows the effect of production-based subsidies. Subsidies effectively grant priority to VGs to some extent and thus lead to less competitive price offers. The price offers in the cases of $r = 20$ and 40 \$/MWh lie in between the price offers in the priority dispatch and economic curtailment cases. For all problem instances we examined, we find that under the economic curtailment policy, IGs reduce their price offers as VG penetration increases. Under $r = 20$ or 40 \$/MWh, whether IGs increase or decrease their price offers depends on instances, but they always lie between the priority dispatch and economic curtailment cases.

Economic curtailment has little effect on the average price, but the impact on price volatility is significant. Figure 5(c) shows that when the VG penetration level is below 30%, average prices under various VG policies are indistinguishable. At higher VG penetration levels, the average price

Figure 5: Effects of Economic Curtailment on the Equilibrium



under economic curtailment is slightly higher because the curtailment reduces the severity of the negative prices. In contrast, the standard deviation of the price drops considerably under economic curtailment, as shown in Figure 5(d), 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.

6.5 Effects of Economic Curtailment on Efficiency and Emission

In this last section, we study the efficiency and emission impact of economic curtailment. The system efficiency is measured by its average operating cost, which is the sum of the actual production cost (not the stated cost) of the eight generators and the oversupply penalty. Table 1 shows the system operating cost under various VG penetration levels and dispatch policies. 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 by Wu and Kapuscinski (2013).

Table 1 also reveals that higher subsidies reduce the amount of curtailment but increase the system operating cost. Interestingly, higher subsidies also increase the cost saving per MWh of 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, the cost saving per MWh of curtailment 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 the amount of production.

An environmental benefit from increasing VG penetration is the reduced CO_2 emission due to

Table 1: Effects of Curtailment on Cost and CO_2 Emission

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 \$/hour)	Priority Dispatch	44.85	36.96	27.04	17.31
	$r = 40$	44.83	36.89	26.82	16.57
	$r = 20$	44.81	36.84	26.70	16.29
	Economic Curtailment	44.79	36.80	26.61	16.11
System cost saving 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 emission with coal-fired IGs (tons/hour)	Priority Dispatch	1287.8	1044.7	729.9	403.1
	$r = 40$	1288.4	1045.9	733.4	411.2
	$r = 20$	1289.2	1047.7	737.4	418.5
	Economic Curtailment	1290.6	1051.7	745.2	432.6
Total CO_2 emission with nuclear IGs (tons/hour)	Priority Dispatch	171.6	146.4	116.3	86.9
	$r = 40$	171.3	145.9	114.7	82.8
	$r = 20$	170.9	145.1	112.9	79.3
	Economic Curtailment	170.4	143.4	109.5	73.2

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

the displacement of the conventional production by the clean VG production. Table 1 confirms that the total CO_2 emission significantly decreases as the 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 1 demonstrates that when IGs are coal-fired generators and FGs are natural gas combustion turbines, economic curtailment increases CO_2 emission, but when IGs are nuclear power generators, economic curtailment reduces the emission.

7. Conclusion

Electricity markets have been gradually evolving toward deregulated structures that intend 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 paper provides new results that address how the competition is affected by generators' (in)flexibility and variable generation. The two most important messages from this paper 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 classic 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 paper 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 future even without subsidies.

References

Anderson, E. J. 2013. On the existence of supply function equilibria. *Mathematical Programming* **140**(2) 323–349.

Anderson, E. J., A. B. Philpott 2002. Using supply functions for offering generation into an electricity market. *Operations Research* **50**(3) 477–489.

Anderson, E. J., X. Hu 2008. Finding supply function equilibria with asymmetric firms. *Operations Research* **56**(3) 697–711.

Baldick, R., R. Grant, E. Kahn 2004. Theory and application of linear supply function equilibrium in electricity markets. *Journal of Regulatory Economics* **25**(2) 143–167.

Berry, C. A., B. F. Hobbs, W. A. Meroney, R. P. O'Neill, W. R. Stewart Jr 1999. Understanding how market power can arise in network competition: A game theoretic approach. *Utilities Policy* **8**(3) 139–158.

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.

Buygi, M. O., H. Zareipour, 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.

Debreu, G. 1952. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences* **38** 886–893.

Ela, E. 2009. Using economics to determine the efficient curtailment of wind energy. National Renewable Energy Laboratory, NREL/TP-550-45071.

Ela, E., 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, D. Corbus 2009. The evolution of wind power integration studies: Past, present, and future. *Proceedings of Power & Energy Society General Meeting, 2009.*

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.

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., S. S. Reynolds 2011. Supply function equilibria with capacity constraints and pivotal suppliers. *International Journal of Industrial Organization* **29**(4) 432–442.

Green, R. 1996. Increasing competition in the British electricity spot market. *Journal of Industrial Economics* **44**(2) 205–216.

Green, R. 1999. The electricity contract market in England and Wales. *Journal of Industrial Economics* **47**(1) 107–124.

Green, R. J., D. M. Newbery 1992. Competition in the British electricity spot market. *Journal of Political Economy* **100**(5) 929–953.

Hart, E. K., E. D. Stoutenburg, 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.

Heyman, D. P., M. J. Sobel 1984. *Stochastic Models in Operations Research*. Vol. 2. McGraw-Hill.

Hobbs, B. F., C. B. Metzler, J.-S. Pang 2000. Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Transactions on Power Systems* **15**(2) 638–645.

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., A. B. Philpott 2012. Supply function equilibria in networks with transport constraints. Working Paper, Research Institute of Industrial Economics.

Holmberg, P., D. Newbery 2010. The supply function equilibrium and its policy implications for wholesale electricity auctions. *Utilities Policy* **18**(4) 209–226.

Klemperer, P. D., M. A. Meyer 1989. Supply function equilibria in oligopoly under uncertainty. *Econometrica* **57**(6) 1243–1277.

Li, G., J. Shi, 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.

Midcontinent Independent System Operator (MISO) 2013. Business Practices Manual: Energy and Operating Reserve Markets. https://www.misoenergy.org/Library/BusinessPracticesManuals.

Renewable Energy Policy Network for the 21st Century (REN21) 2013. *Renewables Global Status Report*. Paris, France.

Rudkevich, A. 1999. Supply function equilibrium in power markets: Learning all the way. Tabors Caramanis & Associates Technical Paper 1299-1702.

Rudkevich, A., M. Duckworth, 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.

Sioshansi, R. 2011. Increasing the value of wind with energy storage. *Energy Journal* **32**(2) 1–29.

Sioshansi, R., 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, 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, J. R. Birge 2001. Equilibrium values in a competitive power exchange market. *Computational Economics* **17**(1) 93–121.

Ventosa, M., Á. Baíllo, A. Ramos, M. Rivier 2005. Electricity market modeling trends. *Energy Policy* **33**(7) 897–913.

Willems, B., I. Rumiantseva, H. Weigt 2009. Cournot versus supply functions: What does the data tell us?. *Energy Economics* **31**(1) 38–47.

Wilson, R. 2008. Supply function equilibrium in a constrained transmission system. *Operations Research* **56**(2) 369–382.

Wu, O. Q., R. Kapuscinski 2013. Curtailing intermittent generation in electrical systems. *Manufacturing & Service Operations Management* **15**(4) 578–595.

Appendix

Proof of Lemma 1. The inverse supply function is defined as $S_k^{-1}(q) = \inf\{p : S_k(p) > q\}$. Under Assumption 2(i) and (ii), $S_k^{-1}(0) = p_k^{\min}$ and $S_k^{-1}(q)$ is continuous and increasing in q for $q \geq 0$. Assumption 2(iii) implies that $S_k^{-1}(q) > 0$ for $q > 0$. Hence, $C_k(q) = \int_0^q S_k^{-1}(x)dx$ is continuously differentiable, convex, and strictly increasing in q for $q \geq 0$.

Consider the optimal allocation problem in (3), rewritten below

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

For any $\overline{q} > 0$, the objective in (A.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 (A.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 $G_+^F = \big\{j \in G^F : C_j'(q_j^*) = p\big\}$. Then, $C^F(q) = \sum_{j \in G_+^F} C_j(q_j^*)$. For $j \notin G_+^F$, we have $q_j^* = 0$ and $C_j'(0) > p$. Then, for sufficiently small $\varepsilon > 0$, we have

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

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

$$C^F(q + \varepsilon) = \sum_{j \in G_+^F} \Big[C_j(q_j^*) + \varepsilon_j C_j'(q_j^*) + o(\varepsilon_j)\Big] = C^F(q) + \varepsilon p + o(\varepsilon),$$

where $o(x)$ is a function $g(x)$ satisfying $g(x)/x \to 0$ as $x \to 0$. Similarly, we can show that $C^F(q) - C^F(q - \varepsilon) = \varepsilon p + o(\varepsilon)$. Hence, $C^F(q)$ is differentiable with derivative $(C^F)'(q) = p > 0$.

Finally, we show $p = (S^F)^{-1}(q)$. For $j \in G_+^F$, we have $p = C_j'(q_j^*) = S_j^{-1}(q_j^*)$ or $S_j(p) = q_j^*$. For $j \notin G_+^F$, we have $C_j'(0) > p$, which implies $S_j^{-1}(0) = p_j^{\min} > p$, which in turn leads to $S_j(p) = 0 = q_j^*$ due to Assumption 2(i). Hence, $S^F(p) = \sum_{j \in G^F} S_j(p) = \sum_{j \in G^F} q_j^* = q$, which leads to $p = (S^F)^{-1}(q)$. Because $S^F(p)$ also satisfies Assumption 2, $(S^F)^{-1}(q)$ is continuous in q. Therefore, $C^F(q)$ is continuously differentiable and $(C^F)'(q) = (S^F)^{-1}(q)$.

Similar results can be shown for IGs' problem in (2), which completes the proof. ∎

Proof of Theorem 1. We first prove that (12) is optimal in the case of $L_t - q^I - W_t \geq 0$. In this case, constraints (10)-(11) 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 (9).

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 (9); (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 (9). Hence, $q_t^{F*} = 0$. We determine q_t^{V*} by

$$\min\big\{ - r\, q_t^V + h(q^I + q_t^V - L_t) \ : \ 0 \leq q_t^V \leq W_t \big\},$$

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 $L_t - q^I + \mu(r) \geq W_t$, then $q_t^{V*} = W_t$. If $L_t - q^I + \mu(r) < 0$, then $q_t^{V*} = 0$. This proves that (12) is optimal.

For any $\overline{q} > \overline{L}$, the objective function in (9) is convex on a closed convex set $\{(q^I, L_t, W_t, q_t^F, q_t^V) : q^I \in [0, \overline{q}],\ L_t \in [\underline{L}, \overline{L}],\ W_t \in [0, K],\ q_t^F \in [0, \overline{q}],\ (10), \text{ and } (11)\}$. 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 2. 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 (16). Using Theorem 1, 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 1 under the four events. The indicators of these events can be written as

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

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 1, we have

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

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

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

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 (A.5) all have the form of $\mathsf{E}\big[g(q^I, L_t, W_t)\mathbf{1}_{L_t \le b(q^I, W_t)}\big]$, 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

$$
\begin{aligned}
\frac{d}{dq^I}\mathsf{E}\Big[g(q^I, L_t, W_t)\mathbf{1}_{L_t \le b(q^I, W_t)}\Big] &= \frac{d}{dq^I}\left[\int_0^K \int_{\underline{L}}^{\underline{L} \vee b(q^I, w) \wedge \overline{L}} g(q^I, l, w) f_t(l, w)\, dl\, dw\right] \\
= \mathsf{E}\left[\frac{\partial g(q^I, L_t, W_t)}{\partial q^I}\mathbf{1}_{L_t \le b(q^I, W_t)}\right] &+ \int_0^K g\big(q^I, b(q^I, w), w\big)\, f_t\big(b(q^I, w), w\big)\frac{\partial b(q^I, w)}{\partial q^I}\mathbf{1}_{b(q^I, w) \in [\underline{L}, \overline{L}]}\, dw. \qquad \text{(A.6)}
\end{aligned}
$$

The last expectation in (A.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

$$
\begin{aligned}
\frac{d}{dq^I}\mathsf{E}\Big[g(q^I, L_t, W_t)\mathbf{1}_{L_t > b(q^I, W_t)}\Big] &= \frac{d}{dq^I}\left[\int_0^K \int_{\underline{L} \vee b(q^I, w) \wedge \overline{L}}^{\overline{L}} g(q^I, l, w) f_t(l, w)\, dl\, dw\right] \\
= \mathsf{E}\left[\frac{\partial g(q^I, L_t, W_t)}{\partial q^I}\mathbf{1}_{L_t > b(q^I, W_t)}\right] &- \int_0^K g\big(q^I, b(q^I, w), w\big)\, f_t\big(b(q^I, w), w\big)\frac{\partial b(q^I, w)}{\partial q^I}\mathbf{1}_{b(q^I, w) \in [\underline{L}, \overline{L}]}\, dw. \qquad \text{(A.7)}
\end{aligned}
$$

Now, applying (A.6)-(A.7) to the derivatives of the expectations in (A.5), we find that the integral term $g\big(q^I, b(q^I, w), w\big)$ in (A.6)-(A.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 (A.4). This leads to

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

where the last equality follows from Lemma 1 and the definition in (14).

Now, consider the optimization problem in (16). If $q^{I*} > 0$, then it must satisfy

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

which is equivalent to $(C^I)'(q^{I*}) = \overline{P}(q^{I*}, S^F)$. Applying Lemma 1, we have $q^{I*} = S^I\big(\overline{P}(q^{I*}, S^F)\big)$.

If $q^{I*} = 0$, then $(C^I)'(0) \geq \overline{P}(0, S^F)$. Because $S^I(p) = 0$ for any $p \leq (C^I)'(0)$, we have $S^I\big(\overline{P}(0, S^F)\big) = 0$. Thus, $q^{I*} = S^I\big(\overline{P}(q^{I*}, S^F)\big)$ holds for $q^{I*} = 0$. This proves equation (17).

Finally, we show $P(q^I, L_t, W_t, S^F)$ has the alternative expression in (18). We verify this by considering separate regions. The inequality in (18) is

$$S^F(p) + W_t\mathbf{1}_{\{p \geq -r\}} - \mu(-p) \;\geq\; L_t - q^I. \tag{A.8}$$

In region A_1, $L_t - q^I \leq -\mu(r)$, and (A.8) clearly holds if $p = p^o \equiv -h'(q^I - L_t)$. Note that $p^o \leq -r$. Thus, for any other price $p_1 < p^o$, the left side of (A.8) becomes $-\mu(-p_1)$, which is strictly less than $L_t - q^I$. Hence, p^o is the minimum price for (A.8) to hold.

In region A_2, $L_t - q^I \in (-\mu(r), W_t - \mu(r))$. If $p = -r$, then (A.8) holds because $W_t - \mu(r) > L_t - q^I$. For any other $p_1 < -r$, (A.8) 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 (A.8) 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 (A.8) also holds with equality: $(L_t - W_t - q^I) + W_t = L_t - q^I$.

Hence, the minimum price p for (A.8) to hold is exactly $P(q^I, L_t, W_t, S^F)$. ∎

Proof of Lemma 2. Recall that the average price $\overline{P}(q^I, \beta^F)$ decreases in q^I, as we discussed after the definition in (15). 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 $\overline{P}_1 \leq 0$.

Equation (25), $q^I - \beta^I\overline{P}(q^I, \beta^F) = 0$, 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:

$$\begin{aligned} \frac{\partial q^I}{\partial \beta_i} &= \frac{\overline{P}}{1 - \beta^I\overline{P}_1} = \frac{q^I}{\beta^I(1 - \beta^I\overline{P}_1)} > 0, \qquad i \in G^I, \\ \frac{\partial q^I}{\partial \beta_j} &= \frac{\beta^I\overline{P}_2}{1 - \beta^I\overline{P}_1} \;<\; 0, \qquad j \in G^F, \end{aligned} \tag{A.9}$$

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 (32). 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 (24) as

$$\overline{P}(q^I,\beta^F) = \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - c_h\int_{q^I-\mu(r)}^{q^I}(q^I-x)f_D(x)\,dx - c_h\int_{-\infty}^{q^I-\mu(r)}(q^I-x)f_L(x)\,dx$$
$$- a_hF_D(q^I) + (a_h-r)F_D(q^I-\mu(r)) + (r-a_h)F_L(q^I-\mu(r)). \tag{A.10}$$

Thus,

$$\overline{P}_2 \equiv \frac{\partial\overline{P}}{\partial\beta^F} = -\frac{1}{(\beta^F)^2}\mathsf{E}\left[(D-q^I)^+\right] \;<\; 0,$$
$$\frac{\partial q^I}{\partial\beta_j} = \frac{\beta^I\overline{P}_2}{1-\beta^I\overline{P}_1} = -\frac{\beta^I\mathsf{E}\left[(D-q^I)^+\right]}{(\beta^F)^2\,(1-\beta^I\overline{P}_1)} \;<\; 0, \qquad j\in G^F. \tag{A.11}$$

This completes the proof. ∎

Proof of Lemma 3. We first bound the average price in (A.10). 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 (A.10) is bounded above by

$$\overline{P}(q^I,\beta^F) < \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - a_hF_D(q^I) + (a_h-r)F_D(q^I-\mu(r)). \tag{A.12}$$

If $a_h \geq r$, then $\mu(r)=0$ and (A.12) 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 (A.12) implies $\overline{P}(q^I,\beta^F) \;<\; \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - a_hF_D(q^I)$. Combining these two cases, we obtain

$$\overline{P}(q^I,\beta^F) \;<\; \frac{1}{\beta^F}\mathsf{E}\left[(D-q^I)^+\right] - \min\{r,a_h\}F_D(q^I). \tag{A.13}$$

We can express and bound $\mathsf{E}\left[(D-q^I)^+\right]$ as follows:

$$\begin{aligned}
\mathsf{E}\left[(D-q^I)^+\right] &= \int_{q^I}^{\infty}(x-\mu_D+\mu_D-q^I)f_D(x)\,dx\\
&= \int_{q^I}^{\infty}\frac{x-\mu_D}{\sqrt{2\pi}\sigma_D}\exp\left(-\frac{(x-\mu_D)^2}{2\sigma_D^2}\right)dx + \int_{q^I}^{\infty}(\mu_D-q^I)f_D(x)\,dx\\
&= \frac{\sigma_D}{\sqrt{2\pi}}\int_{\frac{q^I-\mu_D}{\sigma_D}}^{\infty} y\,\exp\left(-y^2/2\right)dy + (\mu_D-q^I)\overline{F}_D(q^I)\\
&\leq \frac{\sigma_D}{\sqrt{2\pi}} + (\mu_D-q^I)\overline{F}_D(q^I).
\end{aligned} \tag{A.14}$$

The inequalities (A.13) and (A.14) lead to

$$\overline{P}(q^I,\beta^F) \;<\; \frac{\sigma_D}{\sqrt{2\pi}\beta^F} + \frac{(\mu_D-q^I)\overline{F}_D(q^I)}{\beta^F} - \min\{r,a_h\}F_D(q^I). \tag{A.15}$$

Using (25), (27), and (A.15), we have

$$
\begin{aligned}
q^{I\max} &= \beta^{I\max}\overline{P}(q^{I\max}, \beta^{F\min}) \\
&< \beta^{I\max}\left[\frac{\sigma_D}{\sqrt{2\pi}\beta^{F\min}} + \frac{(\mu_D - q^{I\max})\overline{F}_D(q^{I\max})}{\beta^{F\min}} - \min\{r, a_h\}F_D(q^{I\max})\right].
\end{aligned} \tag{A.16}
$$

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 (A.16) 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 3. 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 (A.10) with respect to q^I and using $\mu(r) = (r - a_h)^+/c_h$, we obtain

$$
\begin{aligned}
\overline{P}_1(q^I, \beta^F) \equiv \frac{\partial \overline{P}}{\partial q^I} &= -\frac{1}{\beta^F}\overline{F}_D(q^I) - c_h\left[F_D(q^I) - F_D(q^I - \mu(r)) + F_L(q^I - \mu(r))\right] \\
&\quad - a_h f_D(q^I) + \left[c_h\mu(r) + (a_h - r)\right]\left[f_D(q^I - \mu(r)) - f_L(q^I - \mu(r))\right] \\
&= -\frac{1}{\beta^F}\overline{F}_D(q^I) - c_h\left[F_D(q^I) - F_D(q^I - \mu(r)) + F_L(q^I - \mu(r))\right] \\
&\quad - a_h f_D(q^I) + (a_h - r)^+\left[f_D(q^I - \mu(r)) - f_L(q^I - \mu(r))\right],
\end{aligned} \tag{A.17}
$$

$$
\begin{aligned}
\overline{P}_{11}(q^I, \beta^F) \equiv \frac{\partial^2 \overline{P}}{\partial {q^I}^2} &= \frac{1}{\beta^F} f_D(q^I) - c_h\left[f_D(q^I) - f_D(q^I - \mu(r)) + f_L(q^I - \mu(r))\right] \\
&\quad - a_h f_D'(q^I) + (a_h - r)^+\left[f_D'(q^I - \mu(r)) - f_L'(q^I - \mu(r))\right].
\end{aligned} \tag{A.18}
$$

By Lemma 3, 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 (A.17)-(A.18) 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 (29). To prove its quasi-concavity in β_i, we will show that its derivative $\partial\pi_i/\partial\beta_i$ can cross zero value from above at most once as β_i increases, while holding β_{-i} and β^F constant.

In (29), 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 (A.9). Differentiating (29) 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 (A.18). 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 (A.17) and (A.18), 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 (32), and denote $D = L - W$ and $q^I = Q^I(\beta^I,\beta^F)$, we can write FG's profit function in (31) 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 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 (A.11) and the following fact

$$\frac{\partial}{\partial q^I}\mathsf{E}\left[\big((D-q^I)^+\big)^2\right] = \frac{\partial}{\partial q^I}\int_{q^I}^{\infty}(x-q^I)^2 f_D(x)dx = \int_{q^I}^{\infty}-2(x-q^I)f_D(x)dx = -2\mathsf{E}\big[(D-q^I)^+\big],$$

we obtain

$$\begin{aligned}
\frac{\partial \pi_j}{\partial \beta_j} &= \frac{\beta_{-j}(1-c_j\beta_j)-\beta_j}{(\beta^F)^3}\mathsf{E}\left[\big((D-q^I)^+\big)^2\right] + \frac{\beta_j(2-c_j\beta_j)}{(\beta^F)^2}\mathsf{E}\left[(D-q^I)^+\right]\frac{\beta^I\mathsf{E}\left[(D-q^I)^+\right]}{(\beta^F)^2(1-\beta^I\overline{P}_1)} \\
&= \frac{\beta_j\mathsf{E}\big[((D-q^I)^+)^2\big]}{(\beta^F)^3(1-\beta^I\overline{P}_1)}\Big[Y(\beta_j,\beta_{-j},\beta^I) + \beta^I Z(\beta_j,\beta_{-j},\beta^I)\Big],
\end{aligned}$$

where

$$\begin{aligned}
Y(\beta_j,\beta_{-j},\beta^I) &\stackrel{\text{def}}{=} \Big(\frac{\beta_{-j}}{\beta_j} - (1+c_j\beta_{-j})\Big)(1-\beta^I\overline{P}_1), \\
Z(\beta_j,\beta_{-j},\beta^I) &\stackrel{\text{def}}{=} \frac{2-c_j\beta_j}{\beta_j+\beta_{-j}}\psi(q^I), \\
\psi(q^I) &\stackrel{\text{def}}{=} \frac{\big(\mathsf{E}\left[(D-q^I)^+\right]\big)^2}{\mathsf{E}\left[((D-q^I)^+)^2\right]}.
\end{aligned}$$

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

$$\psi'(q^I) = -\frac{2\mathsf{E}\big[(D-q^I)^+\big]\overline{F}_D(q^I)}{\mathsf{E}\big[\big((D-q^I)^+\big)^2\big]} + \frac{2\big(\mathsf{E}\big[(D-q^I)^+\big]\big)^3}{\big(\mathsf{E}\big[\big((D-q^I)^+\big)^2\big]\big)^2} = -\frac{2\psi(q^I)(\overline{F}_D(q^I)-\psi(q^I))}{\mathsf{E}[(D-q^I)^+]}.$$

Using this derivative and $\partial q^I/\partial\beta_j$ in (A.11), we have

$$\begin{aligned}
\frac{\partial Z}{\partial \beta_j} &= \frac{-c_j\beta^F - (2-c_j\beta_j)}{(\beta^F)^2}\,\psi(q^I) + \frac{2-c_j\beta_j}{\beta^F}\,\psi'(q^I)\frac{\partial q^I}{\partial \beta_j}\\
&= \frac{-c_j\beta^F - (2-c_j\beta_j)}{(\beta^F)^2}\,\psi(q^I) + \frac{2-c_j\beta_j}{\beta^F}\,\frac{2\psi(q^I)(\overline{F}_D(q^I)-\psi(q^I))\beta^I}{(\beta^F)^2(1-\beta^I\overline{P}_1)}\\
&= \frac{\psi(q^I)}{(\beta^F)^2}\Big[-c_j\beta^F - (2-c_j\beta_j)\Big(1-\frac{2(\overline{F}_D(q^I)-\psi(q^I))\beta^I}{\beta^F(1-\beta^I\overline{P}_1)}\Big)\Big].
\end{aligned}$$

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 3). 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. ∎