UNIVERSITY OF ALBERTA
FACULTY OF ARTS
Department of Economics

Working Paper No. 2020-05

Vertical Integration and Capacity Investment in the Electricity Sector

David P. Brown
University of Alberta

David E. M. Sappington
University of Florida

April 2020

Vertical Integration and Capacity Investment in the Electricity Sector

by

David P. Brown* and David E. M. Sappington**

Abstract

We examine the incentives for and the effects of vertical integration in the electricity sector. We find that vertical integration generally reduces retail prices and increases industry capacity investment, consumer surplus, and total welfare. Unilateral vertical integration is profitable, so it arises in equilibrium. However, ubiquitous vertical integration can reduce aggregate industry profit.

Keywords: vertical integration; capacity investment; electricity sector

JEL Codes: L51, L94, Q28, Q40

April 2020

* Department of Economics, University of Alberta, Edmonton, Alberta T6G 2H4 Canada (dpbrown@ualberta.ca).

** Department of Economics, University of Florida, Gainesville, FL 32611 USA USA (sapping@ufl.edu).

Support from the Government of Canada's *Canada First Research Excellence Fund* under the Future Energy Systems Research Initiative is gratefully acknowledged.

1 Introduction

In electricity sectors around the world, competition continues to develop at both the wholesale and retail stages of production, and renewable generation is expanding (Borenstein and Bushnell, 2015; Newbery et al., 2018; Halkos, 2019). Increasing competition and expanding output by intermittent generators raise questions about whether appropriate incentives for investment in dispatchable generation prevail. An important related question is whether the vertical integration of wholesale and retail suppliers of electricity hinders or promotes such investment.

A primary purpose of this research is to determine whether vertical integration promotes investment in generation capacity in the presence of wholesale competition, retail competition, and uncertain net market demand (i.e., market demand net of intermittent renewables that is met by dispatchable generation). We also examine the impact of vertical integration on retail and wholesale electricity prices, consumer surplus, and total industry welfare. In addition, we examine the incentives for vertical integration by industry suppliers. We analyze these issues in a stylized setting with two suppliers of dispatchable generation (G1 and G2) and two retail suppliers of electricity (R1 and R2). We consider three industry structures. Under vertical separation, each firm acts independently, and no firm has any financial interest in any other firm. Under full vertical integration, R1 and G1 are vertically integrated, as are R2 and G2. Under partial vertical integration, R1 and G1 are vertically integrated, whereas R2 and G2 are independent enterprises.[1]

Under all three industry structures, G1 and G2 first choose their capacity investments. R1 and R2 then set their retail prices. Next, after the net market demand for electricity is realized, G1 and G2 decide how much electricity to supply to the wholesale market. Finally, after the market-clearing wholesale price is determined, R1 and R2 purchase the amount of electricity required to serve the demands of their retail customers.

We find that vertical integration generally leads to increased capacity investment and

[1] These three structures all prevail in practice (Kury, 2013; Halkos, 2019).

expanded supply of electricity. It does so because the increased investment and corresponding output supplied by, say, G1 reduces the wholesale price of electricity (w). The reduction in w increases R1's retail profit, which G1 values under vertical integration.

We also find that vertical integration generally promotes lower retail prices for electricity for two reasons. First, a lower retail price increases the demand for electricity, which raises its wholesale price. The increase in w enhances the profit of the wholesale division of the integrated enterprise. Second, when a retail firm (e.g., R1) reduces its retail price, its retail output increases. This increased output ensures that R1's retail profit increases more rapidly as the wholesale price of electricity (w) declines. Consequently, when G1 and R1 are vertically integrated, G1's incentive to reduce w increases. This increased incentive endows G1 with a credible commitment to compete aggressively against G2 when choosing its capacity investment and its subsequent supply of electricity. This commitment serves to increase G1's equilibrium profit, which R1 values under vertical integration.[2]

In part because of this commitment effect, unilateral vertical integration by R1 and G1 generally increases their combined profit above the level of profit they secure under vertical separation (VS). In contrast, the combined profit of (independent firms) R2 and G2 generally declines under partial vertical integration (PVI). Once R1 and G1 have merged their operations, R2 and G2 generally find it more profitable to combine their operations than to compete as independent suppliers. Doing so enhances the commitments of both R2 and G2 to compete more aggressively against their rivals.

Because the combined profit of R1 and G1 is higher under PVI than under VS and because the combined profit of R2 and G2 is higher under full vertical integration (VI) than under PVI, VI arises in the absence of barriers to integration. Expected consumer surplus and expected total welfare generally are both higher under full vertical integration than under VS. However, expected industry profit can be lower under VI than under VS (because of the

[2]Vertical integration can either increase or reduce the wholesale price of electricity in our model. This is the case because vertical integration generally increases both the wholesale demand for electricity and the wholesale supply of electricity.

relatively intense competition that prevails under full vertical integration). Thus, vertical integration can give rise to prisoner's dilemma considerations (e.g., Fudenberg and Tirole, 1991, pp. 9-10) because even though unilateral vertical integration is profitable, ubiquitous vertical integration can systematically reduce profit.

The vertical integration that arises in equilibrium in our model is consistent with the extensive vertical integration that is observed in practice in many jurisdictions. To illustrate, Meade and O'Connor (2011) report particularly pronounced vertical integration in Spain and New Zealand, and substantial vertical integration in regions of Australia, the United States, and the United Kingdom.

We develop and explain our findings as follows. Section 2 describes the key elements of our formal model. Section 3 examines the impact of vertical integration on the wholesale supply of electricity, taking capacity investment as given. Section 4 analyzes the impact of vertical integration on industry investment. Section 5 determines how vertical integration affects retail electricity prices. The complexity of the analysis limits the number of general analytic conclusions that can be established, so Section 6 examines numerical solutions to the problems under consideration. Section 7 summarizes our primary findings and discusses extensions of our work. The Appendix provides the proofs of all formal conclusions.

Before proceeding, we briefly explain the relationship between our work and its predecessors. Several studies (e.g., Pérez, 2007; de Braganca and Daglish, 2017) examine the effects of vertical integration in settings where investment in dispatchable generation capacity is specified exogenously. Other studies examine incentives for capacity investment, taking the prevailing industry structure as given. To illustrate, Murphy and Smeers (2005), Fabra et al. (2011), and Grimm and Zoettl (2013) examine these incentives under vertical separation. Boom (2009) explores the corresponding incentives in a setting where all industry suppliers are vertically integrated.[3]

[3]Buehler and Schmutzler (2008) demonstrate that vertical integration can reduce downstream cost-reducing investment. The reduced investment arises when a firm's equilibrium output declines as the elimination of double marginalization induces a vertically-integrated rival to expand its downstream output.

Our analysis is most closely related to the important work of Boom and Buehler (2020). The authors analyze endogenous capacity investment in the electricity sector under both vertical integration and vertical separation.[4] They do so in a setting where each generator bids a uniform price at which it is willing to supply its entire capacity.[5] This institutional arrangement leaves a vertically-integrated firm susceptible to having to purchase electricity at a high wholesale price if its retail commitment exceeds its capacity. Consequently, vertically-integrated firms set relatively high retail prices in order to restrict retail demand and thereby reduce the likelihood of having to purchase electricity in the wholesale market. In contrast, vertical integration promotes lower retail prices in our model because a low retail price enhances the commitment of the integrated generator to expand its (endogenous, divisible) supply of electricity.

Our finding that vertical integration can sometimes entail prisoner's dilemma considerations is reminiscent of Allaz and Vila (AV) (1993)'s corresponding conclusion regarding forward contracting. Unilateral forward contracting is profitable in AV's model because it induces the wholesale supplier to compete relatively aggressively. As noted above, the lower retail price induced by vertical integration can provide a corresponding commitment effect in our model. However, the nature and details of the commitment effects differ under forward contracting and vertical integration.[6] Furthermore, although ubiquitous vertical integration can sometimes reduce industry profit, it does not always do so.

To facilitate the simultaneous consideration of endogenous industry structure and endogenous generation capacity in the presence of imperfect wholesale competition, imperfect retail competition, and uncertain (net) market demand, we abstract from some elements of the electricity sector that other studies incorporate. Specifically, we do not consider forward

[4]In contrast to Boom and Buehler (2020), we identify the market structure (VS, PVI, or VI) that prevails in equilibrium.

[5]In contrast, we allow each generator to choose the amount of electricity (up to capacity) it will supply in the wholesale market.

[6]The effects in both models reflect the considerations identified in the seminal work of Fershtman and Judd (1987) and Sklivas (1987).

contracting,[7] supplier risk aversion, or risk hedging.[8] As explained further in the concluding discussion, the primary forces that underlie our key qualitative findings seem likely to prevail in the presence of these additional elements of the electricity sector.

2 The Model

We analyze a setting with two wholesale suppliers (G1 and G2) and two retail suppliers (R1 and R2) of electricity. Wholesale supplier (or "generator") Gi can produce electricity at constant unit cost $c_i > 0$ up to capacity $K_i \geq 0$ (for $i \in \{1, 2\}$). Generator Gi's unit cost of capacity is $k_i > 0$.

Retail supplier (or "retailer") Ri's unit cost of supplying electricity to retail customers is $w + c_i^r$ (for $i \in \{1, 2\}$), where w is the wholesale price of electricity and c_i^r is Ri's incremental unit downstream production cost. The demand for Ri's product is:

$$Q_i^r(r_1, r_2) \;=\; a_i - b_i\, r_i + d_i\, r_j \quad \text{for} \;\; i,\, j \in \{1, 2\} \;(j \neq i), \tag{1}$$

where r_i is Ri's retail price, r_j is Rj's retail price, and $a_i > 0$, $b_i > 0$, and $d_i > 0$ are parameters. This demand formulation reflects the fact that, in practice, retail consumers often view electricity supplied by different retail firms to be differentiated products. The perceived differentiation can reflect incumbency advantages, differences in supply reliability, and variations in customer service, for example (Woo et al., 2014; Hortacsu et al., 2017). In standard fashion, we assume $\min\{b_1, b_2\} > \max\{d_1, d_2\}$ to ensure the demand for each firm's product is more sensitive to its own price than to its rival's price. Consumers choose their preferred retail supplier after observing the prices set by both firms.

R1 and R2 always purchase at wholesale price w the amount of electricity required to serve the demand for their product. To introduce meaningful variation in the (net) demand for

[7] Under forward contracting in the electricity sector, a wholesale supplier commits to deliver a specified amount of electricity to a retail firm at a pre-specified price. The commitment is made (via binding contract) before the final demand for electricity is determined. Models in which forward contracting plays a central role include Allaz and Vila (1993), Powell (1993), Green (2003), Bushnell (2007), Bushnell et al. (2008), Gans and Wolak (2012), and Li (2014).

[8] Models in which risk aversion and risk hedging play a central role include Aïd et al. (2011) and Baldursson and Von der Fehr (2007).

(dispatchable) electricity, we assume that large industrial customers also purchase electricity at wholesale price w. The industrial customers' demand for wholesale electricity at price w is:

$$Q^I(w) \;=\; a^I - b^I\, w \,+ \eta\,, \tag{2}$$

where $a^I > 0$ and $b^I > 0$ are parameters and where η is the realization of a stochastic demand state.[9] Equations (1) and (2) imply that the aggregate demand for wholesale electricity is:

$$Q(w) \;=\; Q_1^r(\cdot) + Q_2^r(\cdot) + Q^I(w) \;=\; Q_1^r(\cdot) + Q_2^r(\cdot) + a^I - b^I\, w \,+ \eta\,. \tag{3}$$

The corresponding inverse demand for wholesale electricity is:

$$w(Q) \;=\; b^w[\, a^I + Q_1^r(\cdot) + Q_2^r(\cdot) - Q\,] + \varepsilon \quad \text{where} \;\; b^w \;\equiv\; \frac{1}{b^I} \;\; \text{and} \; \varepsilon \;\equiv\; \frac{\eta}{b^I}\,. \tag{4}$$

$H(\varepsilon)$ will denote the distribution function for the random demand state $\varepsilon \in [\,\underline{\varepsilon}, \overline{\varepsilon}\,]$. $h(\varepsilon)$ will denote the corresponding density function.

Gi's (wholesale) profit is:

$$\pi_i^G \;=\; [\, w - c_i\,]\, q_i - k_i\, K_i \tag{5}$$

where q_i is Gi's output of wholesale electricity. Ri's corresponding (retail) profit is:

$$\pi_i^R \;=\; [\, r_i - c_i^r - w\,]\, Q_i^r(r_1, r_2)\,. \tag{6}$$

The wholesale price of electricity (w) is the price that equates the aggregate demand for and the aggregate supply of wholesale electricity. Aggregate supply is $q_1 + q_2$, the sum of the outputs of G1 and G2. The wholesale suppliers select their outputs after choosing their capacities (K_i) and after ε is realized, so the aggregate demand for wholesale electricity is known.[10] We examine settings in which each retailer serves some customers in equilibrium (so $Q_i^r(\cdot) > 0$ for $i \in (1, 2\}$) and each generator supplies some wholesale electricity in equilibrium (so $q_i > 0$ for $i \in (1, 2\}$) for each realization of ε.

We will compare outcomes under three industrial structures: vertical separation (VS),

[9] $Q^I(\cdot)$ need not literally reflect solely the demand of large industrial customers. More generally, $Q^I(\cdot)$ can be viewed as the net demand for electricity by all entities other than R1 and R2 after accounting for the aggregate supply of electricity from all renewable sources.

[10] Other studies that employ Cournot competition to model the interaction among electricity generators include Ventosa et al. (2005), Brown and Eckert (2017), and Lundin and Tangeras (2020).

full vertical integration (VI), and partial vertical integration (PVI). Under VS, every industry supplier acts independently, and no industry supplier has any financial interest in another supplier. Under VI, R1 and G1 seek to maximize their joint profit ($\pi_1^R + \pi_1^G$). Similarly, R2 and G2 act to maximize their joint profit ($\pi_2^R + \pi_2^G$). Under PVI, R1 and G1 are vertically integrated and so seek to maximize their joint profit ($\pi_1^R + \pi_1^G$). In contrast, R2 and G2 act independently. Specifically, R2 acts to maximize π_2^R and G2 acts to maximize π_2^G.

The timing in the model under all industrial structures is as follows. (See Figure 1.) First, G1 and G2 choose their capacities (K_1 and K_2) simultaneously and independently. Second, R1 and R2 set their retail prices (r_1 and r_2) simultaneously and independently. Third, the demand state (ε) is realized, so the aggregate demand for electricity ($Q(\cdot)$) is determined. Fourth, G1 and G2 choose their outputs (q_1 and q_2) simultaneously and independently. Fifth, after the wholesale price of electricity (w) is determined (by equating demand and supply), R1 and R2 each purchase the electricity required to serve the demand of its retail customers.

G1 and G2 choose K_1 and K_2	R1 and R2 set r_1 and r_2	Demand state ε is realized	G1 and G2 choose q_1 and q_2	w is determined	R1 and R2 purchase Q_1^r and Q_2^r

------|------------------|------------------|---------------|----------------|------------------|-------

Figure 1. Timing in the Model.

3 Wholesale Electricity Supply

We first examine the impact of vertical integration on the wholesale supply of electricity, taking capacity investment as given. The analysis is facilitated by characterizing equilibrium outcomes under all three industrial structures simultaneously. To do so, we introduce the parameter $\alpha_i \in \{0, 1\}$ for $i \in \{1, 2\}$ and we view Gi as acting to maximize $\pi_i^G + \alpha_i \pi_i^R$. We also view R$i$ as acting to maximize $\pi_i^R + \alpha_i \pi_i^G$. VS prevails when $\alpha_1 = \alpha_2 = 0$. VI prevails when $\alpha_1 = \alpha_2 = 1$. PVI prevails when $\alpha_1 = 1$ and $\alpha_2 = 0$.

After choosing its capacity K_i and after observing retail prices (r_1, r_2) and the realized demand for wholesale electricity, Gi chooses q_i to:

$$\underset{q_i \leq K_i}{\text{Maximize}} \quad [\,w(\cdot) - c_i\,]\, q_i + \alpha_i\, [\, r_i - c_i^r - w(\cdot)\,]\, Q_i^r(r_1, r_2). \tag{7}$$

Differentiating expression (7), using equation (4), implies that when it is not capacity-constrained, Gi's reaction function is:[11]

$$q_i \;=\; \frac{1}{2}\left[\, a^I + Q_i^r\,(1+\alpha_i) + Q_j^r\,\right] + \frac{\varepsilon - c_i}{2\, b^w} - \frac{1}{2}\, q_j\,. \tag{8}$$

Equations (4) and (8) underlie the observations in Lemmas 1 – 3. The lemmas refer to $q_i^{nc}(\varepsilon)$ and $w^{nc}(\varepsilon)$, which denote the equilibrium output of Gi and the equilibrium wholesale price of electricity, respectively, when demand state ε is realized and n generators are capacity-constrained ($n \in \{\,0, 1, 2\,\}$).

Lemma 1. *Suppose neither G1 nor G2 is capacity-constrained. Then in equilibrium, for $i, j \in \{1, 2\}$ $(j \neq i)$:*

$$q_i^{0c}(\varepsilon) \;=\; \frac{b^w\left[\, a^I + (1 + 2\,\alpha_i)\, Q_i^r + (1-\alpha_j)\, Q_j^r\,\right] + \varepsilon + c_j - 2\,c_i}{3\, b^w}\,; \;\; \textit{and}$$

$$w^{0c}(\varepsilon) \;=\; \frac{1}{3}\left[\, b^w\left(a^I + [\,1-\alpha_1\,]\, Q_1^r + [\,1-\alpha_2\,]\, Q_2^r \right) + \varepsilon + c_1 + c_2\,\right]. \tag{9}$$

Define ε_0 to the largest realization of ε for which neither generator is capacity-constrained. Also define ε_{12} to be the smallest realization of ε for which both generators are capacity-constrained.[12]

Lemma 2. *Suppose only Gj is capacity-constrained for $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$. Then in equilibrium, for $i, j \in \{1, 2\}$ $(j \neq i)$:*

$$q_j^{1c}(\varepsilon) \;=\; K_j\,;\;\; q_i^{1c}(\varepsilon) \;=\; \frac{1}{2}\left[\, a^I + Q_i^r\,(1+\alpha_i) + Q_j^r\,\right] + \frac{1}{2\, b^w}\,[\,\varepsilon - c_i\,] - \frac{1}{2}\, K_j\,; \;\; \textit{and}$$

$$w^{1c}(\varepsilon) \;=\; \frac{1}{2}\, b^w\left[\, a^I + Q_i^r\,(1-\alpha_i) + Q_j^r\,\right] + \frac{1}{2}\,[\,\varepsilon + c_i - b^w\, K_j\,]\,. \tag{10}$$

[11]The proof of Lemma 1 provides a formal derivation of this conclusion.

[12]For expositional ease, we focus on settings in which $\varepsilon_0, \varepsilon_{12} \in (\,\underline{\varepsilon}, \overline{\varepsilon}\,)$.

Lemma 3. *Suppose G1 and G2 are both capacity-constrained. Then in equilibrium:*

$$q_1^{2c}(\varepsilon) \;=\; K_1\,; \;\; q_2^{2c}(\varepsilon) \;=\; K_2\,; \text{ and } \;\; w^{2c}(\varepsilon) \;=\; b^w \left[\, a^I + Q_1^r + Q_2^r - K_1 - K_2 \,\right] + \varepsilon\,. \qquad (11)$$

Conclusions 1 and 2, which follow directly from Lemmas 1 – 3, explain how changes in the retail demand for electricity affect the wholesale supply of electricity under VS and VI.

Conclusion 1. *Suppose (only) Gj is capacity-constrained when* $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$*. Then under VS, for* $i, j \in \{1, 2\}$ $(j \neq i)$*: (i)* $\frac{\partial q_i^{0c}}{\partial Q_i^r} = \frac{\partial q_i^{0c}}{\partial Q_j^r} = \frac{1}{3}$*; (ii)* $\frac{\partial q_i^{1c}}{\partial Q_i^r} = \frac{\partial q_i^{1c}}{\partial Q_j^r} = \frac{1}{2}$*; (iii)* $\frac{\partial w^{0c}}{\partial Q_i^r} = \frac{\partial w^{0c}}{\partial Q_j^r} = \frac{1}{3}\, b^w$*; (iv)* $\frac{\partial w^{1c}}{\partial Q_i^r} = \frac{\partial w^{1c}}{\partial Q_j^r} = \frac{1}{2}\, b^w$*; and (v)* $\frac{\partial w^{2c}}{\partial Q_i^r} = \frac{\partial w^{2c}}{\partial Q_j^r} = b^w$*.*

Conclusion 2. *Suppose (only) Gj is capacity-constrained when* $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$*. Then under VI, for* $i, j \in \{1, 2\}$ $(j \neq i)$*: (i)* $\frac{\partial q_i^{0c}}{\partial Q_i^r} = 1$ *and* $\frac{\partial q_i^{0c}}{\partial Q_j^r} = 0$*; (ii)* $\frac{\partial q_i^{1c}}{\partial Q_i^r} = 1$ *and* $\frac{\partial q_i^{1c}}{\partial Q_j^r} = \frac{1}{2}$*; (iii)* $\frac{\partial w^{0c}}{\partial Q_i^r} = \frac{\partial w^{0c}}{\partial Q_j^r} = 0$*; (iv)* $\frac{\partial w^{1c}}{\partial Q_i^r} = 0$ *and* $\frac{\partial w^{1c}}{\partial Q_j^r} = \frac{1}{2}\, b^w$*; and (v)* $\frac{\partial w^{2c}}{\partial Q_i^r} = \frac{\partial w^{2c}}{\partial Q_j^r} = b^w$*.*

Conclusions 1 and 2 reflect the following considerations. Under VS, an increase in the retail demand for electricity (Q_1^r or Q_2^r) increases the wholesale demand for electricity, which increases the equilibrium wholesale price of electricity and induces both generators to increase their supply of electricity.

Gi's output (q_i) increases more rapidly as Ri's retail output (Q_i^r) increases when Gi and Ri are vertically integrated than when they act independently. This is the case because as Q_i^r increases, Ri's retail profit increases more rapidly as w declines. (Ri's retail profit margin increases as w declines.) When Gi values Ri's retail profit, Gi responds to the increased benefit of reducing w by expanding its supply of electricity. Specifically, when Gi is not capacity-constrained, it increases q_i at the same rate that Q_i^r increases, which leaves the wholesale price of electricity unchanged, *ceteris paribus*.

When Gi and Ri are vertically-integrated, Gi increases its output more rapidly as Q_i^r increases than as Q_j^r increases.[13] This is the case because Gi values Ri's retail profit, but not

[13] From Conclusion 2, $\frac{\partial q_i^{0c}}{\partial Q_j^r} < \frac{\partial q_i^{0c}}{\partial Q_i^r}$ and $\frac{\partial q_i^{1c}}{\partial Q_j^r} < \frac{\partial q_i^{1c}}{\partial Q_i^r}$ when Gj is capacity-constrained when $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$.

Rj's retail profit, under VI. Therefore, an increase in Q_i^r enhances Gi's incentive to increase q_i in a manner that an increase in Q_j^r does not. Specifically, an increase in Q_i^r increases the rate at which Ri's retail profit increases as w declines, which motivates Gi to reduce w by increasing q_i.

These considerations help to explain how vertical integration affects a generator's supply of electricity.

Conclusion 3. *When Gi is not capacity-constrained, it produces more electricity (thereby reducing the wholesale price of electricity) when Gi is vertically-integrated with Ri than when they are not vertically-integrated, ceteris paribus.*

Conclusion 3 reflects the fact that when Gi values Ri's retail profit, the generator acts to increase this profit. It does so by increasing its output of electricity, which reduces the wholesale price of electricity, and thereby increases Ri's profit margin.

4 Capacity Investment

To determine how vertical integration affects industry investment, observe that Gi values Ri's expected retail profit ($E\left\{\pi_i^R\right\}$) under VI. From equation (6):

$$E\left\{\pi_i^R\right\} = [\, r_i - c_i^r - E\left\{w(\varepsilon)\right\}\,]\, Q_i^r(r_1, r_2)$$

$$\text{where } \; E\left\{w(\varepsilon)\right\} \equiv \int_{\underline{\varepsilon}}^{\varepsilon_0} w^{0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} w^{1c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} w^{2c}(\varepsilon)\, dH(\varepsilon)\,. \tag{12}$$

Lemma 4 identifies the rate at which $E\left\{\pi_i^R\right\}$ varies with Gi's capacity investment, K_i.

Lemma 4.
$$\frac{\partial E\left\{\pi_i^R\right\}}{\partial K_i} = -\, Q_1^r\, \frac{\partial E\left\{w(\varepsilon)\right\}}{\partial K_i} + [\, r_i - E\left\{w(\varepsilon)\right\} - c_i^r\,] \left[-\, b_i\, \frac{\partial r_i}{\partial K_i} + d_j\, \frac{\partial r_j}{\partial K_i} \right] + Q_i^r\, \frac{\partial r_i}{\partial K_i}\,. \tag{13}$$

Lemma 4 indicates that Gi's concern with $E\left\{\pi_i^R\right\}$ under VI introduces two distinct effects of a change in K_i: a wholesale price effect (reflected in the first term following the equality in equation (13)) and a retail price effect (captured by the remaining terms following the equality in equation (13)). Conclusion 4 implies that the wholesale price effect alone leads Gi to increase K_i.

Conclusion 4. $$\frac{\partial E\{w(\varepsilon)\}}{\partial K_i} = \psi_i \left[-\frac{1}{2}\, b^w \right] [\, H(\varepsilon_{12}) - H(\varepsilon_0)\,] - b^w\, [\, 1 - H(\varepsilon_{12})\,] \ < \ 0\,,$$

$$\textit{where} \ \ \psi_i \ = \ \begin{cases} 1 & \textit{if Gi is capacity-constrained when } \varepsilon \in (\varepsilon_0, \varepsilon_{12}) \\ 0 & \textit{if Gj is capacity-constrained when } \varepsilon \in (\varepsilon_0, \varepsilon_{12})\,. \end{cases}$$

An increase in K_i permits an expanded supply of electricity when Gi is capacity-constrained. The increased supply of electricity lowers the wholesale price of electricity, and thereby enhances Ri's profit, *ceteris paribus*.

The direction of the retail price effect is more difficult to characterize because of the retailers' varied considerations when choosing retail prices. These considerations are analyzed next.

5 Retail Electricity Prices.

Equilibrium retail prices are not difficult to characterize implicitly under VS. When this industrial structure prevails, Ri chooses r_i to maximize $E\left\{\pi_i^R\right\}$. From equations (1) and (12), for $i,\, j \in \{1, 2\}$ $(j \neq i)$:

$$\frac{\partial E\left\{\, \pi_i^R \,\right\}}{\partial r_i} \ = \ 0 \ \Leftrightarrow \ [\, a_i - b_i\, r_i + d_i\, r_j\,] \left[1 - \frac{\partial E\{\, w(\varepsilon)\,\}}{\partial\, r_i} \right] \ = \ b_i\, [\, r_i - c_i^r - E\{\, w(\varepsilon)\,\}\,]\,. \tag{14}$$

Lemmas 1 – 3 imply that $E\{\, w(\varepsilon)\,\}$ and $\frac{\partial E\{\, w(\varepsilon)\,\}}{\partial\, r_i}$ are functions of r_1, r_2, and $H(\cdot)$. Consequently, the implicit characterization of retail prices in equation (14) does not readily give rise to closed-forms solutions for the equilibrium values of r_1 and r_2.

Retail prices are considerably more challenging to characterize (even implicitly) when one or both retailers are vertically-integrated. When R_i is vertically-integrated, it chooses r_i to maximize:

$$E\{\pi_i\} \;=\; \int_{\underline{\varepsilon}}^{\varepsilon_0} \Pi_i^{0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \Pi_i^{1c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \Pi_i^{2c}(\varepsilon)\, dH(\varepsilon) \tag{15}$$

where, from equations (5) and (6), for $n \in \{0, 1, 2\}$:

$$\Pi_i^{nc}(\varepsilon) \;=\; [\, w^{nc}(\varepsilon) - c_i\,]\, q_i^{nc}(\varepsilon) - k_i\, K_i + [\, r_i - c_i^r - w^{nc}(\varepsilon)\,]\, Q_i^r\,. \tag{16}$$

$\Pi_i^{nc}(\varepsilon)$ is the equilibrium combined profit of Ri and Gi when ε is realized and n generators are capacity-constrained. Viewing this profit as a function of $q_1^{nc}(\varepsilon)$, $q_2^{nc}(\varepsilon)$, Q_i^r, and r_i,[14] the rate at which $\Pi_i^{nc}(\varepsilon)$ varies with r_i can be expressed as:

$$\frac{d\Pi_i^{nc}(\varepsilon)}{dr_i} \;=\; \frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial q_i^{nc}(\varepsilon)}\, \frac{\partial q_i^{nc}(\varepsilon)}{dr_i} + \frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial q_j^{nc}(\varepsilon)}\, \frac{\partial q_j^{nc}(\varepsilon)}{dr_i} + \frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial Q_i^r}\, \frac{\partial Q_i^r}{\partial r_i} + \frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial r_i}\,. \tag{17}$$

The four terms to the right of the equality in equation (17) capture the impact a change in r_i has on $\Pi_i^{nc}(\cdot)$ by affecting: (i) Gi's output; (ii) Gj's output; (iii) Ri's output; and (iv) Ri's revenue (holding Q_i^r constant) and Rj's output, which affects $\Pi_i^{nc}(\cdot)$ by altering the wholesale price of electricity.[15]

The following four observations help to further characterize the derivatives in equation (17) under VI. First, $\frac{\partial q_i^{0c}(\varepsilon)}{dr_i} = \frac{\partial Q_i^r}{\partial r_i} = -\, b_i$ and $\frac{\partial q_j^{0c}(\varepsilon)}{dr_i} = d_j$, from equations (1) and (9). Second, $\frac{\partial q_1^{2c}(\varepsilon)}{dr_i} = \frac{\partial q_2^{2c}(\varepsilon)}{dr_i} = 0$ for $i \in \{1, 2\}$ because outputs of wholesale electricity do not vary with retail prices when generators are capacity-constrained. Third, equations (1) and (10) imply that for $j \neq i$, $\frac{\partial q_j^{1c}(\varepsilon)}{dr_i} = d_j - \frac{1}{2}\, b_i$ when only Gi is capacity-constrained, whereas $\frac{\partial q_j^{1c}(\varepsilon)}{dr_i} = 0$ when only Gj is capacity-constrained. Fourth, $\frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial q_i^{nc}(\varepsilon)} = 0$ for $i \in \{1, 2\}$ because Gi chooses $q_i^{nc}(\varepsilon)$ to maximize $\Pi_i^{nc}(\varepsilon)$ under VI.

These observations and equations (15) and (17) imply that Ri's choice of r_i under VI is determined by:

[14] $\Pi_i^{nc}(\varepsilon)$ is also a function of K_i. However, K_i is not a function of r_i because Gi chooses K_i before Ri chooses r_i.

[15] Equations (4) and (16) imply that for $i, j \in \{1, 2\}$ $(j \neq i)$: (i) $\frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial q_j^{nc}(\varepsilon)} = \frac{\partial w^{nc}(\varepsilon)}{\partial q_j(\varepsilon)}\, [\, q_i^{nc}(\varepsilon) - Q_i^r\,] = -\, b^w\, [\, q_i^{nc}(\varepsilon) - Q_i^r\,]$; (ii) $\frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial Q_i^r} = \frac{\partial w^{nc}(\varepsilon)}{\partial Q_i^r}\, [\, q_i^{nc}(\varepsilon) - Q_i^r\,] + r_i - c_i^r - w^{nc}(\varepsilon) \;=\; b^w\, [\, q_i^{nc}(\varepsilon) - Q_i^r\,] + r_i - c_i^r - w^{nc}(\varepsilon)$; and (iii) $\frac{\partial\Pi_i^{nc}(\varepsilon)}{\partial r_i} = Q_i^r + \frac{\partial w^{nc}(\varepsilon)}{\partial Q_j^r}\, \frac{\partial Q_j^r}{\partial r_i}\, [\, q_i^{nc}(\varepsilon) - Q_i^r\,] \;=\; b^w\, d_j\, q_i^{nc}(\varepsilon) + [\, 1 - b^w\, d_j\,]\, Q_i^r$.

$$\int_{\underline{\varepsilon}}^{\varepsilon_0} \left[d_j \frac{\partial \Pi_i^{0c}(\varepsilon)}{\partial q_j^{0c}(\varepsilon)} - b_i \frac{\partial \Pi_i^{0c}(\varepsilon)}{\partial Q_i^r} + \frac{\partial \Pi_i^{0c}(\varepsilon)}{\partial r_i} \right] dH(\varepsilon)$$

$$+ \int_{\varepsilon_0}^{\varepsilon_{12}} \left[\psi_j^{1c} \frac{\partial \Pi_i^{1c}(\varepsilon)}{\partial q_j^{1c}(\varepsilon)} - b_i \frac{\partial \Pi_i^{1c}(\varepsilon)}{\partial Q_i^r} + \frac{\partial \Pi_i^{1c}(\varepsilon)}{\partial r_i} \right] dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \left[- b_i \frac{\partial \Pi_i^{2c}(\varepsilon)}{\partial Q_i^r} + \frac{\partial \Pi_i^{2c}(\varepsilon)}{\partial r_i} \right] dH(\varepsilon)$$

$$+ \left[\Pi_i^{0c}(\varepsilon_0) - \Pi_i^{1c}(\varepsilon_0) \right] h(\varepsilon_0) \frac{d\varepsilon_0}{dr_i} + \left[\Pi_i^{1c}(\varepsilon_{12}) - \Pi_i^{2c}(\varepsilon_{12}) \right] h(\varepsilon_{12}) \frac{d\varepsilon_{12}}{dr_i} \;=\; 0 \tag{18}$$

$$\text{where} \quad \psi_j^{1c} \;=\; \begin{cases} d_j - \frac{1}{2} b_i & \text{if G}i \text{ is capacity-constrained for } \varepsilon \in (\varepsilon_0, \varepsilon_{12}) \\ 0 & \text{if G}j \text{ is capacity-constrained for } \varepsilon \in (\varepsilon_0, \varepsilon_{12}) \,. \end{cases}$$

Under VI, Ri values Gi's expected wholesale profit:

$$\begin{aligned} E\left\{ \pi_i^G \right\} \;\equiv\; & \int_{\underline{\varepsilon}}^{\varepsilon_0} \left[w^{0c}(\varepsilon) - c_i \right] q_i^{0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \left[w^{1c}(\varepsilon) - c_i \right] q_i^{1c}(\varepsilon)\, dH(\varepsilon) \\ & + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \left[w^{2c}(\varepsilon) - c_i \right] K_i \, dH(\varepsilon) - k_i\, K_i \,. \end{aligned} \tag{19}$$

Lemma 5 helps to determine how VI affects Ri's choice of r_i by identifying the rate at which Gi's expected wholesale profit changes with r_i.

Lemma 5. $$\frac{\partial E\left\{ \pi_i^G \right\}}{\partial r_i} \;=\; \int_{\underline{\varepsilon}}^{\varepsilon_0} \frac{\partial w^{0c}(\varepsilon)}{\partial r_i}\, q_i^{0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \frac{\partial w^{1c}(\varepsilon)}{\partial r_i}\, q_i^{1c}(\varepsilon)\, dH(\varepsilon)$$

$$+ \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \frac{\partial w^{2c}(\varepsilon)}{\partial r_i} K_i \, dH(\varepsilon) + \int_{\underline{\varepsilon}}^{\varepsilon_0} \left[w^{0c}(\varepsilon) - c_i \right] \frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i}\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \left[w^{1c}(\varepsilon) - c_i \right] \frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i}\, dH(\varepsilon) \,. \tag{20}$$

Lemma 5 indicates that Ri's additional concern with $E\left\{\pi_i^G\right\}$ under VI induces Ri to favor a higher wholesale price of electricity (which increases Gi's wholesale profit margin) and expanded output (q_i) by Gi. Conclusion 5 reports that Ri can secure an increase in Gi's output by reducing r_i.

Conclusion 5. *When Gi and Ri are vertically-integrated: (i)* $\frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i} = -\frac{1}{3}\left[\,3\,b_i - (1-\alpha_j)\,d_j\,\right] < 0$*; and (ii)* $\frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} = -\frac{1}{2}\left[\,2\,b_i - d_j\,\right] < 0$ *if (only) Gj is capacity-constrained when* $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$.

A reduction in r_i increases the demand for wholesale electricity by increasing aggregate retail output ($Q_1^r + Q_2^r$). When it is not capacity-constrained, Gi finds it profitable to increase q_i in response to the increased demand for its wholesale product. Under VI, an increase in Q_i^r also endows Gi with a credible commitment to compete more aggressively against Gj. The larger is Q_i^r, the more rapidly Ri's retail profit increases as its profit margin increases. Therefore, the larger is Q_i^r, the stronger is Gi's incentive to increase q_i in order to reduce w and thereby increase Ri's profit margin. Thus, by reducing r_i, Ri increases Gi's expected wholesale profit by enhancing Gi's commitment to expand its output relatively aggressively in its competition with Gj.

Unlike the impact of r_i on q_i, the impact of r_i on w is indeterminate, as Conclusion 6 reports.

Conclusion 6. *Under VI: (i)* $\frac{\partial w^{0c}(\varepsilon)}{\partial r_i} = 0$ *; (ii)* $\frac{\partial w^{2c}(\varepsilon)}{\partial r_i} = -\,b^w\left[\,b_i - d_j\,\right] < 0$ *; (iii)* $\frac{\partial w^{1c}(\varepsilon)}{\partial r_i} = -\,\frac{b^w}{2}\,b_i < 0$ *when Gi is capacity-constrained for* $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$ *; and (iv)* $\frac{\partial w^{1c}(\varepsilon)}{\partial r_i} = \frac{b^w}{2}\,d_j > 0$ *when Gj is capacity-constrained for* $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$ *.*

Conclusion 6(i) indicates that a change in r_i does not affect w when neither generator is capacity-constrained. Under VI in this case, each generator increases its wholesale output at the same rate that its retail affiliate's demand for electricity increases. (Recall Conclusion 2.) The offsetting changes in the demand for and the supply of electricity leave w unchanged.[16]

Conclusion 6(ii) indicates that a reduction in r_i causes w to increase when both generators are capacity-constrained. A reduction in r_i increases aggregate retail demand ($Q_1^r + Q_2^r$) and

[16] $\frac{\partial w^{0c}(\varepsilon)}{\partial r_i} = \frac{b^w}{3}\,d_j > 0$ if Gi and Ri are vertically-integrated but Gj and Rj are not vertically-integrated. In this case, a reduction in r_i reduces Q_j^r, and Gj does not reduce its supply of electricity to offset this reduction in the demand for electricity. Consequently, the wholesale price of electricity declines.

thereby increases the demand for wholesale electricity. w increases when binding capacity constraints preclude any offsetting increase in the supply of electricity.

Conclusion 6(iii) reports that w also increases as r_i declines when only Gi is capacity-constrained. In this case, Gi does not increase q_i as Q_i^r increases in response to the reduction in r_i. The increased demand for electricity, coupled with no offsetting increase in supply, cause w to increase. The reduction in r_i reduces Q_j^r. However, Gj reduces q_j to offset this source of reduced demand for electricity under VI. (Recall Conclusion 2.)[17]

In contrast, Conclusion 6(iv) indicates that a reduction in r_i causes w to decline when only Gj is capacity-constrained. Under VI in this case, Gi increases q_i to offset the increased demand for electricity that arises when Q_i^r increases due to a reduction in r_i. However, when it is capacity-constrained, Gj does not adjust q_j as Q_j^r declines due to the reduction in r_i. The resulting more pronounced decline in the demand for electricity than in the corresponding supply causes w to decline.

In summary, Ri's concern with $E\left\{\pi_i^G\right\}$ under VI leads Ri to adjust r_i so as to increase q_i and w. A reduction in r_i increases q_i. A reduction is r_i may also increase w, but is not certain to do so.

6 Numerical Solutions

The foregoing analysis suggests that vertical integration often has two primary effects. First, VI often induces a retailer (Ri) to reduce the price it charges for electricity. The reduction in r_i increases Q_i^r, which can increase Gi's wholesale profit by increasing w and by enhancing Gi's commitment to increase q_i aggressively in its competition with Gj. Second, VI often induces a generator (Gi) to increase its capacity investment (K_i) and its supply of electricity (q_i). The increased electricity supply reduces the wholesale price of electricity and thereby increases Ri's retail profit.

[17] $\frac{\partial w^{1c}(\varepsilon)}{\partial r_i} = -\frac{b^w}{2}\left[\, b_i - d_j \,\right] < 0$ if Gi and Ri are vertically-integrated but Gj and Rj are not vertically-integrated. In this case, Gj does not reduce q_j to offset the reduction in Q_j^r induced by a reduction in r_i. This effect alone promotes a reduction in w as r_i declines. However, this effect is outweighed by the (relatively large) increase in Q_i^r (with no offsetting increase in q_i) induced by a reduction in r_i.

We now examine numerical solutions in part to assess the extent to which these effects prevail under circumstances that might plausibly arise in practice.[18] To illustrate some possible such circumstances, we initially consider a *baseline setting.* We subsequently consider alternative settings to determine which outcomes in the illustrative baseline setting persist in other settings of potential interest.

The unit wholesale cost of production ($c_1 = c_2$) in the baseline setting is chosen to approximate the marginal cost (in $/MWh) of generating electricity in a modern combined-cycle gas turbine plant. The heat rate of such a facility is $6,350$ Btu/KWh, or 6.70 GJ/MWh (EIA, 2019a, Table 2).[19] In 2019, the average price of Henry Hub natural gas was $2.597/mmBtu, or $2.46/GJ (EIA, 2020).[20] EIA (2019a) reports variable operations and maintenance (O&M) costs for a combined-cycle natural gas plant to be $3.61/MWh. Consequently, abstracting from environmental regulations, the marginal cost of generation from such a plant (in $/MWh) is the sum of O&M costs and the product of the heat rate and fuel price: $3.61 + 6.70\,[\,2.46\,] = 20.09$, which we round to 20 for expositional simplicity.

The unit capital cost ($k_1 = k_2$) in the baseline setting is chosen to approximate the annualized cost of constructing a combined-cycle gas turbine plant. Lazard (2019, Slide 2) estimates the levelized cost of energy for such a plant to be between 44 and 68 ($/MWh). Taking the midpoint of this range (56) and subtracting the estimated unit wholesale cost of production (20) provides an estimate of 36 for the unit cost of capital, which we round to 35.

The unit retail cost ($c_1^r = c_2^r$) in the baseline setting is selected to reflect: (i) billing and metering costs; plus (ii) transactions costs of contracting with wholesale suppliers, trading charges and fees imposed by the system operator, and the cost of collateral required to

[18]The methodology employed to derive the numerical solutions is described in Appendix B. Brown and Sappington (2020) provides the additional analytic work that underlies the numerical solutions.

[19]There are $947,817$ Btu in a GJ and $1,000$ KWhs in a MWh. Therefore, a heat rate of $6,350$ Btu/KWh can be expressed as $6,350 \left[\frac{1}{947,817} \right] [\,1,000\,] = 6.70$ GJ/MWh.

[20]There are $\frac{1}{1.055}$ mmBtu per GJ. Therefore, a fuel price of $2.597/mmbtu can be expressed as $2.597 \left[\frac{1}{1.055} \right] =$ $2.46/GJ.

comply with financial security requirements. EPCOR (2017) estimates the latter costs to total \$0.87/MWh of retail energy served. This estimate is increased to \$1.0/MWh to account for other costs, including billing and metering costs.

The values of the demand parameters in the baseline setting are determined as follows. First, the lowest demand state ($\underline{\varepsilon}$) is normalized to zero. Second, to help determine the price sensitivity parameters for retail demand ($b_1 = b_2$ and $d_1 = d_2$), the own-price (cross-price) elasticity of retail demand is taken to be -0.25 (0.10).[21] Both of these elasticities are calculated when the retail price reflects the estimated levelized unit cost of producing and supplying electricity ($k_i + c_i + c_i^r = 56$). Third, to help determine the slope of the industrial demand curve ($b^I = \frac{1}{b^w}$), the price elasticity of industrial demand evaluated at estimated unit cost (56) is taken to be -0.33.[22]

Fourth, to help determine the intercepts of the retail demand curves ($a_1 = a_2$), we take the ratio of residential demand to total market demand evaluated at estimated unit cost (56) to be 0.40.[23] Fifth, to help determine the intercept of the industrial demand curve (a^I), wholesale demand under the highest demand state ($\overline{\varepsilon}$) is taken to be $25,000$ when electricity is priced at four times the unit wholesale cost of production (80).[24] Sixth, to help determine the magnitude of $\overline{\varepsilon}$, the ratio of the highest wholesale demand for electricity (when it is priced at four times wholesale marginal cost, 80) to the lowest wholesale demand

[21] These values are consistent with common empirical estimates, e.g., King and Chatterjee (2003), Espey and Espey (2004), Narayan and Smyth (2005), and Paul et al. (2009).

[22] See Labandeira et al. (2017), for example.

[23] EIA (2019b) reports that in the first nine months of 2019, residential and commercial electricity consumption accounted for 39.41% of total electricity consumption. The corresponding proportions were 38.67% in 2017 and 39.24% in 2018.

[24] This peak demand is chosen to reflect the experience of a mid-sized U.S. ISO/RTO. ISO New England experienced a peak demand of $25,559$ MWs in 2018 (ISONE, 2020). New York State experienced a peak demand of $29,699$ MWs in 2017 (NYISO, 2018). We will consider substantial variation in parameter values to account for the fact that peak demand can vary significantly across geographic regions (e.g., $12,029$ MWs in Massachusetts in 2018 (ISONE, 2020) and $71,445$ MWs in Texas in 2019 (ERCOT, 2018)).

The price of wholesale electricity can be far above marginal cost during periods of particularly high demand. To illustrate, this price has increased to the regulated cap of \$9,000 in Texas. We adopt the fairly conservative assumption that price is four times marginal cost during the largest demand shock in the baseline setting. Our comparative static analysis admits a wide range of alternatives.

for electricity (when it is priced at wholesale marginal cost, 20) is assumed to be 2.70.[25]

These considerations provide parameter values of approximately $a^I = 7,388.89$, $b^I = 33.33$, $\underline{\varepsilon} = 0$, $\overline{\varepsilon} = 687.22$, $a_1 = a_2 = 2,129.63$, $b_1 = b_2 = 8.267$, and $d_1 = d_2 = 3.307$. These values are rounded to provide the numbers in Table 1. The baseline setting also assumes that ε has a uniform distribution on $[\,0, 700\,]$.

Parameter	Value		Parameter	Value		Parameter	Value
$c_1 = c_2$	20		$a_1 = a_2$	2,100		a^I	7,500
$k_1 = k_2$	35		$b_1 = b_2$	8		b^I	35
$c_1^r = c_2^r$	1		$d_1 = d_2$	3		$\underline{\varepsilon}$	0
						$\overline{\varepsilon}$	700

Table 1. Parameter Values in the Baseline Setting.

Table 2 records key outcomes in the baseline setting. The first column in the table lists the outcomes of interest, which include the expected wholesale price of electricity ($E\{w\}$), total industry capacity investment ($K_1 + K_2$), expected consumer surplus ($E\{CS\}$),[26] expected total welfare ($E\{CS + \pi\}$),[27] and for $i \in \{1, 2\}$, the expected wholesale output of Gi ($E\{q_i\}$), the expected wholesale profit of Gi ($E\{\pi_i^G\}$), and the expected retail profit of Ri ($E\{\pi_i^R\}$). The last three columns of Table 2 present the relevant outcomes under vertical separation (VS), partial vertical integration (PVI), and full vertical integration (VI), respectively.[28]

[25]ERCOT (2020) reports that in 2019, the ratio of the highest to the lowest hourly ERCOT system-wide demand for electricity was 2.70.

[26]Expected consumer surplus is the sum of the expected surplus of residential consumers ($\sum_{i=1}^{2} \int_{r_i}^{\frac{a_i + d_i r_j}{b_i}} [\, a_i - b_i r_i + d_i r_j \,] \, dr_i$) and industrial customers ($\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{w}^{\frac{a^I}{b^I} + \varepsilon} \left[\, a^L - b^L (\, \widetilde{w} + \varepsilon \,) \right] d\widetilde{w} \, dH(\varepsilon)$).

[27]Expected total welfare is the sum of expected consumer surplus and expected industry profit.

[28]All entries in Table 2 (and in the corresponding tables in the Appendix) are rounded to the nearest whole number. The rounding may produce some apparent discrepancies in adding.

	VS	PVI	VI
r_1	308	239	212
r_2	308	293	212
$E\{w\}$	231	227	224
K_1	$8,489$	$9,208$	$8,919$
K_2	$8,489$	$8,187$	$8,919$
$K_1 + K_2$	$16,978$	$17,395$	$17,838$
$E\{q_1\}$	$6,396$	$7,209$	$6,991$
$E\{q_2\}$	$6,396$	$6,154$	$6,991$
$E\{q_1\} + E\{q_2\}$	$12,791$	$13,363$	$13,982$
$E\left\{\pi_1^G\right\}$	$1,219,163$	$1,332,668$	$1,281,875$
$E\left\{\pi_1^R\right\}$	$42,492$	$11,807$	$-14,024$
$E\left\{\pi_1^G + \pi_1^R\right\}$	$1,261,656$	$1,344,475$	$1,267,851$
$E\left\{\pi_2^G\right\}$	$1,219,163$	$1,152,405$	$1,281,875$
$E\left\{\pi_2^R\right\}$	$42,492$	$30,886$	$-14,024$
$E\left\{\pi_2^G + \pi_2^R\right\}$	$1,261,656$	$1,183,291$	$1,267,851$
$E\{CS\}$	$2,199,751$	$2,289,615$	$2,358,343$
$E\{CS + \pi\}$	$4,723,062$	$4,817,380$	$4,894,045$

Table 2. Outcomes in the Baseline Setting.

Eight features of Table 2 that persist more broadly warrant emphasis. First, retail prices are lowest under VI and highest under VS. As explained in Section 5, vertically-integrated retailers favor relatively low retail prices in order to increase the demand for electricity and thereby increase wholesale profit by raising the expected wholesale price of electricity. Second, industry capacity and expected wholesale output are highest under VI and lowest under VS. As explained in Sections 3 and 4, the elevated capacity and output under VI arise because a vertically-integrated generator values the increase in retail profit that arises when an increased supply of electricity reduces the expected wholesale price.

Third, the combined expected profit of G1 and R1 is higher under PVI than under VS. Consequently, G1 an R1 have an incentive to integrate if G2 and R2 are not integrated. Fourth, under PVI, the increased capacity investment and expected output of G1 induces a reduction in G2's equilibrium investment and expected output below their levels under VS. Fifth, the relatively aggressive behavior of R1 and G1 reduces the combined expected profit of G2 and R2 ($E\left\{\pi_2^G + \pi_2^R\right\}$) under PVI below its level under VS.

Sixth, $E\left\{\pi_2^G + \pi_2^R\right\}$ is higher under VI than under PVI because vertical integration of G2 and R2 induces them to compete more aggressively.[29] Seventh, the relatively low retail prices that arise under VI help to ensure that expected consumer surplus is highest under VI and lowest under VS. Eighth, expected total welfare is highest under VI and lowest under VS.

The expected wholesale price of electricity ($E\{w\}$) is lower under VI than under VS in the baseline setting. However, $E\{w\}$ can be lower under VS than under VI in other plausible settings. This indeterminacy arises because vertical integration tends to increase both the demand for electricity and its supply, and either effect can dominate.

In the baseline setting, the expected profit of each vertically-integrated firm under VI exceeds the combined profit of the relevant generator and retailer under VS.[30] The increased profit reflects in part the increased surplus that arises from the relatively low retail prices of electricity that arise under VI. More generally, the enhanced competitive aggression that arises under VI can cause the expected profit of both integrated firms to fall below the corresponding levels of profit under VS. As Table 3 reports, this is the case, for example, when all parameters are as specified in the benchmark setting except that d_1 and d_2 are both increased from 3 to 3.5. As is evident from equation (1), an increase in $d_1 = d_2$ increases the demand for each retailer's product and renders the demand more sensitive to changes in the rival's price.

[29]Specifically, R2 reduces r_2 because it values the increased expected profit of G2 that arises when an increased demand for electricity increases its expected wholesale price. In addition, G2 increases K_2 and $E\{q_2\}$ because it values the increased expected profit of R2 that arises when an increased supply of electricity reduces its expected wholesale price.

[30]Specifically, $E\left\{\pi_i^G + \pi_i^R\right\}$ is higher under vertical integration than under vertical separation both for $i = 1$ and for $i = 2$.

	VS	PVI	VI
r_1	318	252	221
r_2	318	302	221
$E\{w\}$	232	227	225
K_1	$8,529$	$9,254$	$8,934$
K_2	$8,529$	$8,216$	$8,934$
$K_1 + K_2$	$17,058$	$17,470$	$17,869$
$E\{q_1\}$	$6,442$	$7,268$	$7,024$
$E\{q_2\}$	$6,442$	$6,178$	$7,024$
$E\{q_1\} + E\{q_2\}$	$12,884$	$13,447$	$14,048$
$E\left\{\pi_1^G\right\}$	$1,237,351$	$1,348,092$	$1,291,115$
$E\left\{\pi_1^R\right\}$	$53,942$	$26,091$	$-4,767$
$E\left\{\pi_1^G + \pi_1^R\right\}$	$1,291,292$	$1,374,183$	$1,286,349$
$E\left\{\pi_2^G\right\}$	$1,237,351$	$1,161,314$	$1,291,115$
$E\left\{\pi_2^R\right\}$	$53,942$	$39,788$	$-4,767$
$E\left\{\pi_2^G + \pi_2^R\right\}$	$1,291,292$	$1,201,102$	$1,286,349$
$E\{CS\}$	$2,192,252$	$2,290,749$	$2,361,959$
$E\{CS + \pi\}$	$4,774,836$	$4,866,033$	$4,934,656$

Table 3. Outcomes when $d_1 = d_2 = 3.5$.

Observe that each retailer-generator pair has a unilateral incentive to integrate in the setting of Table 3.[31] However, ubiquitous vertical integration reduces the expected profit of both integrated firms.[32] Thus, vertical integration can introduce the prisoner's dilemma considerations that often arise when an activity enhances the aggression of multiple competitors simultaneously (e.g., Fershtman and Judd, 1987; Sklivas, 1987; Allaz and Vila, 1993).

The eight primary conclusions that arise in the baseline setting arise more generally. The Appendix demonstrates that the conclusions continue to hold as numerous parameters in the baseline setting are increased unilaterally by 20%. Brown and Sappington (2020) document

[31]Specifically, $E\left\{\pi_1^G + \pi_1^R\right\}$ is higher under PVI than under VS and $E\left\{\pi_2^G + \pi_2^R\right\}$ is higher under VI than under PVI.

[32]Specifically, $E\left\{\pi_i^G + \pi_i^R\right\}$ is lower under VI than under VS for $i = 1, 2$.

that the conclusions hold much more broadly. Indeed, the conclusions persist in all of the many variations on the baseline setting that we have considered.

7 Conclusions

We have examined the incentives for and the effects of vertical integration in the electricity sector in the presence of both wholesale and retail competition. We found that vertical integration generally reduces retail prices while increasing industry capacity investment, the wholesale supply of electricity, consumer surplus, and industry welfare. Because vertical integration enhances supplier commitment to compete aggressively, unilateral vertical integration is profitable. However, the bilateral increase in competitive aggression that arises under ubiquitous vertical integration can sometimes reduce industry profit, even as it increases consumer surplus and welfare.

We have analyzed a streamlined model in order to characterize the incentives for and the effects of vertical integration clearly and concisely. However, the basic forces at play in our model seem likely to persist more generally.[33] They likely persist, for example, in the presence of risk aversion. Even as vertical integration helps to reduce the variation in profit introduced by variation in the wholesale price of electricity,[34] vertical integration can simultaneously enhance the commitment of the integrated suppliers to compete relatively aggressively. Thus, it seems likely that vertical integration will arise in equilibrium and serve to reduce retail prices and expand industry investment and the supply of electricity even when industry suppliers are risk averse.

The basic forces that we have identified also seem likely to persist when industry suppliers can undertake forward contracting. As Allaz and Vila (1993) demonstrate, a generator can employ forward contracting to enhance its commitment to compete aggressively. The

[33] We have verified that the qualitative conclusions drawn above persist under alternative formulations of retail demand. They persist, for example, when retailers serve some captive customers and compete (via Hotelling competition) for other customers. They also persist when, as in Klemperer (1987), retail consumers incur a switching cost if they purchase electricity from a new supplier rather than their initial (incumbent) supplier.

[34] Vertical integration can reduce the profit variation associated with variation in the wholesale price of electricity (w) because an increase in w increases wholesale profit as it reduces retail profit.

corresponding commitment value of vertical integration that we have identified may be diminished by the presence of commitment secured via forward contracting. However, it seems that unilateral vertical integration will continue to enhance profit for the reasons we have identified even in the presence of forward contracting. Future research might compare outcomes under vertical integration and forward contracting and analyze the extent to which the two sources of enhanced commitment to compete aggressively are substitutes or complements.[35] Future research might also consider other factors that can affect incentives for vertical integration. Such factors include transactions costs of integration, asymmetric information about production costs and other industry conditions, retail price regulation, and wholesale price restrictions.

[35] Aïd et al. (2011) analyze the effects of forward contracting and vertical integration in a model with risk-averse suppliers. Their analysis might be extended to incorporate imperfect competition among generators and endogenous investment in generation capacity.

Appendix

This Appendix has three components. Appendix A provides the proofs of the Lemmas and Conclusions in the text. Appendix B describes the computational approach that underlies the numerical solutions in Section 6. Appendix C presents additional numerical solutions.

Appendix A. Proofs of Formal Conclusions in the Text

Proof of Lemma 1. (4) and (7) imply that Gi's choice of q_i is determined by:

$$
\begin{aligned}
& w'(Q)\,[\,q_i - \alpha_i\, Q_i^r\,] + w(Q) - c_i \;=\; 0 \\
\Leftrightarrow \quad & -\, b^w\,[\,q_i - \alpha_i\, Q_i^r\,] + b^w\left[\,a^I + Q_i^r + Q_j^r\,\right] - b^w\, q_i - b^w\, q_j + \varepsilon - c_i \;=\; 0 \\
\Leftrightarrow \quad & q_i \;=\; \frac{b^w\left[\,a^I + Q_j^r + Q_i^r\,(1+\alpha_i)\,\right] + \varepsilon - c_i}{2\,b^w} - \frac{1}{2}\,q_j\,.
\end{aligned} \tag{21}
$$

(21) implies that G1's equilibrium output is:

$$
\begin{aligned}
q_1 \;=\; & \frac{b^w\left[\,a^I + Q_2^r + Q_1^r\,(1+\alpha_1)\,\right] + \varepsilon - c_1}{2\,b^w} \\
& -\frac{1}{2}\left[\frac{b^w\left[\,a^I + Q_1^r + Q_2^r\,(1+\alpha_2)\,\right] + \varepsilon - c_2}{2\,b^w} - \frac{1}{2}\,q_1\right] \\
\Rightarrow \quad \frac{3}{4}\,q_1 \;=\; & \frac{1}{4\,b^w}\left\{\,2\,b^w\left[\,a^I + Q_2^r + Q_1^r\,(1+\alpha_1)\,\right] + 2\,\varepsilon - 2\,c_1\right. \\
& \left. -\, b^w\left[\,a^I + Q_1^r + Q_2^r\,(1+\alpha_2)\,\right] - \varepsilon + c_2\,\right\} \\
\Rightarrow \quad q_1^{0c}(\varepsilon) \;=\; & \frac{b^w\left[\,a^I + (1+2\,\alpha_1)\,Q_1^r + (1-\alpha_2)\,Q_2^r\,\right] + \varepsilon + c_2 - 2\,c_1}{3\,b^w}\,.
\end{aligned} \tag{22}
$$

(21) and (22) imply:

$$
\begin{aligned}
q_2 \;=\; & \frac{b^w\left[\,a^I + Q_1^r + Q_2^r\,(1+\alpha_2)\,\right] + \varepsilon - c_2}{2\,b^w} \\
& -\, \frac{b^w\left[\,a^I + Q_1^r\,(1+2\,\alpha_1) + Q_2^r\,(1-\alpha_2)\,\right] + \varepsilon + c_2 - 2\,c_1}{6\,b^w} \\
=\; & \frac{1}{6\,b^w}\left\{\,3\,b^w\left[\,a^I + Q_1^r + Q_2^r\,(1+\alpha_2)\,\right] + 3\,\varepsilon - 3\,c_2\right. \\
& \left. -\, b^w\left[\,a^I + Q_1^r\,(1+2\,\alpha_1) + Q_2^r\,(1-\alpha_2)\,\right] - \varepsilon - c_2 + 2\,c_1\,\right\}
\end{aligned}
$$

$$= \frac{1}{6\,b^w}\,\left\{\, b^w \left[\, 2\,a^I + (2 - 2\,\alpha_1)\,Q_1^r + (2 + 4\,\alpha_2)\,Q_2^r \,\right] + 2\,\varepsilon + 2\,c_1 - 4\,c_2 \,\right\}$$

$$\Rightarrow \quad q_2^{0c}(\varepsilon) \;=\; \frac{b^w \left[\, a^I + (1 + 2\,\alpha_2)\; Q_2^r + (1 - \alpha_1)\,Q_1^r \,\right] + \varepsilon + c_1 - 2\,c_2}{3\,b^w}\,. \tag{23}$$

(4), (22), and (23) imply that the corresponding wholesale price when ε is realized in this case is:

$$\begin{aligned} w^{0c}(\varepsilon) \;&=\; b^w \left[\, a^I + Q_1^r + Q_2^r \,\right] - b^w \left[\, q_1^{0c}(\varepsilon) + q_2^{0c}(\varepsilon) \,\right] + \varepsilon \\ &=\; \frac{1}{3} \left\{\, 3\,b^w \left[\, a^I + Q_1^r + Q_2^r \,\right] + 3\,\varepsilon \right. \\ &\qquad \left. -\; b^w \left[\, 2\,a^I + (2 + \alpha_1)\,Q_1^r + (2 + \alpha_2)\,Q_2^r \,\right] - 2\,\varepsilon + c_1 + c_2 \,\right\} \\ &=\; \frac{1}{3} \left\{\, b^w \left[\, a^I + (1 - \alpha_1)\,Q_1^r + (1 - \alpha_2)\,Q_2^r \,\right] + \varepsilon + c_1 + c_2 \,\right\}\,. \quad \blacksquare \end{aligned} \tag{24}$$

Proof of Lemma 2. Given that Gj's equilibrium output in this case is $q_j^{1c}(\varepsilon) = K_j$, (21) implies that G$i$'s equilibrium output is:

$$q_i^{1c}(\varepsilon) \;=\; \frac{1}{2} \left[\, a^I + Q_j^r + Q_i^r\,(1 + \alpha_i) \,\right] + \frac{1}{2\,b^w} \left[\, \varepsilon - c_i \,\right] - \frac{1}{2}\,K_j\,. \tag{25}$$

(4) and (25) imply that when ε is realized in this case:

$$\begin{aligned} w^{1c}(\varepsilon) \;&=\; b^w \left[\, a^I + Q_i^r + Q_j^r \,\right] - \frac{b^w}{2} \left[\, a^I + Q_j^r + Q_i^r\,(1 + \alpha_i) + K_j \,\right] - \frac{1}{2} \left[\, \varepsilon - c_i \,\right] + \varepsilon \\ &=\; \frac{1}{2}\,b^w \left[\, a^I + Q_i^r\,(1 - \alpha_i) + Q_j^r \,\right] + \frac{1}{2} \left[\, \varepsilon + c_i - b^w\,K_j \,\right]. \quad \blacksquare \end{aligned} \tag{26}$$

Proof of Lemma 3. Because $q_1^{2c}(\varepsilon) = K_1$ and $q_2^{2c}(\varepsilon) = K_2$, (4) implies:

$$w^{2c}(\varepsilon) \;=\; b^w \left[\, a^I + Q_1^r + Q_2^r - K_1 - K_2 \,\right] + \varepsilon\,. \quad \blacksquare \tag{27}$$

Proofs of Conclusions 1 and 2. The Conclusions follow directly from Lemmas 1 – 3. $\blacksquare$

Proof of Conclusion 3. (9) and (10) imply $\frac{\partial q_i^{0c}(\varepsilon)}{\partial \alpha_i} = \frac{2}{3}\,Q_i^r > 0$, $\frac{\partial q_i^{1c}(\varepsilon)}{\partial \alpha_i} = \frac{1}{2}\,Q_i^r > 0$, $\frac{\partial w^{0c}(\varepsilon)}{\partial \alpha_i} = -\,\frac{b^w}{3}\,Q_i^r < 0$, and $\frac{\partial w^{1c}(\varepsilon)}{\partial \alpha_i} = -\,\frac{b^w}{2}\,Q_i^r < 0$. $\blacksquare$

Proof of Lemma 4. Differentiating (12), using (1), provides:

$$\frac{\partial E\left\{\pi_i^R\right\}}{\partial K_i} = \left[\, r_i - E\left\{w(\varepsilon)\right\} - c_i^r \,\right] \left[- b_i \frac{\partial r_i}{\partial K_i} + d_i \frac{\partial r_j}{\partial K_i} \right] + Q_i^r \left[\frac{\partial r_i}{\partial K_i} - \frac{\partial E\left\{w(\varepsilon)\right\}}{\partial K_i} \right]$$

$$+ \; Q_i^r \left\{ \left[\, w^{1c}(\varepsilon_0) - w^{0c}(\varepsilon_0) \,\right] h(\varepsilon_0) \frac{\partial \varepsilon_0}{\partial K_i} + \left[\, w^{2c}(\varepsilon_{12}) - w^{1c}(\varepsilon_{12}) \,\right] h(\varepsilon_{12}) \frac{\partial \varepsilon_{12}}{\partial K_i} \right\}. \quad (28)$$

(13) follows from (28) because the term in $\{\cdot\}$ in (28) is zero since $w^{1c}(\varepsilon_0) = w^{0c}(\varepsilon_0)$ and $w^{2c}(\varepsilon_{12}) = w^{1c}(\varepsilon_{12})$. To prove that these equalities hold, consider the case where Gj is capacity-constrained when $\varepsilon \in (\,\varepsilon_0, \varepsilon_{12}\,)$. (9) implies that in this case:

$$K_j = \frac{b^w \left[\, a^I + (1 + 2\,\alpha_j)\; Q_j^r + (1 - \alpha_i)\, Q_i^r \,\right] + \varepsilon_0 + c_i - 2\,c_j}{3\,b^w}$$

$$\Rightarrow \quad \frac{1}{3\,b^w} \left[\, \varepsilon_0 - 2\,c_j + c_i \,\right] = K_j - \frac{1}{3} \left[\, a^I + (1 + 2\,\alpha_j)\; Q_j^r + (1 - \alpha_i)\, Q_i^r \,\right]$$

$$\Rightarrow \quad \varepsilon_0 = 3\,b^w\, K_j + 2\,c_j - c_i - b^w \left[\, a^I + (1 + 2\,\alpha_j)\; Q_j^r + (1 - \alpha_i)\, Q_i^r \,\right]. \quad (29)$$

(10) implies that in this case:

$$K_i = \frac{1}{2} \left[\, a^I + Q_j^r + Q_i^r\,(1 + \alpha_i) \,\right] + \frac{\varepsilon_{12} - c_i}{2\,b^w} - \frac{1}{2}\,K_j$$

$$\Rightarrow \quad 2\,b^w\,K_i = b^w \left[\, a^I + Q_j^r + Q_i^r\,(1 + \alpha_i) \,\right] + \; \varepsilon_{12} - c_i - b^w\,K_j$$

$$\Rightarrow \quad \varepsilon_{12} = 2\,b^w\,K_i + b^w\,K_j + c_i - b^w \left[\, a^I + Q_i^r\,(\,1 + \alpha_i\,) + Q_j^r \,\right]. \quad (30)$$

(9) and (29) imply that under VI in this case:

$$w^{0c}(\varepsilon_0) = \frac{1}{3} \left[\, b^w\,a^I + c_i + c_j \,\right] + \frac{1}{3} \left[\, 3\,b^w\,K_j + 2\,c_j - c_i \,\right] - \frac{1}{3}\,b^w \left[\, a^I + 3\,Q_j^r \,\right]$$

$$= b^w\,K_j + c_j - b^w\,Q_j^r\,. \quad (31)$$

(10), (29), and (31) imply that under VI in this case:

$$w^{1c}(\varepsilon_0) = \frac{1}{2}\,b^w \left[\, a^I + Q_j^r - K_j \,\right] + \frac{1}{2}\,c_i$$

$$+ \; \frac{1}{2} \left[\, 3\,b^w\,K_j + 2\,c_j - c_i \,\right] - \frac{1}{2}\,b^w \left[\, a^I + 3\,Q_j^r \,\right]$$

$$= -\,b^w\,Q_j^r + b^w\,K_j + c_j = w^{0c}(\varepsilon_0)\,. \quad (32)$$

(11) and (30) imply that under VI in this case:

$$w^{2c}(\varepsilon_{12}) = b^w \left[\, a^I + Q_i^r + Q_j^r - K_i - K_j \,\right] + 2\,b^w\,K_i + b^w\,K_j + c_i$$

$$- \, b^w \left[\, a^I + 2\, Q_i^r + Q_j^r \, \right] \;=\; - \, b^w \, Q_i^r + b^w \, K_i + c_i \, . \tag{33}$$

(10), (30), and (33) imply that under VI in this case:

$$\begin{aligned} w^{1c}(\varepsilon_{12}) \;&=\; \frac{1}{2}\, b^w \left[\, a^I + Q_j^r - K_j \, \right] + \frac{1}{2}\, c_i + b^w \, K_i + \frac{1}{2}\, b^w \, K_j + \frac{1}{2}\, c_i \\ &- \; \frac{1}{2}\, b^w \left[\, a^I + 2\, Q_i^r + Q_j^r \, \right] \;=\; c_i + b^w \, K_i - b^w \, Q_i^r \;=\; w^{2c}(\varepsilon_{12}) \, . \end{aligned} \tag{34}$$

The proof for the case where Gi is capacity-constrained when $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$ is analogous, and so is omitted. ■

Proof of Conclusion 4. From (12):

$$\begin{aligned} \frac{\partial E \left\{ w(\varepsilon) \right\}}{\partial K_i} \;&=\; \int_{\underline{\varepsilon}}^{\varepsilon_0} \frac{\partial w^{0c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\overline{\varepsilon}}} \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon) \\ &+ \; \left[\, w^{0c}(\varepsilon_0) - w^{1c}(\varepsilon_0) \, \right] h(\varepsilon_0) \, \frac{\partial \varepsilon_0}{\partial K_i} + \left[\, w^{1c}(\varepsilon_{12}) - w^{2c}(\varepsilon_{12}) \, \right] h(\varepsilon_{12}) \, \frac{\partial \varepsilon_{12}}{\partial K_i} \, . \end{aligned} \tag{35}$$

(9) – (11) imply that when Gj is capacity-constrained for $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$:

$$\frac{\partial w^{0c}(\varepsilon)}{\partial K_i} \;=\; \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} \;=\; 0 \quad \text{and} \quad \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} \;=\; - \, b^w \, . \tag{36}$$

(32) and (34) imply that when Gj is capacity-constrained for $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$:

$$w^{0c}(\varepsilon_0) \;=\; w^{1c}(\varepsilon_0) \;\; \text{and} \;\; w^{1c}(\varepsilon_{12}) \;=\; w^{2c}(\varepsilon_{12}) \, . \tag{37}$$

(35), (36), and (37) imply that when Gj is capacity-constrained for $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$:

$$\frac{\partial E \left\{ w(\varepsilon) \right\}}{\partial K_i} \;=\; - \, b^w \left[\, 1 - H(\varepsilon_{12}) \, \right] \;<\; 0 \, .$$

(9) – (11) imply that when Gi is capacity-constrained for $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$:

$$\frac{\partial w^{0c}(\varepsilon)}{\partial K_i} \;=\; 0, \quad \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} \;=\; - \, \frac{1}{2} \, b^w, \quad \text{and} \quad \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} \;=\; - \, b^w \, . \tag{38}$$

An analysis that parallels the analysis in (29) – (34) demonstrates that (37) holds when Gi is capacity-constrained for $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$. (35), (37), and (38) imply that in this case:

$$\frac{\partial E \left\{ w(\varepsilon) \right\}}{\partial K_i} \;=\; - \, \frac{1}{2} \, b^w \left[\, H(\varepsilon_{12}) - H(\varepsilon_0) \, \right] - \, b^w \left[\, 1 - H(\varepsilon_{12}) \, \right] \;<\; 0 \, . \;\; ■$$

Proof of Lemma 5. Differentiating (19) provides:

$$\begin{aligned}
\frac{\partial E\left\{\pi_i^G\right\}}{\partial r_i} \;=\; & \int_{\underline{\varepsilon}}^{\varepsilon_0} \left\{ \left[\, w^{0c}(\varepsilon) - c_i \,\right] \frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{0c}(\varepsilon)}{\partial r_i}\, q_i^{0c}(\varepsilon) \right\} dH(\varepsilon) \\
& + \int_{\varepsilon_0}^{\varepsilon_{12}} \left\{ \left[\, w^{1c}(\varepsilon) - c_i \,\right] \frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{1c}(\varepsilon)}{\partial r_i}\, q_i^{1c}(\varepsilon) \right\} dH(\varepsilon) \\
& + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \left\{ \left[\, w^{2c}(\varepsilon) - c_i \,\right] \frac{\partial K_i}{\partial r_i} + \frac{\partial w^{2c}(\varepsilon)}{\partial r_i}\, K_i \right\} dH(\varepsilon) \\
& + \left\{ \left[\, w^{0c}(\varepsilon_0) - c_i \,\right] q_i^{0c}(\varepsilon_0) - \left[\, w^{1c}(\varepsilon_0) - c_i \,\right] q_i^{1c}(\varepsilon_0) \right\} h(\varepsilon_0)\, \frac{d\varepsilon_0}{dr_i} \\
& + \left\{ \left[\, w^{1c}(\varepsilon_{12}) - c_i \,\right] q_i^{1c}(\varepsilon_{12}) - \left[\, w^{2c}(\varepsilon_{12}) - c_i \,\right] K_i \right\} h(\varepsilon_{12})\, \frac{d\varepsilon_{12}}{dr_i}\,. \qquad (39)
\end{aligned}$$

(20) follows from (39) because: (i) $\frac{\partial K_i}{\partial r_i} = 0$; (ii) $\lim\limits_{\varepsilon \searrow \varepsilon_0} q_i^{1c}(\varepsilon) = q_i^{0c}(\varepsilon_0)$; (iii) $w^{1c}(\varepsilon_0) = w^{0c}(\varepsilon_0)$; (iv) $q_i^{1c}(\varepsilon_{12}) = K_i$ when Gi is capacity-constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$; (v) $\lim\limits_{\varepsilon \nearrow \varepsilon_{12}} q_i^{1c}(\varepsilon) = K_i$ when Gj is capacity-constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$; and (vi) $w^{2c}(\varepsilon_{12}) = w^{1c}(\varepsilon_{12})$.

(i) holds because Gi chooses K_i before Ri chooses r_i. (37) and the corresponding discussion in the proof of Conclusion 4 imply that (iii) and (vi) hold regardless of which generator is capacity-constrained when $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$.

To prove that (ii) holds, observe that (9) and (29) imply that under VI:

$$\begin{aligned}
q_i^{0c}(\varepsilon_0) \;=\; & \frac{1}{3}\left[\, a^I + 3\, Q_i^r \,\right] + \frac{1}{3\, b^w}\left[\, c_j - 2\, c_i \,\right] \\
& + \frac{1}{3\, b^w}\left[\, 3\, b^w K_j + 2\, c_j - c_i \,\right] - \frac{1}{3}\left[\, a^I + 3\, Q_j^r \,\right] \\
=\; & Q_i^r - Q_j^r + \frac{1}{b^w}\left[\, c_j - c_i \,\right] + K_j\,. \qquad (40)
\end{aligned}$$

(10), (29), and (40) imply that when Gj is capacity-constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$ under VI:

$$\begin{aligned}
\lim_{\varepsilon \searrow \varepsilon_0} q_i^{1c}(\varepsilon) \;=\; & \frac{1}{2}\left[\, a^I + 2\, Q_i^r + Q_j^r \,\right] - \frac{1}{2\, b^w}\, c_i - \frac{1}{2}\, K_j \\
& + \frac{1}{2\, b^w}\left[\, 3\, b^w K_j + 2\, c_j - c_i \,\right] - \frac{1}{2}\left[\, a^I + 3\, Q_j^r \,\right] \\
=\; & Q_i^r - Q_j^r - \frac{1}{b^w}\, c_i + \frac{1}{b^w}\, c_j + K_j \;=\; q_i^{0c}(\varepsilon_0)\,. \qquad (41)
\end{aligned}$$

The proof that (ii) holds when Gi is capacity-constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$ is analogous, and so is omitted.

To prove that (v) holds when Gj is capacity-constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$, observe that (10) and (30) imply:

$$\begin{aligned}
\lim_{\varepsilon \nearrow \varepsilon_{12}} q_i^{1c}(\varepsilon) &= \frac{1}{2}\left[a^I + 2\,Q_i^r + Q_j^r\right] - \frac{1}{2\,b^w}\,c_i - \frac{1}{2}\,K_j \\
&\quad + \frac{1}{2\,b^w}\left[2\,b^w\,K_i + b^w\,K_j + c_i - b^w\left(a^I + 2\,Q_i^r + Q_j^r\right)\right] \\
&= -\frac{1}{2\,b^w}\,c_i - \frac{1}{2}\,K_j + \frac{1}{2\,b^w}\left[2\,b^w\,K_i + b^w\,K_j + c_i\right] \;=\; K_i\,. \;\blacksquare
\end{aligned} \tag{42}$$

Proof of Conclusion 5.

The Conclusion follows directly from (1), (9), and (10). ■

Proof of Conclusion 6.

The Conclusion follows directly from (1), (9), (10), and (11). ■

Appendix B. Description of the Computational Solution Approach

We employ backward induction to generate numerical solutions to our model. First, closed-form solutions for the wholesale equilibrium output levels, $q_1^*(r_1, r_2, K_1, K_2, \varepsilon)$ and $q_2^*(r_1, r_2, K_1, K_2, \varepsilon)$, are determined for each demand state ε, taking capacity investments and retail prices as given. Second, the retail prices that maximize each retailer's objective (given $q_1^*(\cdot)$, $q_2^*(\cdot)$, K_1, and K_2) are characterized implicitly by the functions:

$$J_1(r_1, r_2, K_1, K_2) = 0 \quad \text{and} \quad J_2(r_1, r_2, K_1, K_2) = 0\,. \tag{43}$$

Third, capacity investments are determined. Generator Gi chooses K_i to maximize its expected payoff, taking K_j as given ($i, j \in \{1, 2\}$, $j \neq i$) and anticipating the retail prices that will ensue. Gi's problem can be written as:

$$\max_{K_i} \; E\left\{ \pi_i^G(r_1^*(K_i, K_j), r_2^*(K_i, K_j), K_i, K_j) \right\} \tag{44}$$

where $r_1^*(K_i, K_j)$ and $r_2^*(K_i, K_j)$ are characterized by (43).

Gi's choice of K_i can be written as a stochastic mathematical program with equilibrium constraints (SMPEC) that treats the retail price conditions in (43) as constraints:[36]

$$\underset{K_i, r_1, r_2}{\text{Maximize}} \; E\left\{ \pi_i^G(r_1, r_2, K_i, K_j) \right\}$$

subject to:

$$\begin{aligned} \mu_{i1}: \qquad & J_1(r_1, r_2, K_1, K_2) = 0\,, \\ \mu_{i2}: \qquad & J_2(r_1, r_2, K_1, K_2) = 0\,, \end{aligned} \tag{45}$$

where μ_{i1} and μ_{i2} denote the relevant Lagrange multipliers.

Because Gi's choice of K_i is represented by the SMPEC in (45), the overall problem is a stochastic equilibrium problem with equilibrium constraints (SEPEC) (DeMiguel and Xu, 2009; Leyffer and Munson, 2010). A solution to this SEPEC requires the simultaneous solution of the two SMPECs.

The first step in solving this SEPEC is to derive an analytic expression for the objective function in (45). Doing so (which is facilitated by assuming ε has a uniform distribution on $[\underline{\varepsilon}, \overline{\varepsilon}\,]$) allows us to employ standard techniques that are used to solve deterministic EPECs.

Following Hu (2002), Hu and Ralph (2007), and DeMiguel and Xu (2009), we characterize the Karush-Kuhn-Tucker conditions associated with (45) for G1 and for G2. Formally, the

[36] The necessary conditions for the optimal choices of q_1 and q_2 do not need to be imposed as constraints because the relevant closed-form solutions, $q_1^*(\cdot)$ and $q_2^*(\cdot)$, are substituted directly into both Gi's objective function and conditions that characterize profit-maximizing retail prices.

Lagrangian function for Gi's SMPEC in (45) is:

$$\mathcal{L}^i \;=\; E\left\{ \pi_i^G(r_1, r_2, K_i, K_j) \right\} + \mu_{i1}\, J_1(r_1, r_2, K_1, K_2) + \mu_{i2}\, J_2(r_1, r_2, K_1, K_2). \tag{46}$$

The corresponding solution to Gi's SMPEC is characterized by:

$$\begin{aligned}
K_i: &\quad \frac{\partial E\{\pi_i^G(r_1, r_2, K_i, K_j)\}}{\partial K_i} + \mu_{i1}\, \frac{\partial J_1(r_1, r_2, K_1, K_2)}{\partial K_i} + \mu_{i2}\, \frac{\partial J_2(r_1, r_2, K_1, K_2)}{\partial K_i} \;=\; 0\,; \\
r_1: &\quad \frac{\partial E\{\pi_i^G(r_1, r_2, K_i, K_j)\}}{\partial r_1} + \mu_{i1}\, \frac{\partial J_1(r_1, r_2, K_1, K_2)}{\partial r_1} + \mu_{i2}\, \frac{\partial J_2(r_1, r_2, K_1, K_2)}{\partial r_1} \;=\; 0\,; \\
r_2: &\quad \frac{\partial E\{\pi_i^G(r_1, r_2, K_i, K_j)\}}{\partial r_2} + \mu_{i1}\, \frac{\partial J_1(r_1, r_2, K_1, K_2)}{\partial r_2} + \mu_{i2}\, \frac{\partial J_2(r_1, r_2, K_1, K_2)}{\partial r_2} \;=\; 0\,; \\
\mu_{i1}: &\quad J_1(r_1, r_2) \;=\; 0\,; \\
\mu_{i2}: &\quad J_2(r_1, r_2) \;=\; 0\,.
\end{aligned} \tag{47}$$

The solution to the SEPEC is characterized by solving the equations in (47) for G1 and G2 simultaneously.[37] There are eight endogenous variables $(K_1, K_2, r_1, r_2, \mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})$. There are also eight unique equations because the last two equations in (47) are the same in the SMPECs of G1 and G2. Thus, we have a square constrained non-linear system in which the number of conditions is equal to the number of endogenous variables. This system can be solved using Newton's Method and the PATH algorithm using the GAMS software (Ferris and Monson, 2000).

We proceed by first adopting Assumption G2C, i.e., G2 becomes capacity-constrained for a strictly smaller realization of ε than G1. If Assumption G2C is violated (because $\varepsilon_0 \geq \varepsilon_{12}$ at the identified solution), then we characterize equilibrium outcomes under Assumption G1C, i.e., G1 becomes capacity-constrained for a strictly smaller realization of ε than G2. If Assumptions G1C and G2C are both violated at the relevant identified solutions, then equilibrium outcomes are characterized by imposing the additional constraint $\varepsilon_0 \leq \varepsilon_{12}$. If Assumption G$i$C is violated but Assumption GjC is not violated ($i, j \in \{1, 2\}$, $j \neq i$), then the solution is as derived under Assumption GjC. In all of the cases considered in the text and in Appendix C, either Assumption G1C or G2C (but not both) is violated at the relevant identified solution.[38]

[37] Hobbs et al. (2000), Hu and Ralph (2007), and Brown and Eckert (2017) employ corresponding approaches to characterize equilibrium outcomes in models of the electricity sector.

[38] Brown and Sappington (2020) identify two cases in which neither Assumption G1C nor Assumption G2C is violated at the relevant identified solutions. Thus, two equilibria arise in these cases. Aggregate industry investment, retail prices, expected consumer surplus, and expected industry profit are all similar in the two equilibria.

Appendix C. Additional Numerical Solutions

Tables A1 – A9 present the equilibrium outcomes that arise when the parameter identified in the title of the table is increased by 20% above its value in the baseline setting. All other parameter values are as specified in the baseline setting. Brown and Sappington (2020) present additional outcomes that arise in different settings. In all of the settings considered here and in Brown and Sappington (2020): (i) retail prices are lowest under VI and highest under VS; (ii) industry capacity and expected wholesale output are highest under VI and lowest under VS; (iii) the combined expected profit of G1 and R1 is higher under PVI than under VS; (iv) G2's equilibrium investment and expected output are lower under PVI than under VS; (v) the combined expected profit of G2 and R2 ($E\left\{\pi_2^G + \pi_2^R\right\}$) is lower under PVI than under VS; (vi) $E\left\{\pi_2^G + \pi_2^R\right\}$ is higher under VI than under PVI; (vii) expected consumer surplus is highest under VI and lowest under VS; and (viii) expected total welfare is highest under VI and lowest under VS.

	VS	PVI	VI
r_1	315	242	213
r_2	315	300	213
$E\{w\}$	244	240	238
K_1	8,979	9,707	9,424
K_2	8,979	8,675	9,424
$K_1 + K_2$	17,958	18,382	18,849
$E\{q_1\}$	6,874	7,695	7,490
$E\{q_2\}$	6,874	6,640	7,490
$E\{q_1\} + E\{q_2\}$	13,748	14,334	14,980
$E\left\{\pi_1^G\right\}$	1,394,107	1,520,935	1,471,864
$E\left\{\pi_1^R\right\}$	36,710	546	-27,744
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,430,817	1,521,481	1,444,120
$E\left\{\pi_2^G\right\}$	1,394,107	1,325,928	1,471,864
$E\left\{\pi_2^R\right\}$	36,710	25,001	-27,744
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,430,817	1,350,929	1,444,120
$E\{CS\}$	2,556,147	2,647,457	2,714,440
$E\{CS + \pi\}$	5,417,781	5,519,867	5,602,680

Table A1. Outcomes when $a^I = 9,000$.

	VS	PVI	VI
r_1	300	236	211
r_2	300	286	211
$E\{w\}$	219	215	213
K_1	9,573	10,279	9,985
K_2	9,573	9,271	9,985
$K_1 + K_2$	19,146	19,550	19,970
$E\{q_1\}$	7,107	7,909	7,680
$E\{q_2\}$	7,107	6,858	7,680
$E\{q_1\} + E\{q_2\}$	14,214	14,766	15,361
$E\left\{\pi_1^G\right\}$	1,277,930	1,379,694	1,328,493
$E\left\{\pi_1^R\right\}$	48,070	21,623	-2,505
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,326,000	1,401,316	1,325,988
$E\left\{\pi_2^G\right\}$	1,277,930	1,212,589	1,328,493
$E\left\{\pi_2^R\right\}$	48,070	36,495	-2,505
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,326,000	1,249,084	1,325,988
$E\{CS\}$	2,315,281	2,402,919	2,472,230
$E\{CS + \pi\}$	4,967,281	5,053,319	5,124,206

Table A2. Outcomes when $b^I = 42$.

	VS	PVI	VI
r_1	321	244	213
r_2	321	304	213
$E\{w\}$	253	249	247
K_1	9,860	10,592	10,312
K_2	9,860	9,556	10,312
$K_1 + K_2$	19,719	20,148	20,625
$E\{q_1\}$	7,169	7,998	7,805
$E\{q_2\}$	7,169	6,942	7,805
$E\{q_1\} + E\{q_2\}$	14,338	14,940	15,611
$E\left\{\pi_1^G\right\}$	1,566,560	1,702,076	1,654,970
$E\left\{\pi_1^R\right\}$	33,097	-6,729	-36,911
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,599,657	1,695,347	1,618,059
$E\left\{\pi_2^G\right\}$	1,566,560	1,497,666	1,654,970
$E\left\{\pi_2^R\right\}$	33,097	21,507	-36,911
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,599,657	1,519,173	1,618,059
$E\{CS\}$	2,905,455	2,997,866	3,063,095
$E\{CS + \pi\}$	6,104,769	6,212,386	6,299,212

Table A3. Outcomes when $\bar{\epsilon} = 840$.

	VS	PVI	VI
r_1	337	266	239
r_2	314	298	217
$E\{w\}$	233	227	225
K_1	8,574	9,419	9,113
K_2	8,574	8,210	8,948
$K_1 + K_2$	17,147	17,629	18,062
$E\{q_1\}$	6,477	7,433	7,200
$E\{q_2\}$	6,477	6,170	7,027
$E\{q_1\} + E\{q_2\}$	12,953	13,603	14,227
$E\left\{\pi_1^G\right\}$	1,247,411	1,374,067	1,319,421
$E\left\{\pi_1^R\right\}$	78,878	49,057	17,396
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,326,289	1,423,125	1,336,816
$E\left\{\pi_2^G\right\}$	1,247,411	1,157,990	1,290,314
$E\left\{\pi_2^R\right\}$	47,840	36,001	-8,965
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,295,250	1,193,991	1,281,348
$E\{CS\}$	2,194,191	2,318,553	2,390,117
$E\{CS + \pi\}$	4,815,730	4,935,669	5,008,282

Table A4. Outcomes when $a_1 = 2,520$.

	VS	PVI	VI
r_1	273	202	180
r_2	300	285	205
$E\{w\}$	228	226	224
K_1	8,438	9,161	8,894
K_2	8,438	8,159	8,899
$K_1 + K_2$	16,876	17,320	17,793
$E\{q_1\}$	6,328	7,148	6,949
$E\{q_2\}$	6,328	6,134	6,954
$E\{q_1\} + E\{q_2\}$	12,656	13,282	13,903
$E\left\{\pi_1^G\right\}$	1,190,825	1,318,275	1,271,192
$E\left\{\pi_1^R\right\}$	16,462	-25,707	-44,730
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,207,287	1,292,567	1,226,463
$E\left\{\pi_2^G\right\}$	1,190,825	1,145,137	1,272,067
$E\left\{\pi_2^R\right\}$	36,601	24,654	-19,394
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,227,426	1,169,792	1,252,672
$E\{CS\}$	2,215,750	2,276,420	2,343,088
$E\{CS + \pi\}$	4,650,463	4,738,779	4,822,223

Table A5. Outcomes when $b_1 = 9.6$.

	VS	PVI	VI
r_1	321	251	221
r_2	310	295	217
$E\{w\}$	232	227	225
K_1	8,518	9,269	8,946
K_2	8,518	8,203	8,912
$K_1 + K_2$	17,036	17,472	17,858
$E\{q_1\}$	6,429	7,293	7,046
$E\{q_2\}$	6,429	6,163	6,982
$E\{q_1\} + E\{q_2\}$	12,859	13,456	14,028
$E\left\{\pi_1^G\right\}$	1,232,420	1,349,613	1,293,531
$E\left\{\pi_1^R\right\}$	57,083	26,158	-4,690
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,289,504	1,375,771	1,288,842
$E\left\{\pi_2^G\right\}$	1,232,420	1,156,686	1,283,434
$E\left\{\pi_2^R\right\}$	44,753	32,967	-9,235
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,277,174	1,189,653	1,274,199
$E\{CS\}$	2,194,329	2,298,038	2,360,792
$E\{CS + \pi\}$	4,761,006	4,863,462	4,923,833

Table A6. Outcomes when $d_1 = 3.6$**.**

	VS	PVI	VI
r_1	308	242	215
r_2	308	294	214
$E\{w\}$	232	228	225
K_1	8,236	8,931	8,652
K_2	8,621	8,323	9,032
$K_1 + K_2$	16,857	17,254	17,683
$E\{q_1\}$	6,319	7,107	6,890
$E\{q_2\}$	6,434	6,197	7,024
$E\{q_1\} + E\{q_2\}$	12,752	13,304	13,913
$E\left\{\pi_1^G\right\}$	1,152,563	1,258,471	1,210,072
$E\left\{\pi_1^R\right\}$	42,072	13,621	-11,510
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,194,635	1,272,092	1,198,562
$E\left\{\pi_2^G\right\}$	1,235,856	1,170,876	1,300,196
$E\left\{\pi_2^R\right\}$	42,072	30,913	-12,955
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,277,928	1,201,789	1,287,241
$E\{CS\}$	2,184,188	2,269,717	2,336,144
$E\{CS + \pi\}$	4,656,751	4,743,598	4,821,947

Table A7. Outcomes when $k_1 = 42$**.**

	VS	PVI	VI
r_1	311	248	221
r_2	311	298	214
$E\{w\}$	237	233	231
K_1	7,991	8,666	8,383
K_2	8,732	8,434	9,164
$K_1 + K_2$	16,723	17,100	17,547
$E\{q_1\}$	5,911	6,676	6,454
$E\{q_2\}$	6,623	6,389	7,236
$E\{q_1\} + E\{q_2\}$	12,534	13,065	13,690
$E\left\{\pi_1^G\right\}$	1,052,537	1,150,955	1,104,860
$E\left\{\pi_1^R\right\}$	39,701	13,778	-10,201
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,092,238	1,164,733	1,094,659
$E\left\{\pi_2^G\right\}$	1,300,924	1,234,488	1,371,482
$E\left\{\pi_2^R\right\}$	39,701	29,365	-18,883
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,340,625	1,263,853	1,352,599
$E\{CS\}$	2,124,684	2,205,271	2,272,591
$E\{CS + \pi\}$	4,557,547	4,633,858	4,719,849

Table A8. Outcomes when $c_1 = 40$**.**

	VS	PVI	VI
r_1	308	239	212
r_2	308	293	212
$E\{w\}$	231	227	224
K_1	8,489	9,207	8,918
K_2	8,489	8,187	8,919
$K_1 + K_2$	16,977	17,394	17,837
$E\{q_1\}$	6,395	7,208	6,990
$E\{q_2\}$	6,395	6,154	6,991
$E\{q_1\} + E\{q_2\}$	12,791	13,362	13,982
$E\left\{\pi_1^G\right\}$	1,219,097	1,332,557	1,281,743
$E\left\{\pi_1^R\right\}$	42,382	11,694	-14,114
$E\left\{\pi_1^G + \pi_1^R\right\}$	1,261,479	1,344,250	1,267,630
$E\left\{\pi_2^G\right\}$	1,219,097	1,152,417	1,281,900
$E\left\{\pi_2^R\right\}$	42,517	30,904	-14,006
$E\left\{\pi_2^G + \pi_2^R\right\}$	1,261,613	1,183,322	1,267,894
$E\{CS\}$	2,199,770	2,289,510	2,358,267
$E\{CS + \pi\}$	4,722,862	4,817,082	4,893,791

Table A9. Outcomes when $c_1^r = 1.2$**.**

References

Aïd, R., Chemla, G., Porchet, A., & Touzi, N. (2011). Hedging and vertical integration in electricity markets. *Management Science*, 57, 1438-1452.

Allaz, B., & Vila, J. (1993). Cournot competition, forward markets, and efficiency. *Journal of Economic Theory*, 59, 1-16.

Baldursson, F., & von der Fehr, N. (2007).. Vertical integration and long-term contracts in risky markets. University of Iceland discussion paper.

Boom, A. (2009). "Vertically integrated firms' investments in electricity generating capacities, *International Journal of Industrial Organization*, 27, 544-551.

Boom, A., & Buehler, S. (2020). Vertical structure and the risk of rent extraction in the electricity industry. *Journal of Economics & Management Strategy*, 29, 210-237.

Borenstein, S., & Bushnell, J. (2015). The US electricity industry after 20 years of restructuring. *Annual Review of Economics*, 7, 437-463.

Brown, D., & Eckert, A. (2017). Electricity market mergers with endogenous forward contracting. *Journal of Regulatory Economics*, 51, 269-310.

Brown, D., & Sappington, D. (2020). Technical appendix to accompany 'Vertical integration and capacity investment in the electricity sector.' Available at: https://uofa.ualberta.ca/arts/about/people-collection/david-brown.

Buehler, S., & Schmutzler, A. (2008). Intimidating competitors – Endogenous vertical integration and downstream investment in successive oligopoly. *International Journal of Industrial Organization*, 26, 247-265.

Bushnell, J. (2007). Oligopoly equilibria in electricity contract markets. *Journal of Regulatory Economics*, 32, 225-245.

Bushnell, J., Mansur, E., & Saravia, C. (2008). Vertical arrangements, market structure, and competition: An analysis of restructured electricity markets. *American Economic Review*, 98, 237-266.

de Braganca, G., & Daglish, T. (2017). Investing in vertical integration: Electricity retail market participation, *Energy Economics*, 67, 355-365.

DeMiguel, V., & Xu, H. (2009). A stochastic multiple-leader Stackelberg model: Analysis, computation, and application. *Operations Research*, 57, 1220-1235.

EIA. (2019a). Cost and performance characteristics of new generation technologies. *Annual Energy Outlook 2019*, Energy Information Administration. Available at: https://www.eia.gov/outlooks/aeo/assumptions/pdf/table_8.2.pdf.

EIA. (2019b). *Energy Consumption by Sector, December 2019.* Energy Information Administration. Available at: www.eia.gov/totalenergy/data/monthly/pdf/sec2_3.pdf.

EIA. (2020). Henry Hub natural gas spot price. *Natural Gas Data*, Energy Information Administration. Available at: https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm.

EPCOR. (2017). 2018 – 2021 *Energy Price Setting Plan.* EPCOR Energy Alberta GP Inc. Alberta Utilities Commission Proceeding 22357.

ERCOT. (2018). ERCOT system performs well during summer of 2018. *Press Release.* Available at: http://www.ercot.com/content/wcm/lists/144927/2018_Summer_Performance_One_Pager_FINAL.pdf

ERCOT. (2020). *Hourly Load Data Archives.* Accessed: February 20, 2020. Available at: http://www.ercot.com/gridinfo/load/load_hist.

Espey, J., & Espey, M. (2004). Turning on the lights: A meta-analysis of residential electricity demand elasticities. *Journal of Agricultural and Applied Economics*, 36, 65-81.

Fabra, N., von der Fehr, N.H., & de Frutos, M. (2011). Market design and investment incentives, *The Economic Journal*, 121, 1340-1360.

Ferris, M., & Munson, T. (2000). Complementarity problems in GAMS and the PATH solver, *Journal of Economic Dynamics and Control*, 24, 165-188.

Fershtman, C., & Judd, K. (1987). Equilibrium incentives in oligopoly. *American Economic Review*, 77, 927-940.

Fudenberg, D., & Tirole, J. (1991). *Game Theory.* Cambridge, MA: MIT Press.

Gans, J., & Wolak, F. (2012). A comparison of ex ante versus ex post vertical market power: Evidence from the electricity supply industry. NBER Working Paper. Available at: https://conference.nber.org/conferences/2007/si2007/IO/wolak.pdf.

Green, R. (2003). Retail competition and electricity contracts. Cambridge Working Papers in Economics. Available at: https://www.repository.cam.ac.uk/handle/1810/385.

Grimm, V., & Zoettl, G. (2013). Investment incentives and electricity spot market competition, *Journal of Economics & Management Strategy*, 22, 832-851.

Halkos, G. (2019). Examining the level of competition in the energy sector. *Energy Policy*, 134, Article 110951.

Hobbs, B., Metzler, C., & Pang, J. (2000). Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Transactions Power Systems*, 15, 637-645.

Hortacsu, A., Madanizadeh, S., & Puller, S. (2017). Power to choose? An analysis of consumer inertia in the residential electricity market, *American Economic Journal: Economic Policy*, 9, 192-226.

Hu, X. (2002). *Mathematical Programs with Complementarity Constraints and Game Theory in Electricity Markets.* Ph.D. Dissertation, Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia.

Hu, X., & Ralph, D. (2007). Using EPECs to model bilevel games in restructured electricity markets with locational prices. *Operations Research*, 55, 809-827.

ISONE. (2020). *FCM Annual System Peak Day, Hour, and Load. Independent System Operator of New England.* Accessed: March 7, 2020. Available at: https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/ann-sys-peak-day-hr-load.

King, C., & Chatterjee, S. (2003). Predicting California demand response: How do consumers react to hourly price? *Public Utilities Fortnightly*, 141, 27-32.

Klemperer, P. (1987). Markets with consumer switching costs. *Quarterly Journal of Economics*, 102, 375-394.

Kury, T. (2013). Price effects of independent transmission system operators in the United States electricity market. *Journal of Regulatory Economics*, 43, 147-167.

Labandeira, X., Labeaga, J., & Lopez-Otero, X. (2017). A meta-analysis on the price elasticity of energy demand. *Energy Policy*, 102, 549-568.

Lazard. (2019). *Lazard's Levelized Cost of Energy Analysis – Version 13.0.* Available at: https://www.lazard.com/media/451086/lazards-levelized-cost-of-energy-version-130-vf.pdf.

Leyffer, S., & Munson, T. (2010). Solving multi-leader-common-follower games. *Optimisation, Methods, & Software*, 25, 601-623.

Li, Y. (2014). Vertical structure and forward contract in electricity market. Working Paper 2014-117, Department of Research, Ipag Business School. Available at: https://ideas.repec.org/p/ipg/wpaper/2014-117.html.

Lundin, E., & Tangeras, T. (2020). Cournot competition in wholesale electricity markets: The Nordic power exchange, Nord Pool, *International Journal of Industrial Organization*, 68, Article Number 102536.

Meade, R., & O'Connor, S. (2011). Comparison of long-term contracts and vertical integration in decentralized electricity markets, in Glachant, J., Finon, D. & de Hauteclocque, A. (eds.) *Competition, Contracts, and Electricity Markets.* Cheltenham, UK: Edward Elgar Publishing.

Murphy, F., & Smeers, Y. (2005). Generation capacity expansion in imperfectly competitive restructured electricity markets. *Operations Research*, 53, 646-661.

Narayan, P., & Smyth, R. (2005). The residential demand for electricity in Australia: An application of the bounds testing approach to cointegration. *Energy Policy*, 33, 467-474.

Newbery, D., Pollitt, M., Ritz, R., & Strielkowski, W. (2018). Market design for a high-renewables European electricity system. *Renewable and Sustainable Energy Reviews*, 91, 695-707.

NYISO. (2018). *2018 Power Trends. New York's Dynamic Power Grid.* New York Independent System Operator. Available at: https://www.nyiso.com/documents/20142/2223020/2018-Power-Trends.pdf/4cd3a2a6-838a-bb54-f631-8982a7bdfa7a.

Paul, A., Myers, E., & Palmer, K. (2009). A partial adjustment model of U.S. electricity demand by region, season, and sector. Resources for the Future Discussion Paper RFF DP 08-50. Available at: https://www.rff.org/documents/rff-dp-08-50.pdf.

Pérez, E. (2007). A model of vertical integration and investment in generation capacity in electricity markets: The case of the bidding game. *Socio-Economic Planning Sciences*, 41, 272-290.

Powell, A. (1993). Trading forward in an imperfect market: The case of electricity in Britain. *The Economic Journal*, 103, 444-453.

Sklivas, S. (1987). The strategic choice of managerial incentives. *RAND Journal of Economics*, 18, 452-458.

Ventosa, M., Baillo, A., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends, *Energy Policy*, 33, 897-913.

Woo, C. K., Sreedharan, P., Hargreaves, J., Kahrl, F., Wang, J., & Horowitz, I. (2014). A review of electricity product differentiation, *Applied Energy*, 114, 262-272.

Department of Economics, University of Alberta
Working Paper Series

2020-04: Residential Electricity Pricing in Texas's Competitive Retail Market – **Brown, D.,** Tsai, C., Woo, C., Zarnikau, J., Zhu, S.
2020-03: Competition in Higher Education – Kaganovich, M., Sarpca, S., **Su, X.**
2020-02: Misallocation across Establishment Gender – **Ranasinghe, A.**
2020-01: Does Locational Marginal Pricing Impact Generation Investment Location Decisions? An Analysis of Texas's Wholesale Electricity Market – **Brown, D.,** Zarnikau, J., Woo, C.
2019-16: A Unified Explanation of Trade Liberalization Effects Across Models of Imperfect Competition – **Alfaro, M.**, Lander, D.
2019-15: Analyzing Firm Behaviour in Restructured Electricity Markets: Empirical Challenges with a Residual Demand Analysis – **Brown, D., Eckert, A.**
2019-14: LinkedIn(to) Job Opportunities: Experimental Evidence from Job Readiness Training – **Wheeler, L.**, Garlick, R., Johnson, E., Shaw, P., Gargano, M.
2019-13: Entry Preemption by Domestic Leaders and Home-Bias Patterns: Theory and Empirics – **Alfaro, M.**
2019-12: How Much are Good Schools Worth? Evidence from School Acquisitions in Beijing – **Su, X.,** Yu, H.
2019-11: Enumerating Rights: More is Not Always Better – Ball, S., **Dave, C.,** Dodds, S.
2019-10: Motivating the Optimal Procurement and Deployment of Electric Storage as a Transmission Asset – **Brown, D.,** Sappington, D.
2019-09: Pricing Patterns in Wholesale Electricity Markets: Unilateral Market Power or Coordinated Behavior? – **Brown, D., Eckert, A.**
2019-08: Son Preference and Child Under Nutrition in the Arab Countries: Is There a Gender Bias Against Girls? – **Sharaf, M.,** Rashad, A., Mansour, E.
2019-07: How Local Economic Conditions Affect School Finances, Teacher Quality, and Student Achievement: Evidence from the Texas Shale Boom – **Marchand, J.,** Weber, J.
2019-06: Subjects in the Lab, Activists in the Field: Public Goods and Punishment – **Dave, C.,** Hamre, S., Kephart, C., Reuben, A.
2019-05: Fiscal Austerity in Emerging Market Economies – **Dave, C.,** Ghate, C., Gopalakrishnan, P., Tarafdar, S.
2019-04: Inequality and Trade Policy: Pro-Poor Bias of Contemporary Trade Restrictions – **Ural Marchand, B.**
2019-03: Determinants of Locational Patenting Behavior of Canadian Firms – **Eckert, A., Langinier, C., Zhao, L.**
2019-02: The Microeconomics of New Trade Models – **Alfaro, M.**
2019-01: Universal Childcare for the Youngest and the Maternal Labour Supply – Kunze, A., **Liu, X.**
2018-19: On the Benefits of Behind-the-Meter Rooftop Solar and Energy Storage: The Importance of Retail Rate Design – **Boampong, R., Brown, D.**
2018-18: The Value Premium During Flights – **Galvani, V.**
2018-17: Asymmetric Information, Predictability and Momentum in the Corporate Bond Market – **Galvani, V., Li, L.**
2018-16: The Momentum Effect for Canadian Corporate Bonds – **Galvani, V., Li, L.**
2018-15: Green Technology and Patents in the Presence of Green Consumers – **Langinier, C.,** Ray Chaudhuri, A.