# 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 often reduces retail prices and increases industry capacity investment, consumer surplus, and total welfare. Unilateral vertical integration often is profitable, and so 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

July 2021

* 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).

We thank the editor, the coeditor, and two anonymous referees for extremely helpful comments. 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 examine whether vertical integration promotes investment in generation capacity in the presence of imperfect wholesale competition, imperfect 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 consider the incentives for vertical integration by industry suppliers. We explore 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 often leads to increased capacity investment and ex-

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

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

We also find that vertical integration of R1 and G1 often promotes lower retail prices for electricity. It does so for two reasons. First, a lower retail price increases the demand for electricity, which raises the wholesale price. The increase in $w$ enhances G1's wholesale profit, which R1 values under vertical integration. Second, when 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, which enhances G1's incentive to reduce $w$. This increased incentive endows G1 with a credible commitment to compete aggressively against G2 when choosing its supply of electricity. This commitment serves to enhance 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 often 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 often is lower under partial vertical integration (PVI) than under VS. Once R1 and G1 have merged their operations, R2 and G2 often 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 often is higher under PVI than under VS and because the combined profit of R2 and G2 often is higher under full vertical integration (VI) than under PVI, VI often arises in the absence of barriers to integration. Expected consumer surplus and expected total welfare typically 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 relatively intense competition that prevails under full vertical integration).

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

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.[3,4]

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,[5] and substantial vertical integration in regions of Australia,[6] the United States,[7] and the United Kingdom.[8]

Our analysis contributes to the literature in part by considering both endogenous capacity investment and endogenous vertical industry structure. 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

[3] Unilateral vertical integration is profitable in part because it eliminates double marginalization (Spengler, 1950). Despite this source of increased profit, ubiquitous vertical integration can reduce the equilibrium profit of both integrated enterprises by enhancing incentives for aggressive competition.

[4] The fact that vertical integration can entail prisoner's dilemma considerations is reminiscent of Allaz and Vila (AV) (1993)'s corresponding conclusion regarding forward contracting, which arises when a generator agrees before choosing its output level to subsequently deliver a specified level of output at a fixed price. Unilateral forward contracting is profitable in AV's model because it induces the wholesale supplier to compete relatively aggressively. 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. Furthermore, although ubiquitous vertical integration can sometimes reduce industry profit, it does not always do so.

[5] As of January 2020, 86% of the electricity produced in New Zealand was distributed to customers by generation-owned retailers (EMI, 2020).

[6] In Australia, the fraction of generation capacity held by vertically integrated firms has increased from less than 10% in 2004 to nearly 80% in 2016 (Frontier Economics, 2017, Figure 4).

[7] In 2017, incumbent vertically integrated suppliers served more than half of retail customers in Illinois, Massachusetts, and Ohio (EIA, 2018).

[8] As of 2016, the six largest suppliers of electricty in the UK were all vertically integrated. These suppliers provided approximately 90% of retail supply and accounted for approximately 70% of the wholesale supply of electricity (CMA, 2016, p. 3). ENMAX, a vertically-integrated supplier, served more than half of electricity customers in Alberta, Canada between 2012 and 2015 (Brown and Eckert, 2017b).

incentives in a setting where all industry suppliers are vertically integrated.[9]

Our analysis complements the important work of Boom and Buehler (2020), who analyze endogenous capacity investment in the electricity sector under both VI and VS. Boom and Buehler (BB)'s model differs from ours in two primary respects. First, BB assume that all consumers purchase electricity from the retail supplier that charges the lowest price. The assumption of product homogeneity facilitates analytic characterization of equilibrium retail prices. However, we allow for imperfect retail price competition because, in practice, retail consumers often purchase electricity from incumbent suppliers even when competing retail suppliers charge lower prices (e.g., Woo et al., 2014; Hortaçsu et al., 2017; Deller et al., 2021).

Second, BB assume the wholesale market for electricity functions as a uniform price auction in which each generator always bids its entire capacity at a single unit price (e.g., von der Fehr and Harbor, 1993; Fabra et al., 2006). Supply is accepted first from the generator that submits the lowest bid. All supply that is bid is accepted as long as demand exceeds supply. The equilibrium wholesale price is: (i) the lowest bid that equates demand and supply; or (ii) the highest bid if aggregate demand exceeds aggregate supply at this bid price. Rationing is imposed if demand exceeds supply at the highest bid price.

In contrast, we allow generators to choose the fraction of their installed capacity that they wish to supply to the wholesale market. We also assume that large industrial customers adjust their consumption of electricity as the wholesale price changes. This formulation avoids rationing of retail customers, which is rare in practice.[10] This formulation also allows us to capture the fact that generators do not always supply their entire capacity to the wholesale market in practice.[11] In addition, our Cournot formulation of wholesale market

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

[10] Abstracting from major events such as hurricanes, electricity service interruptions in the U.S. averaged 2.3 hours per household in 2019 (EIA, 2021). Houser et al. (2017) estimate that between 2012 and 2016, less than 0.01% of all service interruption time reflected rationing caused by generation capacity constraints.

[11] Empirical studies find that generators bid less than their full capacity in practice (e.g., Borenstein et al.,

competition enjoys empirical support.[12]

Under the Bertrand wholesale competition in BB's model, each vertically integrated firm is highly motivated to avoid excess demand for its product, i.e., retail demand in excess of installed capacity. If a supplier ("supplier 1") faces excess demand, it must purchase electricity in the wholesale market to fulfil its retail obligation. When rival supplier 2 recognizes that supplier 1 faces excess demand, supplier 2 can increase its bid price to the level that fully exhausts supplier 1's rent. To limit the incidence of such rent extraction, each vertically integrated supplier acts to reduce the likelihood that it will face excess demand. A generator does so by undertaking a relatively large level of capacity investment and setting a relatively high retail price (to reduce the amount of electricity its retail customers seek to purchase).

Corresponding full rent extraction does not arise in our model with Cournot wholesale competition. The absence of such a harsh penalty for excess demand limits the incentive to suppress retail consumption by setting a high retail price. In addition, a low retail price enhances the commitment of a vertically integrated generator to compete aggressively when choosing its wholesale output. Thus, vertical integration often promotes lower retail prices in our model, whereas it induces higher retail prices in BB's model.

Like BB, we find that vertical integration often promotes expanded capacity investment. However, the enhanced investment does not serve to limit the incidence of full rent extraction in our model. Instead, increased capacity investment admits increased output, which reduces the wholesale price, thereby increasing the downstream profit margin. A vertically integrated generator values this source of increased downstream profit.

The imperfect retail price competition and the Cournot wholesale quantity competition we analyze allow us to capture realistic elements of industry practice and examine new con-

2002; Joskow and Kahn, 2002; Puller, 2007; Bushnell et al., 2008). Doing so can increase wholesale profit by increasing the market-clearing wholesale price of electricity.

[12] Bushnell et al. (2008) find that, accounting for all relevant generator forward commitments, observed outcomes in wholesale electricity markets are predicted reasonably well by models of Cournot competition. Other studies that employ Cournot competition to model the interaction among electricity generators include Ventosa et al. (2005), Brown and Eckert (2017a), and Lundin and Tangeras (2020).

siderations of interest (e.g., the enhanced wholesale commitment that lower retail prices can provide). However, the complexity these features add to the model preclude us from providing a systematic analytic characterization of the impact of vertical integration on industry outcomes.[13] Therefore, we employ a combination of analytic conclusions and calibrated numerical solutions to assess the overall impact of vertical integration on industry outcomes. We identify a broad range of conditions under which the identified effects of vertical integration prevail.

The analysis proceeds 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 considers the key qualitative effects of vertical integration on retail electricity prices. Section 5 reviews the corresponding effects on industry investment. Section 6 presents the calibrated numerical solutions that complement our analytic conclusions. Section 7 summarizes our primary findings and discusses extensions of our work. The Appendix provides the proofs of all formal conclusions.

## 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") G$i$ can produce electricity at constant unit cost $c_i > 0$ up to capacity $K_i \geq 0$ (for $i \in \{1, 2\}$). Generator G$i$'s unit cost of capacity is $k_i > 0$.

Retail supplier (or "retailer") R$i$'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 R$i$'s incremental

[13] The consideration of imperfect retail price competition complicates the characterization of retail price reaction functions. This characterization is further complicated because (as is the case in virtually all retail electricity markets in practice) the final demand for electricity is not known at the time retail prices are set. Consequently, in our model of Cournot wholesale competition, the impact of a change in a retail price varies according to which of four (endogenous) regimes prevails when the generators choose outputs: neither G1 nor G2 is capacity constrained, both generators are constrained, only G1 is constrained, or only G2 is constrained. As Conclusion 6 (below) reports, a higher retail price can either increase or reduce wholesale profit, depending on which of these regimes prevails.

The consideration of imperfect retail price competition also complicates the determination of the impact on retail profit of a retail price reduction induced by expanded capacity investment under vertical integration. (See the discussion in Section 5.)

unit downstream production cost. The demand for R$i$'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 R$i$'s retail price, $r_j$ is R$j$'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; Hortaçsu et al., 2017; Deller et al., 2021). 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 each purchase at wholesale price $w$ the amount of electricity it requires to serve the demand for its product. 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 $\eta$ is the realization of a stochastic demand state.[14] 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.

G$i$'s (wholesale) profit is:

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

where $q_i$ is G$i$'s output of wholesale electricity. R$i$'s corresponding (retail) profit is:

[14] $Q^I(\cdot)$ need not reflect only 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.

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

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 $\varepsilon$ is realized, so the aggregate demand for wholesale electricity is known. 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 $\varepsilon$.

We will analyze outcomes under three industrial structures: vertical separation (VS), 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 $\pi_2^R$ and G2 acts to maximize $\pi_2^G$.

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 non-cooperatively. Second, R1 and R2 set their retail prices ($r_1$ and $r_2$) simultaneously and non-cooperatively. Third, the demand state ($\varepsilon$) 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 non-cooperatively. Fifth, after the wholesale price of electricity ($w$) is determined (by equating demand and supply), R1 and R2 each purchase the electricity required to fulfil the demand of its retail customers.

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

**Figure 1. Timing in the Model.**

Equations (5) and (6) identify the many ways in which vertical integration can directly affect the profit of a generator (G$i$) and retail supplier (R$i$). Vertical integration can affect: (i) the wholesale profit margin (by altering the wholesale price of electricity, $w$); (ii) wholesale output, $q_i$; (iii) capacity investment, $K_i$; (iv) the retail profit margin (by affecting the wholesale ($w$) and retail ($r_i$) prices of electricity); and (v) retail output $Q_i^r$ (by changing retail prices $r_i$ and $r_j$.). We now proceed to assess these varied effects of vertical integration.

# 3 Wholesale Electricity Supply

We begin by characterizing industry outcomes in the third stage of the game where each generator chooses the amount of electricity it supplies in the wholesale market. These output choices are made after capacity investments $(K_1, K_2)$ are implemented, after retail prices $(r_1, r_2)$ are set, and after the state of demand $(\varepsilon)$ is realized. The characterization of output choices is facilitated by analyzing 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 G$i$ 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 state of demand $(\varepsilon)$, G$i$ 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, G$i$'s reaction function is:[15]

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

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 G$i$ and the equilibrium wholesale price of electricity, respectively, when demand state $\varepsilon$ is realized and $n$ generators are capacity

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

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 $\varepsilon_0$ to the largest realization of $\varepsilon$ for which neither generator is capacity constrained. Also define $\varepsilon_{12}$ to be the smallest realization of $\varepsilon$ for which both generators are capacity constrained.[16]

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

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

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

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

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

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) $\frac{\partial w^{2c}}{\partial Q_i^r} = \frac{\partial w^{2c}}{\partial Q_j^r} = b^w$.*

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

**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.

G$i$'s output ($q_i$) increases more rapidly as R$i$'s retail output ($Q_i^r$) increases when G$i$ and R$i$ are vertically integrated than when they act independently. This is the case because as $Q_i^r$ increases, R$i$'s retail profit increases more rapidly as $w$ declines, *ceteris paribus*.[17] When G$i$ values R$i$'s retail profit, G$i$ responds to the increased benefit of reducing $w$ by expanding its supply of electricity. Specifically, when G$i$ 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. These considerations lead G$i$ to increase its output more rapidly as $Q_i^r$ increases than as $Q_j^r$ increases when G$i$ and R$i$ are vertically integrated.[18]

These observations 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 G$i$ values R$i$'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 R$i$'s profit margin.[19]

[17] R$i$'s retail profit margin increases as $w$ declines.

[18] From Conclusion 2, $\frac{\partial q_i^{0c}}{\partial Q_j^r} < \frac{\partial q_i^{0c}}{\partial Q_i^r}$ and $\frac{\partial q_i^{1c}}{\partial Q_j^r} < \frac{\partial q_i^{1c}}{\partial Q_i^r}$ if G$j$ is capacity constrained when $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$. These derivatives reflect the fact that G$i$ values R$i$'s retail profit, but not R$j$'s retail profit, under VI.

[19] Thus, expanded retail output under vertical integration in our model affects wholesale output decisions

## 4 Retail Electricity Prices.

We now analyze the second stage of the game where the retail firms set their prices after capacity investments have been undertaken but before the demand state and wholesale outputs of electricity are determined. Our primary concern is how VI affects the prices the retail firms set.

Under VI, R$i$ values both its expected downstream profit, $E\{\pi_i^R\}$, and G$i$'s expected upstream profit, $E\{\pi_i^G\}$. These expected profits are, for $l \in \{G, R\}$:

$$E\left\{\pi_i^l\right\} \;=\; \int_{\underline{\varepsilon}}^{\varepsilon_0} \Pi_i^{l0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \Pi_i^{l1c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \Pi_i^{l2c}(\varepsilon)\, dH(\varepsilon) \tag{12}$$

where, for $n \in \{0,1,2\}$: $\Pi_i^{Rnc}(\varepsilon) \;\equiv\; [\, r_i - w^{nc}(\varepsilon) - c_i^r \,]\, Q_i^r(r_1, r_2)\,;$ and

$$\Pi_i^{Gnc}(\varepsilon) \;\equiv\; [\, w^{nc}(\varepsilon) - c_i \,]\, q_i^{nc}(\varepsilon) - k_i\, K_i\,. \tag{13}$$

Therefore, under VI, R$i$'s choice of $r_i$ is determined by:

$$\frac{\partial E\left\{\pi_i^G\right\}}{\partial r_i} + \frac{\partial E\left\{\pi_i^R\right\}}{\partial r_i} \;=\; 0\,. \tag{14}$$

As is apparent from equation (14), VI can affect R$i$'s choice of $r_i$ by affecting the rate at which expected retail profit and expected wholesale profit vary with $r_i$. The latter rate is considered in Lemma 4.[20]

**Lemma 4.** $$\frac{\partial E\left\{\pi_i^G\right\}}{\partial r_i} \;=\; \int_{\underline{\varepsilon}}^{\varepsilon_0} \left\{ [\, w^{0c}(\varepsilon) - c_i \,]\, \frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{0c}(\varepsilon)}{\partial r_i}\, q_i^{0c}(\varepsilon) \right\} dH(\varepsilon)$$

much like forward commitments do in models with forward contracting (e.g., Allaz and Vila, 1993). In the latter models, an *ex ante* commitment to deliver more output at a fixed price reduces the amount of a generator's output that is sold at the wholesale price. This reduced exposure to the wholesale price reduces the rate at which the generator's profit declines as the wholesale price declines. The resulting reduced concern with reductions in the wholesale price induces generators to increase their outputs. In similar fashion, expanded retail output in our model induces a vertically integrated generator to increase its output because the resulting reduction in the wholesale price enhances retail profit more rapidly the higher is retail output.

[20] The rate at which R$i$'s expected retail profit increases with $r_i$ is the difference between the rate at which its retail revenue increases ($\frac{\partial}{\partial r_i}\,(\, r_i\, Q_i^r(\cdot)\,)$) and the rate at which its expected retail cost increases ($\frac{\partial}{\partial r_i}\, E\{\, [\, w(\cdot) + c_i^r \,]\, Q_i^r(\cdot)\,\}$). The latter rate includes the impact of $r_i$ on the expected wholesale price, $E\{\, w(\cdot)\,\}$. These equilibrium rates can vary aross vertical industry structures because equilibrium retail and wholesale prices can vary across structures.

$$+ \int_{\varepsilon_0}^{\varepsilon_{12}} \left\{ \left[ w^{1c}(\varepsilon) - c_i \right] \frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{1c}(\varepsilon)}{\partial r_i} \, q_i^{1c}(\varepsilon) \right\} dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \frac{\partial w^{2c}(\varepsilon)}{\partial r_i} \, K_i \, dH(\varepsilon) \, . \quad (15)$$

Expression (15) follows from differentiating expression (13), noting that G$i$'s output does not vary with $r_i$ when G$i$ is capacity constrained.[21] Lemma 4 indicates that R$i$'s additional concern with G$i$'s upstream profit under VI induces R$i$ to favor: (i) expanded output by G$i$ whenever the upstream profit margin ($w^{\bullet}(\varepsilon) - c_i$) is positive; and (ii) a higher wholesale price of electricity, which increases G$i$'s wholesale profit margin.

Conclusion 4 reports that R$i$ can ensure G$i$'s output never declines, and often increases, under VI by reducing $r_i$.

**Conclusion 4.** *When G$i$ and R$i$ are vertically integrated: (i)* $\frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i} = -\frac{1}{3}\left[\, 3\, b_i - (1-\alpha_j)\, d_j \,\right] < 0$*; (ii)* $\frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} = -\frac{1}{2}\left[\, 2\, b_i - d_j \,\right] < 0$ *if (only) G$j$ is capacity constrained when* $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$*; (iii)* $\frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} = 0$ *if G$i$ is capacity constrained when* $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$*; and (iv)* $\frac{\partial q_i^{2c}(\varepsilon)}{\partial r_i} = 0$*.*

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, G$i$ 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 G$i$ with a credible commitment to compete more aggressively against G$j$. The larger is $Q_i^r$, the more rapidly R$i$'s retail profit increases as its profit margin increases. Therefore, the larger is $Q_i^r$, the stronger is G$i$'s incentive to increase $q_i$ in order to reduce $w$ and thereby increase R$i$'s profit margin, *ceteris paribus*. Thus, by reducing $r_i$, R$i$ can increase G$i$'s expected wholesale profit by enhancing G$i$'s commitment to expand its output relatively aggressively in its competition with G$j$.

Unlike the impact of $r_i$ on G$i$'s output, the impact of $r_i$ on the wholesale price is not systematic, as Conclusion 5 reports.

[21] Formally, $\frac{\partial q_i^{2c}(\varepsilon)}{\partial r_i} = 0$. Similarly, $\frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} = 0$ if G$i$ is capacity constrained when $\varepsilon \in (\, \varepsilon_0, \varepsilon_{12} \,)$.

**Conclusion 5.** *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 5(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.[22]

Conclusion 5(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 thereby increases the demand for wholesale electricity. $w$ increases when binding capacity constraints preclude any offsetting increase in the supply of electricity.

Conclusion 5(iii) reports that $w$ also increases as $r_i$ declines when only G$i$ is capacity constrained. In this case, G$i$ 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, G$j$ reduces $q_j$ to offset this source of reduced demand for electricity under VI. (Recall Conclusion 2.)[23]

In contrast, Conclusion 5(iv) indicates that a reduction in $r_i$ causes $w$ to decline when only G$j$ is capacity constrained. Under VI in this case, G$i$ increases $q_i$ to offset the increased demand for electricity that arises when the reduction in $r_i$ causes $Q_i^r$ to increase. However, when it is capacity constrained, G$j$ does not adjust $q_j$ as $Q_j^r$ declines due to the reduction in $r_i$. Consequently, the demand for electricity declines by more than the supply declines, causing $w$ to decline.

---

22 $\frac{\partial w^{0c}(\varepsilon)}{\partial r_i} = \frac{b^w}{3}\, d_j > 0$ if G$i$ and R$i$ are vertically integrated but G$j$ and R$j$ are not vertically integrated. In this case, a reduction in $r_i$ reduces $Q_j^r$, and G$j$ does not reduce its supply of electricity to offset this reduction in the demand for electricity. Consequently, the wholesale price of electricity declines.

23 $\frac{\partial w^{1c}(\varepsilon)}{\partial r_i} = -\, \frac{b^w}{2} \left[\, b_i - d_j \,\right] < 0$ if G$i$ and R$i$ are vertically integrated but G$j$ and R$j$ are not vertically integrated. In this case, G$j$ 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 $r_i$.

Conclusions 4 and 5 imply that a reduction in $r_i$ enhances G$i$'s upstream profit under VI except possibly when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$. This finding is recorded formally in Conclusion 6.

**Conclusion 6.** *Suppose Gi's upstream profit margin is nonnegative for all* $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$. *Then* $\frac{\partial \Pi_i^{G0c}(\varepsilon)}{\partial r_i} < 0$ *and* $\frac{\partial \Pi_i^{G2c}(\varepsilon)}{\partial r_i} < 0$ *under VI. Furthermore,* $\frac{\partial \Pi_i^{G1c}(\varepsilon)}{\partial r_i} < 0$ *under VI if Gi is capacity constrained for* $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$.

Conclusions 4 – 6 suggest that, holding capacity investment constant, R$i$ often will reduce the price it charges for electricity as the value that R$i$ places on G$i$'s upstream profit ($\alpha_i$) increases. R$i$ always does so in the *symmetric setting* where $\alpha_1 = \alpha_2$, $a_1 = a_2$, $b_1 = b_2$, $c_1 = c_2 \equiv c$, $c_1^r = c_2^r$, $d_1 = d_2$, and in equilibrium $K_1 = K_2$, $r_1 = r_2$, and $q_1^{nc}(\varepsilon) = q_2^{nc}(\varepsilon)$ for all $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ and for $n \in \{0, 2\}$.[24] It will never be the case that exactly one generator is capacity constrained in the symmetric setting, so $H(\varepsilon_{12}) - H(\varepsilon_0) = 0$.[25] Therefore, Conclusion 6 implies that a reduction in $r_i$ always increases G$i$'s upstream profit under VI in this setting. Furthermore, although $r_j$ can increase as $\alpha_i$ increases in this setting,[26] the average retail price ($\frac{1}{2}[r_1 + r_2]$) declines as $\alpha_i$ increases.

**Conclusion 7.** $\left.\frac{\partial r_i}{\partial \alpha_i}\right|_{K_i, K_j \ constant} < 0$ *and* $\left.\frac{\partial (r_i + r_j)}{\partial \alpha_i}\right|_{K_i, K_j \ constant} < 0$ *in the symmetric setting.*

Observe that:

[24] We assume the standard sufficient conditions for a maximum are satisfied in the symmetric setting. Formally, let $L^i$ denote R$i$'s objective function and let subscripts here denote partial derivatives. We assume that for $i, j \in \{1, 2\}$ ($j \neq i$), $L^i_{r_i r_i} < 0$, $L^j_{r_j r_j} < 0$, and $L^i_{r_i r_i} L^j_{r_j r_j} > L^i_{r_i r_j} L^j_{r_j r_i}$.

[25] Conceivably, the enterprises might choose different capacities, wholesale outputs, or retail prices even when their demand and cost parameters are symmetric. The symmetric setting restricts attention to instances in which these endogenous variables are symmetric in equilibrium. PVI cannot arise in the symmetric setting because $\alpha_1 = \alpha_2$ in this setting.

[26] $\left.\frac{\partial r_j}{\partial \alpha_i}\right|_{K_i, K_j \ \text{constant}}$ can be positive or negative in the symmetric setting. R$j$ may find it most profitable to reduce $r_j$ as R $i$ reduces $r_i$ in response to an increase in $\alpha_i$. However, the reduction in $r_i$ increases the demand for electricity and thereby increases the equilibrium expected wholesale price of electricity. The higher expected input cost can induce R$j$ to increase $r_j$, especially when product homogeneity is limited (i.e., when $d_1 = d_2$ is close to 0).

$$\frac{dr_i}{d\alpha_i} \;=\; \left.\frac{\partial r_i}{\partial \alpha_i}\right|_{K_i, K_j \text{ constant}} + \frac{dr_i}{dK_i}\,\frac{dK_i}{d\alpha_i} + \frac{dr_i}{dK_j}\,\frac{dK_j}{d\alpha_i}\,. \tag{16}$$

Equation (16) and Conclusion 7 imply that if generators do not reduce their capacity investments as $\alpha_i$ increases (so $\frac{dK_i}{d\alpha_i} \geq 0$ and $\frac{dK_j}{d\alpha_i} \geq 0$) and if retail prices decline as capacity investments increase (so $\frac{dr_i}{dK_i} < 0$ and $\frac{dr_i}{dK_j} < 0$), then vertical integration will ensure lower retail prices in the symmetric setting. We show in Section 5 that retail prices do indeed decline as capacity investments increase in the symmetric setting.

## 5 Capacity Investment

To analyze the generators' capacity investments, recall that under VI, G$i$ acts to maximize the sum of its expected upstream profit and R$i$'s expected downstream profit, as specified in equations (12) and (13). Formally, G$i$'s choice of $K_i$ under VI is determined by:

$$\frac{\partial E\left\{\pi_i^G\right\}}{\partial K_i} + \frac{\partial E\left\{\pi_i^R\right\}}{\partial K_i} \;=\; 0\,. \tag{17}$$

As is apparent from equation (17), VI can affect G$i$'s choice of $K_i$ by affecting the rate at which expected wholesale profit and expected retail profit vary with $K_i$. Lemma 5 identifies some of the effects of VI on G$i$'s choice of $K_i$ by identifying the rate at which R$i$'s expected downstream profit changes with $K_i$.[27] The lemma refers to:

$$E\left\{w(\varepsilon)\right\} \;\equiv\; \int_{\underline{\varepsilon}}^{\varepsilon_0} w^{0c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} w^{1c}(\varepsilon)\, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} w^{2c}(\varepsilon)\, dH(\varepsilon)\,. \tag{18}$$

**Lemma 5.** $$\frac{dE\left\{\pi_i^R\right\}}{dK_i} \;=\; -\,Q_i^r\,\frac{\partial E\left\{w(\varepsilon)\right\}}{\partial K_i}$$
$$+\;\left[\,r_i - E\left\{w(\varepsilon)\right\} - c_i^r\,\right]\left[-\,b_i\,\frac{dr_i}{dK_i} + d_j\,\frac{dr_j}{dK_i}\right] + Q_i^r\,\frac{dr_i}{dK_i}\,. \tag{19}$$

Lemma 5 indicates that the value G$i$ places on $E\left\{\pi_i^R\right\}$ under VI introduces two distinct effects of a change in $K_i$: a wholesale price effect (reflected in the first term following the

[27] G$i$'s expected wholesale profit increases with $K_i$ at the rate $\frac{\partial}{\partial K_i}\,E\left\{\,[\,w(\cdot) - c_i\,]\,q_i(\cdot)\,\right\} - k_i$. This equilibrium rate can vary across vertical industry structures because wholesale prices and outputs can vary across structures.

equality in equation (19)) and a retail price effect (captured by the remaining terms following the equality in equation (19)). The wholesale price effect captures the impact of a change in $K_i$ on R$i$'s expected retail profit through its impact on the wholesale price (holding retail prices constant). The retail price effect captures the impact of a change in $K_i$ on R$i$'s expected retail profit through its impact on retail prices (holding the expected wholesale price constant).[28]

Conclusion 8 implies that the wholesale price effect of a change in $K_i$ encourages G$i$ to increase $K_i$.

**Conclusion 8.** $$\frac{\partial E\{w(\varepsilon)\}}{\partial K_i} = \psi_i \left[ -\frac{1}{2} b^w \right] [\, H(\varepsilon_{12}) - H(\varepsilon_0)\,] - b^w [\, 1 - H(\varepsilon_{12})\,] < 0\,,$$

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

An increase in $K_i$ permits an expanded supply of electricity when G$i$ is capacity constrained. The increased supply of electricity lowers the wholesale price of electricity, and thereby enhances R$i$'s retail profit, *ceteris paribus*. Because G$i$ values R$i$'s retail profit under VI, the wholesale price effect encourages G$i$ to increase $K_i$.

Tedious calculations admit a characterization of components of the retail price effect in the symmetric setting, as Conclusion 9 reports.

**Conclusion 9.** *$\frac{dr_i}{dK_i} < 0$ and $\frac{dr_j}{dK_i} < 0$ for $i, j \in \{1, 2\}$ ($j \neq i$) in the symmetric setting under both VS and VI.*

Increased capacity reduces the expected wholesale price. Under VS, the lower wholesale price reduces the cost of a key input for retailers, thereby reducing profit-maximizing retail prices. The same effect arises under VI. In addition, a lower retail price increases the wholesale demand for electricity. The associated increase in the wholesale price enhances upstream profit, which retail suppliers value under VI.

[28] As is evident from equation (17), G$i$ also considers the effects of $K_i$ on its expected upstream profit, and these effects can vary with the prevailing vertical structures in the industry.

Expanded retail output induced by lower retail prices increases retail profit when the retail profit margin is positive. (See the terms in square brackets in equation (19).) However, the lower retail price reduces retail profit, holding retail output constant. (See the last term in equation (19).) Therefore, the retail price effect is not readily signed, which makes it difficult to determine the impact of vertical integration on capacity investment analytically.[29] However, numerical solutions can be employed to derive additional insight regarding the effects of vertical integration on capacity investment and on prices, outputs, earnings, and welfare.

# 6 Numerical Solutions

The foregoing analysis suggests that vertical integration can affect the incentives of retailers and generators in two primary ways. First, VI can encourage a retailer (R$i$) to reduce the price it charges for electricity. The reduction in $r_i$ increases $Q_i^r$, which can enhance G$i$'s wholesale profit by increasing $w$ and by strengthening G$i$'s commitment to increase $q_i$ aggressively in its competition with G$j$. Second, VI can encourage a generator (G$i$) 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 R$i$'s retail profit, *ceteris paribus.*

We now examine numerical solutions in part to assess the extent to which vertical integration promotes lower retail prices and expanded capacity investment under circumstances that might plausibly arise in practice.[30] To illustrate some possible such circumstances, we initially consider a *baseline setting.* We subsequently consider alternative settings to assess

[29] Suppose G$l$ seeks to maximize its upstream profit plus the fraction $\alpha_l$ of R$l$'s downstream profit ($l \in \{i, j\}$). Then the impact of VI on G$i$'s investment given $\alpha_j$ can be written as $K_i(1, \alpha_j) - K_i(0, \alpha_j) = \int_0^1 \frac{dK_i(\alpha_i, \alpha_j)}{d\alpha_i} \, d\alpha_i$, where $K_i(\alpha_i, \alpha_j)$ is G$i$'s equilibrium capacity investment given $\alpha_i$ and $\alpha_j$. Standard techniques provide analytic expressions for $\frac{dK_i}{d\alpha_i}$ and $\frac{dK_j}{d\alpha_i}$. However, these expressions are not readily signed because they encompass the impact of vertical integration on the incentives of retailers and generators as well as the effects of a generator's capacity investment on retail and wholesale prices and outputs (including the indeterminate retail price effect).

[30] The methodology employed to derive the numerical solutions is described in Appendix B. Brown and Sappington (2021) reviews the additional analytic work that underlies the numerical solutions.

the robustness of our findings.

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).[31] In 2019, the average price of Henry Hub natural gas was \$2.597/mmBtu, or \$2.46/GJ (EIA, 2020).[32] 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. We round this estimate 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 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

[31] 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.

[32] 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.

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).[33] 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 $-0.33$.[34]

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.[35] 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).[36] 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 for electricity (when it is priced at wholesale marginal cost, 20) is assumed to be 2.70.[37]

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

[33] 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).

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

[35] 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.

[36] 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.

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

that $\varepsilon$ 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 ($E\{CS\}$),[38] expected total welfare ($E\{CS + \pi\}$),[39] and for $i \in \{1, 2\}$, G$i$'s expected wholesale output ($E\{q_i\}$) and profit ($E\{\pi_i^G\}$), and R$i$'s expected retail profit ($E\{\pi_i^R\}$). 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.[40]

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 4, vertically integrated retailers often 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 5, elevated capacity and output can arise under VI because a vertically integrated generator values the increase in retail profit that arises when an increased supply of electricity reduces the expected wholesale price.

[38] 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 ($\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{w}^{\frac{a^I}{b^I} + \varepsilon} [a^L - b^L(\widetilde{w} + \varepsilon)] \, d\widetilde{w} \, dH(\varepsilon)$).

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

[40] 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.**

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.[41] 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 and thereby secure increased expected

[41] We abstract from bargaining frictions and transactions costs and assume that G$i$ and R$i$ merge whenever doing so increases their joint profit. In principle, G$i$ and R$i$ can choose ownership shares to achieve any desired distribution of the incremental profit their merger generates.

profit.[42] 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.[43] 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.[44] 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 equation (1) indicates, an increase in $d_1 = d_2$ renders each retailer's demand more sensitive to changes in its rival's price.

In the setting that underlies Table 3, each retailer-generator pair has a unilateral incentive to integrate.[45] However, ubiquitous vertical integration reduces the expected profit of

[42] 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 G2 values the increased expected profit of R2 that arises when an increased supply of electricity reduces the expected wholesale price.

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

[44] Vertical integration can enhance competitive aggression and increase capacity investment and wholesale output even when vertical integration does not eliminate a double marginalization problem. Specifically, suppose each retail supplier always sets its retail price equal to its expected unit cost (i.e., $r_i = E\{w\} + c_i^r$ for $i \in \{1, 2\}$). Numerical solutions reveal that unilateral vertical integration by G1 and R1 is profitable in the baseline setting (and more generally) in this case. Furthermore, under PVI, G1 undertakes more capacity investment and produces more output than does G2. Because G1 values R1's retail profit under PVI and because $r_1$ is fixed at the time G1 sets $q_1$, G1 increases its output in order to reduce the wholesale price and thereby increase R1's retail profit margin. G2 and R2 find it profitable to integrate once G1 and R1 have done so in the baseline setting. In equilibrium, the combined expected profit of R$i$ and G$i$ ($i \in \{1, 2\}$) is lower under VI than under VS in the baseline setting when retail prices reflect expected production costs. See Part IV in Brown and Sappington (2021) for additional details.

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

both integrated firms.[46] 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).

| | 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 $d_1 = d_2 = 3.5$.**

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 (2021) document 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.

under PVI.

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

It remains to explain how the private gains from vertical integration vary with model parameters. Figure 2 illustrates how these gains vary with the cross-price sensitivity of retail demand (i.e., with $d_1 = d_2 \equiv d$), holding all other parameters at their values in the baseline setting. The upper curve in Figure 2 (labelled "PVI") depicts the percentage increase in the combined expected profit of R1 and G1 ("R1-G1") under PVI (where only R1 and G1 are integrated) relative to their profit under VS. Formally, this measure of R1-G1's gain from unilateral integration is $100\left[\frac{\Pi_1^P - \Pi_1^S}{\Pi_1^S}\right]$ where $\Pi_1^P = E\left\{\pi_1^G + \pi_1^R\right\}$ under PVI and $\Pi_1^S = E\left\{\pi_1^G + \pi_1^R\right\}$ under VS. The lower curve in Figure 2 (labelled "VI") depicts the corresponding percentage increase in $E\left\{\pi_1^G + \pi_1^R\right\}$ under VI relative to its value under VS. Formally, this measure of R1-G1's gain from ubiquitous integration is $100\left[\frac{\Pi_1^V - \Pi_1^S}{\Pi_1^S}\right]$ where $\Pi_1^V = E\left\{\pi_1^G + \pi_1^R\right\}$ under VI. The variable on the horizontal axis in Figure 2 is $\%\Delta\, d \equiv 100\left[\frac{d - d^B}{d^B}\right]$, where $d^B$ is 3, the value of $d$ in the baseline setting.

Four elements of Figure 2 (and Table 2) warrant emphasis. First, R1-G1's profit increases moderately (i.e., by approximately 6.65%) when they alone vertically integrate in the baseline setting (where $d = d^B$, so $\%\Delta\, d = 0$). Second, the increase in R1-G1's expected profit when they integrate and R2 and G2 also integrate is *de minimis* (approximately 0.49%) in the baseline setting. Third, R1-G1's gain from unilateral and ubiquitous integration both decline as the cross-price retail demand sensitivity ($d$) increases. Fourth, R1-G1's gain from integration declines more rapidly as $d$ increases under ubiquitous integration than under unilateral integration (i.e., the lower curve in Figure 2 is more steeply sloped than the upper curve).

[ **Figure 2 Here** ]

These observations reflect the following considerations. As $d$ increases, retail demand shifts outward and becomes more sensitive to the rival's price. (Recall equation (1).) The increased cross-price sensitivity promotes lower retail prices. The associated reduction in R1-G1's retail profit declines with $d$ under unilateral vertical integration (due to R1-G1's asymmetric commitment to compete more aggressively) but increases with $d$ under ubiquitous vertical integration (due to the particularly intense retail competition that arises under bilateral integration). The increase in retail output ($Q_1^r$) induced by an increase in $d$ enhances G1's incentive to increase its capacity investment ($K_1$) under PVI and VI (in order to reduce the wholesale price, thereby enhancing R1's retail profit). Under the asymmetric enhanced incentive that arises under PVI, R1-G1 expands $K_1$ considerably, inducing G2 to reduce $K_2$. The associated substantial increase in R1-G1's wholesale profit more than offsets the reduction in retail profit, rendering unilateral vertical integration profitable. In contrast, ubiquitous vertical integration reduces R1-G1's profit (relatively rapidly) as $d$ increases above its level in the baseline setting. The aggressive wholesale and retail competition that arises when R1-G1 and R2-G2 are both vertically integrated and that intensifies as $d$ increases more than dissipates the increased surplus that arises from lower retail prices and expanded capacity investment.

Appendix D explains how changes in other model parameters affect the private gains from vertical integration. To illustrate, the private gains from unilateral and ubiquitous vertical integration both increase as the own-price retail demand sensitivity ($b$) increases. This is the case because as $b$ increases, retail demand contracts (holding prices constant) and becomes more sensitive to own-price changes, thereby promoting lower retail prices, output, and profit. The reduction in retail profit is more pronounced under PVI and VI than under VS because vertical integration enhances incentives to compete aggressively. The reduction in retail output reduces the wholesale demand for electricity, which reduces wholesale profit. The decline in wholesale profit is less pronounced under VI and PVI than under VS because vertical integration enhances incentives for capacity investment. On balance, the difference between wholesale profit under vertical integration and vertical separation increases more rapidly as $b$ increases than the corresponding difference in retail profit declines. Consequently, the private gains from from unilateral and ubiquitous integration increase as $b$ increases.

In contrast, the private gains from vertical integration decline as the price sensitivity of industrial demand ($b^I$) increases. This is the case because as $b^I$ increases, generators become more intent on securing a relatively low wholesale price. Consequently, commitments to compete relatively aggressively (which promote lower retail prices and thus increased retail output which places upward pressure on the wholesale price) become less lucrative.

## 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 often 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 often 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.

The basic forces that underlie these findings in our model seem likely to persist more generally.[47] They likely persist, for example, under other formulations of wholesale market competition. We have considered Cournot competition in part to avoid the multiplicity of equilibria that arise when generators choose supply functions rather than output levels.[48] Klemperer and Meyer (1989) demonstrate that models of supply function competition admit a continuum of equilibria in which aggregate output can vary between the equilibrium outputs in Cournot and Bertrand models of competition. Thus, relative to the range of possible supply function equilibria, the equilibrium in our Cournot model might be viewed as one in which generators have relatively pronounced market power.[49] This market power and associated profitability may enhance incentives for capacity investment. However, the key

[47] We have verified that these forces 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 from their initial (incumbent) supplier.

[48] A supply function specifies how the output a firm will supply varies with the prevailing price for the output.

[49] Willems et al. (2009) find that Cournot models and models with supply functions explain similar fractions of observed price variation in the German electricity sector. The authors "therefore suggest using Cournot models for short-term analysis, since these models can accommodate additional market details, such as network constraints" (p. 38).

considerations that determine investment decisions in our model seem likely to underlie investment decisions in models where generators choose supply functions rather than output levels.[50]

The key forces at play in our model also seem likely to persist in the presence of risk aversion.[51] Even as vertical integration helps to reduce the variation in profit introduced by changes in the wholesale price of electricity,[52] vertical integration can simultaneously enhance the commitment of the integrated suppliers to compete relatively aggressively. Thus, vertical integration seems likely to arise in equilibrium and to reduce retail prices and expand capacity investment and electricity supply even when industry suppliers are risk averse.[53]

The primary forces that we have identified also seem likely to persist when industry suppliers can undertake forward contracting.[54] As Allaz and Vila (1993) demonstrate, a generator can employ forward contracting to enhance its commitment to compete aggressively. The presence of commitment secured via forward contracting could diminish the commitment value of vertical integration that we have identified. However, even in the presence of forward contracting, unilateral vertical integration seems likely to enhance profit for the reasons we have identified.[55] Future research might examine the extent to which

[50] Recall that Boom and Buehler (2020) also find that vertical integration enhances capacity investment in the presence of Bertrand wholesale competition.

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

[52] 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.

[53] For analytic ease, we have abstracted from uncertainty about retail demand. Such uncertainty could reduce incentives for vertical integration when suppliers are risk averse. A low demand realization reduces retail profit, holding the wholesale price constant. The low demand realization also promotes a lower wholesale price, which serves to increase retail profit, *ceteris paribus*. The countervailing effect of the reduced wholesale price is attractive to a risk averse retail supplier under vertical separation. Under vertical integration, the reduced wholesale price serves to reduce wholesale profit, which compounds the reduction in retail profit caused by the low demand realization. The correlated profit reductions can reduce the attraction of vertical integration to risk averse suppliers.

[54] Recall that under forward contracting, a wholesale supplier commits to deliver a specified amount of electricity to a retail customer 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).

[55] Furthermore, the extent to which a generator is vertically integrated typically is readily observed. In

the enhanced commitments to compete aggressively engendered by forward contracting and vertical integration are substitutes or complements.[56]

contrast, a generator's forward commitments may be less apparent, and the commitment effect of a forward commitment varies with its observability (Hughes and Kao, 1997). Consequently, in practice, a generator's commitment to compete aggressively may be enhanced more by vertical integration than by forward contracting.

[56] 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 four 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 D illustrates how the private gains from vertical integration vary with model parameters.

## Appendix A. Proofs of Formal Conclusions in the Text

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

$$\begin{aligned} & w'(Q)\,[\,q_i - \alpha_i\,Q_i^r\,] + w(Q) - c_i \;=\; 0 \\ \Leftrightarrow \quad & -\,b^w\,[\,q_i - \alpha_i\,Q_i^r\,] + b^w\left[\,a^I + Q_i^r + Q_j^r\,\right] - b^w\,q_i - b^w\,q_j + \varepsilon - c_i \;=\; 0 \\ \Leftrightarrow \quad & q_i \;=\; \frac{b^w\left[\,a^I + Q_j^r + Q_i^r\,(1+\alpha_i)\,\right] + \varepsilon - c_i}{2\,b^w} - \frac{1}{2}\,q_j\,. \end{aligned} \tag{20}$$

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

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

$$\Rightarrow \quad q_1^{0c}(\varepsilon) \;=\; \frac{b^w\left[\,a^I + (1+2\,\alpha_1)\,Q_1^r + (1-\alpha_2)\,Q_2^r\,\right] + \varepsilon + c_2 - 2\,c_1}{3\,b^w}\,. \tag{21}$$

(20) and (21) imply:

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

$$\Rightarrow \quad q_2^{0c}(\varepsilon) \;=\; \frac{b^w\left[\,a^I + (1+2\,\alpha_2)\;Q_2^r + (1-\alpha_1)\,Q_1^r\,\right] + \varepsilon + c_1 - 2\,c_2}{3\,b^w}\,. \tag{22}$$

(4), (21), and (22) imply that the corresponding wholesale price when $\varepsilon$ is realized in this case is:

$$
\begin{aligned}
w^{0c}(\varepsilon) &= b^w \left[ a^I + Q_1^r + Q_2^r \right] - b^w \left[ q_1^{0c}(\varepsilon) + q_2^{0c}(\varepsilon) \right] + \varepsilon \\
&= \frac{1}{3} \{ 3\, b^w \left[ a^I + Q_1^r + Q_2^r \right] + 3\, \varepsilon \\
&\qquad - b^w \left[ 2\, a^I + (2+\alpha_1)\, Q_1^r + (2+\alpha_2)\, Q_2^r \right] - 2\, \varepsilon + c_1 + c_2 \} \\
&= \frac{1}{3} \left\{ b^w \left[ a^I + (1-\alpha_1)\, Q_1^r + (1-\alpha_2)\, Q_2^r \right] + \varepsilon + c_1 + c_2 \right\} . \ \blacksquare \qquad (23)
\end{aligned}
$$

**Proof of Lemma 2**. Because G$j$'s equilibrium output in this case is $q_j^{1c}(\varepsilon) = K_j$, (20) implies that G$i$'s equilibrium output is:

$$
q_i^{1c}(\varepsilon) = \frac{1}{2} \left[ a^I + Q_j^r + Q_i^r\,(1+\alpha_i) \right] + \frac{1}{2\, b^w} \left[ \varepsilon - c_i \right] - \frac{1}{2}\, K_j . \qquad (24)
$$

(4) and (24) imply that when $\varepsilon$ is realized in this case:

$$
\begin{aligned}
w^{1c}(\varepsilon) &= b^w \left[ a^I + Q_i^r + Q_j^r \right] - \frac{b^w}{2} \left[ a^I + Q_j^r + Q_i^r\,(1+\alpha_i) + K_j \right] - \frac{1}{2} \left[ \varepsilon - c_i \right] + \varepsilon \\
&= \frac{1}{2}\, b^w \left[ a^I + Q_i^r\,(1-\alpha_i) + Q_j^r \right] + \frac{1}{2} \left[ \varepsilon + c_i - b^w\, K_j \right] . \ \blacksquare \qquad (25)
\end{aligned}
$$

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

$$
w^{2c}(\varepsilon) = b^w \left[ a^I + Q_1^r + Q_2^r - K_1 - K_2 \right] + \varepsilon . \ \blacksquare \qquad (26)
$$

**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 (13) provides:

$$
\begin{aligned}
\frac{\partial E\left\{\pi_i^G\right\}}{\partial r_i} &= \int_{\underline{\varepsilon}}^{\varepsilon_0} \left\{ \left[ w^{0c}(\varepsilon) - c_i \right] \frac{\partial q_i^{0c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{0c}(\varepsilon)}{\partial r_i}\, q_i^{0c}(\varepsilon) \right\} dH(\varepsilon) \\
&\quad + \int_{\varepsilon_0}^{\varepsilon_{12}} \left\{ \left[ w^{1c}(\varepsilon) - c_i \right] \frac{\partial q_i^{1c}(\varepsilon)}{\partial r_i} + \frac{\partial w^{1c}(\varepsilon)}{\partial r_i}\, q_i^{1c}(\varepsilon) \right\} dH(\varepsilon)
\end{aligned}
$$

$$+ \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \left\{ \left[ w^{2c}(\varepsilon) - c_i \right] \frac{\partial K_i}{\partial r_i} + \frac{\partial w^{2c}(\varepsilon)}{\partial r_i} K_i \right\} dH(\varepsilon)$$

$$+ \left\{ \left[ w^{0c}(\varepsilon_0) - c_i \right] q_i^{0c}(\varepsilon_0) - \left[ w^{1c}(\varepsilon_0) - c_i \right] q_i^{1c}(\varepsilon_0) \right\} h(\varepsilon_0) \frac{d\varepsilon_0}{dr_i}$$

$$+ \left\{ \left[ w^{1c}(\varepsilon_{12}) - c_i \right] q_i^{1c}(\varepsilon_{12}) - \left[ w^{2c}(\varepsilon_{12}) - c_i \right] K_i \right\} h(\varepsilon_{12}) \frac{d\varepsilon_{12}}{dr_i} . \tag{27}$$

(15) follows from (27) 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 G$i$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$; (v) $\lim\limits_{\varepsilon \nearrow \varepsilon_{12}} q_i^{1c}(\varepsilon) = K_i$ when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$; and (vi) $w^{2c}(\varepsilon_{12}) = w^{1c}(\varepsilon_{12})$.

(i) holds because G$i$ chooses $K_i$ before R$i$ chooses $r_i$. To prove that (iii) and (vi) hold, consider the case where G$j$ is capacity constrained when $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$. (9) implies that in this case:

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

$$\Rightarrow \varepsilon_0 = 3b^w K_j + 2c_j - c_i - b^w \left[ a^I + (1 + 2\alpha_j) Q_j^r + (1 - \alpha_i) Q_i^r \right] . \tag{28}$$

(10) implies that in this case:

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

$$\Rightarrow \varepsilon_{12} = 2b^w K_i + b^w K_j + c_i - b^w \left[ a^I + Q_i^r (1 + \alpha_i) + Q_j^r \right] . \tag{29}$$

(9) and (28) imply that under VI in this case:

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

$$= b^w K_j + c_j - b^w Q_j^r . \tag{30}$$

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

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

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

$$= -b^w Q_j^r + b^w K_j + c_j = w^{0c}(\varepsilon_0) . \tag{31}$$

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

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

$$- \; b^w \left[\, a^I + 2\, Q_i^r + Q_j^r \,\right] \;=\; -\, b^w\, Q_i^r + b^w\, K_i + c_i\,. \tag{32}$$

(10), (29), and (32) imply that under VI in this case:

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

$$-\; \frac{1}{2}\, b^w \left[\, a^I + 2\, Q_i^r + Q_j^r \,\right] \;=\; c_i + b^w\, K_i - b^w\, Q_i^r \;=\; w^{2c}(\varepsilon_{12})\,. \tag{33}$$

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

To prove that (ii) holds, observe that (9) and (28) imply that under VI:

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

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

$$=\; Q_i^r - Q_j^r + \frac{1}{b^w} \left[\, c_j - c_i \,\right] + K_j\,. \tag{34}$$

(10), (28), and (34) imply that when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$ under VI:

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

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

$$=\; Q_i^r - Q_j^r - \frac{1}{b^w}\, c_i + \frac{1}{b^w}\, c_j + K_j \;=\; q_i^{0c}(\varepsilon_0)\,. \tag{35}$$

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

To prove that (v) holds when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$, observe that (10) and (29) imply:

$$\lim_{\varepsilon \nearrow \varepsilon_{12}} \, q_i^{1c}(\varepsilon) \;=\; \frac{1}{2} \left[\, a^I + 2\, Q_i^r + Q_j^r \,\right] - \frac{1}{2\, b^w}\; c_i - \frac{1}{2}\, K_j$$

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

$$=\; -\, \frac{1}{2\, b^w}\, c_i - \frac{1}{2}\, K_j + \frac{1}{2\, b^w} \left[\, 2\, b^w\, K_i + b^w\, K_j + c_i \,\right] \;=\; K_i\,. \;\; \blacksquare \tag{36}$$

**Proof of Conclusion** 4. The Conclusion follows directly from (1), (9), and (10). ■

**Proof of Conclusion 5**. The Conclusion follows directly from (1), (9), (10), and (11). ■

**Proof of Conclusion 6**. (13) implies that for $n \in \{0, 1, 2\}$:

$$\frac{\partial \Pi_i^{Gnc}(\varepsilon)}{\partial r_i} = \left[ w^{nc}(\varepsilon) - c_i \right] \frac{\partial q_i^{nc}}{\partial r_i} + q_i^{nc} \frac{\partial w^{nc}(\varepsilon)}{\partial r_i} . \qquad (37)$$

Conclusions 4 and 5 imply that $\frac{\partial q_i^{0c}}{\partial r_i} < 0$ and $\frac{\partial w^{0c}(\varepsilon)}{\partial r_i} = 0$ under VI. Therefore, (37) implies that $\frac{\partial \Pi_i^{G0c}(\varepsilon)}{\partial r_i} < 0$ under VI when $w^{0c}(\varepsilon) > c_i$.

Conclusions 4 and 5 imply that $\frac{\partial q_i^{2c}}{\partial r_i} = 0$ and $\frac{\partial w^{2c}(\varepsilon)}{\partial r_i} < 0$ under VI. Therefore, (37) implies that $\frac{\partial \Pi_i^{G2c}(\varepsilon)}{\partial r_i} < 0$ under VI when $w^{2c}(\varepsilon) > c_i$.

Conclusions 4 and 5 imply that $\frac{\partial q_i^{1c}}{\partial r_i} = 0$ and $\frac{\partial w^{2c}(\varepsilon)}{\partial r_i} < 0$ under VI when only G$i$ is capacity constrained. Therefore, (37) implies that $\frac{\partial \Pi_i^{G1c}(\varepsilon)}{\partial r_i} < 0$ in this case when $w^{1c}(\varepsilon) > c_i$. ■

**Proof of Conclusion 7**.[57] Given $\alpha_i$, R$i$ seeks to maximize:

$$\begin{aligned} L^i \equiv & \int_{\underline{\varepsilon}}^{\varepsilon_0} \left\{ \alpha_i \left[ w^{0c}(\varepsilon) - c_i \right] q_i^{0c}(\varepsilon) + \left[ r_i - c_i^r - w^{0c}(\varepsilon) \right] Q_i^r(r_i, r_j) \right\} dH(\varepsilon) \\ & + \int_{\varepsilon_0}^{\varepsilon_{12}} \left\{ \alpha_i \left[ w^{1c}(\varepsilon) - c_i \right] q_i^{1c}(\varepsilon) + \left[ r_i - c_i^r - w^{1c}(\varepsilon) \right] Q_i^r(r_i, r_j) \right\} dH(\varepsilon) \\ & + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \left\{ \alpha_i \left[ w^{2c}(\varepsilon) - c_i \right] q_i^{2c}(\varepsilon) + \left[ r_i - c_i^r - w^{2c}(\varepsilon) \right] Q_i^r(r_i, r_j) \right\} dH(\varepsilon) - \alpha_i \, k_i \, K_i . \qquad (38) \end{aligned}$$

(38) implies that equilibrium choices of $r_i$ and $r_j$ are determined by:

$$\frac{\partial L^i}{\partial r_i} \equiv L^i_{r_i} = 0 \quad \text{and} \quad \frac{\partial L^j}{\partial r_j} \equiv L^j_{r_j} = 0 . \qquad (39)$$

Define $L^i_{xy} \equiv \frac{\partial^2 L^i}{\partial x \partial y}$. Differentiating (39) provides:

$$\begin{bmatrix} L^i_{r_i r_i} & L^i_{r_i r_j} \\ L^j_{r_j r_i} & L^j_{r_j r_j} \end{bmatrix} \begin{bmatrix} dr_i \\ dr_j \end{bmatrix} = - \begin{bmatrix} L^i_{r_i \alpha_i} \\ L^j_{r_j \alpha_i} \end{bmatrix} d\alpha_i . \qquad (40)$$

[57] Brown and Sappington (2021) provides a detailed proof of this Conclusion.

Cramer's Rule implies that under the maintained conditions:

$$\frac{dr_i}{d\alpha_i} = \frac{\begin{vmatrix} -L^i_{r_i\alpha_i} & L^i_{r_ir_j} \\ 0 & L^j_{r_jr_j} \end{vmatrix}}{\begin{vmatrix} L^i_{r_ir_i} & L^i_{r_ir_j} \\ L^j_{r_jr_i} & L^j_{r_jr_j} \end{vmatrix}} \stackrel{s}{=} -L^i_{r_i\alpha_i} L^j_{r_jr_j} \stackrel{s}{=} L^i_{r_i\alpha_i}. \tag{41}$$

Tedious calculations using (38) and Lemmas 1 – 3 imply that under the maintained conditions:

$$\begin{aligned} L^i_{r_i\alpha_i} = & -\frac{1}{9}\int_{\underline{\varepsilon}}^{\varepsilon_0} \Big\{ 2\,[\,(1+\alpha)\,b - (1-\alpha)\,d\,]\,\big[\,b^w a^I + \varepsilon - c\,\big] \\ & \qquad + 2\,b^w Q^r_i\,[\,1-\alpha\,]\,[\,(2+3\alpha)\,b - (1+\alpha)\,d\,] \\ & \qquad + 2\,b^w Q^r_j\,[\,1-\alpha\,]\,[\,(1+\alpha)\,b - (1-\alpha)\,d\,] \Big\}\, dH(\varepsilon) \\ & - \int_{\varepsilon_{12}}^{\overline{\varepsilon}} b^w\,[\,b-d\,]\,q^{2c}_i(\varepsilon)\,dH(\varepsilon) \;<\; 0\,. \end{aligned} \tag{42}$$

(9) implies that to ensure the generators produce positive output under VS in the symmetric setting (for all nonnegative values of $Q^r_1(\cdot)$ and $Q^r_2(\cdot)$), it must be the case that $b^w a^I + \varepsilon > c$. Therefore, the first of the three terms in the $\int_{\underline{\varepsilon}}^{\varepsilon_0}(\cdot)$ integral in (42) is strictly positive. The other two terms in the integral are nonnegative. Furthermore, $q^{2c}_i(\varepsilon) \geq 0$. Therefore, the inequality in (42) holds. Consequently, (41) implies that $\frac{\partial r_i}{\partial \alpha_i}\Big|_{K_i, K_j \text{ constant}} < 0$.

The proof that $\frac{\partial(\,r_i + r_j\,)}{\partial \alpha_i}\Big|_{K_i, K_j \text{ constant}} < 0$ is similar, and so is omitted. ■

**Proof of Lemma 5**. Differentiating (12), using (13), provides:

$$\begin{aligned} \frac{\partial E\left\{\pi^R_i\right\}}{\partial K_i} = & \,[\,r_i - E\{w(\varepsilon)\} - c^r_i\,]\left[-b_i\frac{dr_i}{dK_i} + d_i\frac{dr_j}{dK_i}\right] + Q^r_i\left[\frac{dr_i}{dK_i} - \frac{\partial E\{w(\varepsilon)\}}{\partial K_i}\right] \\ & + Q^r_i\left\{\big[\,w^{1c}(\varepsilon_0) - w^{0c}(\varepsilon_0)\,\big]\,h(\varepsilon_0)\frac{d\varepsilon_0}{dK_i} + \big[\,w^{2c}(\varepsilon_{12}) - w^{1c}(\varepsilon_{12})\,\big]\,h(\varepsilon_{12})\frac{d\varepsilon_{12}}{dK_i}\right\}. \end{aligned} \tag{43}$$

(19) follows from (43) because the term in $\{\cdot\}$ in (43) is zero since $w^{1c}(\varepsilon_0) = w^{0c}(\varepsilon_0)$ and $w^{2c}(\varepsilon_{12}) = w^{1c}(\varepsilon_{12})$, as established in the proof of Lemma 4. ■

**Proof of Conclusion 8**. Differentiating (18) provides:

$$\frac{\partial E\{w(\varepsilon)\}}{\partial K_i} = \int_{\underline{\varepsilon}}^{\varepsilon_0} \frac{\partial w^{0c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon) + \int_{\varepsilon_0}^{\varepsilon_{12}} \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon) + \int_{\varepsilon_{12}}^{\overline{\varepsilon}} \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} \, dH(\varepsilon)$$

$$+ \left[ w^{0c}(\varepsilon_0) - w^{1c}(\varepsilon_0) \right] h(\varepsilon_0) \frac{d\varepsilon_0}{dK_i} + \left[ w^{1c}(\varepsilon_{12}) - w^{2c}(\varepsilon_{12}) \right] h(\varepsilon_{12}) \frac{d\varepsilon_{12}}{dK_i} . \quad (44)$$

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

$$\frac{\partial w^{0c}(\varepsilon)}{\partial K_i} = \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} = 0 \quad \text{and} \quad \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} = -\, b^w . \quad (45)$$

(31) and (33) imply that when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$:

$$w^{0c}(\varepsilon_0) = w^{1c}(\varepsilon_0) \;\; \text{and} \;\; w^{1c}(\varepsilon_{12}) = w^{2c}(\varepsilon_{12}) . \quad (46)$$

(44), (45), and (46) imply that when G$j$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$:

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

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

$$\frac{\partial w^{0c}(\varepsilon)}{\partial K_i} = 0, \quad \frac{\partial w^{1c}(\varepsilon)}{\partial K_i} = -\frac{1}{2} b^w, \quad \text{and} \quad \frac{\partial w^{2c}(\varepsilon)}{\partial K_i} = -\, b^w . \quad (47)$$

An analysis that parallels the analysis in (28) – (33) demonstrates that (46) holds when G$i$ is capacity constrained for $\varepsilon \in (\varepsilon_0, \varepsilon_{12})$. (44), (46), and (47) imply that in this case:

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

**Proof of Conclusion 9**.[58] Let $\widetilde{\pi}^{Ri}$ denote R$i$'s objective when it fully values its own retail profit and derives benefit $\alpha_i$ from each dollar of G$i$'s upstream profit. Formally:

$$\widetilde{\pi}^{Ri} \equiv E\{ \alpha_i [ ( w(\cdot) - c_i ) q_i - k_i K_i ] + [ r_i - c_i^r - w(\cdot) ] Q_i^r(r_i, r_j) \} . \quad (48)$$

Given $\alpha_i$ and $\alpha_j$, the equilibrium values of $r_i$ and $r_j$ ($i, j \in \{1, 2\}$ $(j \neq i)$) are determined by:

$$\widetilde{\pi}^{Ri}_{r_i} \equiv \frac{\partial \widetilde{\pi}^{Ri}}{\partial r_i} = 0 \quad \text{and} \quad \widetilde{\pi}^{Ri}_{r_j} \equiv \frac{\partial \widetilde{\pi}^{Rj}}{\partial r_j} = 0 . \quad (49)$$

Letting $\widetilde{\pi}^{Ri}_{xy}$ denote $\frac{\partial^2 \widetilde{\pi}^{Ri}}{\partial x \partial y}$, expression (49) implies:

$$\begin{bmatrix} \widetilde{\pi}^{Ri}_{r_i r_i} & \widetilde{\pi}^{Ri}_{r_i r_j} \\ \widetilde{\pi}^{Rj}_{r_j r_i} & \widetilde{\pi}^{Rj}_{r_j r_j} \end{bmatrix} \begin{bmatrix} dr_i \\ dr_j \end{bmatrix} = - \begin{bmatrix} \widetilde{\pi}^{Ri}_{r_i K_i} \\ \widetilde{\pi}^{Ri}_{r_j K_i} \end{bmatrix} dK_i . \quad (50)$$

[58] Brown and Sappington (2021) provides a detailed proof of this Conclusion.

Equation (50) and Cramer's Rule imply:

$$\frac{dr_i}{dK_i} = \frac{1}{Z^R}\left[\widetilde{\pi}^{Ri}_{r_i r_j}\,\widetilde{\pi}^{Ri}_{r_j K_i} - \widetilde{\pi}^{Rj}_{r_j r_j}\,\widetilde{\pi}^{Ri}_{r_i K_i}\right] \text{ and } \frac{dr_j}{dK_i} = \frac{1}{Z^R}\left[\widetilde{\pi}^{Rj}_{r_j r_i}\,\widetilde{\pi}^{Ri}_{r_i K_i} - \widetilde{\pi}^{Ri}_{r_i r_i}\,\widetilde{\pi}^{Rj}_{r_j K_i}\right]$$

$$\text{where } Z^R \equiv \widetilde{\pi}^{Ri}_{r_i r_i}\,\widetilde{\pi}^{Rj}_{r_j r_j} - \widetilde{\pi}^{Ri}_{r_i r_j}\,\widetilde{\pi}^{Rj}_{r_j r_i}\,. \tag{51}$$

(51) implies:

$$\frac{dr_i}{dK_i} \stackrel{s}{=} \widetilde{\pi}^{Ri}_{r_i r_j}\,\widetilde{\pi}^{Ri}_{r_j K_i} - \widetilde{\pi}^{Rj}_{r_j r_j}\,\widetilde{\pi}^{Ri}_{r_i K_i}\,. \tag{52}$$

Tedious calculations using (48) and Lemmas 1 – 3 reveal that in the symmetric setting where $b_1 = b_2 \equiv b$, $d_1 = d_2 \equiv d$, and $\alpha_1 = \alpha_2 \equiv \alpha$:

$$\widetilde{\pi}^{Ri}_{r_i K_i} = -\,b^w\,[\,\alpha + b\,]\,[\,1 - H(\varepsilon_{12})\,] < 0 \quad \text{and} \quad \widetilde{\pi}^{Rj}_{r_j K_i} = -\,b\,b^w\,[\,1 - H(\varepsilon_{12})\,] < 0$$

$$\Rightarrow \quad \left|\,\widetilde{\pi}^{Ri}_{r_i K_i}\right| \geq \left|\,\widetilde{\pi}^{Rj}_{r_j K_i}\right|. \tag{53}$$

(52) and (53) imply:

$$\frac{dr_i}{dK_i} \stackrel{s}{=} \widetilde{\pi}^{Rj}_{r_j r_j}\left|\,\widetilde{\pi}^{Ri}_{r_i K_i}\right| - \widetilde{\pi}^{Ri}_{r_i r_j}\left|\,\widetilde{\pi}^{Rj}_{r_j K_i}\right| \stackrel{s}{=} \frac{\left|\,\widetilde{\pi}^{Ri}_{r_i K_i}\right|}{\left|\,\widetilde{\pi}^{Rj}_{r_j K_i}\right|}\,\widetilde{\pi}^{Rj}_{r_j r_j} - \widetilde{\pi}^{Ri}_{r_i r_j}\,. \tag{54}$$

Tedious calculations reveal that under VI in the symmetric setting:

$$\widetilde{\pi}^{Rj}_{r_j r_j} = -\,2\,b - 2\,b^w\,b\,[\,b - d\,]\,[\,1 - H(\varepsilon_{12})\,] < 0 \text{ and}$$

$$\widetilde{\pi}^{Ri}_{r_i r_j} = d - b^w\,[\,b - d\,]^2\,[\,1 - H(\varepsilon_{12})\,]\,. \tag{55}$$

If $\widetilde{\pi}^{Ri}_{r_i r_j} \geq 0$, then (54) and (55) imply $\frac{dr_i}{dK_i} < 0$. If $\widetilde{\pi}^{Ri}_{r_i r_j} < 0$, then (54) and (55) imply $\frac{dr_i}{dK_i} < 0$ if $\left|\,\widetilde{\pi}^{Rj}_{r_j r_j}\right| > \left|\,\widetilde{\pi}^{Ri}_{r_i r_j}\right| \Leftrightarrow b^w\,[\,b - d\,]\,[\,b + d\,]\,[\,1 - H(\varepsilon_{12})\,] + 2\,b + d > 0$. This inequality holds, so $\frac{dr_i}{dK_i} < 0$.

The corresponding proofs that: (i) $\frac{dr_i}{dK_i} < 0$ under VS in the symmetric setting; and (ii) $\frac{dr_j}{dK_i} < 0$ under the identified conditions are similar, and so are omitted. ■

## 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 $\varepsilon$, 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:

$$J_1(r_1, r_2, K_1, K_2) \;=\; 0 \quad \text{and} \quad J_2(r_1, r_2, K_1, K_2) \;=\; 0\,. \tag{56}$$

Third, capacity investments are determined. Generator G$i$ 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. G$i$'s problem can be written as:

$$\max_{K_i} \;\; E\left\{\, \pi_i^G(r_1^*(K_i, K_j), r_2^*(K_i, K_j), K_i, K_j)\,\right\} \tag{57}$$

where $r_1^*(K_i, K_j)$ and $r_2^*(K_i, K_j)$ are characterized by (56).

G$i$'s choice of $K_i$ can be written as a stochastic mathematical program with equilibrium constraints (SMPEC) that treats the retail price conditions in (56) as constraints:[59]

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

subject to:

$$\begin{aligned} \mu_{i1}: &\qquad J_1(r_1, r_2, K_1, K_2) \;=\; 0\,, \\ \mu_{i2}: &\qquad J_2(r_1, r_2, K_1, K_2) \;=\; 0\,, \end{aligned} \tag{58}$$

where $\mu_{i1}$ and $\mu_{i2}$ denote the relevant Lagrange multipliers.

Because G$i$'s choice of $K_i$ is represented by the SMPEC in (58), 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 (58). Doing so (which is facilitated by assuming $\varepsilon$ 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 (58) for G1 and for G2. Formally, the

[59] 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 G$i$'s SMPEC in (58) is:

$$\mathcal{L}^i \;=\; E\left\{\, \pi_i^G(r_1, r_2, K_i, K_j)\,\right\} + \mu_{i1}\, J_1(r_1, r_2, K_1, K_2) + \mu_{i2}\, J_2(r_1, r_2, K_1, K_2). \tag{59}$$

The corresponding solution to G$i$'s SMPEC is characterized by:

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

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

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

$$\mu_{i1}: \quad J_1(r_1, r_2) \;=\; 0\,;$$

$$\mu_{i2}: \quad J_2(r_1, r_2) \;=\; 0\,. \tag{60}$$

The solution to the SEPEC is characterized by solving the equations in (60) for G1 and G2 simultaneously.[60] 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 (60) 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 $\varepsilon$ 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 $\varepsilon$ 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 G$j$C is not violated ($i, j \in \{1, 2\}$, $j \neq i$), then the solution is as derived under Assumption G$j$C. 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.[61]

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

[61] Brown and Sappington (2020b) 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 (2021) present additional outcomes that arise in different settings. In all of the settings considered here and in Brown and Sappington (2021): (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 $\bar{\varepsilon} = 840$**.**

| | 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 $c_1^r = 1.2$**.**

## Appendix D. Illustrating the Private Gains from Vertical Integration

Figures A1 – A4 below parallel Figure 2 in the text, illustrating how the percentage increase in R1 and G1's combined expected profit ($\%\Delta\Pi_1$) varies with the percentage increase in the model parameter specified on the horizontal axis.[62]

Figure A1 indicates that as the own-price sensitivity of retail demand ($b_1 = b_2 \equiv b$) increases, R1-G1's gains from unilateral integration and ubiquitous integration both increase, for the reasons explained in the text.

Figure A2 reports that as retail demand curves shift outward (i.e., as $a_1 = a_2 \equiv a$ increases), R1-G1's gain from unilateral integration increases whereas their gain from ubiquitous integration declines. As $a$ increases, industry surplus increases. A unilateral commitment to compete aggressively enables R1-G1 to capture some of the increased surplus, so R1-G1's gain from unilateral vertical integration increases as $a$ increases. In contrast, under ubiquitous vertical integration, the increased surplus induces increased competitive aggression that more than dissipates the increment in surplus. Consequently, R1-G1's gain from unilateral vertical integration declines as $a$ increases.

Figure A3 demonstrates that as the price sensitivity of industrial demand ($b^I$) increases, R1-G1's gains from unilateral and ubiquitous integration both decline, for the reasons explained in the text.

Figure A4 indicates that as the industrial demand curve shifts outward (i.e., as $a^I$ increases), R1-G1's gain from unilateral integration declines (modestly) whereas their gain from ubiquitous integration increases. An increase in industrial demand increases the wholesale price, thereby reducing retail profit. The reduction in retail profit is more pronounced under PVI and (especially more pronounced under) VI than under VS. This is the case because less of the cost increase is passed on to retail customers in the form of higher retail prices under the relatively intense competition that prevails under PVI and (especially under) VI.

Wholesale profit increases as industrial demand increases under PVI, VI, and VS. Under all three industrial structures, the increase in wholesale profit exceeds the reduction in retail profit as $a^I$ increases. However, aggregate profit increases more slowly under PVI than under VS as $a^I$ increases. In contrast, aggregate profit increases more rapidly under VI than under VS as $a^I$ increases.

Changes in $c_1^r = c_2^r \equiv c^r$, $c_1 = c_2 \equiv c$, and $k$ have relatively little impact on R1-G1's gains from vertical integration. As the retail cost component $c^r$ increases, retail prices

[62] Recall that all parameters other than the one identified on the horizontal axis assume their values in the baseline setting. For parameter $\zeta \in \{b,\ a,\ b^I,\ a^I\}$, the variable on the horizontal axis is $\%\Delta\zeta \equiv 100\left[\frac{\zeta - \zeta^B}{\zeta^B}\right]$, where $\zeta^B$ denotes the value of $\zeta$ in the baseline setting.

increase whereas retail outputs and profits decline under VS, PVI, and VI. The rates at which R1-G1's aggregate profit declines as $c^r$ increases are very similar under all three industry structures, so R1-G1's gains from vertical integration vary little as $c^r$ changes.[63]

As $c$ increases, R1-G1's gain from expanded capacity investment declines because the increased output engendered by expanded capacity is more costly to produce. Therefore, R1-G1's gain from committing to compete more aggressively (in part by expanding capital investment) declines. However, the increase in $c$ renders the rival generator more accommodating by reducing its opportunity cost of reducing its capacity investment. These two countervailing effects ensure that R1-G1's gains from vertical integration change little as $c$ changes.[64]

Vertical integration promotes expanded capacity investment. As $k$ increases, the private gain from committing to expand capacity declines. However, the increase in $k$ also renders the rival generator more accommodating. These two countervailing effects ensure that the private gains from vertical integration change little as $k$ changes.[65]

[63] As $c^r$ varies from 40% below to 40% above its value in the baseline setting, the percentage increase in R1-G1's profit under PVI (VI) relative to their profit under VS varies from 6.57 to 6.53 (0.48 to 0.54).

[64] As $c$ varies from 40% below to 40% above its value in the baseline setting, the percentage increase in R1-G1's profit under PVI (VI) relative to their profit under VS varies from 6.61 to 6.52 (0.45 to 0.53).

[65] As $k$ varies from 40% below to 40% above its value in the baseline setting, the percentage increase in R1-G1's profit under PVI (VI) relative to their profit under VS varies from 6.54 to 6.53 (0.24 to 0.68).

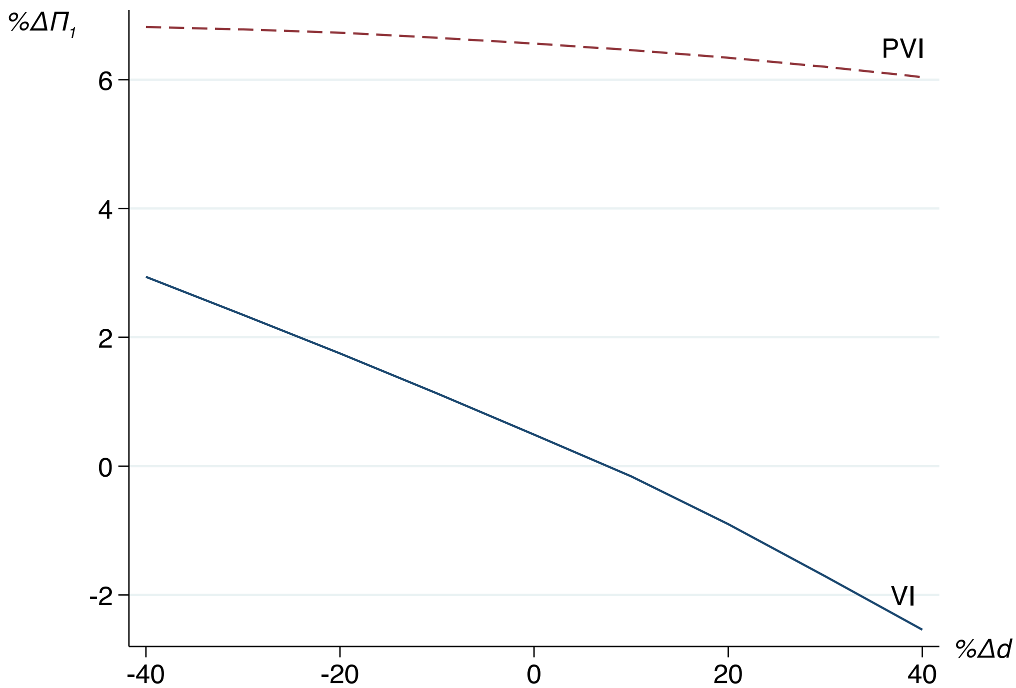


**Figure 2. The Effect of *d* on the Profitability of Integration.**

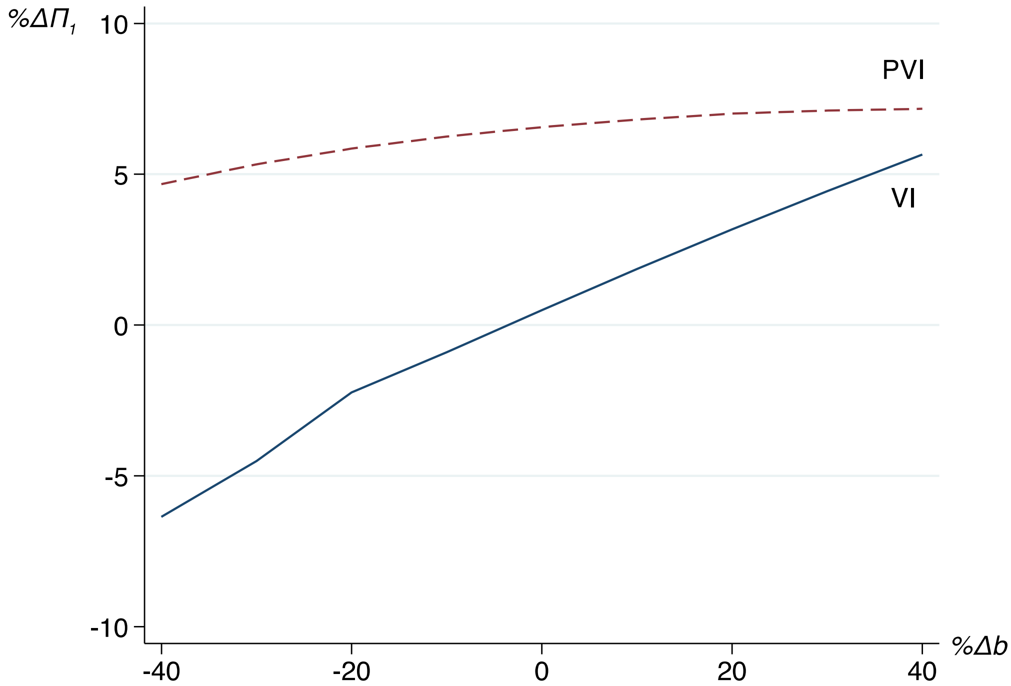


**Figure A1. The Effect of *b* on the Profitability of Integration.**

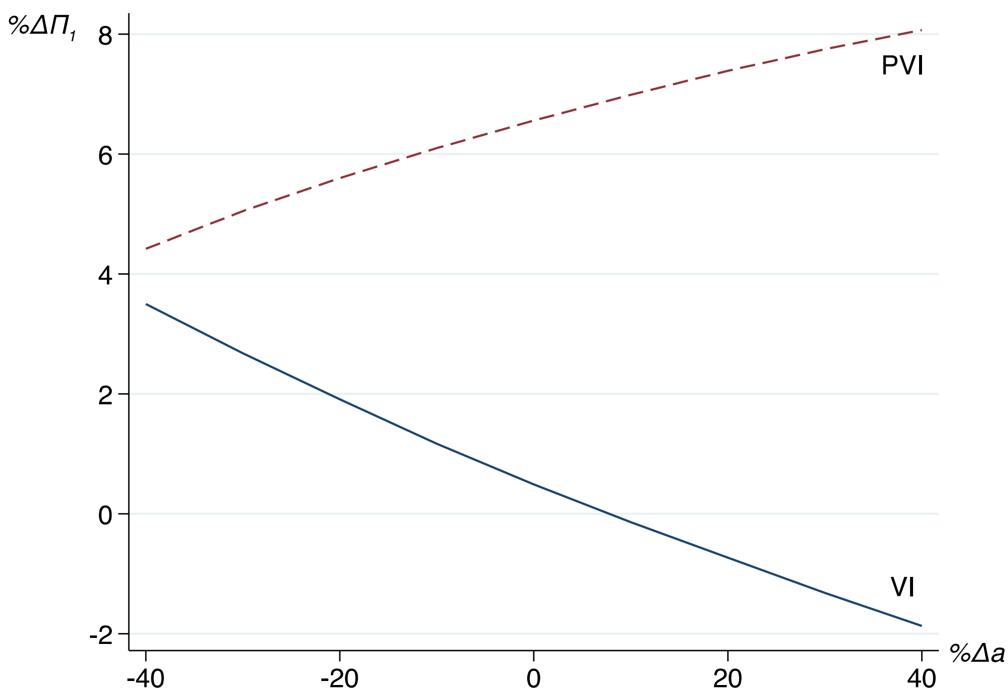


**Figure A2. The Effect of *a* on the Profitability of Integration.**

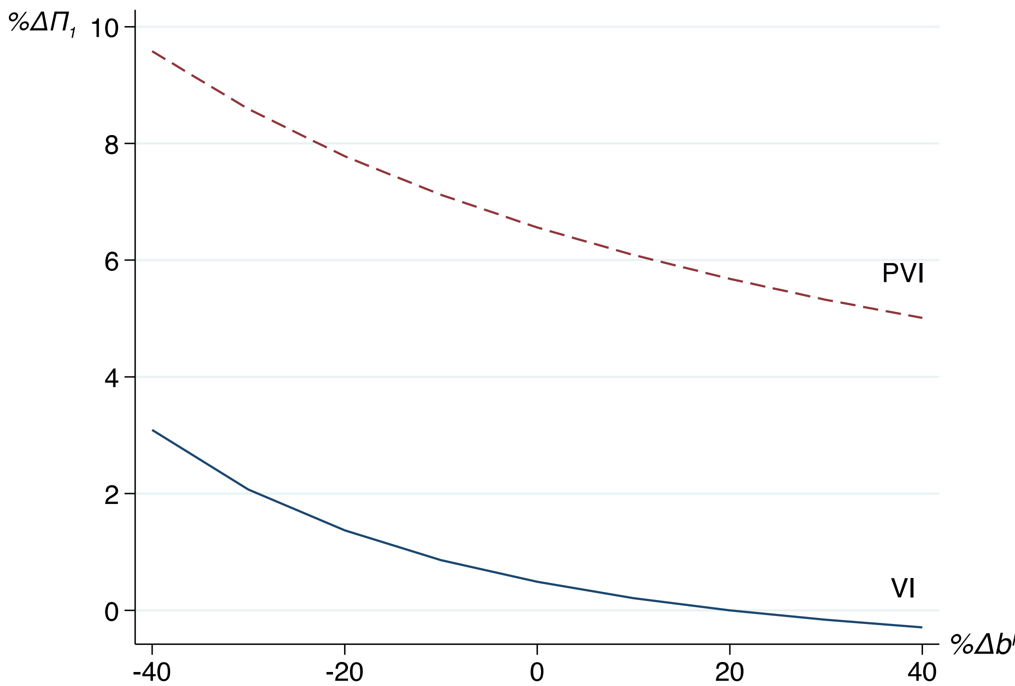


**Figure A3. The Effect of $b^I$ on the Profitability of Integration.**

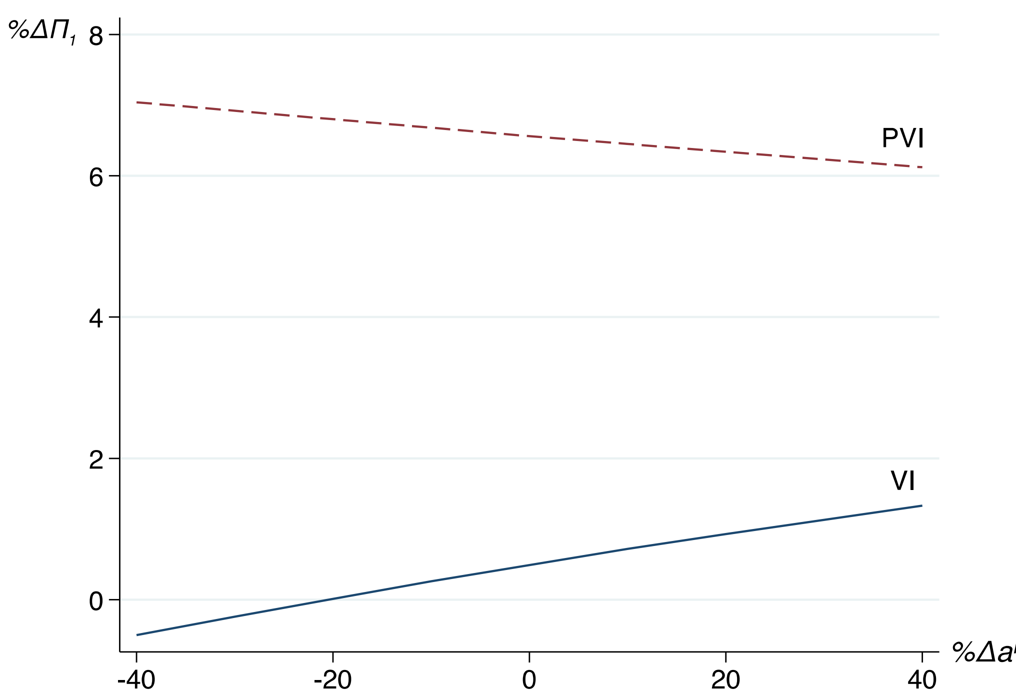


**Figure A4. The Effect of $a^I$ on the Profitability of Integration.**

## 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.

Borenstein, S., Bushnell, J., & Wolak, F. (2002). Measuring market inefficiencies in California's restructured wholesale electricity market. *American Economic Review*, 95, 1376-1405.

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

Brown, D., & Eckert, A. (2017b). The effect of default rates on retail competition and pricing decisions of competitive retailers: The case of Alberta. *Energy Policy*, 118, 298-311.

Brown, D., & Sappington, D. (2021). 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.

CMA. (2016). Energy market investigation: Final report. Competition & Markets Authority. Available at: https://assets.publishing.service.gov.uk/media/5773de34e5274a0da3000113/final-report-energy-market-investigation.pdf.

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

Deller, D. Giulietti, M., Loomes, G., Waddams Price, C., Moniche, A., & Jeon, J. (2021). Switching energy suppliers: It's not all about the money. *The Energy Journal*, 42, 1-16.

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

EIA. (2018). Electricity residential retail choice participation has declined since 2014 peak. Energy Information Administration. Available at: https://www.eia.gov/todayinenergy/detail.php?id=37452.

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.

EIA. (2021). Annual electric power industry report. *EIA Form-861 Detailed Data Files.* Energy Information Administration. Available at: https://www.eia.gov/electricity/data/eia861.

EMI. (2020). Vertical integration trends. *Electricity Market Information.* Electricity Authority Te Mana Hiko. Available at: https://www.emi.ea.govt.nz/Wholesale/Reports/BLKL4U.

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.

Fabra, N., von der Fehr, N.H., & Harbord, D. (2006). Designing electricity auctions. *RAND Journal of Economics*, 37, 23-46.

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.

Frontier Economics. (2017). Effects of vertical integration on capacity bidding behaviour. Available at: https://www.accc.gov.au/system/files/AGL%20%28Frontier%20report%29.pdf.

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.

Hortaçsu, 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.

Houser, T., Larsen, J., & Marsters, P. (2017). The real electricity reliability crisis. *Rhodium Group Note*. Available at: https://rhg.com/research/the-real-electricity-reliability-crisis-doe-nopr.

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.

Hughes, J., & Kao, J. (1997). Strategic forward contracting and observability. *International Journal of Industrial Organization*, 16, 121-133.

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.

Joskow, P., & Kahn, E. (2002). A quantitative analysis of pricing behavior in California's wholesale electricity market during Summer 2000: The final word. *Energy Journal*, 23, 1-35.

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.

Klemperer, P., & Meyer, M. (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica*, 57, 1243-1277.

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 Noridic 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 eystem. *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.

Puller, S. (2007). Pricing and firm conduct in California's deregulated electricity market. *Review of Economics and Statistics*, 89, 75-87.

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

Spengler, J. (1950). Vertical integration and antitrust policy. *Journal of Political Economy*, 58, 347-352.

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

von der Fehr, N.H., & Harbord, D. (1993). Spot market competition in the UK electricity industry. *Economic Journal*, 103, 531-546.

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

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.