The Competitive Effects of Linking Electricity Markets Across Space and Time*

Thomas P. Tangerås† and Frank A. Wolak‡

July 26, 2018

Abstract

We show that a common regulatory mandate in electricity markets that employ locational marginal pricing (LMP) requiring all electricity retailers to purchase their wholesale electricity at the same quantity-weighted average of these prices can improve the performance of imperfectly competitive wholesale electricity markets. Linking locational markets strengthens the incentive for vertically integrated firms to participate in the retail market, which increases competition in the short-term wholesale market. Simulations based on a stylized model of an electricity supply industry find economically significant price reductions are likely for actual markets that employ this regulatory mandate. Our results imply that a policy designed to address equity considerations associated with implementing a locational marginal pricing market design can also enhance wholesale market efficiency.

Key Words: Electricity markets, equity, market design, market performance, market power, vertical integration.

JEL: C72, D43, G10, G13, L13

*We thank Jan Abrell, Eirik Gaard Kristiansen, Thomas Olivier-Léautier, Robert Ritz, Bert Willems, participants at "The Performance of Electricity Markets" conference (2014) in Waxholm, Norio X (2016) in Reykjavik, the 8th Swedish Workshop on Competition and Public Procurement Research (2016) in Stockholm, MEC (2017) in Mannheim and seminar audiences at Aalto University, Stanford University and University of Cambridge for their comments. This research was conducted within the "Economics of Electricity Markets" research program at IFN. Financial support from Jan Wallanders och Tom Hedelius Stiftelse (Tangerås), the Swedish Competition Authority (Tangerås) and the Swedish Energy Agency (Tangerås and Wolak) is gratefully acknowledged.

†Research Institute of Industrial Economics (IFN) P.O. Box 55665, 10215 Stockholm, Sweden, e-mail: thomas.tangeras@ifn.se. Associated researcher Energy Policy Research Group (EPRG), University of Cambridge. Visiting scholar Stanford University.

‡Program on Energy and Sustainable Development and Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072, e-mail: wolak@stanford.edu.

1 Introduction

Many European countries and regions of the United States have ambitious renewable energy goals. The intermittent nature of wind and solar energy, the two major renewable resources deployed to achieve these goals, creates significant operational challenges for wholesale electricity market operators. Locational marginal pricing (LMP) or nodal pricing, is generally acknowledged to be the wholesale market design that can most efficiently manage these operational challenges.[1]

Locational marginal prices and generation unit-level energy sales are computed by minimizing the as-offered cost of serving demand at all locations in the transmission network subject to all relevant operating constraints on generation units and the transmission network. This process can give rise to thousands of different locational marginal prices within the geographic footprint of the wholesale market during each pricing period, which provides the market operator with the maximum the flexibility to manage the operational challenges of increasing amounts of intermittent renewable energy.

The locational marginal price (LMP) at location or node n in the transmission network is equal to the increase in the minimized as-offered cost of withdrawing an additional megawatt-hour (MWh) at node n. Therefore, if all suppliers submit their generation unit's marginal cost as its offer price to the wholesale market operator, the resulting LMPs are the economically efficient price signal for each location in the transmission network.[2]

Despite these market efficiency benefits of LMP pricing, it has been extremely difficult for regulators in the United States to charge electricity consumers a price that reflects the LMP at their location in the transmission network.[3] The potential of an LMP market design to set different prices for consumers at different locations in the transmission network has been a major barrier to the adoption of this market design in European countries, despite the likely substantial market efficiency benefits from doing so.

Arguments against requiring that retailers pay a wholesale price that reflects the LMP at the customer's location in the transmission network are typically based on the view that it is

[1] PJM Interconnection, California Independent System Operator (ISO), ISO-New England, New York ISO, Midcontinent ISO, and Electricity Reliability Council of Texas (ERCOT) all use locational marginal pricing for their day-ahead and real-time markets. Singapore and New Zealand also employ locational marginal pricing market design for their real-time markets.

[2] Specifically, if these LMPs are set at each node in the transmission network, then all suppliers will find it unilaterally profit-maximizing to produce at a level of output that minimizes the total variable cost of serving demand at all locations in the transmission network.

[3] Wolak (2011) finds that the transition from a zonal to a nodal market design in California was associated with annual savings in the variable cost of serving demand of over 100 million dollars annually.

unfair to charge customers in major load centers higher prices than customers withdrawing energy from other nodes in the transmission network, even though LMPs in load centers are typically higher than prices at other nodes. In most all LMP markets, these equity concerns have been addressed by a regulatory mandate that requires all loads to pay a wholesale price equal to the quantity-weighted average of the locational marginal prices at all nodes in a retailer's service territory or the entire geographic area of the wholesale market. For example, all retail customers in the service territory of each of the three investor-owned utilities in California pay a wholesale price equal to the quantity-weighted average of LMPs at all load withdrawal nodes in that utility's service territory. All customers of each utility face the same Load Aggregation Point (LAP) wholesale price regardless of where they are located in their utility's service territory. Singapore operates a nodal pricing market and all loads purchase their wholesale electricity at the Uniform Singapore Electricity Price (USEP) which is equal to the quantity-weighted average of the LMPs at all load withdrawal points in Singapore.[4]

This paper demonstrates that the regulatory mandate that all loads must to purchase their wholesale electricity at a quantity-weighted average of nodal prices can improve market performance in imperfectly competitive wholesale electricity markets. Recent experience in a number of Western European countries illustrates the policy relevance of our findings. For example, in Germany there is a surplus of renewable energy production in the North (because this is where the conditions for solar and wind electricity production are the most favorable) despite the fact that a large share of demand is located in the South. Germany currently comprises a single price area with Austria and Luxembourg. Introducing more spatial granularity in wholesale electricity pricing would provide the market operator with more tools to manage the operational challenges associated an increasing amount of intermittent renewable generation resources.[5] Other Western European markets face similar operational challenges with integrating significant intermittent renewable resources that could be more efficiently addressed with greater spatial granularity in wholesale electricity pricing. Our results imply that increasing the number of price areas for producers while maintaining a uniform wholesale energy purchase price for retailers based on the weighted average of those area prices, would improve congestion management, maintain price equality for consumers and improve

[4]This regulatory mandate does not imply all customers must pay the same retail electricity price each hour, only that all electricity retailers must purchase their wholesale energy at this quantity-weighted average wholesale price each hour.

[5]See, for instance, Egerer et al. (2016) and references therein. The Agency for the Cooperation of Energy Regulators recently proposed to break out Austria as a separate price area (ACER, 2015), but the involved parties turned it down.

short-term market efficiency.

Our basic insight is that linking M local markets through a single retail contract in which consumers pay the quantity-weighted average of the locational short-term prices over the M markets for their wholesale energy and the L producers in each local market face the local short-term price, increases the competitiveness of all short-term markets beyond the level that would exist if there were M independent markets where retail customers in each local market purchased their wholesale energy at the local short-term price.

The competitive benefits of linking retail markets across space can be substantial. In a symmetric Cournot model, linking together three local duopoly markets ($M = 3$, $L = 2$) through integrated retail markets causes the short-term price-cost margin to drop by 44% in each local market. We also construct a model of a more realistic industry structure with market asymmetries that depend on transmission network capacity constraints. The competitive effects are weaker but still substantial: Prices drop between 5 and 20 percent. Notably, price effects are purely the result of integrating retail markets and do not rely on any changes in market structure: Our results are derived for a fixed number of producers and exogenous transmission constraints.

The starting point of our analysis is the observation that virtually all large electricity retailers in formal wholesale markets are vertically integrated between generation and retailing. The equilibrium degree of vertical integration that emerges from our model balances two opposing forces. First, increased vertical integration and fixed-price retail sales commits the firm to more aggressive behavior in the short-term market that triggers a strategic response which causes competitors to reduce their output. The short-term market profit increases as a consequence of this output contraction. Second, increased fixed-price retail sales reduces the retail price and thereby retail profit.

Some of the retail price effect spills over to other local markets when retail markets are linked. For historical reasons, vertically-integrated firms typically have a significant presence only in a few local markets, usually their historical service territory during the former vertically-integrated monopoly regime. If so, they will not fully internalize the negative retail price effects of their retail sales across all local markets. Therefore, in equilibrium vertically-integrated firms sell a larger share of their output in the retail market when retail markets are linked compared to the case when these markets are independent. This increased vertical integration in turn improves short-term market performance.

We also explore the implications of linking markets over time using a fixed-price long-term contract for electricity. For the case of a long-term contract, the M markets are different versions of the same market over time, and the L suppliers are the generation unit owners in those M markets. In this case, it is reasonable to assume that all suppliers participate in all M markets. The weighted average of the short-term market price-cost margins is found to be the same under long-term as short-term retail contracts. However, if firms do not produce in all periods, because of scheduled maintenance, long-term contracts improve short-term market performance compared to to short-term contracts.[6]

Vertical integration between generation and retailing in a single market is formally equivalent to a producer selling in the forward market under the assumptions of our model. Allaz and Vila (1993) are the first to demonstrate the pro-competitive effects of forward contracting in a model with a single short-term market.[7] Our main contribution to this literature is to show how linking retail markets across geographical locations through a uniform wholesale purchase price for all consumers improves performance in all local wholesale markets.[8] Mahenc and Salanié (2004) find forward contracting to reduce market performance if firms compete in prices in the spot market instead of in quantities as assumed here. Holmberg (2011) establishes conditions under which forward contracting improves market performance when firms compete in supply functions. These results indicate that the competitive effects of linking retail markets are sensitive to the mode of competition in the short-term market. Cournot competition has been used by empirical researchers to model to strategic interaction among suppliers in many wholesale markets for electricity, including California, New England and PJM (Bushnell et al., 2008), the Midwest market (Mercadal, 2016), the German market (Willems et al., 2009) and the Nordic market (Lundin and Tangerås, 2017).

The rest of the paper is organized as follows. Section 2 introduces the baseline model and establishes the pro-competitive effects of vertical integration on short-term market performance. Section 3 contains our main result showing that linking retail markets has an additional positive effect on market performance unless all firms are active in all local markets. Section 4 shows the fundamental difference between linking retail markets versus forward contract markets on the performance of short-term markets. Section 5 considers the case of both linked retail and forward markets. Section 6 investigates the implications of long-term

[6] The welfare comparison between a retail contract that is linked both across space and time and one that is independent in both dimensions would be ambiguous. Linking contracts across space improves market performance because prices go down, but linking markets across time reduces welfare because of increased price variability. The choice between a combined contract and a fully independent contract would revolve around the degree of market asymmetry. We have not explored this combined case in any detail because the welfare implications of linking contracts are so clear in each of the two dimensions.

[7] A difference between vertical integration and forward contracting is that firms are likely to be better informed about competitors' retail sales than forward contracting positions.

[8] Green and Le Coq (2010) show that increasing the contract length (linking electricity markets across time) has ambiguous effects on the ability to sustain collusion. We consider unilateral market power and thus leave the question of how different market designs affect collusion for future research.

retail contracts. In Section 7, we incorporate transmission network constraints and perform numerical simulations of the model. Section 8 concludes with a discussion of the implications of our results for design of wholesale electricity markets.

2 The baseline model: Spatially independent markets

We consider a model of $M \geq 1$ local electricity markets. There are $L_m \geq 1$ firms that possess market power in market $m \in \mathcal{M} = \{1, ..., M\}$. These firms are vertically integrated between production and retail. We also assume that each local market has a number of independent retailers and producers that behave competitively. All local markets are physically disconnected from one another in the sense that there is no flow of electricity between them. This assumption hugely simplifies the analysis, but also reveals a key insight: There can be market performance gains from linking markets even if there is no actual trade of goods between them. In Section 7, we allow the degree to which local markets are disconnected to be affected by stochastic transmission network constraints.[9]

We analyze a two-stage game in which firms compete in the retail market in the first-stage and in the short-term market in the second stage. This setup captures in a simple manner the fact that consumers usually are on fixed-price long-term retail contracts that span multiple production periods, whereas production and consumption are short-term decisions. We expand the model to consider long-term retail contracts in Section 6. In this section we show that our result holds for even single-period retail contracts.

Electricity is a homogeneous good with highly price inelastic demand. Hence, we assume total demand to be constant and equal to D_m in local market m. All electricity is sold at a uniform price r_m in the retail market m and a uniform price p_m in short-term market m.

In the first stage, each vertically integrated firm $l \in \{1, ..., L_m\}$ inelastically supplies $k_{lm} \geq 0$ megawatt hours (MWh) of electricity to retail market m, taking the retail supply $K_{-lm} = \sum_{i \neq l} k_{im}$ of the other $L_m - 1$ vertically integrated firms as given. Independent retailers cover the residual demand $D_m - K_m$, where $K_m = k_{lm} + K_{-lm}$ is the total amount of retail energy sold by vertically integrated suppliers in m. These firms have no production of their own and purchase their electricity in the short-term market at price p_m. We assume that this short-term price is the (constant) marginal cost of the independent retailers selling electricity. We also assume the retail market is perfectly competitive, which implies that the

[9]See Holmberg and Philpott (2018) and references therein for illustrations of the complications caused by analyzing supplier behavior in imperfectly competitive electricity markets with endogenous network constraints.

market-clearing price in the retail market is $r_m = p_m$.

In the second stage, each vertically integrated firm l in market m observes K_{-lm} and decides how much additional electricity, $\tilde{q}_{lm} - k_{lm}$, to inelastically supply to the short-term market at constant marginal cost c_{lm}, taking the production $\tilde{Q}_{-lm} = \sum_{i \neq l} \tilde{q}_{im}$ of the other firms with the ability to exercise unilateral market power as given. The demand facing independent producers in the short-term market equals the demand $D_m - K_m$ by the independent retailers minus the short-term market supply $\tilde{Q}_m - K_m$ of the L_m vertically integrated firms, where $\tilde{Q}_m = \tilde{q}_{lm} + \tilde{Q}_{-lm}$. The competitive fringe supplies electricity at a linear marginal cost $b_m Q_{fringe}$, so the market-clearing short-term price equals $p_m = b_m(D_m - \tilde{Q}_m)$. If we let $a_m = b_m D_m$, then the inverse demand curve facing the L_m firms with the ability to exercise unilateral market power in short-term market m becomes $P_m(\tilde{Q}_m) = a_m - b_m\tilde{Q}_m$ as a function of their total production $\tilde{Q}_m$. We later refer to the ratio $K_m/\tilde{Q}_m$ as the *degree of vertical integration* in local market m.

By construction, our baseline model represents a generalization of the two-stage game of forward contracting in Allaz and Vila (1993) to the case of $M \geq 1$ markets with $L_m \geq 1$ producers in each market. Instead of forward contracting, we consider vertical integration between production and retail markets. Assuming quantity competition in the retail market subject to a zero retailing profit pricing condition is consistent with the basic model by Allaz and Vila (1993) and greatly simplifies our analysis.[10] We solve for the unique subgame-perfect equilibrium to this two-stage game by backward induction.

Equilibrium in the short-term market The earnings of producer l in market m consists of its retail revenue plus the revenue from sales in the short-term wholesale market:

$$r_m k_{lm} + p_m(\tilde{q}_{lm} - k_{lm})$$

Because the vertically integrated firms have the ability to exercise unilateral market power, they take into account the effect of their behavior on the short-term price when they decide how much to produce in the second period. Hence, the profit in the second stage can be written as

$$(r_m - P_m(\tilde{Q}_m))k_{lm} + (P_m(\tilde{Q}_m) - c_{lm})\tilde{q}_{lm} \tag{1}$$

as a function of producer l's output, $\tilde{q}_{lm}$. The first term is the retail profit because $P_m(\tilde{Q}_m)$ represents the opportunity cost of the electricity supplied in the retail market. Hence, the

[10] Appendix A.5 shows that the fundamental results of the paper continue to hold if retail competition is in prices instead.

retail position k_{lm} is essentially a forward contract that clears against the short-term price. The second term in the above equation is the profit from sales in the short-term market.

The first-order condition for profit maximization is (quantities without the tilde denote equilibrium values throughout):

$$-P'_m(Q_m)k_{lm} + P_m(Q_m) - c_{lm} + P'_m(Q_m)q_{lm} = 0. \tag{2}$$

The first term is due to vertical integration and represents the reduction in the opportunity cost of the electricity sold in the retail market. The remaining terms constitute the usual trade-off in a Cournot oligopoly between the benefit of a higher output against the cost of a lower price. Production is independent of the retail price r_m because the retail revenue $r_m k_{lm}$ is sunk at the second stage. Solving this linear equation system for the L_m firms in market m yields:

Lemma 1 *Assume that K_m is sufficiently small so that all L_m producers have positive production in equilibrium. The markup of the short-term price over the average marginal production cost $c_m = \frac{1}{L_m}\sum_{l=1}^{L_m} c_{lm}$ of the L_m vertically integrated firms in short-term market m is given by*

$$p_m(K_m) - c_m = P_m(Q(K_m)) - c_m = \frac{a_m - c_m}{L_m + 1} - \frac{b_m K_m}{L_m + 1}. \tag{3}$$

The corresponding equilibrium production of firm l is

$$q_{lm}(k_{lm}, K_{-lm}) = \frac{1}{L_m + 1}\frac{a_m - c_m}{b_m} + \frac{c_m - c_{lm}}{b_m} + \frac{L_m k_{lm}}{L_m + 1} - \frac{K_{-lm}}{L_m + 1} \tag{4}$$

and the sum of the production of all other firms in market m:

$$Q_{-lm}(k_{lm}, K_{-lm}) = \frac{L_m - 1}{L_m + 1}\frac{a_m - c_m}{b_m} - \frac{c_m - c_{lm}}{b_m} - \frac{L_m - 1}{L_m + 1}k_{lm} + \frac{2K_{-lm}}{L_m + 1}. \tag{5}$$

An increase in the volume of electricity sold in the retail market mitigates a firm's incentive to exercise unilateral market power in the short-term market, $\frac{\partial q_{lm}}{\partial k_{lm}} = \frac{L_m}{L_m+1} > 0$, but also triggers a strategic response by which all other firms reduce their production, $\frac{\partial Q_{-lm}}{\partial k_{lm}} = -\frac{L_m-1}{L_m+1} < 0$. The direct effect is stronger than the strategic effect, so the total effect of an increase in k_{lm} on the short-term price is negative: $p'_m(K_m) = -\frac{b_m}{L_m+1} < 0$.

Equilibrium in the retail market The market-clearing price in retail market m equals

$$r_m(K_m) = p_m(K_m) = P_m(Q_m(K_m)), \tag{6}$$

and the first-stage profit of firm l is

$$(r_m(K_m) - p_m(K_m))k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm}).$$

The marginal effect on profit of increasing k_{lm} can be written as

$$\underbrace{r'_m(K_m)k_{lm}}_{\text{Marginal retail profit}} \underbrace{-(p_m(K_m) - c_{lm})\frac{\partial Q_{-lm}(k_{lm}, K_{-lm})}{\partial k_{lm}}}_{\text{Strategic effect in short-term market}} \tag{7}$$

after invoking equilibrium condition (6) and the short-term market first-order condition (2). The firm's optimal amount of retail sales trades off the cost of a marginal reduction in retail profit against the strategic benefit of the marginal reduction in competitors' production. The following result generalizes Proposition 2.3 in Allaz and Vila (1993) of symmetric duopoly to asymmetric oligopoly:

Lemma 2 *Assume that the M local electricity markets are spatially independent. The volume of retail contracts sold by the L_m vertically integrated firms in local market $m \in \mathcal{M}$ equals*

$$K_m^I = L_m \frac{L_m - 1}{L_m^2 + 1} \frac{a_m - c_m}{b_m}$$

in an interior equilibrium. The average equilibrium markup in the short-term market equals

$$p_m^I - c_m = \frac{a_m - c_m}{L_m^2 + 1}.$$

Proof. *Set (7) equal to zero, substitute in the marginal effects $r'_m(K_m) = -\frac{b_m}{L_m+1}$ and $\frac{\partial Q_{-lm}}{\partial k_{lm}} = -\frac{L_m-1}{L_m+1}$, sum up over all L_m firms and use the definition of c_m to get*

$$-\frac{b_m K_m^I}{L_m + 1} + L_m \frac{L_m - 1}{L_m + 1}(p_m^I - c_m) = 0.$$

Add $\frac{a_m - c_m}{L_m+1}$ to both sides of this expression, and use $p_m^I - c_m = \frac{a_m - c_m}{L_m+1} - \frac{b_m K_m^I}{L_m+1}$ to solve for $p_m^I - c_m$. Plug this expression into the first-order condition to solve for K_m^I. Hence, the only possible interior equilibrium is the one specified in the lemma. The equilibrium exists because the profit of each firm l is strictly concave in k_{lm}. ■

Market performance under vertical integration A strategic incentive makes it individually rational for producers to vertically integrate into the retail market with the purpose of committing to aggressive behavior in the short-term market (marginal profit (7) is strictly positive if $K_m = 0$). A prisoners' dilemma arises by which all firms with the ability to exercise unilateral market power are vertically integrated in equilibrium.

Notwithstanding the pro-competitive effects of vertical integration, the equilibrium short-term price still remains excessive from a welfare viewpoint because firms' degree of vertical integration is incomplete in equilibrium:

$$\frac{K_m^I}{Q_m(K_m^I)} = \frac{L_m - 1}{L_m} < 1.$$

Hence, it would be socially valuable to reinforce producers' incentives to participate in the retail market. The key insight of this paper is that linking retail markets creates such an incentive.

3 Linking retail markets across space

Let the M local markets be collected in one regional market in the sense that each retailer must sell all wholesale electricity at the same price to all its retail customers regardless of where they are located in the region. Retailers then buy electricity at the price

$$p(\mathbf{K}) = \sum_{m=1}^{M} \frac{D_m}{D} p_m(K_m) \tag{8}$$

in the short-term market, where $D = \sum_{m=1}^{M} D_m$ is total demand, and $\mathbf{K} = (K_1, ..., K_m, ..., K_M)$ is the vector of retail contracts sold by all vertically integrated firms in each of the M local markets. The price p defines a quantity-weighted average of all M short-term prices, where each short-term price p_m is weighted by the size of local market m relative to the size of the whole market, measured in terms of MWh consumed.

Under this market design, independent retailers' marginal cost of supplying electricity to retail market m equals p instead of p_m as in the case of spatially independent retail markets. Under the maintained assumption that these retailers offer retail contracts at marginal cost, the market-clearing retail price r equals p at all M locations. The total revenue of firm l in market m then equals:

$$rk_{lm} + p_m(\tilde{q}_{lm} - k_{lm}).$$

The retail revenue rk_{lm} is sunk independently of how the retail price is determined when the L_m vertically integrated firms in market m decide how much electricity to bid into the short-term market. Hence, the short-term equilibrium in market m is still characterized by equations (3)-(5) as functions of the firms' retail positions.

3.1 Producers active in one location

Assume that each vertically integrated firm is active in one location in the sense that it sells retail contracts and has generation capacity in a single local market. As noted earlier, geographical concentration of production and retailing activity is realistic in electricity markets where many companies are former monopolists with local production capacity and distribution networks connected to local consumers.[11] We consider the effects of producers operating in multiple locations below.

The first-stage profit of firm l in market m is given by

$$(r(\mathbf{K}) - p_m(K_m))k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm}) \tag{9}$$

when retail markets are linked. The marginal effect on profit of increasing k_{lm} is

$$\underbrace{\frac{\partial r(\mathbf{K})}{\partial K_m}k_{lm} + r(\mathbf{K}) - p_m(K_m)}_{\text{Marginal retail profit}} \underbrace{-(p_m(K_m) - c_{lm})\frac{\partial Q_{-lm}(k_{lm}, K_{-lm})}{\partial k_{lm}}}_{\text{Strategic effect in short-term market}}. \tag{10}$$

Solving the set of linear first-order equations yields (the proof is in Appendix A.1):

Lemma 3 *Assume that a regional market links the M local electricity markets through a common retail price that is the quantity-weighted average of all short-term prices in all M markets. The average equilibrium markup in short-term market m equals*

$$p_m^R - c_m = \frac{\frac{1}{L_m}\frac{D_m}{D}\frac{a_m - c_m}{L_m+1}}{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1}}\frac{1}{\Psi_m} + \sum_{i=1}^{M}\frac{D_i}{D}\frac{(\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1})(c_m - c_i) + \frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1} - \frac{1}{L_i}\frac{D_i}{D}\frac{a_i-c_i}{L_i+1}}{(1 - \frac{1}{L_i}\frac{D_i}{D} - \frac{L_i-1}{L_i+1})(\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1})}\frac{1}{\Psi_m}$$

in an interior equilibrium if each vertically integrated firm is active in one local market, where

$$\Psi_m = \sum_{i=1}^{M}\frac{D_i}{D}\frac{\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1}}{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1}}\frac{1 - \frac{1}{L_m}\frac{D_m}{D} - \frac{L_m-1}{L_m+1}}{1 - \frac{1}{L_i}\frac{D_i}{D} - \frac{L_i-1}{L_i+1}}.$$

[11] Owning generation units in the same local markets that the vertically integrated firm serves load, provides a physical hedge against locational price difference between where a retailer produces or purchases wholesale energy and where it sells this energy to final consumers.

The equilibrium exists if and only if $L_m + 1 \geq \frac{D}{D_m}$ *for all* $m \in \mathcal{M}$.

Because retail markets are linked, market performance in each local market m depends not only on the demand and cost characteristics of that particular market, but also on the conditions in the other local markets that form the regional market. For instance, short-term market m tends to be less competitive if either marginal production costs are higher or the residual demand facing the vertically integrated firms is less price elastic (b_m is relatively large) than in the other markets.

Contrary to the case with spatially independent markets, the model with linked retail markets does not guarantee the existence of an equilibrium at the retail stage. But equilibrium non-existence is a general problem that does not arise specifically as a result of linking retail markets. To see this, rewrite the first-stage profit as

$$(p_m(K_m) - c_{lm})(q_{lm}(k_{lm}, K_{-lm}) - k_{lm}) + (\hat{r}_m(\mathbf{K}) - c_{lm})k_{lm},$$

where we allow firm l's retail price in market m to depend on the vector of retail positions more generally through an arbitrary price function $\hat{r}_m(\mathbf{K})$. The first part of the profit function is always strictly convex in k_{lm} no matter how the retail price is determined:

$$\frac{\partial^2}{\partial k_{lm}^2}(p_m(K_m) - c_{lm})(q_{lm}(k_{lm}, K_{-lm}) - k_{lm}) = \frac{2b_m}{(L_m + 1)^2} > 0. \tag{11}$$

Hence, the total first stage profit is well-behaved if and only if $(\hat{r}_m(\mathbf{K}) - c_{lm})k_{lm}$ is concave enough in k_{lm} to dominate the convexity of the first term. This condition is satisfied in Allaz and Vila (1993) because of full pass-through of the short-term price to the retail price:

$$\frac{\partial^2}{\partial k_{lm}^2}(\hat{r}_m(\mathbf{K}) - c_{lm})k_{lm} = 2p'_m(K_m) = -\frac{2b_m}{L_m + 1}.$$

It is never satisfied with zero pass-through, as would be the case under fixed-price regulation of the retail price, because then $(\hat{r}_m(\mathbf{K}) - c_{lm})k_{lm}$ would be linear in k_{lm}. In general, pass-through must be sufficiently high for the profit function to be well-behaved.

With linked retail markets, pass-through is high in local market m if there is a sufficient amount of vertically integrated firms in that market:

$$\frac{2b_m}{(L_m + 1)^2} + \frac{\partial^2(\hat{r}_m(\mathbf{K}) - c_{lm})k_{lm}}{\partial k_{lm}^2} = -\frac{2b_m}{L_m + 1}[\frac{D_m}{D} - \frac{1}{L_m + 1}].$$

Hence, the existence condition stated at the end of Lemma 3.

The effects on market performance of linking retail markets across space By comparing the first-stage marginal profit expressions (7) and (10), we first see that the retail price is less sensitive to an increase in individual firms' retail sales when local markets are linked through a regional retail price relative to the case when local markets are spatially independent,

$$\frac{\partial r(\mathbf{K})}{\partial K_m} = \frac{D_m}{D} p'_m(K_m) > p'_m(K_m) = r'_m(K_m), \tag{12}$$

because the short-term price in market m now constitutes only a fraction $\frac{D_m}{D}$ of the retail price. Some of the retail price effect of a higher retail supply in market m spills over to the other local markets under linked retail contracts. This negative externality on retail profit tends to push integrated firms in all local markets to increase sales in the retail market beyond what they would do if local markets were independent, which improves market performance in *all* local short-term markets.

The retail markup $r - p_m$ in (10) can either be positive or negative depending on whether m is a low-price or high-price location, which either reinforces or mitigates firms' incentives to participate in the retail market. Thus, linking electricity markets across space unambiguously improves short-term market competition in low price areas, but may have an ambiguous effect on market performance in high price areas. The following key proposition establishes the competitive effect of linking electricity markets across space:

Proposition 1 *Linking M local electricity markets through a regional market that establishes a common retail price equal to the quantity-weighted average of all short-term prices, causes p_m^R, the short-term price in local market m, to change by*

$$p_m^I - p_m^R = \frac{1 - \frac{D_m}{D}}{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m - 1}{L_m + 1}} \frac{L_m - 1}{L_m + 1}(p_m^I - c_m) + \frac{r^R - p_m^R}{\frac{1}{L_m}\frac{D_m}{D_x} + \frac{L_m-1}{L_m+1}}.$$

compared to p_m^I, the short-term price for the case of spatially independent markets, if all vertically integrated firms are active in one local market. Short-term market performance improves in all local markets in $\mathcal{M}$ if these are oligopolistic and sufficiently similar in terms of demand, cost characteristics and market structure (D_m, b_m, c_m and $L_m \geq 2$ are similar for all $m \in \mathcal{M}$).

Proof. *Set (10) equal to zero, substitute in the marginal effects $\frac{\partial r}{\partial K_m} = -\frac{D_m}{D}\frac{b_m}{L_m+1}$ and $\frac{\partial Q_{-lm}}{\partial k_{lm}} = -\frac{L_m-1}{L_m+1}$, sum up over all L_m firms and use the definition of c_m to get:*

$$-\frac{1}{L_m}\frac{D_m}{D}\frac{b_m}{L_m+1}K_m^R + r^R - p_m^R + \frac{L_m - 1}{L_m + 1}(p_m^R - c_m) = 0.$$

Add $\frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1}$ to both sides of this condition and use (3) to solve for the markup:

$$p_m^R - c_m = \frac{\frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1}}{\frac{1}{L_m}\frac{D_m}{D}+\frac{L_m-1}{L_m+1}} - \frac{r_x^R - p_m^R}{\frac{1}{L_m}\frac{D_m}{D}+\frac{L_m-1}{L_m+1}}.$$

Subtract this expression from $p_m^I - c_m$ characterized Lemma 2 to obtain the expression in Proposition 1. The equilibrium retail price

$$r^R = \sum_{i=1}^{M}\frac{D_i}{D}p_i^R$$

is close to p_m^R for all $m \in \mathcal{M}$ if all local markets are sufficiently similar. In this case, the second term on the right-hand side of the expression in Proposition 1 is dominated by the first term, so that $p_m^I > p_m^R$ for all $m \in M_x$. ■

Market integration can improve market performance in all product markets even if one cannot trade in physical goods between them. All production and consumption is cleared locally in this model. Integration of retail markets is enough to improve incentives for competitive behavior by vertically integrated suppliers in all local product markets.

In the special case of perfect symmetry across all local markets, so that $a_m = a$, $b_m = b$, $c_m = c$ and $L_m = L$ for all m, index

$$H(L,M) = 1 - \frac{p^R - c}{p^I - c} = \frac{L(L-1)(M-1)}{ML(L-1)+L+1}. \qquad (13)$$

measures relative market performance between linked and independent markets by the associated percentage reduction in the short-term price-cost margin. This index depends only on the number L of firms with market power in each market and the number M of local markets that are linked. Conveniently, the demand and cost characteristics cancel out because of symmetry and the linear structure of the model.

Figure 1 plots $H(L, L+1)$, the competitive effects of linking the maximum number of local markets consistent with equilibrium. The number L of vertically integrated firms in each local market places an upper bound to the number $M \leq L+1$ of local markets that can be linked while preserving equilibrium existence. Linking retail markets nevertheless has substantial quantitative welfare effects because the incremental effects on wholesale competition are the strongest for small M. The incremental pro-competitive effect is the largest when local

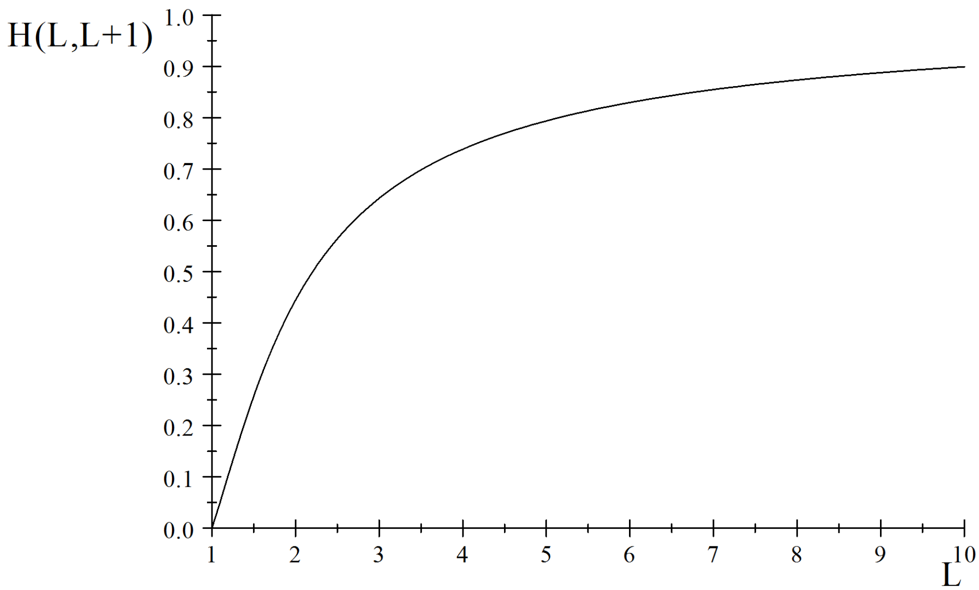

Figure 1: The competitive effects of linking retail markets across space

competition is weak, i.e. for small L. Linking three duopoly markets causes the average markup to drop by nearly half (44 percent) in each market. Linking four triopoly markets generates a 20 percent additional improvement in market performance. In reality, markets are asymmetric and market concentration likely to change depending on transmission network capacity constraints. In Section 7, we extend the model to include these features and show that the quantitative effects are smaller, but still substantial.

3.2 Producers active in multiple locations

Assume now that vertically integrated firms can have production facilities in more than one local market. Firm l chooses its retail portfolio $\mathbf{k}_l = (k_{l1}, ..., k_{lm}, ..., k_{lM})$ across the local markets to maximize profit

$$\sum_{m=1}^{M} \beta_{lm}[(r(\mathbf{K}) - p_m(K_m))k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm})],$$

where β_{lm} is an indicator function taking the value 1 if firm l owns generation capacity in local market m and 0 if not. The marginal effect on profit of increasing k_{lm} is

$$\sum_{i=1}^{M} \frac{\partial r(\mathbf{K})}{\partial K_m} \beta_{li} k_{li} + \beta_{lm}[r(\mathbf{K}) - p_m(K_m) - (p_m(K_m) - c_{lm}) \frac{\partial Q_{-lm}(k_{lm}, K_{-lm})}{\partial k_{lm}}]. \tag{14}$$

Compared to the case in which producers are active in only one market, the firm now takes into account the spill-over effects of the retail price reduction in the other markets in which it is present.

By the linearity of the marginal profit functions, one could solve explicitly for the equilibrium level of retail sales in each local market for each firm. To obtain results that are easier to interpret, we here impose perfect symmetry on the model, similar to the case considered in (13). Let there be L vertically integrated firms with the ability to exercise unilateral market power, each of which has production facilities in $\bar{M}$ of the M local markets. Firm locations are symmetric, with the same number $S = L\frac{\bar{M}}{M}$ of firms active in each local market. Then (the proof is in Appendix A.2):

Lemma 4 *Assume that a regional market links M local and symmetric electricity markets through a common retail price that is the quantity-weighted average of all M short-term prices. The markup in each short-term market equals*

$$p^R - c = \frac{a - c}{\frac{M-\bar{M}}{\bar{M}} S(S-1) + S^2 + 1} \tag{15}$$

in a symmetric interior equilibrium if each vertically integrated firm is active in $\bar{M} \geq 1$ local markets. This equilibrium exists only if $S + 1 \geq \frac{M}{\bar{M}}$. The condition is also sufficient if firms are restricted to symmetric retail positions.

The convexity problem of the first-stage profit function established in (11) for the case when firms have production in a single local market becomes more severe under multi-market presence. We show in Appendix A.2 that it would always be a profitable deviation under the conditions of Lemma 4 for a vertically integrated producer to concentrate its retail sales in *one* of its local markets and reduce retail sales proportionally in the other local markets. However, such an asymmetric deviation could be difficult to achieve under symmetric market conditions. The retail price is the same across the regional market. Hence, the firm is equally likely to attract retail customers from all local markets. To build an asymmetric market share, the firm would then have to accept customers from one location, but systematically refuse to sell retail contracts to customers from all other locations. A regulatory rule prohibiting such location-based discrimination in the retail market would serve to even out market shares in a symmetric, linked competitive retail market. This rule would not bind and therefore entail no welfare loss in symmetric equilibrium.

The effects on market performance of linking retail markets Multi-market presence implies that vertically integrated firms account for a larger share of the spill-over effects of the retail price into the other local markets. This increased internalization of retail price effects softens the incentive to participate in the retail market. Subtracting p^R defined in Lemma 4 from p^I yields:

Proposition 2 *Linking M local and symmetric electricity markets through a regional market that establishes a common retail price equal to the quantity-weighted average of all short-term prices, causes short-term prices in each local market to fall by*

$$\frac{p^I - p^R}{p^I - c} = \frac{(M - \bar{M})S(S-1)}{MS(S-1) + \bar{M}(S+1)} \geq 0$$

if each vertically integrated firm is active in $\bar{M}$ local markets.

Linking electricity markets drives down short-term market prices in all symmetric equilibria unless all firms have production in all local markets so that $\bar{M} = M$.

4 Linking forward markets across space

In Section 3, we studied vertical integration and the effects of linking of local electricity markets through a common retail contract. We consider now forward contracting and analyze instead the consequences of linking local markets through a forward contract that clears against the quantity-weighted average $\sum_{m=1}^{M} \frac{D_m}{D} p_m$ of the short-term wholesale market prices. Such contracts are common for instance in the Nordic market, where a standard forward contract clears against the market-wide system price. US LMP markets often create trading hub prices, which are quantity-weighted averages of LMPs in a sub-region of the market. Because these hub prices are typically less volatile than any individual component LMP, market participants manage their short-term price risk by purchasing forward contracts that clear against these trading hub prices.

Our main result in this section demonstrates that linking retail and forward markets have fundamentally different effects on short-term market performance. We only detail the case of linked forward markets here because the analysis of spatially independent forward contracts is formally equivalent to the analysis in Section 2.

Equilibrium in the short-term market Producers with the ability to exercise unilateral market power take into account how their forward contract position affects their output choice in the short-term market.[12] The second-stage profit of producer l in market m thus becomes

$$(f - \sum_{i=1}^{M} \frac{D_i}{D} P_i(\tilde{Q}_i))k_{lm} + (P_m(\tilde{Q}_m) - c_{lm})\tilde{q}_{lm},$$

where f is the delivery price of the composite forward contract, k_{lm} here refers to the volume of forward contracts sold by firm l in market m, and we assume that firms with market power are active in one local market only. The first-order condition for the firm's quantity choice reads:

$$-\frac{D_m}{D} P'_m(Q_m)k_{lm} + P_m(Q_m) - c_{lm} + P'_m(Q_m)q_{lm} = 0,$$

which differs from the case of spatially independent forward markets, see equation (2), by an increase in production now having a relatively smaller positive effect on forward profit because of the term $\frac{D_m}{D}$. Therefore, holding the forward contract quantity constant, the short-term market behavior by the vertically integrated firms is less competitive than under spatially independent forward markets, which implies the following result:

Lemma 5 *Assume that M local markets are linked through a forward contract that clears against the quantity-weighted average of the short-term prices in those local markets. Assume also that K_m is sufficiently small that all producers have positive production in equilibrium. The average equilibrium markup is given by*

$$p_m(\frac{D_m}{D} K_m) - c_m = \frac{a_m - c_m}{L_m + 1} - \frac{b_m}{L_m + 1} \frac{D_m}{D} K_m$$

in short-term market m. The corresponding equilibrium production of firm l is

$$q_{lm}(\frac{D_m}{D} k_{lm}, \frac{D_m}{D} K_{-lm}) = \frac{1}{L_m + 1} \frac{a_m - c_m}{b_m} + \frac{c_m - c_{lm}}{b_m} + \frac{L_m}{L_m + 1} \frac{D_m}{D} k_{lm} - \frac{1}{L_m + 1} \frac{D_m}{D} K_{-lm}$$

[12] Wolak (2000) demonstrates the empirical relevance of this mechanism for a large supplier in an Australian wholesale electricity market.

and of all other firms in local market m:

$$Q_{-lm}(\frac{D_m}{D}k_{lm}, \frac{D_m}{D}K_{-lm}) = \frac{L_m - 1}{L_m + 1}\frac{a_m - c_m}{b_m} - \frac{c_m - c_{lm}}{b_m} - \frac{L_m - 1}{L_m + 1}\frac{D_m}{D}k_{lm} + \frac{2}{L_m + 1}\frac{D_m}{D}K_{-lm}.$$

Note that the strategic effect is weaker than in equations (4) and (5) because an increase in k_{lm} now has a smaller effect on output q_{lm}.

Equilibrium in the forward market Assuming perfect foresight makes the forward price satisfy the equation:

$$f(\mathbf{K}) = \sum_{i=1}^{M} \frac{D_i}{D} p_i(\frac{D_i}{D}K_i)$$

in a perfectly competitive forward market. Firm l in market m chooses k_{lm} to maximize the first stage profit

$$(f(\mathbf{K}) - \sum_{i=1}^{M} \frac{D_i}{D} p_i(\frac{D_i}{D}K_i))k_{lm} + (p_m(\frac{D_m}{D}K_m) - c_{lm})q_{lm}(\frac{D_m}{D}k_{lm}, \frac{D_m}{D}K_{lm}). \tag{16}$$

The marginal effect

$$\frac{\partial f(\mathbf{K})}{\partial K_m}k_{lm} - (p_m(\frac{D_m}{D}K_m) - c_{lm})\frac{\partial Q_{-lm}(\frac{D_m}{D}k_{lm}, \frac{D_m}{D}K_{lm})}{\partial k_{lm}} \tag{17}$$

on profit of increasing forward sales k_{lm} differs in two respects from the marginal profit (7) when forward markets are spatially independent. On the one hand, an increase in the volume of forward contracts sold by an individual firm has a relatively smaller effect on the forward contract price when forward markets are linked. This effect tends to drive up forward sales and reinforce short-term market competition. On the other hand, the strategic effect is weaker under linked forward markets, which tends to soften competition for any given volume of forward contracts. As it turns out, the two marginal effects exactly offset one another compared to the case of spatially independent forward markets:

Proposition 3 *Assume that M local markets are linked through a forward contract that clears against the quantity-weighted average of the short-term prices in those local markets. The volume of retail contracts sold by the L_m firms with market power in local market m equals*

$$K_m^F = \frac{D}{D_m}L_m\frac{L_m - 1}{L_m^2 + 1}\frac{a_m - c_m}{b_m} = \frac{D}{D_m}K_m^I$$

in interior equilibrium if each firm is active in one local market. The average markup in short-term market m equals

$$p_m^F - c_m = \frac{a_m - c_m}{L_m^2 + 1} = p_m^I - c_m.$$

Proof. *Set (17) equal to zero, substitute in the marginal effects $\frac{\partial f}{\partial K_m} = -(\frac{D_m}{D})^2 \frac{b_m}{L_m+1}$ and $\frac{\partial Q_{-lm}(\frac{D_m}{D}k_{lm}, \frac{D_m}{D}K_{lm})}{\partial k_{lm}} = -\frac{D_m}{D}\frac{L_m-1}{L_m+1}$, sum up over all L_m firms and divide through by $L_m \frac{D_m}{D}$ to get the first-order condition*

$$-\frac{1}{L_m}\frac{b_m}{L_m+1}\frac{D_m}{D}K_m^F + \frac{L_m - 1}{L_m + 1}(p_m^F - c_m) = 0.$$

Add $\frac{1}{L_m}\frac{a_m - c_m}{L_m+1}$ to both sides of the equation and solve for the equilibrium markup $p_m^F - c_m$ above. Plug this expression into the first-order condition to solve for K_m^F. ■

The effects on market performance of linking forward vs retail markets Linking forward markets through a contract that clears against a quantity-weighted average of wholesale prices has an effect on the firm's short-term market behavior that does not arise when retail markets are linked by requiring the firm to sell wholesale electricity to its retail customers at the quantity-weighted average of locational prices. The fundamental difference between the two market designs is illustrated by comparing the first-stage profit expressions (16) and (9). For a firm located in market m, a linked forward contract clears against the quantity-weighted average of short-term locational prices. A linked retail contract essentially clears against the local short-term price at location m, p_m. This difference softens the strategic benefit so much under forward contracting that the total competitive effect of linking markets vanishes.

5 Combined retail and forward markets

In this Section, we extend our model of vertically integrated firms selling their output to retail customers to allow them to sell forward financial contracts as way to commit to more aggressive behavior in the short-term market. For simplicity, we assume that firms in local market m sell forward contracts that clear against the local short-term price p_m.[13]

[13] Assuming instead that forward contracts clear against the quantity-weighted average short-term price as in Section 4 would not change the results.

Equilibrium when markets are spatially independent The total revenue of firm l in market m equals:

$$r_m k_{rlm} + p_m(\tilde{q}_{lm} - k_{rlm}) + (f_m - p_m)k_{flm} = r_m k_{rlm} + f_m k_{flm} + p_m(\tilde{q}_{lm} - k_{lm}),$$

where k_{rlm} (k_{flm}) refers to the volume of retail (forward) contracts sold by l in market m, f_m is the delivery price in forward market m, and $k_{lm} = k_{rlm} + k_{flm}$ is the total volume of first-stage commitments.

The first stage revenues $r_m k_{rlm}$ and $f_m k_{flm}$ are sunk at the second stage, so the production decision of firm l depends only on the total first-stage volume k_{lm}. Hence, the equilibrium in the short-term market can be characterized exactly as in Lemma 1. The profit of firm l in market m therefore equals

$$r_m(K_m)k_{rlm} + f_m(K_m)k_{flm} - p_m(K_m)k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm}),$$

where the retail price in market m and the forward price in market m are both equal to the short-term price:

$$r_m(K_m) = f_m(K_m) = p_m(K_m).$$

Because the second-stage profit only depends on the firm's total first-stage commitment k_{lm}, firm l divides any given volume k_{lm} across the retail and the forward market to maximize the first-stage revenue. But the retail price and the forward price are identical, so the firm is indifferent between participating in the two markets. Hence, it is optimal to set $k_{rlm} = k_{lm}$ and $k_{flm} = 0$ and then maximize total profit over k_{lm}, which brings us back to the analysis in Section 2. Hence, a forward market clearing against the local short-term price p_m adds nothing to competition when markets are spatially independent.

Equilibrium when retail markets are linked This section considers the case of linked retail markets with the option of vertically-firms to sell fixed-price forward contracts that clear against their local short-term price. The total profit of firm l can be written as

$$r(\mathbf{K})k_{rlm} + f_m(K_m)k_{flm} - p_m(K_m)k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm})$$

if retail markets are linked and the firm participates in one single local market. The firm is no longer indifferent between participating in the two markets in the general case where the retail price differs from the local spot price. A firm in a low-price location, $f_m = p_m < r$, would participate entirely in the retail market, whereas a firm in a high price location, $f_m = p_m > r$, would sell forward contracts. The following characterization is straightforward.[14]

Proposition 4 *Assume that a regional market links M local electricity markets through a common retail price that is the quantity-weighted average of all short-term prices. Assume also that firms can sell forward contracts that clear against the local spot price. The average equilibrium markups can then be the characterized by*

$$p_m^{RF} - c_m = \frac{a_m - c_m}{L_m^2 + 1} = p_m^I - c_m \text{ for } p_m^{RF} \geq r^{RF} \tag{18}$$

$$p_m^{RF} - c_m = (1 - \frac{1 - \frac{D_m}{D}}{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m - 1}{L_m + 1}} \frac{L_m - 1}{L_m + 1})(p_m^I - c_m) - \frac{r^{RF} - p_m^{RF}}{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m - 1}{L_m + 1}} \text{ for } p_m^{RF} < r^{RF} \tag{19}$$

if all vertically integrated firms are active in one local market.

Linking retail markets has no effect on short-term competition in high price locations, but has positive consequences for competition in low price locations. The markup in low-price locations is a fraction (less than one) of the markup in that location under independent retail markets less a factor that depends on the difference between the regional retail price r^{RF} and the local price p_m^{RF}. Therefore, even if suppliers have the option to sign fixed-price forward contracts that clear against their local short-term price, market performance is non-decreasing in all markets as a result of linking retail markets through the requirement that all retail electricity sell at regional price, r^{RF}.

6 Linking retail markets across time

Consider now the case of linking retail markets over time through a long-term retail contract. Assume that there is a single local market, M production periods and L vertically integrated producers with market power. We allow demand D_m to fluctuate across time, but assume it to be deterministic. Total demand across the M periods is given by $D = \sum_{m=1}^{M} D_m$. Denote by k_{lm} the retail obligation of firm l in period m and K_m the total retail obligations of all L firms in period m.

The output q_{lm} of producer l that maximizes period m profit is independent of the retail price r_m that period, so the short-term price, producer l's equilibrium quantity and all

[14] The existence proof is complicated by kinks in the profit functions. We consider in Appendix A.3 the case of two local markets and show that sufficient conditions for existence, in addition to local concavity, are that the high-price area has a higher demand than the low-price area and that any cost advantage by firms in the high price area is sufficiently small.

other producers' output in period m are given by Lemma 1. A sequence of short-term retail contracts is equivalent to a set of spatially independent retail contracts, so Lemma 2 characterizes the equilibrium also under short-term retail contracts.

6.1 Fixed-price retail contracts

Let firms supply long-term retail contracts in period 1 that are valid for M periods. In this section we assume that customers pay the same price for every MWh of electricity consumed during the contract period, consistent with the fixed-price retail contracts customers sign in many competitive retail markets. Under this contract, firm l cannot tailor its retail volume k_{lm} optimally period by period.[15] Instead, we let each firm decide the average per-period volume k_l of electricity to supply to the retail market through this type of contract. Its retail position in period m therefore equals $k_{lm} = \frac{D_m}{D} M k_l$.

The total retail profit of firm l equals

$$\sum_{m=1}^{M} k_{lm}(r - p_m(K_m)) = \sum_{m=1}^{M} \frac{D_m}{D} M k_l (r - p_m(K_m))$$

where r is the retail price, and the discount rate set is equal to one. Perfect competition in the retail market and perfect foresight drive the retail profit down to zero, in which case the equilibrium price of the long-term contract becomes:

$$r(\mathbf{K}) = \sum_{m=1}^{M} \frac{D_m}{D} p_m(K_m)$$

as a function of the retail profile $\mathbf{K} = (K_1, ..., K_m, ...K_M)$ of the vertically integrated firms.

In period 1, firm l maximizes the total profit

$$\sum_{m=1}^{M} [(r(\mathbf{K}) - p_m(K_m))k_{lm} + (p_m(K_m) - c_{lm})q_{lm}(k_{lm}, K_{-lm})]$$

over k_l. The marginal effect on profit of increasing k_l equals

$$M \sum_{m=1}^{M} \frac{D_m}{D} [\sum_{t=1}^{M} \frac{\partial r(\mathbf{K})}{\partial K_m} k_{lt} + r(\mathbf{K}) - p_m(K_m) - (p_m(K_m) - c_{lm}) \frac{\partial Q_{-lm}(k_{lm}, K_{-lm})}{\partial k_{lm}}]. \quad (20)$$

[15] We consider the implications of flexible long-term retail contracts below.

The expression inside the square brackets is similar to the one of vertically integrated firms being active in multiple (all) local markets, see equation (14), where a local market here is the same market repeated over time. The period-level marginal profit is then aggregated across all M periods because firms are constrained to sell the same average volume k_l of retail electricity every period, instead of tailoring contracts individually for each period.

Proposition 5 *Assume that firms sell fixed-price retail contracts that cover M periods. The average volume of electricity sold in the retail market each period by the L firms with market power equals*

$$K^T = L\frac{L-1}{L^2+1}\frac{\frac{1}{M}\sum_{m=1}^{M} D_m(a_m - c_m)}{\sum_{m=1}^{M} a_m \frac{D_m}{D}} \tag{21}$$

in an interior equilibrium. The price-cost margins in the short-term market satisfy

$$\sum_{m=1}^{M} \frac{D_m}{D}(p_m^T - c_m) = \sum_{m=1}^{M} \frac{D_m}{D}\frac{a_m - c_m}{L^2+1} = \sum_{m=1}^{M} \frac{D_m}{D}(p_m^I - c_m). \tag{22}$$

Proof. *By setting (20) equal to zero and repeating the same steps as in the previous proofs, we obtain the aggregate first-order condition*

$$-\frac{1}{L}\sum_{m=1}^{M} D_m \frac{b_m K_m^T}{L+1} + \frac{L-1}{L+1}\sum_{m=1}^{M} D_m(p_m^T - c_m) = 0.$$

after simplification. Adding and subtracting $\frac{a_m - c_m}{L+1}$ *under the first sum allows us to solve for (22). Replacing* $p_m^T - c_m$ *in (22) with* $\frac{a_m-c_m}{L+1} - \frac{b_m}{L+1}\frac{D_m}{D}MK^T$ *yields* K^T *given by (21). This solution represents an equilibrium because the second-derivative with respect to* k_l *of* l*'s total profit function is negative:* $\frac{-2L}{(L+1)^2}\sum_{m=1}^{M} a_m < 0$. ■

Note that the price-cost margin in any given period can be higher or lower under long-term than short-term retail contracts depending on how the demand and cost characteristics vary over time. Proposition 5 adapted to this context demonstrates that long-term contracts have no effect on average market performance. The average markup in the spot market is the same under both types of retail contracts if each period m is weighted by relative consumption $\frac{D_m}{D}$.[16] The reason is because firms in this case produce in all periods and are constrained to

[16] However, contracting over time affects market efficiency. We show in Appendix A.4 that total production costs can be smaller or higher under long-term than short-term retail contracts depending on market heterogeneity. Specifically, long-term contracts yield comparatively small total costs if the main source of heterogeneity relates to cost differences across firms with the ability to exercise unilateral market power.

sell the same average volume of electricity in the retail market in all periods.[17]

6.2 Planned Outages

Assume now that firms do not produce in all periods. Specifically, each firm must shut down its production exactly one period for maintenance reasons. To minimize the number of simultaneous outages, the system operator takes one firm off-line every period m. Scheduling is done prior to market transactions, so it is common knowledge at stage 0 which generators will be on-line in all periods. Let there be $L+1$ vertically integrated firms, and assume that there are $M = L$ symmetric production periods. This means that there are L firms producing in every period m.

Maintenance does not affect production by the remaining firms, so firm l produces $q(k_{lm}, K_{-lm})$ when active in period m, the other firms produce $Q_{-l}(k_{lm}, K_{-lm})$, and the spot price is $p(K_m)$. The equilibrium under short-term contracting remains unaffected, so the total retail position of the vertically integrated firms equals K^I, and the equilibrium price-cost margin is $p^I - c$ in every period.

When it comes to the long-term retail contracts, assume that these are fully flexible in the sense that firm l can set k_{lm} independently of k_{lt} for the entire contracting period. Maintain the assumption that there is no discounting of profit. Under these circumstances, long-term retail contracts are formally equivalent to a spatial contract, with each firm being active in all local markets but one. Hence, the equilibrium markup is characterized by Lemma 4 with $N = M - 1$. In light of Proposition 2, the following result is straightforward:

Proposition 6 *Assume that firms sell flexible long-term retail contracts that cover M periods and that outages are planned so that each firm shuts down its production exactly one period. Assume also that markets are symmetric. Then, the price-cost margins in the spot market are smaller under long-term retail contracts compared to the case of short-term contracts.*

Planned outages imply that some of the retail price effects of a larger retail position spill over to periods during which the firm is not active. In this case, a long-term retail contract is more competitive than a sequence of short-term retail contracts.

Instead, short-term contracting is relatively more efficient if most of the heterogeneity stems from differences in demand across markets.

[17]The analytical analogue to the fixed-price contract in a spatial setting is a situation where firms own generation units in all local markets. Moreover, firms cannot individually target local retail markets. Instead, they attract retail customers from each local market m in proportion to the relative size $\frac{D_m}{D}$ of market m.

6.3 Random Outages

Assume that everything is as in Section 6.2, except now the system operator takes one unit off-line every period m by randomizing among the generators. The probability that generator l is on-line in period m is equal to $\mu = \frac{M-1}{M}$. This randomization is done prior to production taking place, but in such a manner that producers do not know the exact period they will be off-line when they sell long-term retail contracts. However, it is common knowledge that there are L units on-line in every period.

Let $\mathbf{k}_l = (k_{l1}, ..., k_{lm}, ..., k_{lM})$ be the retail profile of firm l. The equilibrium production levels $q(k_{lm}, \tilde{K}_{-lm})$, $Q_{-l}(k_{lm}, \tilde{K}_{-lm})$ and the short-term prices $p(\tilde{K}_{-lm} + \beta_{lm}k_{lm})$ in period m are qualitatively the same as in Lemma 1, except $\tilde{K}_{-lm} = \sum_{i \neq l} \beta_{im}k_{im}$ are the retail positions of those generators other than l are *active* in period m, whereas $K_m = \sum_{l=1}^{L+1} k_{lm}$ is the aggregate forward positions of all $L+1$ producers that period.

We assume that all uncertainty about outages is resolved before firms choose their short-term retail positions in each period. This means that the equilibrium is exactly the same as in the case of independent markets characterized in Section 2: The price-cost margin in the short-term market equals $p^I - c$ in every period.

The expected net present value of a long-term retail contract is

$$\sum_{m=0}^{M} (r - \mu E[p(\tilde{K}_{-lm} + k_{lm})] - (1-\mu)p(\tilde{K}_{-lm}))D,$$

where E is the period 1 expectations operator with respect to other firms than l being operative, conditional on firm l being operative. Perfect competition in the retail market and risk-neutrality drive the expected retail margin to zero, so

$$r(\mathbf{k}) = \frac{1}{M+1} \sum_{m=0}^{M} (\mu E[p(\tilde{K}_{-lm} + k_{lm})] + (1-\mu)p(\tilde{K}_{-lm})),$$

where $\mathbf{k} = (\mathbf{k}_1, ..., \mathbf{k}_l, ..., \mathbf{k}_{L+1})$.

Producer l chooses $\mathbf{k}_l$ to maximize the total expected profit

$$\sum_{m=0}^{M} E[(r(\mathbf{k}) - \mu p(\tilde{K}_{-lm} + k_{lm}) - (1-\mu)p(\tilde{K}_{-lm}))k_{lm} + \mu(p(\tilde{K}_{-lm} + k_{lm}) - c)q(k_{lm}, \tilde{K}_{-lm})].$$

The marginal benefit of increasing k_{lm} is

$$\sum_{t=0}^{M} \frac{\partial r(\mathbf{k})}{\partial k_{lm}} k_{lt} + \mu(r(\mathbf{k}) - E[p(\tilde{K}_{-lm} + k_{lm})]) - \frac{\partial Q_{-l}(k_{lm}, \tilde{K}_{-lm})}{\partial k_{lm}} \mu(E[p(\tilde{K}_{-lm} + k_{lm})] - c),$$

where $\frac{\partial r(\mathbf{k})}{\partial k_{lm}} = \frac{\mu}{M+1} p'(\tilde{K}_{-lm} + k_{lm})$. The ex ante probability of a spot price change in period m is only μ because of stochastic outages. However, each firm still internalizes all expected marginal externalities because of the linearity of the marginal effects. Hence, the outcome is the same as under short-term contracts:

Proposition 7 *Assume that firms sell flexible long-term retail contracts that cover M periods and that outages are random. Assume also that markets are symmetric. Then, the price-cost margins in the spot market are the same under long-term and short-term retail contracts in stationary equilibrium.*

Proof. *If all firms sell the same amount of retail contracts and the equilibrium is stationary, $k_{lm}^T = k^T$ for all l and m, it follows that $E[p(\tilde{K}_{-lm}^T + k_{lm}^T)] = p(\tilde{K}_{-lm}^T) = p(Lk_m^T) = p(Lk^T)$. Hence, $r^T = p(Lk^T)$, which implies the first-order condition*

$$\mu[-\frac{b}{L+1} k^T + \frac{L-1}{L+1}(p(Lk^T) - c)] = 0,$$

with solution $Lk^T = K^I$. The equilibrium can be sustained under a restriction to symmetric deviations. ■

The result demonstrates that random outages alone are not a reason to link retail markets across time because market performance is independent of the probability μ of outages in a symmetric equilibrium.

7 Transmission network constraints

Thus far we have taken the market structure as given and analyzed how load aggregation pricing of retail demand affects market performance. However, a firm's ability to influence prices in the short-term wholesale market depends on the extent to which the transmission network is capacity constrained. Transmission network constraints shrink the size of the geographic market that a given supplier competes in, which increases the ability of the supplier to exercise unilateral market power. The extent and number of transmission constraints in a wholesale electricity market varies throughout the year, although they tend to bind more during peak demand hours. In this section, we randomly increase and decrease the number of competitors and market demand each strategic wholesale supplier faces in order to quantify in a realistic manner the impact of load aggregation pricing on competition and prices in the short-term wholesale market with transmission network constraints.

At every node $n \in \mathcal{N} = \{1, ..., n, ..., N\}$ in the system there is one generation unit with constant marginal cost c_n and the ability to exercise unilateral market power, plus a competitive fringe that supplies electricity to the short-term market at linear marginal cost $b_n Q_{nfringe}$. There are L firms with market power. Each firm $l \in \mathcal{L} = \{1, 2..., L\}$ owns the non-fringe generation unit in a subset $\mathcal{N}_l$ of the nodes, i.e. $\cup_{l \in \mathcal{L}} \mathcal{N}_l = \mathcal{N}$.

We model network constraints as a probability distribution $(\phi_1, ..., \phi_j, ..., \phi_J)$, $\phi_j > 0 \forall j$, $\sum_{j=1}^{J} \phi_j = 1$ over J different market structures realized after firms participate in the retail market, but before they compete in the short-term market. We allow for the possibility that network constraints are the results of demand fluctuations: Demand at node $n \in \mathcal{N}$ is D_{nj} under market structure j. This model implies that firms can neither alleviate transmission network congestion nor induce it by their actions. We are trying to capture the market size-shrinking impact of transmission congestion without adding the complexity of wholesale market participants causing congestion. This model implies that congestion is random from the perspective of retailers, which is seems to be a realistic assumption. We solve the model by backward induction.

Short-term markets Market structure j is a partitioning $\mathcal{M}_j = \{\mathcal{M}_{1j}, ..., \mathcal{M}_{mj}, ..., \mathcal{M}_{M_j j}\}$ of the N nodes in the network into M_j short-term local markets with the same locational marginal price (LMP) at all nodes within the local market: $p_{nj} = p_{mj} \forall n \in \mathcal{M}_{mj}$. In other words, network capacity is sufficient to handle equilibrium trade flows within each local market, but bottlenecks prevent price equalization across local markets.

Let $\mathcal{L}_{mj}$ be the subset of firms with market power that own non-fringe generation units in market mj. If we let $\tilde{Q}_{mj}$ be their total production, then

$$P_{mj}(\tilde{Q}_{mj}) = a_{mj} - b_{mj}\tilde{Q}_{mj},\ a_{mj} = b_{mj}(D_{mj} - Imp_{mj}),\ \frac{1}{b_{mj}} = \sum_{n \in \mathcal{M}_{mj}} \frac{1}{b_n}$$

returns the inverse demand curve facing firms with market power in local market mj. Summing up across nodes yields demand $D_{mj} = \sum_{n \in \mathcal{M}_{mj}} D_{nj}$ in local market mj. Imp_{mj} measures net imports into local market mj from surrounding local markets and could be positive or negative depending on whether mj is import or export constrained.[18]

Our assumptions of constant marginal cost and no capacity constraints imply that the marginal cost of firm $l \in \mathcal{L}_{mj}$ is $c_{lmj} = \min\{c_n\}_{n \in \mathcal{N}_l \cap \mathcal{M}_{mj}}$. The average marginal production cost of the L_{mj} large firms with market power in market mj equals $c_{mj} = \frac{1}{L_{mj}} \sum_{l \in \mathcal{L}_{mj}} c_{lmj}$.

Firms with market power are vertically integrated. Let k_{lnj} be the volume consumed by firm l's retail customers at node n under market structure j. The total retail consumption of the firm's consumers is $k_{lmj} = \sum_{n \in \mathcal{M}_{mj}} k_{lnj}$ in local market mj. Summing up over the L_{mj} firms gives the total retail volume $K_{mj} = \sum_{l \in \mathcal{L}_{mj}} k_{lmj}$ delivered by the vertically integrated firms with market power in market mj.

As the market structure is given when firms participate in the short-term market, the short-term equilibrium in local market mj is characterized by equations (3)-(5) subject to an appropriate adjustment of notation, so that $p_{mj}(K_{mj})$ refers to the short-term price (LMP) at all nodes in local market mj, $q_{lmj}(k_{lmj}, K_{-lmj})$ is the total production of firm $l \in \mathcal{L}_{mj}$, $Q_{-lmj}(k_{lmj}, K_{-lmj})$ is the production of the other firms with market power in market mj, and $K_{-lmj} = K_{mj} - k_{lmj}$ is the total retail obligation of those other firms.

Retail markets Every firm $l \in \mathcal{L}$ decides before the market structure is realized the volume of retail contracts to supply at each node $n \in \mathcal{N}$ in the system. Independent retailers cover the difference between consumption D_{nj} and the total retail supply $K_{nj} = \sum_{l=1}^{L} k_{lnj}$ by vertically integrated firms at every node n for every market structure j.

Let $\mathcal{X} = \{\mathcal{X}_1, ..., \mathcal{X}_x, ..., \mathcal{X}_X\}$ be a partitioning of the N nodes into X service territories or load aggregation points. One can think of each service territory as a retail market because a regulatory mandate requires retailers to serve all demand within territory x at the same retail price no matter which node $n \in \mathcal{X}_x$ consumers are located at. By that same mandate, independent retailers purchase electricity at the quantity-weighted average of the short-term prices at all nodes $n \in \mathcal{X}_x$. If $D_{xmj} = \sum_{n \in \mathcal{X}_x \cap \mathcal{M}_{mj}} D_{nj}$ is the consumption in territory x that takes place in local market mj, and $D_{xj} = \sum_{n \in \mathcal{X}_x} D_{nj}$ is total consumption in service territory x, then

$$p_{xj}(\mathbf{K}_j) = \sum_{m=1}^{M_j} \frac{D_{xmj}}{D_{xj}} p_{mj}(K_{mj})$$

is the wholesale price of electricity independent retailers in territory x pay for their electricity under market structure j, where $\mathbf{K}_j = (K_{1j}, ..., K_{mj}, ..., K_{M_j j})$ identifies vertically integrated firms' aggregate retail supply in every local market under market structure j. Independent

[18]The underlying assumption is that export capacity is always bid inelastically into neighboring markets at price zero and therefore dispatched first.

retailers' expected unit cost of serving retail demand in service territory x then equals

$$\sum_{j=1}^{J} \phi_j \frac{D_{xj} - K_{xj}}{D_x - K_x} p_{xj}(\mathbf{K}_j)$$

if $K_{xj} = \sum_{n \in \mathcal{X}_x} K_{nj}$ is the total retail volume served by vertically integrated firms in service territory x under market structure j, $D_x = \sum_{j=1}^{J} \phi_j D_{xj}$ is the expected total demand and $K_x = \sum_{j=1}^{J} \phi_j K_{xj}$ the expected volume supplied by vertically integrated firms in that service territory.

As in Section 6, retail offers at every node $n \in \mathcal{N}$ are fixed-price contracts that permit consumers at that node to use as much electricity as they want at a uniform price per kWh that is independent of the market structure j. It is reasonable to assume that vertically integrated firms can tailor retail sales to specific nodes n, but are unable to differentiate across market structures j. To accommodate this inflexibility, we assume that firm l sells retail contracts specifying the expected volume $\mathbf{k}_l = (k_{11}, ..., k_{ln}, ..., k_{lN})$ at each of the N nodes, but the actual retail volume k_{lnj} depends on market structure j. Because all consumers in service territory x pay the same retail price, we assume that retail demand is correlated in the sense that $k_{lnj} = \frac{D_{xj}}{D_x} k_{ln} \forall n \in \mathcal{X}_x$.

We let vertically integrated firms face a direct cost of selling retail contracts. Such retail costs may differ across the nodes in the system. In particular, incumbent firms are assumed to have lower costs of selling retail contracts at nodes where they own generation capacity. These firms are typically the incumbent retailer at these locations and therefore have lower customer acquisition and servicing costs than they do at other locations. We therefore assume firm l's retail cost at node n be $\frac{1}{2}(1-\gamma_{ln})\varepsilon k_{ln}^2$, $\varepsilon \geq 0$, where $\gamma_{ln} = 1$ if firm l owns the non-fringe generation unit at n, otherwise $\gamma_{ln} = 0$.

Assuming that independent retailers supply retail contracts at expected unit cost, the inverse retail demand facing the firms with market power in service territory x can be written as

$$r_x(\mathbf{k}) = \sum_{j=1}^{J} \sum_{m=1}^{M_j} \frac{\phi_j D_{xmj}}{D_x} p_{mj}(K_{mj}) \tag{23}$$

as a function of the vertically integrated firms' retail positions $\mathbf{k} = (\mathbf{k}_1, ..., \mathbf{k}_l, ..., \mathbf{k}_L)$.

Equilibrium The expected profit of firm $l \in \mathcal{L}$ can be written as

$$\sum_{j=1}^{J}\sum_{m=1}^{M_j}\phi_j\beta_{lmj}\{\sum_{x=1}^{X}\sum_{n=1}^{N}\delta_{nx}y_{nmj}(p_{xj}-p_{mj})k_{lnj}+(p_{mj}-c_{lmj})q_{lmj}\}-\frac{\varepsilon}{2}\sum_{n=1}^{N}(1-\gamma_{ln})k_{ln}^2.$$

The first term inside the curly brackets is the retail profit gross of direct retail costs, the second is the profit in the short-term market. The variable δ_{nx} is an indicator function that takes the value 1 if node n belongs to service territory x and the value 0 if not. Retail position k_{ln} affects retail profit in market mj if and only if node n is contained in that market. Consequently, we define the indicator function $y_{nmj} = 1$ if $n \in \mathcal{M}_{mj}$, but $y_{nmj} = 0$ otherwise. Firm l earns a profit in short-term market mj if and only if it has a generation unit in that market. To account for ownership, we let $\beta_{lmj} = 1$ if $l \in \mathcal{L}_{mj}$ and $\beta_{lmj} = 0$ otherwise. The final term accounts for the variable cost of firm l associated with the retail position k_{ln} in market n.

The marginal effect

$$\sum_{j=1}^{J}\sum_{m=1}^{M_j}\frac{\phi_j D_{xj}}{D_x}y_{nmj}\{\underbrace{\sum_{h=1}^{N}\sum_{i=1}^{M_j}A_{hmj}\beta_{lij}y_{hij}k_{lhj}p'_{mj}+\beta_{lmj}(p_{xj}-p_{mj})-(1-\gamma_{ln})\varepsilon k_{ln}}_{\text{Marginal retail profit}} \underbrace{-\beta_{lmj}(p_{mj}-c_{lmj})\frac{\partial Q_{-lmj}}{\partial k_{lmj}}}_{\text{Strategic effect in short-term market}}\} \tag{24}$$

on firm l's expected profit of increasing $k_{ln}, n \in \mathcal{X}_x$, has the two well-known terms. The first is the marginal effect on retail profit associated with the increase in retail volume at node n, and the second is the marginal expected effect on competition in the short-term market. Firm l can influence retail prices in service territories $x \neq \hat{x}$ by selling more in the retail market in service territory $\hat{x}$ because x and $\hat{x}$ can have overlapping short-term markets. This cross-territorial effect is captured by the variable

$$A_{hmj} = \sum_{x=1}^{X}\delta_{hx}\frac{D_{xmj}}{D_{xj}} = \sum_{x=1}^{X}\frac{\sum_{n=1}^{N}\delta_{hx}\delta_{nx}y_{nmj}D_{nj}}{\sum_{n=1}^{N}\delta_{nx}D_{nj}} \in [\frac{D_{xmj}}{D_{xj}}, 1].$$

The strategic effect depends on market structure j because of the effect on market concentration in local market mj. Specifically, $-\frac{\partial Q_{-lmj}}{\partial k_{lmj}} = \frac{L_{mj}-1}{L_{mj}+1}$ is larger when there are more local competitors L_{mj}. In other words, firms have a stronger incentive to participate in the retail market if the short-term market is more integrated.

7.1 Spatially independent markets

In this polar case of $X = N$, every single node is a separate service territory. Consequently, retail prices can potentially differ from node to node. Independent retailers' wholesale costs are equal to the LMPs, $p_{xj} = p_{nj} = p_{mj}$, and $A_{\hat{n}mj} = y_{\hat{n}mj}$. Using also $\frac{\partial Q_{-lmj}}{\partial k_{lmj}} = -\frac{L_{mj}-1}{L_{mj}+1}$ and $p'_{mj} = \frac{-b_{mj}}{L_{mj}+1}$ in marginal profit above, we get the first-order-condition

$$\sum_{j=1}^{J}\sum_{m=1}^{M_j}\frac{\phi_j D_{nj}}{D_n}\beta_{lmj}y_{nmj}\{(p^I_{mj}-c_{lmj})\frac{L_{mj}-1}{L_{mj}+1}-\frac{b_{mj}}{L_{mj}+1}k^I_{lmj}\}-(1-\gamma_{ln})\varepsilon k^I_{ln}+\lambda^I_{ln}=0 \quad (25)$$

for firm l's equilibrium retail position k^I_{ln}, where

$$p^I_{mj}-c_{mj}=\frac{a_{mj}-c_{mj}-b_{mj}\sum_{l=1}^{L}\beta_{lmj}k^I_{lmj}}{L_{mj}+1},\ \ k^I_{lmj}=\sum_{n=1}^{N}\frac{D_{nj}}{D_n}y_{nmj}k^I_{ln}. \quad (26)$$

The first term inside the curly brackets of (25) is the strategic effect, and the second is the marginal reduction in retail profit. The third term is firm l's marginal retail cost at node n. The final term, $\lambda^I_{ln} \geq 0$, is a Kuhn-Tucker variable associated with the non-negativity constraint $k^I_{ln} \geq 0$, i.e. $\lambda^I_{ln}k^I_{ln} = 0$.

7.2 One spatially linked retail market

This is the other polar case of a single service territory: $X = 1$. Here, all consumers pay the same retail price no matter where they are located in the network. This is the model applied for instance in Singapore. Using $A_{\hat{n}mj} = \frac{D_{mj}}{D_j}$, we get the first-order condition

$$\begin{aligned}\sum_{j=1}^{J}\sum_{m=1}^{M_j}\frac{\phi_j D_j}{D}y_{nmj}\{\beta_{lmj}[(p^R_{mj}-c_{lmj})\frac{L_{mj}-1}{L_{mj}+1}+p^R_j-p^R_{mj}]\\ -\frac{D_{mj}}{D_j}\frac{b_{mj}}{L_{mj}+1}\sum_{i=1}^{M_j}\beta_{lij}k^R_{lij}\}-(1-\gamma_{ln})\varepsilon k^R_{ln}+\lambda^R_{ln}=0\end{aligned} \quad (27)$$

for firm l's choice $k_{ln}^R \geq 0$, where $\lambda_{ln}^R \geq 0$, $\lambda_{ln}^R k_{ln}^R = 0$ and

$$p_j^R = \sum_{m=1}^{M_j} \frac{D_{mj}}{D_j} p_{mj}^R,\ p_{mj}^R - c_{mj} = \frac{a_{mj} - c_{mj} - b_{mj} \sum_{l=1}^{L} \beta_{lmj} k_{lmj}^R}{L_{mj} + 1},\ k_{lmj}^R = \sum_{n=1}^{N} \frac{D_j}{D} y_{nmj} k_{ln}^R. \quad (28)$$

If we compare the marginal retail price effect in (27) with that in (25), we see that linking retail markets reduces the downward price effect because the LMP in local market mj only constitutes a relatively small share, $\frac{D_{mj}}{D_j}$, of retailers' wholesale electricity cost in the former case. But firm l internalizes a larger share of the retail price effect if the firm owns generation units in more local markets.

7.3 Multiple service territories

This is the intermediary case with more than one service territory, but fewer territories than nodes: $1 < X < N$. Well-known examples of electricity markets with this structure are California ISO and PJM. Firm l's first-order condition equals

$$\begin{aligned} &\sum_{j=1}^{J} \sum_{m=1}^{M_j} \frac{\phi_j D_{xj}}{D_x} y_{nmj} \{\beta_{lmj}[(p_{mj}^X - c_{lmj}) \frac{L_{mj} - 1}{L_{mj} + 1} + p_{xj}^X - p_{mj}^X] \\ &- \frac{b_{mj}}{L_{mj} + 1} \sum_{\hat{n}=1}^{N} \sum_{i=1}^{M_j} A_{\hat{n}mj} \beta_{lij} y_{\hat{n}ij} k_{l\hat{n}j}^X\} - (1 - \gamma_{ln}) \varepsilon k_{ln}^X + \lambda_{ln}^X = 0 \end{aligned} \quad (29)$$

for its choice $k_{ln}^X \geq 0$, $n \in \mathcal{X}_x$ in this most general case, where $\lambda_{ln}^X \geq 0$, $\lambda_{ln}^X k_{ln}^X = 0$ and

$$\begin{aligned} p_{xj}^X = \sum_{m=1}^{M_j} \frac{D_{xmj}}{D_{xj}} p_{mj}^X,\ p_{mj}^X - c_{mj} &= \frac{a_{mj} - c_{mj} - b_{mj} \sum_{l=1}^{L} \beta_{lmj} k_{lmj}^X}{L_{mj} + 1}, \\ k_{lmj}^X = \sum_{n=1}^{N} y_{nmj} k_{lnj}^X,\ k_{lnj}^X &= \sum_{x=1}^{X} \frac{D_{xj}}{D_x} \delta_{nx} k_{ln}^X. \end{aligned} \quad (30)$$

7.4 Simulation results

We now provide simulation results for the above model to see in particular how network constraints and the number of service territories affect market performance. We consider a model with $n = 4$ nodes and $L = 4$ vertically integrated firms with market power. Each controls the output of one non-fringe generation unit: Firm 1 owns the unit at node 1, and so forth.

Table 1: Short-term market configurations in a 4 firm/node network

Quadropoly	Duopolies	Monopoly/triopoly	Monopolies/duopoly	Monopolies
{1,2,3,4}	{1,2}{3,4}	{1}{2,3,4}	{1}{2}{3,4}	{1}{2}{3}{4}
	{1,3}{2,4}	{2}{1,3,4}	{1}{3}{2,4}	
	{1,4}{2,3}	{3}{1,2,4}	{1}{4}{2,3}	
		{4}{1,2,3}	{2}{3}{1,4}	
			{2}{4}{1,3}	
			{3}{4}{1,2}	

All nodes within a {} belong to the same local short-term market.

Table 1 shows the 15 possible market configurations of nodes in this system. By the ownership structure, these local markets differ in terms of market concentration. The first configuration occurs when there are no bottlenecks and is that of a fully integrated quadropoly market. The other polar case is the one in which network constraints are so severe that every node is a local monopoly market. In between are configurations with less severe network bottlenecks. There are three configurations with two dupolies in each configuration, four in which one node is a monopoly and the other three nodes form a triopoly, and six configurations where two nodes each constitute a local monopoly and the remaining two are integrated to form a duopoly. In our benchmark specification, we assume that all market configurations are equally likely and thus occur with probability $\phi_j = \frac{1}{15} \forall j$.

Table 2: Spatially independent markets

Firms\Nodes	1	2	3	4
1	1.36	1.11	1.18	1.11
2	1.13	1.31	1.13	1.06
3	1.18	1.11	1.36	1.11
4	1.13	1.06	1.13	1.31
Short-term prices p_n^I	3.58	3.38	3.58	3.38

$D_1 = D_3 = 10.5$, $D_2 = D_4 = 9.5$, $b = c = 1$, $\varepsilon = 0.1$, $\phi_j = \frac{1}{15} \forall j$.

Table 2 reports equilibrium outcomes for the case of spatially independent markets for a simulation in which demand equals $D_1 = D_3 = 10.5$ in nodes 1, 3 and $D_2 = D_4 = 9.5$ in nodes 2, 4. The other parameters are symmetric across all nodes: $b = c = 1$ for all n, and

$\varepsilon = 0.1$. The first four rows show the retail positions for each of the four firms. The diagonal entries in the matrix thus represent the volume of retail contracts sold at the particular node in which the firm owns generation capacity. Firms sell comparatively more retail contracts in the two larger markets 1 and 3. But the interesting thing to note is that firms sell retail contracts also at off-diagonal nodes, i.e. where they do not own generation capacity. They do so despite there being a marginal retail cost ($\varepsilon > 0$) associated with serving retail demand at off-diagonal nodes, which does not arise at the own node. To firm n, retail positions at nodes n and $\hat{n} \neq n$ are perfect substitutes in market configurations where the two nodes belong to the same local short-term market, but not when the nodes belong to different local markets. In particular, its retail position at node n causes firm n to behave more competitively in the short-term market even when transmission constraints yield firm n a *monopoly position* at node n. The retail position at node $\hat{n}$ has no effect on the short-term price at node n in that case. Firm n thus has an incentive to take retail positions also at nodes other than n to reduce the problem of firm n "competing against itself" whenever it is a local monopoly at node n. The final row in Table 2 shows the average retail price at the different nodes: $p_n^I = \sum_{j=1}^{J} \phi_j p_{nj}^I$. Imperfect competition drives the average short term price above marginal cost $c = 1$, and the average price is higher at the high demand nodes.

Table 3: One spatially linked retail market

Firms\Nodes	1	2	3	4
1	5.11	0.00	0.00	0.00
2	0.00	3.07	0.00	0.00
3	0.00	0.00	5.11	0.00
4	0.00	0.00	0.00	3.07
Short-term prices p_n^R	2.67	2.90	2.67	2.90

$D_1 = D_3 = 10.5$, $D_2 = D_4 = 9.5$, $b = c = 1$, $\varepsilon = 0.1$, $\phi_j = \frac{1}{15} \forall j$.

Table 3 displays the equilibrium outcomes for the same parameter values, but when there is one single spatially linked retail market, so that retail prices always are the same at all four nodes. If we compare results with those of Table 2, we first see that firms with market power now sell relatively more retail contracts at nodes where they own generation capacity, i.e. along the diagonal. This is precisely the pro-competitive effect of linking retail markets that is the key mechanism in this model. Our second finding is that firms now do not take any retail positions at nodes where they do not own generation capacity. Recall from eq. (27)

that the retail markup under market configuration mj is the difference between the volume-weighted average short-term price p_j^R and the local short-term price p_{mj}^R. If mj is a high price location, $p_{mj}^R > p_j^R$, then firms make a retail loss from selling retail contracts at that node under that particular configuration of the short-term market. The incentive to reduce such retail deficits coupled with the marginal cost of offering off-diagonal contracts, leads firms to sell retail contracts exclusively at their own node in the case of one single linked retail market. If we compare prices, we see that increased retail volumes lead to price decreases at all nodes compared to Table 2. The pro-competitive effect in fact is so strong that it drives the equilibrium prices at the high demand nodes below those at the low demand nodes (this effect is even more pronounced for when market asymmetries a larger).

Table 4: Two service territories: $\{1,3\}\{2,4\}$.

Firms\Nodes	1	2	3	4	1	2	3	4
1	5.19	0.62	0.00	0.62	4.15	1.62	0.00	0.39
2	1.29	2.07	1.29	0.00	1.62	4.15	0.39	0.00
3	0.00	0.62	5.19	0.62	0.00	1.53	0.53	1.33
4	1.29	0.00	1.29	2.07	1.53	0.00	1.33	0.53
Short-term prices p_n^X	2.55	3.01	2.55	3.01	2.91	2.91	3.50	3.50

$D_1 = D_3 = 10.5,\ D_2 = D_4 = 9.5\ [D_1 = D_2 = 10.5,\ D_3 = D_4 = 9.5]$, $b = c = 1$, $\varepsilon = 0.1$, $\phi_j = \frac{1}{15} \forall j$.

Table 4 shows the outcomes when the short-term markets at nodes 1 and 3 are linked into one service territory (load aggregation point) and nodes 2 and 4 are linked to form another. It shows results from two simulation exercises. The first four columns have the results of the simulation in which parameter configurations are the same as in the previous two tables. The competitive effect of linking retail markets drives up the diagonal retail positions above those in Table 2. The effects are reinforced in the high demand nodes 1 and 3 and muted in the low demand nodes 2 and 4 compared to Table 3. The marginal spill-over effects into other local markets are smaller when each service territory consists of fewer nodes, but retail prices are closer to the short-term prices, which drives up retail positions and explains the results on the diagonal. Looking at the off-diagonal positions, there is more variation now compared to the other market designs. By taking an off-diagonal position $\hat{n}$ firm 1 achieves the strategic benefit of reducing other firms' output whenever node $\hat{n}$ is integrated with node 1 without limiting its monopoly power when 1 constitutes a local short-term market of its own. But firm 1 also suffers a retail loss at node $\hat{n}$ whenever that node happens to be a high price node, for instance when $\hat{n}$ is a local monopoly market. The first effect dominates for nodes 2 and 4 because they belong to a different service territory than node 1 (similar to the effects in Table 2). The second effect dominates for node 3 because this node belongs to the same service territory as node 1 (similar to the effects in Table 3). Reviewing the price effects, we see that prices go down in the two high demand nodes 1 and 2 compared to the case with spatially independent markets in Table 2, but the effect is muted compared to the case when there is one spatially dependent retail market in Table 3. The average short-term prices are lower at the high demand nodes compared to the two other market designs. Prices in the low demand nodes are between the outcomes with spatially independent markets and a market design with one single retail market (this last result is overturned when demand asymmetries are larger).

The last four columns of Table 4 show the simulation results where we assume that $D_1 = D_2 = 10.5$ and $D_2 = D_4 = 9.5$. The two simulations show the difference between linking nodes with similar demands compared to nodes with dissimilar demands into common service territories. In this second simulation, retail positions are concentrated to the largest node within each service territory. Retail positions at low demand nodes are substantially reduced, even below the level in Table 2. Consumers at high demand nodes still benefit from the two service territories, whereas consumers in low demand nodes would be comparatively better off with a market design under which all markets are spatially independent.

We undertook the above simulations assuming a uniform probability distribution of market configurations. In our second simulation, we analyze the sensitivity of short-term prices to changes in this probability distribution. One can think of this as an analysis of the effect of transmission constraints on prices under the three different market designs. We let the market be fully integrated (quadrupoly) with probability $1 - \phi$. With probability ϕ the system is split up into multiple local markets because of binding transmission constraints. To simplify the exposition, we assume a uniform probability distribution across the remaining configurations, i.e. each occurs with probability $\frac{\phi}{14}$. Also, we let demand be the same at all nodes and equal to $D = 10$. We compare the three different market designs by constructing for each of them a volume- and probability-weighted average short-term price. In the general case of X service territories, this average short-term price is:

$$p^X = \sum_{j=1}^{J} \sum_{m=1}^{M_j} \frac{\phi_j D_{mj}}{D} p_{mj}^X.$$

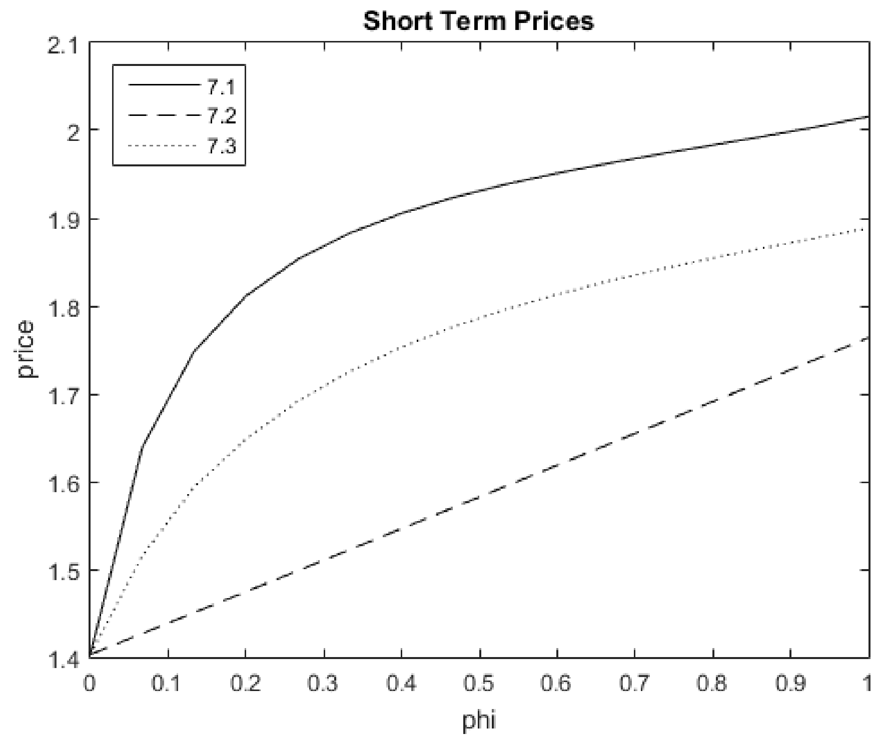

Figure 2: Average short-term prices given spatial independence (7.1), one single retail market (7.2), two service territories (7.3).

Figure 2 plots the average short-term prices for $\phi \in [0, 1]$. The fully drawn line is p^I, the dotted line is p^X, and the dashed line shows p^R. For ϕ close to zero, short-term prices are nearly always the same at all nodes. Linking electricity markets has little effect on competition in this case, and prices therefore are similar under all three market designs. Market concentration increases in the short-term market as transmission constraints become more important (ϕ increases). Prices then go up under all three designs. But price increases are larger in magnitude when markets are spatially independent compared to the other designs, and these marginal effects are particularly strong for small ϕ. Linking retail markets therefore limits the anti-competitive effects of transmission constraints already when congestion occurs rather infrequently. For instance, when there is a 20 percent chance of congestion ($\phi = 0.2$) prices are $\frac{p^I - p^R}{p^I} = 19$ percent lower when there is one spatially dependent market relative to the case with spatial independence. The corresponding price reduction is $\frac{p^I - p^X}{p^I} = 9$ percent when there are two service territories. For $\phi = 1$, the numbers are $\frac{p^I - p^R}{p^I} = 13$ percent and $\frac{p^I - p^X}{p^I} = 6$ percent. Of course, the absolute price reduction is larger when ϕ is larger because p^I then is larger.

Insights from the simulation exercise Linking retail markets by collecting nodes in one or several service territories, or load aggregation points, has substantial and robust effects on competition in a model with transmission constraints and asymmetric demand. In our four-node simulation exercise, short-term prices decrease between 13 and 19 percent when the retail market consists one single service territory compared to the case of spatially independent markets. The pro-competitive effect is weaker when there is more than one service territory, but short-term prices still decrease between 6 and 9 percent when there are two service territories compared to the benchmark case when there are as many local retail markets (four) as there are nodes in the system (four).

The pro-competitive effects of linking local markets are substantially stronger stronger if the nodes within each service territory have similar characteristics compared to the case when they are more dissimilar.

8 Concluding policy discussion

A key problem with market performance in restructured electricity markets is the high degree of market concentration that sometimes arises when transmission constraints divide a single market into smaller local markets with only a few producers in each. Increasing the share of intermittent renewable generation resources typically increases the frequency and number of transmission and other operating constraints that bind, which can increase the ability of strategic suppliers to exercise unilateral market power. Improving competition in each local market through entry or market integration is problematic in many electricity markets because of economic and/or political barriers to large supplier entry and transmission network investments to expand the size of the geographic market.

We propose a solution to the problem of market power that neither involves supplier entry nor network investment. A market design where generation units are paid the short-term locational marginal price (LMP), but all consumers pay the same wholesale price for their electricity equal to the quantity-weighted average of LMPs at all load withdrawal nodes, can substantially increase short-term market performance compared to the case when local markets are independent.

This market rule is also likely to facilitate retailer entry into more local markets by vertically integrated retailers because it mitigates a major source of risk they face in entering a local market where they do not own generation units: spatial price risk between where they own generation units and this local market. This spatial price risk has led many vertically integrated firms to focus their retailing efforts on the local markets where they own generation units to avoid such risk. Moreover, an effective entry deterrence strategy by vertically-integrated retailers with generation units in the same local market as their retail customers is to use their ability to exercise unilateral market power to spike the local wholesale price and effectively eliminate any retail profit margin a new entrant without local generation capacity could earn from selling retail electricity.[19] Requiring all retailers to purchase the wholesale electricity sold to final consumers at a quantity-weighted average of all LMPs significantly limits the spatial price risk any supplier faces from entering any local market, which should increase the competitiveness of retail markets in particular.

Our findings are highly relevant to the debate about future designs for wholesale electricity markets in Europe. Local imbalances in the supply and demand of electricity owing to increasing shares of intermittent electricity production have increased concerns about the security of electricity supply in situations with substantial shortfalls of renewable electricity production. Such security of supply problems would be reduced if prices are set based to local supply and demand conditions and all relevant constraints on operating generation units and the transmission network were explicitly priced as in a system with locational marginal prices (LMPs), as first discussed in Bohn et al. (1984).

As noted in the Introduction, charging consumers different local prices has so far been politically difficult, particularly in Europe. Critics argue that it is unreasonable for some consumers to pay more for electricity than others just because the former happen to live at a location with an "under-supply" of electricity. Such unfairness arguments received a lot of public attention following the division of the Swedish day-ahead market into four price areas in 2011. Previously, Sweden had constituted a single price area. Consumers in the southernmost price area were particularly upset because they would now have to pay a systematically higher price than consumers in the other three price areas. A superior solution from a political and market efficiency viewpoint based on our results, would have been to introduce the four price areas for producers, but to maintain a single price area across Sweden for consumers based on the quantity-weighted average of prices across all locations.

Our results argue that the potential economic gains to retail electricity consumers are substantial from European Union countries transitioning to an LMP market design and implementing our regulatory mandate for the purchase price of wholesale energy. According

[19]Consistent with such entry deterrence incentives, Wolak (2009) presents empirical evidence demonstrating that over the sample period he studies, the four large vertically integrated retailers in the New Zealand wholesale electricity market concentrated their retailing activities in the regions where they owned generation units.

to Eurostat, annual electricity consumption in the EU countries in 2016 was 2,786,137,000 MWh. If the shift to LMP pricing and implementing our single wholesale purchase price reduces average wholesale prices by 2 Euros/MWh, which is at the low end of the values computed in Section 7, this implies an annual savings to EU electricity consumers of over 5 billion Euros.

Appendix

A.1 Proof of Lemma 3

Characterization Substitute (12) into (10), use the marginal effects $p'_m = -\frac{b_m}{L_m+1}$ and $\frac{\partial Q_{-lm}}{\partial k_{lm}} = -\frac{L_m-1}{L_m+1}$, sum up the first-order conditions over all L_m firms in market m and divide through by L_m to get the aggregate equilibrium condition im market m:

$$-\frac{1}{L_m}\frac{D_m}{D}\frac{b_m K_m^R}{L_m+1} + (\frac{L_m-1}{L_m+1} - 1)(p_m^R - c_m) + r^R - c_m = 0.$$

Add and subtract $\frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1}$ from the left-hand side above and use (3) to get

$$(\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1} - 1)(p_m^R - c_m) + r^R - c_m - \frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1} = 0. \quad (31)$$

Subtract the first-order condition for market m from the one for market i to get

$$p_i^R - c_i = \frac{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1} - 1}{\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1} - 1}(p_m^R - c_m) + \frac{c_i - c_m + \frac{1}{L_i}\frac{D_i}{D}\frac{a_i-c_i}{L_i+1} - \frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1}}{\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1} - 1}.$$

Multiply through by $\frac{D_i}{D}$ and sum up over all $i \in \mathcal{M}$ markets:

$$\begin{aligned} r^R - c_m = & \sum_{i=1}^M \frac{D_i}{D}\frac{\frac{1}{L_m}\frac{D_m}{D} + \frac{L_m-1}{L_m+1} - 1}{\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1} - 1}(p_m^R - c_m) \\ + & \sum_{i=1}^M \frac{D_i}{D}\frac{(\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1})(c_i - c_m) + \frac{1}{L_i}\frac{D_i}{D}\frac{a_i-c_i}{L_i+1} - \frac{1}{L_m}\frac{D_m}{D}\frac{a_m-c_m}{L_m+1}}{\frac{1}{L_i}\frac{D_i}{D} + \frac{L_i-1}{L_i+1} - 1}. \end{aligned}$$

Substitute for $r^R - c_m$ in eq. (31) and solve to get the expressions in Lemma 3.

Existence Differentiate firm l's marginal profit function (10) with respect to k_{lm}, using the expressions for p'_m and $\frac{\partial Q_{-lm}}{\partial k_{lm}}$ to get the second-order condition:

$$-\frac{2b_m}{L_m+1}(\frac{D_m}{D}-\frac{1}{L_m+1}) \leq 0.$$

This condition is satisfied for all firms if and only if the conditions of Lemma 3 are satisfied for all markets m.■

A.2 Proof of Lemma 4

Characterization Substitute (12) into (14) and use the marginal effects $p'(K_m) = -\frac{b}{S+1}$ and $\frac{\partial Q_{-lm}}{\partial k_{lm}} = -\frac{S-1}{S+1}$ to get the first-order condition

$$\frac{S-1}{S+1}(p_m^R - c) - \frac{b}{S+1}\frac{D-K_m^R}{MD-K^R}\sum_{i=1}^{M}\beta_{li}k_{li}^R = (1+\frac{\sum_{i=1}^M \beta_{li}k_{li}^R}{MD-K^R}(p_m^R - r^R),$$

where k_{li}^R is the equilibrium retail position of vertically integrated producer l in local market i, K_m^R is equilibrium retail supply in local market m aggregated across all L vertically integrated producers, $K^R = \sum_{m=1}^M K_m^R$ is the supply of retail contracts aggregated across all L vertically integrated firms and M local markets, and r^R is the equilibrium retail price. Multiply both sides by β_{ln}, sum up over all L firms and use $p_m^R - c = \frac{a-c-bK_m^R}{S+1}$ to get the aggregate first-order condition in market m

$$\begin{aligned}
&(S\frac{S-1}{S+1} + \frac{D-K_m^R}{MD-K^R}\frac{\sum_{l=1}^L\sum_{i=1}^M \beta_{lm}\beta_{li}k_{li}^R}{K_m^R})(p_m^R - c)\\
&= \frac{a-c}{S+1}\frac{D-K_m^R}{MD-K^R}\frac{\sum_{l=1}^L\sum_{i=1}^M \beta_{li}\beta_{ln}k_{li}^R}{K_m^R} + (S+\frac{\sum_{l=1}^L\sum_{i=1}^M \beta_{li}\beta_{lm}k_{li}^R}{MD-K^R})(p_n^R - r^R).
\end{aligned}$$

Assume that the equilibrium is symmetric across firms and markets, i.e. $k_{lm}^R = k^R = \frac{K^R}{MS}$ for all l and m. Then $p_m^R = r^R = p^R$, so we get

$$(S\frac{S-1}{S+1} + \frac{1}{MS}\sum_{l=1}^L\sum_{i=1}^M \beta_{lm}\beta_{li})(p^R - c) = \frac{a-c}{S+1}\frac{1}{MS}\sum_{l=1}^L\sum_{i=1}^M \beta_{li}\beta_{lm}.$$

Use $\sum_{m=1}^M \beta_{lm} = \bar{M}$ and $\sum_{l=1}^L \beta_{lm} = S$ to get (15) after simplification.

Necessity Let all firms except possibly l be at their symmetric equilibrium retail positions k^R. Then, Sk^R is the volume of retail contracts sold by vertically integrated producers in all local markets where l has no production capacity, $(S-1)k^R + k_{lm}$ is the corresponding volume of retail contracts in a local market m where l does have production capacity, and $(MS - \bar{M})k^R + \bar{M}\bar{k}_l$ is the total volume of retail contracts sold by all vertically integrated producers in all local markets, where $\bar{k}_l = \frac{1}{\bar{M}}\sum_{l\in M_l} k_{lm}$ is the average retail position of firm l in the subset M_l of markets in which it is present. Consider a unilateral deviation by l from k^R to $k_{lm} = 0$ in all local markets where it has production capacity. This deviation profit equals

$$\pi(0) = \bar{M}(p((S-1)k^R) - c)q(0, (S-1)k^R) = \frac{\bar{M}}{b}(p((S-1)k^R) - c)^2,$$

whereas the profit at the proposed equilibrium equals

$$\pi^R = \bar{M}(p^R - c)(q(k^R, (S-1)k^R) - k^R) + (p^R - c)\bar{M}k^R = \frac{\bar{M}}{b}(p^R - c)^2 + (p^R - c)\bar{M}k^R,$$

where we have used $b(q(k_{lm}, K_{-lm}) - k_{lm}) = p(K_m) - c$. The net profitability

$$\pi(0) - \pi^R = b\hat{M}\frac{M - \hat{M}(S+1)}{M}(\frac{k^R}{S+1})^2 \tag{32}$$

of this deviation is strictly positive if $S+1 < \frac{M}{\hat{M}}$. Hence, k^R can be sustained as a symmetric equilibrium only if $S+1 \geq \frac{M}{\hat{M}}$.

Sufficiency under symmetric deviations Consider a symmetric deviation by l to k in all local markets in which it has production capacity. The deviation profit $\tilde{\pi}(k)$ of firm l can be written as

$$\frac{\tilde{\pi}(k)}{\bar{M}} = (\tilde{r}(k) - p((S-1)k^R + k))k + (p((S-1)k^R + k) - c)q(k, (S-1)k^R),$$

in this case, with the retail price given by

$$\tilde{r}(k) = \frac{\bar{M}}{M}p((S-1)k^R + k) + \frac{M - \bar{M}}{M}p^R.$$

The second-order condition

$$\tilde{\pi}''(k) = -\frac{2\bar{M}b}{S+1}\frac{\bar{M}(S+1) - M}{M(S+1)} = -\frac{2\bar{M}b}{S+1}[\frac{S}{S+1} - \frac{M - \bar{M}}{M}] \leq 0$$

is satisfied if $S+1 \geq \frac{M}{\bar{M}}$.

Asymmetric deviations Assume that $\bar{M} \geq 2$ and consider an asymmetric deviation by l from k^R in all markets to $k^R+\varepsilon$ in one local market and $k^R - \frac{\varepsilon}{\bar{M}-1}$ in all other local markets. This deviation profit equals

$$\begin{aligned}\tilde{\pi}(\varepsilon) &= \tilde{r}(\varepsilon)\bar{M}k^R - p(K^R+\varepsilon)(k^R+\varepsilon) - p(K^R - \tfrac{\varepsilon}{\bar{M}-1})((\bar{M}-1)k^R - \varepsilon) \\ &\quad + (p(K^R+\varepsilon)-c)q(k^R+\varepsilon,(S-1)k^R) + (\bar{M}-1)(p(K^R-\tfrac{\varepsilon}{\bar{M}-1})-c)q(k^R-\tfrac{\varepsilon}{\bar{M}-1},(S-1)k^R),\end{aligned}$$

where

$$\tilde{r}(\varepsilon) = \frac{1}{M}p(K^R+\varepsilon) + \frac{\bar{M}-1}{M}p(K^R - \frac{\varepsilon}{\bar{M}-1}) + \frac{M-\bar{M}}{M}p^R = r^R$$

is the corresponding retail price. The marginal deviation profit becomes

$$\tilde{\pi}'(\varepsilon) = \frac{2b}{S+1}\frac{\bar{M}}{\bar{M}-1}\varepsilon$$

after simplification, which implies that $\tilde{\pi}(\varepsilon)$ is strictly convex in ε and reaches its global minimum at $\tilde{\pi}(0) = \pi^R$. Hence, there always exists a strictly profitable *asymmetric* deviation from the symmetric equilibrium candidate k^R.■

A.3 Sufficient existence conditions for Proposition 4

Consider a region consisting of 2 local markets. Assume that location 1 is the high price market when these are spatially independent, $p_1^I > p_2^I$, as well as linked, $p_1^{RF} > p_2^{RF}$. Let all L_m firms in local market m have the same marginal production cost c_m, with $c_1 \geq c_2$. Assume also that demand is higher in region 1 than 2, $D_1 > D_2$. Assume finally that the number of firms in each market is large in the sense that:

$$L_1 > \frac{D_2}{D_1},\ \frac{L_2^2+1}{2L_2} > \frac{D_1}{D_2}. \tag{33}$$

By the assumption that 1 is the high price market, Proposition 4 yields $p_1^{RF} = p_1^I$ and

$$p_2^{RF} - c_2 = \frac{\frac{1}{L_2}\frac{L_2^2+1}{L_2+1}\frac{D_2}{D}(p_2^I - c_2) - \frac{D_1}{D}(p_1^{RF} - p_2^{RF})}{\frac{L_2-1}{L_2+1} + \frac{1}{L_2}\frac{D_2}{D}}, \tag{34}$$

where $D = D_1 + D_2$ is total demand in the region. By manipulating the two equilibrium conditions for p_1^{RF} and p_2^{RF} we get:

$$p_1^{RF} - p_2^{RF} = \frac{\frac{L_2-1}{L_2+1}\frac{D_1}{D}(p_2^I - c_2) + (\frac{L_2-1}{L_2+1} + \frac{1}{L_2}\frac{D_2}{D})(p_1^I - p_2^I)}{\frac{L_2-1}{L_2+1} + \frac{1}{L_2}\frac{D_2}{D} - \frac{D_1}{D}} > 0. \tag{35}$$

The numerator on the right-hand side of (35) is strictly positive by $p_2^I > c_2$ and the assumption that $p_1^I > p_2^I$. The denominator is strictly positive by (33). Hence, $p_1^{RF} > p_2^{RF}$ is consistent.

Turning to the existence of equilibrium, (33) implies that the local second-derivatives

$$\frac{-2b_m L_m}{(L_m+1)^2} \text{ and } \frac{-2b_m}{L_m+1}(\frac{D_m}{D} - \frac{1}{L_m+1})$$

of the profit function with respect to the own retail position are negative independently of whether location m is a high price or low price market.

Consider first the incentive for firm l in location 2 to deviate from k_2^{RF}, assuming all other firms play the proposed equilibrium strategies. Let $\tilde{k}_2$ solve $p_2((L_2-1)k_2^{RF} + \tilde{k}_2) = p_1^{RF}$:

$$\tilde{k}_2 = k_2^{RF} - \frac{L_2+1}{b_2}(p_1^{RF} - p_2^{RF}).$$

By local strict concavity of the profit function, k_2^{RF} is l's best reply in the domain $k_{l2} \geq \tilde{k}_2$. By the same token, marginal profit is positive in the entire domain $[0, \tilde{k}_2)$ if and only if marginal profit evaluated at $\tilde{k}_2$ is non-negative:

$$\frac{L_2-1}{L_2+1}(p_1^{RF} - c_2) - \frac{b_2\tilde{k}_2}{L_2+1} = \frac{1}{L_2}\frac{L_2^2+1}{L_2+1}(p_2^{RF} - c_2) + \frac{2L_2}{L_2+1}(p_1^{RF} - p_2^{RF}) - \frac{1}{L_2}\frac{a_2 - c_2}{L_2+1} \geq 0.$$

Substituting in the equilibrium expression (34), we get the condition:

$$p_1^{RF} - p_2^{RF} \geq \frac{\frac{L_2-1}{L_2+1}D_1(p_2^I - c_2)}{\frac{(L_2-1)^2}{L_2^2+1}D_1 + \frac{2}{L_2+1}(L_2 D_2 - D_1)}.$$

By virtue of (35), we arrive at the condition:

$$p_1^I - p_2^I \geq -\frac{\frac{1}{L_2}\frac{(L_2-1)^2}{L_2+1}D_1[L_2(L_2-1)\frac{D_1}{D} + (L_2^2+1)\frac{D_2}{D}](p_2^I - c_2)}{((L_2-1)^2 D_1 + 2\frac{L_2^2+1}{L_2+1}(L_2 D_2 - D_1))(\frac{L_2-1}{L_2+1} + \frac{1}{L_2}\frac{D_2}{D})},$$

which holds under the assumptions of this appendix. Hence, k_2^{RF} is a global best-reply for a firm with production in location 2.

Consider next the incentives of firm l in location 1 to deviate from k_1^F. Let $\tilde{k}_1$ be the solution to $p_1((L-1)k_1^{RF}+\tilde{k}_1)=p_2^{RF}$:

$$\tilde{k}_1 = k_1^{RF} + \frac{L_1+1}{b_1}(p_1^{RF}-p_2^{RF}).$$

First, k_1^{RF} is $l1$'s best reply in the domain $k_{l1} \leq \tilde{k}_1$ by local strict concavity of the profit function. Suppose now that $k_{l1} > \tilde{k}_1$, so that 1 becomes the low-price location. The profit of firm l equals

$$\frac{D_2}{D}(p_2^{RF} - p_1((L_1-1)k_1^{RF}+k_{l1}))k_{l1} + (p_1((L_1-1)k_1^{RF}+k_{l1}) - c_1)q_1(k_{l1},(L_1-1)k_1^{RF})$$

in this case, and the corresponding marginal profit is

$$\frac{D_1}{D}\frac{-b_1 k_{l1}}{L_1+1} + \frac{D_2}{D}(p_2^{RF} - p_1((L_1-1)k_1^{RF}+k_{l1})) + \frac{L_1-1}{L_1+1}(p_1((L_1-1)k_1^{RF}+k_{l1}) - c_1).$$

Evaluated at $k_{l1} = \tilde{k}_1$, this marginal profit becomes

$$\frac{L_1-1}{L_1+1}\frac{D_2}{D}(p_1^I - c_1) - (\frac{D_1}{D} + \frac{L_1-1}{L_1+1})(p_1^{RF}-p_2^{RF}).$$

By local strict concavity, marginal profit is strictly negative for all $k_{l1} > \tilde{k}_1$ if and only if

$$p_1^{RF} - p_2^{RF} \geq \frac{\frac{L_1-1}{L_1+1}\frac{D_2}{D}}{\frac{D_1}{D}+\frac{L_1-1}{L_1+1}}(p_1^I - c_1).$$

By substituting in (35), we can write this condition as

$$p_1^I - p_2^I + L_2\frac{L_2-1}{L_2^2+1}\left[\frac{\frac{2L_1}{L_1+1}\frac{D_1-D_2}{D}}{\frac{L_1-1}{L_1+1}+\frac{D_1}{D}}(p_1^I-c_1) + \frac{D_1}{D_2}(c_1-c_2)\right]$$
$$\geq -\frac{\frac{L_1-1}{L_1+1}\frac{D_1 D_2}{D^2} + \frac{1}{L_2}(\frac{2}{L_1+1}+L_2-1)(\frac{D_2}{D})^2}{\frac{1}{L_2}\frac{L_2^2+1}{L_2+1}(\frac{L_1-1}{L_1+1}+\frac{D_1}{D})\frac{D_2}{D}}(p_1^I - c_1).$$

The right-hand side of this expression is strictly negative. The expression on the left-hand side is strictly positive by $p_1^I > p_2^I$, $D_1 > D_2$ and $c_1 \geq c_2$. In particular, the last two conditions can be relaxed. Hence, k_1^{RF} represents a best-reply for producers active in location 1, which establishes existence.

A.4 Effects of fixed-price retail contracts on production costs

The total production cost in period m equals

$$TC_m = \sum_{l=1}^{L} c_{lm} q_{lm}(k_{lm}, K_{-lm}) + \frac{b_m}{2}(D_m - Q_m(K_m))^2$$

as a function of the retail positions in period m. By the definition of $q_{lm}(k_{lm}, K_{-lm})$ in eq. (4) and because $p_m(K_m) = a_m - b_m Q_m(K_m)$, we can write this cost as:

$$TC_m = \sum_{l=1}^{L}(c_{lm} - c_m)k_{lm} + \frac{1}{2b_m}(p_m(K_m) - c_m)^2 + \frac{c_m}{b_m}(a_m - \frac{c_m}{2}) - \frac{1}{b_m}\sum_{l=1}^{L}(c_{lm} - c_m)^2.$$

The difference in production cost under short-term contracting compared to the case when firms sign fixed-price long-term retail contracts therefore equals

$$TC_m^I - TC_m^T = \sum_{l=1}^{L}(c_{lm} - c_m)(k_{lm}^I - k_{lm}^T) + \frac{1}{2b_m}[(p_m^I - c_m)^2 - (p_m^T - c_m)^2].$$

It is straightforward to verify that

$$k_{lm}^I = \frac{K_m^I}{L} + (L_m - 1)\frac{c_m - c_{lm}}{b_m} \text{ and } k_{lm}^T = \frac{K_m^T}{L} + \frac{D}{D_m}(L_m - 1)\frac{c_m - c_{lm}}{b_m}$$

By substituting these two expressions into the cost difference above and aggregating over all periods, we can write the total cost difference as

$$\sum_{m=1}^{M}(TC_m^I - TC_m^T) = \Delta C + \Delta P.$$

The term

$$\Delta C = (L - 1)\sum_{m=1}^{M}\frac{D - D_m}{a_m}\sum_{l=1}^{L}(c_{lm} - c_m)^2$$

is the part of the cost difference that can be traced down to *within-market* differences between firms with market power. It is strictly positive unless $c_{lm} = c_m$ for all L firms in every period m. The term

$$\Delta P = \sum_{m=1}^{M}\frac{1}{2b_m}[(p_m^I - c_m)^2 - (p_m^T - c_m)^2]$$

is the part of the cost differential that can be attributed to *between-market* differences. As the Lemma below shows, ΔP is strictly negative unless $\frac{a_m}{c_m}$ remains constant in all contract periods m. The total cost difference can be positive or negative, depending on which of ΔC and ΔP dominates. Long-term contracting leads to a reduction in total production costs if the main source of heterogeneity is differences between firms. Short-term contracting is more efficient if instead most of the heterogeneity stems from differences between markets.

Lemma 6 *The difference in the variability of price-cost margins under short-term versus long-term contracting, as defined by ΔP, is strictly negative unless $\frac{a_m}{c_m} = \frac{a_t}{c_t}$ in all contract periods m, t, in which case $\Delta P = 0$.*

Proof: Let $\hat{M}$ be the subset of periods with maximal ratio $\frac{a_m}{c_m}$. We first show that a marginal reduction in demand $-D_m$ in each of the periods in $\hat{M}$ causes ΔP to increase. Substitute in

$$p_m^I - c_m = \frac{b_m D_m - c_m - b_m K_m^I}{L+1} = \frac{b_m D_m - c_m}{L^2+1},\ p_m^T - c_m = \frac{b_m D_m - c_m - b_m D_m \frac{MK^T}{D}}{L+1} \quad (36)$$

into ΔP and perform a total differentiation:

$$d\Delta P = \sum_{m\in\hat{M}} [(1 - \frac{MK^T}{D})\frac{p_m^T - c_m}{L+1} - \frac{p_m^I - c_m}{L^2+1}](-dD_m) + \frac{1}{L+1}\sum_{t=1}^{M} D_t(p_t^T - c_t) d\frac{MK^T}{D}.$$

Substitute in

$$d\frac{MK^T}{D} = -\frac{\sum_{m\in\hat{M}} a_m(\frac{K_m^I}{D_m} + L\frac{L-1}{L^2+1} - 2\frac{MK^T}{D})dD_m}{\sum_{t=1}^{M} a_t D_t}$$

into $d\Delta P$ to obtain:

$$d\Delta P = \sum_{m\in\hat{M}} b_m[(1 - \frac{MK^T}{D})\frac{p_m^T - c_m}{L+1} - \frac{p_m^I - c_m}{L^2+1} - a_m\frac{L}{L+1}(\frac{1}{L}\frac{K_m^I}{D_m} + \frac{L-1}{L^2+1} - 2\frac{MK^T}{DL})\frac{\sum_{t=1}^{M} D_t(p_t^T - c_t)}{\sum_{t=1}^{M} a_t D_t}](-dD_m).$$

By utilizing the expressions in (36), we can rewrite the above expression as:

$$d\Delta P = \frac{L}{L+1}\sum_{m\in\hat{M}} a_m \left[\frac{L}{L+1}(\frac{L-1}{L^2+1} - \frac{MK^T}{DL})(\frac{L-1}{L^2+1} - \frac{MK^T}{DL} - \frac{1}{L}\frac{c_m}{a_m})\right.$$
$$\left. + (\frac{1}{L}\frac{K_m^I}{D_m} + \frac{L-1}{L^2+1} - 2\frac{MK^T}{DL})\left(\frac{1}{L^2+1} - \frac{\sum_{t=1}^M D_t(p_t^T - c_t)}{\sum_{t=1}^M a_t D_t}\right)\right](-dD_m).$$

Finally, use

$$1 - \frac{L^2+1}{L-1}\frac{MK^T}{DL} = 1 - (L^2+1)\frac{\sum_{t=1}^M D_t(p_t^T - c_t)}{\sum_{t=1}^M a_t D_t} = \frac{\sum_{t=1}^M c_t D_t}{\sum_{t=1}^M a_t D_t}$$

and

$$\frac{1}{L}\frac{K_m^I}{D_m} - \frac{MK^T}{DL} = \frac{L-1}{L^2+1}\frac{\sum_{t=1}^M D_t(\frac{a_m}{c_m} - \frac{a_t}{c_t})\frac{c_m c_t}{a_m}}{\sum_{t=1}^M a_t D_t}$$

to obtain

$$d\Delta P = \frac{L(L-1)}{(L+1)^2}\frac{2L^2+L+1}{(L^2+1)^2}\frac{\sum_{t=1}^M c_t D_t}{\sum_{t=1}^M a_t D_t}\frac{\sum_{t=1}^M \sum_{m\in\hat{M}} c_m c_t D_t(\frac{a_m}{c_m} - \frac{a_t}{c_t})(-dD_m)}{\sum_{t=1}^M a_t D_t}.$$

By the construction of $\hat{M}$, $\frac{a_m}{c_m} = \frac{a_t}{c_t}$ for all $m, t \in \hat{M}$ and $\frac{a_m}{c_m} > \frac{a_t}{c_t}$ for all $m \in \hat{M}$ and $t \notin \hat{M}$. This implies $d\Delta P < 0$, unless $\frac{a_m}{c_m}$ is the same in all M periods, in which case $d\Delta P = 0$. It follows that any shaving of the peak periods (defined in terms of those with maximal $\frac{a_m}{c_m}$) by $-dD_m$ causes ΔP to increase. Hence, ΔP is maximized when $\frac{a_m}{c_m} = \lambda$ in every period m. If this is the case, then

$$p_m^I - c_m = \frac{a_m - c_m}{L^2+1} = \frac{\lambda - 1}{L^2+1}c_m$$

and

$$p_m^T - c_m = \frac{a_m - c_m - b_m\frac{D_m}{D}MK^T}{L+1} = \frac{\lambda - 1 - \lambda\frac{MK^T}{D}}{L+1}c_m = \frac{\lambda-1}{L^2+1}c_m,$$

where we have substituted in

$$K^T = L\frac{L-1}{L^2+1}\frac{\frac{1}{M}\sum_{t=1}^M D_t(\lambda-1)c_t}{\sum_{t=1}^M \lambda c_t \frac{D_t}{D}} = \frac{D}{M}L\frac{L-1}{L^2+1}\frac{\lambda-1}{\lambda}$$

into $p_m^T - c_m$ and simplified. As spot prices here are the same under long-term and short-term contracting in all periods, it follows that $\Delta P = 0$ if $\frac{a_m}{c_m} = \lambda$ in every period m. Hence, $\Delta P \leq 0$, with strict inequality in the generic case $\frac{a_m}{c_m} \neq \frac{a_t}{c_t}$ for some periods m and t.■

A.5 A model of retail competition in prices

Let retail competition be in prices instead of quantities as in the main text. The purpose is to show that the mode of retail competition does not affect the main result of a pro-competitive effect of linking retail markets across space. Let each vertically integrated firm be active in one single local market. Assume that all local markets are symmetric.

Let the retail demand facing firm l in market m be

$$k_{lm} = k(r_{lm}, r_{-lm}) = \alpha - r_{lm} + (1-\sigma)r_{-lm}$$

depending on the firm's retail price r_{lm} and the average price of all vertically integrated retailers other than l in local market m: $r_{-lm} = \frac{1}{L-1}\sum_{i \neq l} r_{im}$. The degree of horizontal differentiation is increasing in the parameter $\sigma \in [0,1]$, with $\sigma = 1$ signifying monopoly prices. Alternatively, one can write retail demand as

$$k(r_{lm}, r_m) = \alpha - (1 + \frac{1-\sigma}{L-1})r_{lm} + \frac{L}{L-1}(1-\sigma)r_m,$$

where $r_m = \frac{1}{L}\sum_{i=1}^{L} r_{im}$ is the average retail price of *all* vertically integrated producers in market m, including l. Summing up over all L firms yields the total volume of retail sales $K_m = L\alpha - L\sigma r_m$ by the L vertically integrated producers in market m.

Assume that the M retail markets are linked and that retailers compete in markups θ_{lm} over the quantity-weighed average price $r(\mathbf{K})$ defined in (8): $r_{lm} = \theta_{lm} r(\mathbf{K})$. Independent retail markets is a special case where $M = 1$. Let $\theta_m = \frac{1}{L}\sum_{l=1}^{L} \theta_{lm}$ be the average retail markup in local market m. The M demand functions

$$K_m = L\alpha - L\sigma r(\mathbf{K})\theta_m$$

implicitly define total retail demand in market m, $K_m(\boldsymbol{\theta})$, depending on the vector $\boldsymbol{\theta} = (\theta_1, .., \theta_m, .., \theta_M)$ of average retail markups in all M markets,

$$k(\theta_{lm}, \boldsymbol{\theta}) = \alpha + r(\mathbf{K}(\boldsymbol{\theta}))(\frac{L}{L-1}(1-\sigma)\theta_m - (1 + \frac{1-\sigma}{L-1})\theta_{lm}),$$

defines the individual retail demand facing firm l in market m, and

$$K_{-lm}(\theta_{lm}, \boldsymbol{\theta}) = K_m(\boldsymbol{\theta}) - k(\theta_{lm}, \boldsymbol{\theta})$$

is the demand of the other $L-1$ firms in market m.

A total differentiation of retail demand yields

$$\frac{\partial K_m}{\partial \theta_m}\frac{1}{L} = -(1 - \frac{L\sigma\theta_m \frac{\partial r(\mathbf{K})}{\partial K_m}}{1+\sum_{i=1}^{M} L\sigma\theta_i \frac{\partial r(\mathbf{K})}{\partial K_i}})\sigma\theta_m r(\mathbf{K})$$

and

$$\frac{\partial K_n}{\partial \theta_m}\frac{1}{L} = \frac{L\sigma\theta_n \frac{\partial r(\mathbf{K})}{\partial K_m}}{1+\sum_{i=1}^{M} L\sigma\theta_i \frac{\partial r(\mathbf{K})}{\partial K_i}}\sigma\theta_m r(\mathbf{K}) \text{ for } n \neq m.$$

For future reference:

$$\sum_{n=1}^{M} \frac{\partial r(\mathbf{K})}{\partial K_n}\frac{\partial K_n}{\partial \theta_m}\frac{1}{L} = \frac{-\frac{\partial r(\mathbf{K})}{\partial K_m}\sigma\theta_m r(\mathbf{K})}{1+L\sigma\sum_{i=1}^{M}\theta_i \frac{\partial r(\mathbf{K})}{\partial K_i}}$$

is the marginal effect on the linked retail price of an increase in θ_{lm}.

Firm l sets its markup θ_{lm} to maximize the total profit

$$(\theta_{lm} r(\mathbf{K}(\boldsymbol{\theta})) - p(K_m(\boldsymbol{\theta})))k(\theta_{lm},\boldsymbol{\theta}) + (p(K_m(\boldsymbol{\theta})) - c)q(k(\theta_{lm},\boldsymbol{\theta}), K_{-lm}(\theta_{lm},\boldsymbol{\theta}))$$

taking the markups of all other firms in market m as given. The profit maximizing markup is a trade-off between the strategic spot-market effect and the marginal effect on retail profit:

$$\begin{aligned}&(p(K_m)-c)[\frac{\partial q}{\partial k_{lm}}(\frac{\partial k}{\partial \theta_{lm}} + \frac{\partial k}{\partial \theta_m}\frac{1}{L}) + \frac{\partial q}{\partial K_{-lm}}(\frac{\partial K_{-lm}}{\partial \theta_{lm}} + \frac{\partial K_{-lm}}{\partial \theta_m}\frac{1}{L}) - Q'(K_m)\frac{\partial K_m}{\partial \theta_m}\frac{1}{L}]\\ &+ (r + \theta_{lm}\sum_{n=1}^{M}\frac{\partial r}{\partial K_n}\frac{\partial K_n}{\partial \theta_m}\frac{1}{L})k_{lm} + (\theta_{lm} r - p(K_m))(\frac{\partial k}{\partial \theta_{lm}} + \frac{\partial k}{\partial \theta_m}\frac{1}{L})\end{aligned}$$

Set the marginal profit to zero, let $\theta_{lm} = \theta$ for all firms in all markets, invoke the functional forms from Lemma 1 and the above comparative statics results on the retail demand to obtain the symmetric equilibrium markup θ as a solution to $Z(\theta, M) = 0$, where

$$\begin{aligned} Z(\theta, M) = \; & (1 - \tfrac{M-1}{M}\tfrac{b}{L+1}L\sigma\theta)\sigma\theta(p(K(\theta)) - c)\tfrac{2}{L+1} \\ + \; & (1 - \tfrac{b}{L+1}(L - \tfrac{\theta}{M})\sigma\theta)\tfrac{K(\theta)}{L} - (\theta p(K(\theta)) - c)(1 - \tfrac{b}{L+1}(L - \tfrac{\sigma\theta}{M})\sigma\theta)\end{aligned}$$

and

$$\frac{K(\theta)}{L} = \frac{\alpha - (c + \frac{a-c}{L+1})\sigma\theta}{1 - L\frac{b}{L+1}\sigma\theta}.$$

is the firm-level retail volume. Let $\theta^R(M)$ be the smallest solution to $Z(\theta, M) = 0$, i.e. $\theta^R(M)$

is the most competitive retail equilibrium when retail markets are linked. Observe that

$$Z(\theta,1)-Z(\theta,M)=\frac{b}{L+1}\frac{M-1}{M}\sigma^2\theta^2[\frac{K(\theta)}{\sigma L}+\frac{2L}{L+1}(p(K(\theta))-c)-\theta p(K(\theta))+c]$$

is positive for σ sufficiently close to zero. In this case, $Z(\theta,1)>0$ for all $\theta\leq\theta^R(M)$ and therefore all solutions $Z(\theta^I,1)=0$ satisfy $\theta^I>\theta^R(M)$ for $M\geq 2$. Under the reasonable assumption that $\frac{a+Lc}{bL}>\alpha$ so that the retail sales are decreasing in the markup,

$$\frac{K'(\theta)}{L}=-\frac{\sigma}{L+1}\frac{a+Lc-\alpha Lb}{(1-L\frac{b}{L+1}\sigma\theta)^2}<0,$$

it follows that retail sales are higher, $K(\theta^R(M))>K(\theta^I)$, and spot market competition therefore more intense when retail markets are linked compared to the case of independent retail markets.

References

ACER (2015): Opinion on the compliance of NRAs' decisions approving methods of cross-border capacity allocation in the CEE region. No 09-2015.

Allaz, Blaise and Jean-Luc Vila (1993): Cournot competition, forward markets and efficiency. *Journal of Economic Theory* 59, 1-16.

Bohn, Roger E., Michael C. Caramanis, and Fred C. Schweppe (1984): Optimal pricing in electrical networks over space and time. *Rand Journal of Economics* 15, 360-376.

Bushnell, James B., Erin T. Mansur and Celeste Saravia (2008): Vertical arrangements, market structure, and competition: An analysis of restructured U.S. electricity markets. *The American Economic Review* 98, 237-266.

Egerer Jonas, Jens Weibezahn and Hauke Hermann (2016): Two price zones for the German electricity market: Market implications and distributional effects. *Energy Economics* 59, 365-381.

Green, Richard and Chloé Le Coq (2010): The lengths of contracts and collusion. *International Journal of Industrial Organization* 28, 21-29.

Holmberg, Pär (2011): Strategic forward contracting in the wholesale electricity market. *Energy Journal* 31, 169-202.

Holmberg, Pär and Andy Philpott (2018): On supply-function equilibria in radial networks. *European Journal of Operational Research*, forthcoming..

Lundin, Erik and Thomas P. Tangerås (2017): Cournot competition in wholesale electricity markets: The Nordic power exchange, Nord Pool. IFN WP 1191.

Mahenc, Philippe and François Salanié (2004): Softening competition through forward trading. *Journal of Economic Theory* 116, 282-293.

Mercadal, Ignacia (2016): Dynamic competition and arbitrage in electricity markets: The role of financial players. Manuscript.

Willems, Bert, Ina Rumiantseva and Hannes Weigt (2009): Cournot versus supply functions: What does the data tell us? *Energy Economics* 31, 38-47.

Wolak, F. A. (2000). An empirical analysis of the impact of hedge contracts on bidding behavior in a competitive electricity market. *International Economic Journal*, 14(2), 1-39.

Wolak, Frank A. (2009): An assessment of the performance of the New Zealand wholesale electricity market. Report for the New Zealand Commerce Commission.

Wolak, Frank A., (2011): Measuring the benefits of greater spatial granularity in short-term pricing in wholesale electricity markets. *The American Economic Review* 101, 247-252.