ENERGY INSTITUTE AT HAAS

Energy Institute WP 298

Market Size and Market Power: Evidence from the Texas Electricity Market

Matt Woerman

December 2018

Market Size and Market Power: Evidence from the Texas Electricity Market

Matt Woerman*

December 2018

Abstract

Economic theory tells us that market structure is the primary determinant of a firm's ability to exercise market power. However, it is challenging to empirically estimate the causal effect of market structure on market power because a firm rarely experiences exogenous variation in its market's structure. In this paper, I exploit a novel source of exogenous variation in market size within the Texas electricity market — congestion of electricity transmission lines due to ambient temperature shocks — to estimate the causal effect of market size on the exercise of market power. When transmission lines congest, this statewide market splits into smaller localized markets. I find that a 10% reduction in market size causes firms to more than double markups. The direction of this effect is consistent with a model of oligopoly competition in which firms set markups in response to residual demand, which is less elastic in a smaller market. My results imply that the markups induced by transmission congestion at high temperatures generate $7.1–21.5 million of deadweight loss annually. These markups also create large transfers — $2.1 billion per year — from consumers to producers, which raise important equity concerns.

*University of California, Berkeley; Department of Agricultural and Resource Economics and Energy Institute at Haas. Email: matt.woerman@berkeley.edu. I thank Meredith Fowlie, Lucas Davis, and Maximilian Auffhammer for their invaluable guidance and advice. I also thank Peter Berck, Susanna Berkouwer, Joshua Blonz, Severin Borenstein, Fiona Burlig, James Gillan, Nick Hagerty, Solomon Hsiang, Larry Karp, Erin Kelley, Gregory Lane, John Loeser, Jeffrey Perloff, Louis Preonas, Gordon Rausser, James Sallee, Joseph Shapiro, Deirdre Sutula, Itai Trilnick, Sofia Villas-Boas, Reed Walker, Catherine Wolfram, and seminar participants at UC Berkeley, Energy Institute at Haas, AERE Summer Conference, Heartland Environmental and Resource Economics Workshop, and Columbia Interdisciplinary PhD Workshop in Sustainable Development for thoughtful comments and feedback. Click here for the online appendix.

1 Introduction

Economists have long been concerned by a firm's ability to exercise market power because of the inefficiencies it imposes on that market and the economy more broadly. The primary consequence of market power is deadweight loss due to inefficient prices and levels of production (Harberger 1954), but additional consequences also exist. For example, firms that face less competition are slower to innovate (Aghion et al. 2005) and may provide lower quality (Gaynor and Town 2011). In fact, many recent macroeconomic trends, from a decrease in labor force participation to a slowdown in aggregate productivity, are consistent with increases in market power that have been observed globally in recent decades (De Loecker and Eeckhout 2017, 2018).

With such wide-reaching consequences, it is important to understand the factors that influence noncompetitive firm behavior. Economic theory tells us the primary factor is market structure (Cournot 1838), but it is challenging to empirically estimate the effect of market structure on market power. Both are equilibrium outcomes, making the causal nature of their relationship difficult to establish. In the absence of exogenous variation in market structure, empirical work on this topic has typically relied on assumptions about firm behavior or the nature of competition for identification.

In this paper, I use a novel source of exogenous variation in market size to estimate the causal effect of market size on market power in the Texas electricity market. Electricity is moved from producers to consumers over a grid of transmission lines, but the carrying capacity of transmission lines is limited. Transmission congestion splits the market into smaller "local markets," shrinking the set of rivals with which each power plant competes. For example, in the afternoon, a transmission constraint binds and a corner of the grid splits off from the full market into its own local market with higher prices than the rest of the grid; by that evening, the congestion subsides and the entire market is once again full integrated. Theory tells us that, for a power power plant in this corner of the grid, the incentive and ability to exercise market power vary greatly between these two times — when competing with fewer plants to supply electricity, a profit-maximizing firm will exercise market power by increasing markups. To test this theory, I exploit these exogenous but predictable changes in market size and estimate how they affect firm behavior.

Electricity markets provide a useful context to study market power because they simplify many of the common challenges to identification and estimation. First, electricity is a homogeneous good, which removes issues of product differentiation. Second, demand is perfectly inelastic and observed, so demand estimation is not required. Third, power plants have simple production functions with marginal costs that are easily calculated using available data. Fourth, deregulated electricity markets clear through a centralized auction mechanism with specified rules, so the game being played is common to all firms and known to the researcher. For these reasons, many prior studies of market power have focused on electricity markets,[1] which I build on for this work.

I use detailed microdata on power plant-specific pricing behavior and local market structure, which are available in the data-rich Texas electricity market. I observe every power plant's offer curve — a function that indicates how much the plant will produce at each price — at 15-minute intervals. I also have data on power plant operations and fuel prices, which I use to calculate every plant's marginal cost. I combine these offer curves and cost data to calculate plant-specific markups every 15 minutes. I also observe the presence of transmission congestion and, hence, market size at 5-minute intervals. Due to the complex nature of the transmission grid and the presence of nodal pricing[2] in this market, a power plant's market size can vary greatly, both cross-sectionally and over time.

In order to estimate the causal effect of local market size on markups, I leverage exogenous variation in transmission congestion. In particular, I exploit the fact that transmission capacity, and the resulting transmission congestion, is impacted by ambient air temperature, and I instrument for local market size with temperatures throughout the transmission grid. Transmission capacity is constrained by thermal limits on transmission lines, so less electricity can be transmitted at higher ambient temperatures. Hot temperatures also increase the consumption of electricity in Texas due to demand for indoor cooling. Thus, at high temperatures, more electricity is demanded, but the electrical grid has less capacity to transmit, resulting in congestion and local markets. This congestion, however, may extend well beyond the hottest areas of the grid. Due to its network structure, a temperature shock at any point in the grid can alter transmission throughout the entire network, resulting in local markets to form in geographically distant parts of the grid. By instrumenting with distant temperatures throughout the grid, I can control flexibly for the local temperature at a power plant to account for all local market conditions that vary with temperature. Then my

1. See Green and Newbery (1992), Borenstein and Bushnell (1999), Wolfram (1999), Borenstein, Bushnell, and Wolak (2002), Hortaçsu and Puller (2008), and Bushnell, Mansur, and Saravia (2008), among many others.

2. The Texas electricity market uses nodal pricing, meaning prices are determined at every node in the network to reflect the marginal value of electricity at that point in the grid. With no binding constraints, prices are uniform at all power plants. When transmission constraints bind, nodal prices diverge to reflect the shadow value of relevant constraints. This form of pricing is in contrast to zonal pricing, in which the market is divided into several zones and only inter-zonal transmission constraints are considered when setting market-clearing prices. Thus, a nodal market has orders of magnitude more possible configurations of local markets than does a comparable zonal market.

identifying assumption is that, after controlling for local temperature, the remaining variation in other temperatures only affect markups through their impact on transmission congestion.

To select from many possible instruments — hourly temperature at 75 weather stations throughout Texas — and functional forms, I use machine learning techniques to create a power plant-specific prediction model that forecasts each power plant's local market size. I find that, on average, a plant's local market size is best predicted by a nonparametric function of temperature at a single weather station, rather than a parametric function of multiple temperatures. Thus, I match each power plant to the weather station that best predicts its local market size and instrument with a nonparametric function of the matched temperature. This procedure is similar to the instrument selection method of Belloni et al. (2012). This prediction exercise also mimics the forecast that firms must make when submitting offer curves in advance of the market clearing; I show that roughly 45% of variation in local market size can be explained by a simple prediction model using data available to all firms when they submit offers.

I next show how local market size and markups vary with matched temperature. Using high-dimensional fixed effects to control for unobservables, and controlling flexibly for the local temperature, I find a power plant's local market shrinks as the temperature at the plant's matched weather station exceeds 30–35°C (86–95°F); at the extreme, a plant's local market is 9–12% smaller on average when the matched temperature exceeds 40°C (104°F), compared to moderate temperatures. The power plant's markups increase over the same range of matched temperatures; at temperatures over 40°C (104°F), markups are on average 135–195% larger than at moderate temperatures.

I then use two-stage least squares to find the causal effect of market size on markups. I estimate that when a power plant's local market shrinks by 10% of the mean market size, the firm increases markups at the plant by 120–160% of the mean value, or by more than 3.5 times the average marginal cost of a fossil fuel-fired power plant. This result is consistent with a model of firm behavior in which the firm sets markups at a power plant in response to residual demand at the plant. When the plant's local market is smaller, it faces residual demand that is less elastic, so the firm exercises market power by increasing markups. The magnitude of this result is large, showing that firms in this market respond to even a relatively small change in market size by substantially increasing the exercise of market power.

I finally calculate the welfare effects of this source of market power. To do so, I consider a counterfactual in which firms do not increase markups in response to transmission congestion induced by high temperatures. I assume these markups are passed through into higher average retail electricity prices in equilibrium, which has two main effects. First, although demand is inelastic in the real-time market, it has some elasticity in the long run; higher retail prices reduce consumption in all hours, which generates \$7.1–21.5 million of deadweight loss annually. Second, consumers pay an extra \$2.1 billion per year for the electricity they consume, which is purely a transfer to producers and raises important equity concerns. I further calculate how these effects will change as a result of climate change. I find that warming of 2.5°C (4.5°F) will roughly double the welfare effects of markups induced by transmission congestion at high temperatures. These results show that long-run planning and investment in this sector would benefit greatly by considering the competitiveness effects of climate change.

This paper makes four main contributions to the economics literature. First, these results contribute to a literature on the relationship between market structure and competition. Empirical studies on this topic date back to the work of Bain (1951), who estimated the correlation between profits and industry concentration in U.S. manufacturing; Bresnahan (1989) provides a summary of additional work in this "structure-conduct-performance paradigm." More recent empirical work uses models of firm entry and exit to estimate the competitiveness of industries (Bresnahan and Reiss 1991). These studies, however, typically require strong assumptions, such as how firms compete or how the number of firms in the market affects firm profitability (Berry and Reiss 2007).

In this study, by contrast, I exploit an exogenous source of market structure variation, from which I estimate the causal effect of market size on market power using weaker assumptions than these previous approaches. Bergquist (2017) provides experimental evidence of this causal effect by incentivizing the entry of intermediaries in Kenyan agricultural markets. She finds no effects on market outcomes and infers that the entrants easily join the collusive agreements that existed among incumbents. These results differ greatly from my estimates and suggest that this effect varies by settings. My work is the first to estimate this effect under weak assumptions in an industrialized setting.

Second, this paper contributes to a literature on market power and strategic firm behavior in wholesale electricity markets. Green and Newbery (1992), Borenstein and Bushnell (1999), Wolfram (1999), Borenstein, Bushnell, and Wolak (2002), Hortaçsu and Puller (2008), and Bushnell, Mansur, and Saravia (2008), among many others,[3] document the presence of market power in wholesale electricity markets, quantify the costs of this market power, and characterize the type of competition

3. See Wolfram (1998), Puller (2007), Fabra and Reguant (2014), Reguant (2014), Ito and Reguant (2016), and Hortaçsu et al. (2017) for more examples in this literature.

that exists in the market. This is one of the first papers, along with McDermott (2018), to document how market size or market structure affects noncompetitive behavior in these markets. This is also one of the first papers, along with Mercadal (2018), to emphasize the unique way in which local markets and the resulting market power form in nodal electricity markets.

Third, this work contributes to a literature on the role and value of transmission infrastructure in wholesale electricity markets. Borenstein, Bushnell, and Stoft (2000) and Joskow and Tirole (2000) provide theoretical support for the competitiveness effects of transmission infrastructure. In empirical work, Davis and Hausman (2016) exploit the closure of a nuclear power plant to value inter-regional transmission and document suggestive evidence of market power during hours of transmission congestion. Most closely related to this work, Ryan (2017) uses structural estimation to simulate an expansion of inter-regional transmission in the Indian electricity market. This is the first paper to directly estimate the effect of transmission congestion, and the resulting local markets, on firm behavior and the exercise of market power.

Finally, this paper contributes to the growing literature that examines the effects of climate on many economic outcomes, including effects on energy consumption (Auffhammer and Mansur 2014; Auffhammer, Baylis, and Hausman 2017; Auffhammer 2018), agriculture (Schlenker, Hanemann, and Fisher 2005; Deschênes and Greenstone 2007; Schlenker and Roberts 2009), productivity and growth (Dell, Jones, and Olken 2012; Burke, Hsiang, and Miguel 2015), conflict (Hsiang, Meng, and Cane 2011; Hsiang, Burke, and Miguel 2013), and mortality (Deschênes and Greenstone 2011). This is the first study to find climate change effects on market structure and competition, as well as the first to find effects on the supply of electricity.

2 Texas electricity market

The wholesale electricity market in Texas is managed by the Electricity Reliability Council of Texas (ERCOT). ERCOT is the independent system operator (ISO) that schedules power plants to supply electricity and coordinates the use of the transmission grid to deliver that electricity to consumers. Although representing only one state,[4] this market includes more than 610 individual generating units and roughly 25 million end-use consumers, with more than 46,500 miles of high-voltage transmission lines connecting producers and consumers (ERCOT 2018). In the years 2011–2014, nearly $12 billion was transacted annually in ERCOT's real-time wholesale market.

4. The ERCOT territory covers roughly 75% of the geographic area of the state of Texas, but this area accounts for approximately 90% of the electricity consumed in the state.

ERCOT is tasked with maintaining a reliable supply of electricity to consumers throughout its territory. Unlike the markets for other commodities, the wholesale market for electricity must be operated by an ISO due to the unique nature of electricity. First, storage of electricity is prohibitively expensive, so production of electricity must perfectly meet demand in real time to ensure an uninterrupted supply. Second, transmission of electricity is subject to the physical constraints of the electrical grid, such as thermal limits that constrain the quantity of electricity that can be transmitted. Hence, production of electricity must also occur at the correct locations to ensure the transmission grid is not overloaded and compromised.

A second mandate of ERCOT is to provide this reliable supply of electricity at lowest cost. To achieve this goal, ERCOT uses an auction mechanism in which each firm that produces electricity submits an offer curve — a function that indicates how much the plant will produce at each price — for each power plant it owns. ERCOT awards production to the power plants with the lowest-cost offers, subject to the constraints described above. When any of these constraints bind, the marginal cost of electricity can vary across the electrical grid. ERCOT calculates this marginal cost at more than 1000 nodes throughout the grid and assigns each a locational marginal price (LMP) that reflects the shadow value of relevant constraints.

The Texas electricity market is unique among electricity markets in the United States for two main reasons, both of which make this market an attractive setting for my study. First, the ERCOT grid is its own interconnection, meaning there is little capacity for trade with adjacent electricity markets. Thus, I do not consider imports or exports in this study.[5] Second, ERCOT operates an energy-only market[6] with relatively loose limits on prices[7] and firm behavior.[8] These factors reduce the possibility that my estimates are attenuated due to overlapping markets or regulatory restrictions.

5. By contrast, an important component of the Borenstein, Bushnell, and Wolak (2002) study of the California electricity market is the import of electricity into the state because California is one of several markets within the larger Western Interconnection.

6. Many electricity markets are accompanied by a capacity market that pays owners of power plants for their capacity, in addition to actual production, to ensure the market has adequate generation resources.

7. ERCOT began this study period with a price cap of $3,000 per MWh, which it increased to $9,000 per MWh by the end of this study period. Many other electricity markets, by contrast, have price caps of $1,000 per MWh.

8. Most electricity markets have procedures to mitigate excessively high offer prices. ERCOT does so only in the most extreme cases, which I describe in Appendix A.

2.1 Real-time market

The focus of this study is ERCOT's real-time wholesale electricity market.[9] This market clears every five minutes and determines which power plants produce electricity and how transmission is managed to ensure reliable delivery of electricity to end-use consumers. If a producer wants to generate electricity, it must participate in the real-time market by submitting an offer curve.[10] This curve specifies how much electricity a power plant will produce at every price, and the firm submits a separate curve for each power plant it owns. An offer curve must be weakly increasing in price and piecewise linear with up to 35 vertices.[11] The top panel of Figure 2 shows an example offer curve. I observe every offer curve submitted to ERCOT, which allows me to directly observe pricing behavior in this market without making additional assumptions about firm behavior.

On the demand side of the real-time market, purchasers of wholesale electricity are primarily retail providers that purchase electricity to distribute to their end-use consumers. These retail providers contract with their customers to provide electricity at retail rates that rarely vary with the wholesale price in the real-time market.[12] As a result, demand is considered to be inelastic in real time and the market clears on the supply side; ERCOT dispatches power plants to meet demand at lowest cost, subject to the physical constraints of the transmission grid. The resulting prices are equal to the marginal cost of electricity production at each node of the grid, as discussed below. The bottom panel of Figure 2 depicts the simplest example, when no constraints bind; the full market clears where the aggregate offer curve intersects the vertical demand curve, and all producers receive the market-clearing price of roughly $40 per MWh.

A firm offering to produce electricity in this market has great flexibility in its strategy throughout the course of a day; a power plant may have a different offer curve for each hour of the day. Additionally, a firm may adjust its offer curves up to the beginning of the preceding hour,[13] which allows the firm to refine its strategy as new information becomes available as the market time

9. Firms offering to produce electricity in the real-time market may have existing financial positions due to long-term bilateral contracts or ERCOT's day-ahead market. These positions are purely financial, however; physical delivery of electricity is determined only in the real-time market. See Appendix A for details on these other markets.

10. Alternatively, a firm can set an output schedule for a power plant, which specifies a quantity to produce; ERCOT assigns the power plant to produce the specified quantity regardless of the market-clearing price. This option is used infrequently.

11. Firms submit the price-quantity pairs that define the vertices of offer curves, and the offer curve is linearly interpolated from these points. Each curve can have up to 35 vertices, but, in practice, firms often use many fewer points when specifying offer curves.

12. Retail providers may offer time-of-use plans with rates that vary over the course of a day and seasonally. Although these plans are typically designed to follow the general price patterns observed in the wholesale real-time market, the rates paid by the consumers are set when the contract is signed and do not respond in real time to the wholesale market.

13. For example, offer curves for the 2–3 p.m. operating hour can be adjusted up to 1 p.m.

approaches. Much of the information a firm may consider when submitting its offer curves is known to all participants or is forecastable over this time frame. For example, firms have knowledge of the marginal cost of all power plants,[14] as well as which plants are offline at any point in time. Demand is highly forecastable, particularly within an hour or two; in fact, ERCOT publishes hourly demand forecasts beginning one week in advance. Transmission congestion and the formation of local markets are also forecastable, which I discuss next and model in Section 5.

2.2 Transmission congestion and local markets

An important component of ensuring a reliable supply of electricity is ERCOT's management of the transmission grid. Transmission lines provide a simple and inexpensive way to move electricity around the ERCOT region. However, thermal constraints limit the maximum carrying capacity of these lines. Transmission of electricity generates heat in the lines, and the amount of heat produced is proportional to the quantity of electricity being transmitted. When transmission lines overheat they can fail, which not only renders that transmission line inoperable, but can also cause cascading failures throughout the electrical grid.[15] As a result, it is critically important that the ISO ensures the dispatch of power plants, and the subsequent flow of electricity to consumers, does not cause a transmission line to exceed its operable limit.

To account for these constraints while still meeting electricity demand as efficiently as possible, ERCOT implements locational marginal pricing. That is, each location, or node, throughout the electrical grid can have its own price that equals the marginal cost of electricity at that point in the network. For example, when the lowest-cost dispatch of power plants does not violate any constraints, the market clears as a uniform-price auction and all LMPs are equal. However, if this dispatch would force a transmission line to exceed its carrying capacity, ERCOT must deviate from the lowest-cost dispatch, and LMPs diverge as some LMPs incorporate the shadow value of the binding constraints. Figure 3 displays each of these scenarios. In the top panel, no constraints bind and all producers receive the market-clearing price of roughly $40 per MWh. In the bottom panel,

14. The marginal cost of producing electricity at a given power plant is primarily the cost of fuel, or the product of the plant's heat rate — the rate at which the plant converts fuel into electricity — and the fuel price. Data on heat rates are publicly available, as are spot prices for common fuels.

15. Transmission lines experience thermal expansion, causing the lines to lengthen and sag as the internal temperature increases. If lines expand excessively, they can contact other transmission equipment or nearby trees, causing the system to short circuit or, in an extreme case, create sparks that could start a fire. Other components of the transmission system are also put under greater stress at higher temperature, increasing the incidence of component failure in both the short-run and long-run.

however, a constraint limits the flow of electricity into the southern portion of the grid, causing the LMPs in this region increase substantially, to more than $3,500 per MWh.

When transmission lines become congested, the ERCOT region is no longer fully integrated, as indicated by the divergence of LMPs. Any additional demand in the higher-priced region cannot be met with additional production in the lower-priced region but instead must be met locally. Thus, these power plants in the higher-priced region do not compete on the margin with power plants in the lower-price region, only with other plants in the same region. I refer to each of these regions as a "local market." In other words, the local market is the set of power plants within the ERCOT region at large that share a common LMP and compete to supply electricity within that region. When no transmission lines are congested, every power plant's local market is the entire ERCOT market, as depicted in the fully integrated market in the top panel of Figure 3. At the other extreme, when transmission lines are heavily congested, a very small local market forms and its LMP differs greatly from nearby local markets, as in the bottom panel of Figure 3.

By changing the number and kinds of power plants that compete with one another, the formation of local markets changes the ability of any one firm to exercise market power within a local market. In particular, when a local market contains fewer firms or less production capacity, economic theory suggests that market will be subject to greater market power, as I show in Section 3. With fewer power plants in its local market, the residual demand facing any one plant is less elastic, providing greater opportunity to exercise market power and influence the resulting LMP. I exploit this variation in local markets to estimate the effect of market size on markups.

Because the physical limits on transmission are thermal in nature, the carrying capacity of a transmission line is affected by the ambient air temperature; at high temperatures, a transmission line can carry less electricity while staying within the operable thermal limits. For example, an increase from 38°C to 43°C (100°F to 109°F) reduces transmission capacity by 7-8% (Sathaye et al. 2013). Additionally, a substantial portion of electricity demand in the ERCOT region is for indoor cooling and, to a lesser extent, heating, so the ambient temperature drives the quantity of electricity being transmitted throughout the electrical grid. At high temperatures, more electricity is consumed, but the electrical grid has less capacity to transmit electricity, resulting in higher rates of transmission congestion and the formation of smaller local markets. In Section 5 I show that the formation of local markets is forecastable and develop a predictive model of each power plant's local market as a function of ambient temperature. This suggests a firm is able to forecast the local market for each of its power plants and adjust their offer curves accordingly. I exploit the exogenous and quasi-experimental variation in local markets that is induced by temperature in order to causally identify the effect of market size on markups.

2.3 Power plants and firms

The Texas electricity market consists of 73 firms that operate 390 power plants[16] to supply electricity during my study period. Table 1 shows summary statistics for these plants by fuel type. Four types of plants — natural gas, coal, renewables, and nuclear — account for nearly all of the capacity and production in this market. Each type of power plant has unique technical characteristics that influence how it is offered to supply electricity and how it may affect the ability of other firms to exercise market power, which I describe in more detail in Sections 3 and 4. Figure 1 displays a map of units by fuel type, showing that plants are generally spread throughout the state, other than renewable plants, which cluster in areas with the greatest wind resources.

Table 1 also shows summary statistics by firm for the eight largest firms in this market. By many traditional metrics, this market does not appeared concentrated to the level that market power should be a concern. The three largest firms have markets shares of 20%, 15%, and 10%, respectively, when measured by production, and even less when measured by capacity; 60 firms have a market share of less than 1% by either metric. The Herfindahl–Hirschman Index for this industry is 0.059 when measured by capacity and 0.086 when measured by production, neither of which implies an uncompetitive market.[17] Traditional measures of competitiveness, however, are poorly suited for this industry because of the unique nature of electricity, including the lack of storage and the constrained nature of the grid on which the market operates.

3 Theoretical framework

To formalize the behavior of firms in ERCOT's real-time wholesale electricity market, I develop a model of oligopoly competition in this setting. This model closely follows that of Ryan (2017), who also models firm behavior under transmission constraints, but in the Indian electricity market. I

16. I use the terms "power plant" and "unit" interchangeably to describe the cross-sectional unit of my analysis, which I keep as disaggregated as possible while still allowing me to accurately join different data sources. In many cases, a single facility comprises two or more "units" by my definition. For example, if a coal-fired power plant has two separate boilers, each connected to its own generator, which can be operated independent of one another, and data are reported separately for each boiler-generator pair, then I consider this facility to be two separate units.

17. The Herfindahl–Hirschman Index (HHI) is a measure of market concentration, calculated as the sum of squared market share. There is no universally agreed upon threshold to define an industry as competitive, but rules of thumbs would generally consider an industry with HHI less than 0.1 to be competitive.

adapt some notation from Hortaçsu and Puller (2008), who study a previous iteration of the Texas electricity market. I first generate an optimal markup rule and then show how optimal markups change in response to changes in market size.

3.1 Optimal markup rule

Each firm submits an offer curve, $S_i(p)$, for each power plant it owns, where i denotes plant. This offer curve specifies how much the plant will produce at each price and is defined by a vector of offer prices, $\mathbf{b}_i$, and a vector of offer quantities, $\mathbf{q}_i$.[18] The cost to produce at the plant is $C_i(q)$. The profit earned at plant i is $\pi_i = S_i(p_i)p_i - C_i(S_i(p_i))$, where p_i is the LMP that resolves at plant i.

Firms submit offer curves in the face of several sources of uncertainty. First, the market may be fully integrated or transmission congestion may split the market into two or more local markets. Firms are able to partially forecast transmission congestion,[19] but not perfectly, so local markets are stochastic from the perspective of a firm submitting offer curves. I denote the set of local markets as $\mathcal{M}$. Second, demand is also highly predictable but ultimately stochastic. I denote demand in market m as D_m, which I assume is perfectly inelastic.[20] Third, firms do not know the offer curves of other firms. Because this is a repeated game, firms may have an expectation of other firms' offer curves. Firms also typically know which plants are online and ready to produce. They do not, however, have full information on other offer curves. A final consideration of a firm when setting offer curves is any forward position that the firm has.[21] I denote the price and quantity of firm f's forward position in market m as P^F_{fm} and Q^F_{fm}, respectively. Although a firm knows its forward positions, the distribution of those positions over local markets is uncertain because transmission congestion is stochastic.

Each firm maximizes its expected profit summed over all of its plants and existing forward positions:

$$\max_{\{S_i | i \in \mathcal{I}_f\}} \mathbb{E}\left[\sum_{i \in \mathcal{I}_f} \left(S_i(p_i)p_i - C_i(S_i(p_i))\right) + \sum_{m \in \mathcal{M}} (P^F_{fm} - p_m)Q^F_{fm}\right] \tag{1}$$

where p_m is the common LMP that resolves in market m. The final term is the profit or loss from forward positions, which the firm effectively must unwind in the real-time market.

18. I place no restrictions on offer curves in this model. ERCOT gives firms great flexibility in setting offer curves, as I describe in Section 2.

19. See Section 5 for a prediction model of local market size.

20. This assumption is common in the literature and reflects the reality of wholesale electricity demand. See Section 2 for a brief discussion of demand in ERCOT's wholesale electricity market.

21. See Appendix A for a discussion of bilateral contracts and the day-ahead market. I consider these positions to be exogenous from the real-time market.

The market clears by setting the price in each local market such that production equals demand net of flows between markets:

$$\sum_{i \in \mathcal{I}_m} S_i(p_m) = D_m + T_m$$

where $\mathcal{I}_m$ is the set of plants in market m and T_m is net transmission flows out of market m.[22] Rewriting this condition from the perspective of plant i:

$$S_i(p_m) = D_m + T_m - \sum_{\substack{j \in \mathcal{I}_m \\ j \neq i}} S_j(p_m) \tag{2}$$

Note that the right-hand side of this equation is the residual demand of plant i, which I denote $R_i(p_m)$.

To get the firm's first-order condition, I make two assumptions. First, every plant's marginal cost is constant in quantity up to the plant's capacity constraint. I denote plant i's marginal cost as c_i. Second, I assume no constraints are marginally binding. This assumption implies that, on the margin, offer price does not affect whether a constraint binds. I also define Q_{fm} to be the production by firm f in market m net of the firm's forward position in the market. That is, $Q_{fm}(p_m) = \sum_{i \in \mathcal{I}_f \cup \mathcal{I}_m} S_i(p_m) - Q^F_{fm}$.

Taking the derivative of the firm's objective function, from Equation (1), with respect to plant i's kth offer price, b_{ik}, and substituting in the residual demand, from Equation (2), yields the first-order condition:[23]

$$b_{ik} - c_i = \frac{\mathbb{E}[Q_{fm}(b_{ik})]}{-\partial\, \mathbb{E}[R_i(b_{ik})]/\partial p_m} \tag{3}$$

where firm f is the firm that owns plant i. This result is the optimal markup rule for plant i's kth offer quantity, similar to the Lerner index. The numerator is the firm's expected production that will be paid market m's real-time price when this offer price clears the market; when a firm has more inframarginal production in a market, it has a greater incentive to exercise market power. The denominator is the negative of the slope of the plant's expected residual demand when this offer price clears the market; when residual demand is less elastic, a firm has a greater ability to exercise market power.

22. Electricity only flows from lower-priced markets to higher-priced markets, so T_m does depend on prices. However, local markets are defined by binding constraints, so marginal changes in a market's price will not change flows of electricity between local markets. Thus, I consider T_m to be inelastic on the margin.

23. See Appendix B for a full derivation of this first-order condition.

3.2 Comparative statics

From the optimal markup rule in Equation (3), I derive how markups change in response to changes in the market size, which I then estimate in Section 6. I denote the local market size for plant i as K_i, which I measure as the total production capacity in the local market. In my empirics, I consider two additional measures: the number of units in the local market and the number of firms in the local market; theoretical results are comparable for all three measures.

The percent change in markup, $b_{ik} - c_i$, due to a change in the local market size is:[24]

$$\frac{d(b_{ik} - c_i)/dK_i}{b_{ik} - c_i} = \frac{\partial\,\mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i}{\mathbb{E}[Q_{fm}(b_{ik})]} - \frac{\partial^2\,\mathbb{E}[R_i(b_{ik})]/\partial p_m \partial K_i}{\partial\,\mathbb{E}[R_i(b_{ik})]/\partial p_m} \tag{4}$$

The first term is the percent change in the optimal markup rule's numerator, and the second term is the percent change in the optimal markup rule's denominator. Thus, the direction of the effect depends on whether the firm expects a change in market size to have a greater percent effect on its own production in the local market — more production gives the firm a stronger incentive to exercise market power — or on the slope of expected residual demand — less elastic residual demand gives the firm a greater ability to exercise market power. If the first term dominates, then a reduction in the local market size would yield smaller markups. If the second term dominates, then a reduction in the the local market size would cause markups to increase.

In theory, this effect could go in either direction. In practice, however, the second term is likely to dominate in my empirical setting. As I show in Section 4, transmission congestion occurs much more often when the ambient temperature is high. Hot temperatures also result in a higher level of demand, meaning the market clears further out on the aggregate offer curve. This outer portion of the aggregate offer curve tends to become very inelastic. When the market is operating under these conditions, even small changes in local market size may cause large changes in the slope of residual demand. Thus, I expect to see the second term to dominate, resulting in a negative effect — firms set larger markups when the local market size is reduced.

I further examine how this effect varies depending on the types of capacity in the local market. I first consider how a plant's markups are differently effected by a change in local market capacity owned by competitors, K_i^{comp}, and a change in local market capacity owned by its own firm, K_i^{own}.

24. See Appendix B for a full derivation of all comparative statics.

Ownership has an important effect on the first term of the comparative static in Equation (4):

$$\frac{\partial\, \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i^{comp}}{\mathbb{E}[Q_{fm}(b_{ik})]} < \frac{\partial\, \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i}{\mathbb{E}[Q_{fm}(b_{ik})]} < \frac{\partial\, \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i^{own}}{\mathbb{E}[Q_{fm}(b_{ik})]}$$

Changes in the capacity of competitors have no direct effect on a firm's own production, whereas changes in its own capacity have large effects. Thus, the first term dominates most often when considering a firm's own capacity and least often when considering competitors:

$$\frac{d(b_{ik}-c_i)/dK_i^{comp}}{b_{ik}-c_i} < \frac{d(b_{ik}-c_i)/dK_i}{b_{ik}-c_i} < \frac{d(b_{ik}-c_i)/dK_i^{own}}{b_{ik}-c_i} \tag{5}$$

In words, the effect of competitors on markups is more negative than the average effect. On the other hand, the effect of a firm's own capacity is less negative and may even change the sign of the effect — a firm may set lower markups when it has less capacity in a local market.

I finally examine how different types of plants affect markups. I consider renewables and nuclear plants to be nondispatchable because they almost always produce at maximum capacity; these plants have low marginal costs and typically submit offer curves that are almost perfectly elastic. I consider all other plants to be dispatchable because their production tends to follow the market-clearing price; these plants have higher marginal costs and typically submit offer curves that slope upward. Plant type has an important effect on the second term of the comparative static in Equation (4):

$$\frac{\partial^2\, \mathbb{E}[R_i(b_{ik})]/\partial p_m \partial K_i^{nd}}{\partial\, \mathbb{E}[R_i(b_{ik})]/\partial p_m} < \frac{\partial^2\, \mathbb{E}[R_i(b_{ik})]/\partial p_m \partial K_i}{\partial\, \mathbb{E}[R_i(b_{ik})]/\partial p_m} < \frac{\partial^2\, \mathbb{E}[R_i(b_{ik})]/\partial p_m \partial K_i^{disp}}{\partial\, \mathbb{E}[R_i(b_{ik})]/\partial p_m}$$

where K_i^{nd} is nondispatchable capacity and K_i^{disp} is dispatchable capacity. Changes in nondispatchable capacity only shift residual demand, whereas changes in dispatchable capacity rotate residual demand, due to the elasticity of offer curves submitted for each group. Thus, the second term dominates most often when considering dispatchable capacity and least often when considering nondispatchable capacity:

$$\frac{d(b_{ik}-c_i)/dK_i^{disp}}{b_{ik}-c_i} < \frac{d(b_{ik}-c_i)/dK_i}{b_{ik}-c_i} < \frac{d(b_{ik}-c_i)/dK_i^{nd}}{b_{ik}-c_i} \tag{6}$$

In words, the effect of dispatchable capacity on markups is more negative than the average effect, whereas the effect of nondispatchable capacity is less negative than average.

4 Data

To estimate the effect of market size on markups, I construct a novel panel dataset at the power plant-by-hour level for the years 2011–2014. This dataset has four important features: offer curves, marginal costs, local market sizes, and temperatures. I discuss each of these data sources and describe how I construct the main variables of interest.

4.1 Offer curves

I observe all offer curves that firms submit to ERCOT's real-time wholesale electricity market. These data are from ERCOT's daily 60-Day SCED Disclosure Reports.[25] Each report includes the sets of price-quantity pairs that trace out every unit's offer curve at 15-minute intervals. From these offer curves I directly observe the pricing strategy of firms at every power plant, and, when combined with my calculation of marginal, I use them to calculate markups. Figure 4 shows several example offer curves, including one for each of the main fuel types.

The natural gas and coal units share a similar shape. At low quantities, these units are offered at low or negative prices for dynamic reasons; many of these power plants are costly to shut down and start up again, so a firm is willing to accept a price below marginal cost for a short period of time to avoid incurring these fixed costs. The middle range of quantities is offered at a price roughly equal to marginal cost. At higher quantities, however, offer prices are above marginal cost and increase with quantity in a way that is not consistent with marginal cost pricing.

The wind and nuclear units are each offered at negative prices, but for different reasons. The wind unit has no fuel cost, so its marginal cost is approximately zero. It does, however, generate a federal production tax credit when it produces electricity. Thus, the opportunity cost of this unit is not zero but rather the negative of the tax credit, and the unit's entire capacity is offered at this price. Nuclear units, on the other hand, are designed to operate at full capacity, and deviating from full capacity is costly. Thus, the owner of a nuclear unit will accept negative prices to ensure the unit continues to produce at full capacity.

The final two example offer curves in Figure 4 show the same natural gas unit during two different hours. In both hours, the unit has a marginal cost of roughly \$35 per MWh. The offer prices at higher quantities, however, exceed \$150 per MWh in one example and \$800 per MWh in the other. These offer curves are not consistent with the marginal cost of the unit. They are,

25. The 60-Day SCED Disclosure Reports are so-called because the release of each report is delayed 60 days to reduce the ability of firms to exercise market power by best-responding to observed offers of competitors.

however, consistent with the exercise of market power by marking up prices well above marginal cost, which is the behavior I seek to identify in this study.

4.2 Marginal cost

The primary component of a power plant's variable cost is the cost of fuel.[26] I calculate each unit's variable fuel cost as the product of the unit's average heat rate — the rate at which the unit converts fuel into electricity — and the fuel price. Additionally, many fossil fuel-fired units are subject to emissions regulations and must purchase permits to cover their level of emissions. For these units I also calculate the variable cost of emissions, which I calculate as the product of the unit's average emissions rate and the permit price.[27] I sum these two variable costs to calculate marginal cost. I assume units have constant marginal cost up to the maximum capacity of the unit, as is common in the literature.

Data on fuel input and emissions are from the Environmental Protection Agency's (EPA) Continuous Emissions Monitoring System (CEMS), and data on net generation[28] are from ERCOT's 60-Day SCED Disclosure Reports. I combine these data to calculate average heat rates and average emissions rates.[29] Fuel prices are from several sources: S&P Global Market Intelligence, the Energy Information Administration's (EIA) Form 923, and Deutch et al. (2009); natural gas prices vary daily and regionally, coal prices vary monthly, and nuclear prices vary annually. Emissions prices are from S&P Global Market Intelligence and the EPA's market progress reports; emissions prices vary daily or monthly. As described above, I combine these prices with heat rates and emissions rates to calculate each unit's constant marginal cost.

With data on both offer prices and marginal cost, I calculate markup as the difference between price and cost. In most settings, markup is simply a single value. In this market, however, firms submit offer prices for every level of production, so markup is a function of the level of production. To normalize production quantities across power plants of different sizes, I use capacity factor — the plant's production as a percent of maximum possible production — as the measure of production. Figure 5 shows the average offer curve and average marginal cost, as a function of capacity factor,

26. All power plants have a small variable cost of operations and maintenance. These costs are unobserved so I do not include them in my marginal cost calculation. In my empirical specification, unit-level fixed effects will account for these costs.

27. Units face regulations for emissions of SO_2, NO_x, or both. I perform this calculation separately for each pollutant and sum over the pollutants for which the plant is subject to regulation.

28. Net generation is the amount of electricity that is output to the grid. This measure is in contrast to gross generation, which is the total amount of electricity produced, some of which is consumed within the power plant itself.

29. See Appendix C for a more detailed discussion of these calculations, including how I impute missing data.

for all units and hours in my dataset, as well as the averages for each of the four primary types of units.[30]

In order to analyze markups, which are functions rather than single values, I select several points on each unit's offer curve at which I calculate the markup and estimate the effect of market size: 85%, 75%, and 65% of nameplate capacity. A power plant typically cannot offer 100% of its nameplate capacity to the market because some of its production is consumed within the plant, so it is common to observe a plant offer only 90% of its capacity.[31] Additionally, operating near full capacity can reduce the efficiency of the unit[32] and increase the need for maintenance in the long run. Production above 85% of capacity may be subject to one of both of these issues, so I do not consider markups at these high levels of production. On the other hand, below 65% of nameplate capacity, coal-fired units have offer prices below marginal cost on average, due to the plant dynamics I describe above. Thus, markups at lower levels of production are likely to be driven by these operational considerations and not the potential to exercise market power.

4.3 Local market size

I infer local market size from the price data in ERCOT's real-time price reports. These reports give LMPs at every node in the transmission network for every clearance of the market, which occurs at five-minute intervals. As an example, these are the data that I use to create Figure 3. As discussed in Section 2, any binding constraints cause a discrete change in LMPs equal to the shadow value of the constraint. I consider the transmission grid between two power plants to be unconstrained if the LMPs at those plants are within $1 per MWh or 1%, whichever is greater.[33] I then define a local market to be the set of all plants that can be connected by unconstrained transmission.[34] For example, in the bottom panel of Figure 3, I define one local market to be the lighter points in the

30. Each power plant's offer curve must be weakly increasing, but a firm can offer less than the full capacity at any of its plants. In fact, most plants cannot offer 100% of their capacity to the market because some production is consumed at the plant. As a result, the sample changes slightly as quantity increases, particularly as quantity exceeds 90%. For this reason, offer curves may appear to slope downward on average even though each unit-level curve is weakly increasing.

31. The amount of production consumed at the plant is often at least 5% and typically closer to 10%. For example, Puller (2007) assumes a plant is operating at "full capacity" when it produces at 90% of nameplate capacity.

32. Bushnell and Wolfram (2005) show that heat rate increases when a unit exceeds the 98th percentile of its output but is relatively constant at more moderate levels of production.

33. Due to the complex nature of this network, LMPs sometimes differ by a few cents, and it is not clear if these small price differences represent real constraints of the system. By imposing the requirement that LMPs must exceed $1 per MWh and 1%, I construct a conservative measure of local market size.

34. For example, suppose LMPs are $30.00, $30.50, $31.25 at plants A, B, and C, respectively. I consider A and B to be connected by unconstrained transmission, and I also consider B and C to be connected by unconstrained transmission. Because A and C can be connected through B, I define all three plants to share a local market, even though A and C have LMPs that differ by more than $1.00.

southern portion of the grid, and I define a second local market to be the darker points throughout the rest of the grid. With local markets defined, I sum the production capacity, count the number of units, and count the number of firms within a local market to construct three measures of its size.

4.4 Temperature

Temperature data are from the National Oceanic and Atmospheric Administration's (NOAA) Integrated Surface Database. This database includes hourly — and occasionally more frequent — weather recordings from 35,000 weather stations worldwide. I limit my sample to weather stations in the state of Texas with minimal missing data.[35]

As described in Section 2, temperature is an important driver of the formation of local markets. Figure 6 depicts this relationship. The top panel shows the fraction of time that the market is either fully integrated or split into local markets as a function of the mean temperature in the state.[36] At most temperatures, the market is integrated a majority of the time. At mean temperatures above 30°C (86°F), however, the formation of local markets is much more prevalent, occurring roughly 75% of the time. The bottom panel of Figure 6 shows the frequency with which the market splits into a certain number of local markets. In particular, it is more common for the market to split into seven or more local markets when the state experiences high temperatures. Thus, when the mean temperature is greater than 30°C (86°F), not only are local markets more likely to form, but they are typically more numerous and, hence, smaller in size.

4.5 Additional data

I supplement these data with several additional data sources. Unit characteristics, such as fuel type, capacity, and location, are from the EIA's Form 860. Ownership data are from S&P Global Market Intelligence. System-wide demand data, which I use in my prediction models, are from ERCOT's 48-Hour Aggregate Load Summary.

35. A weather station must have data for at least 97.5% of the hours during my study period and no more than 60 consecutive hours of missing data. See Appendix C for a description of how I impute missing temperatures for the weather stations in my sample.

36. I calculate mean temperature as the unweighted mean of all 75 weather stations in the state.

5 Empirical strategy

The goal of this analysis is to examine how markets size affects competition. The most straightforward method to estimate the role of market size on market competitiveness is to simply regress a measure of competition on a measure of market size. For example, in this analysis, that would involve regressing markups on one of my measures of market size: market capacity, number of units in the market, or number of firms in the market. This method, however, has several empirical problems, both in general and because of this specific context.

First, in the general case, both market competitiveness and market size are equilibrium outcomes that are co-determined and depend importantly on underlying market characteristics. In the absence of exogenous variation in market size, any attempt to causally estimate this effect will suffer from simultaneity bias. Second, in this specific case, the complex nature of the transmission grid means that markups can effect the definition of a market. Offer curves and markups determine where electricity is produced and how it is transmitted to consumers, which may result in transmission congestion and the formation of local markets.

This setting features another potential problem; a firm must submit its offer curves to the system operator one to two hours before the market clears, before the firm has perfect information about the market in which it will operate. If the local market size was determined entirely at random after firms had submitted offer curves, then markups would be uncorrelated with the ultimate market size. In this case, the simple regression of markups on market size would yield no significant effect, even if a firm would have exercised market power had it known the size of the market in which it would be operating.

To overcome the potential endogeneity of market size and to isolate predictable variation in local market size, I use an instrumental variables approach, instrumenting for market size with contemporaneous temperature. In order for temperature to be a valid instrument, however, temperature must be uncorrelated with markups other than through its effect on local market size. If this is not true, then the estimated effect of local market size on markups is confounded with any other channels through which temperature may influence the markup decisions of firms.

The primary concern in this setting is that the temperature near a power plant may affect the characteristics of that plant's local market through more channels than just the size of the market. In particular, the level of demand in this market is strongly driven by temperature due to indoor cooling and heating. If an hour of extreme heat both induces transmission congestion and increases demand, then local markets are smaller and clear further out on the aggregate offer curve, corresponding to a less elastic supply of electricity. Both effects decrease the elasticity of residual demand faced by a plant in this local market, which increases the incentive to exercise market power. But an empirical strategy that instruments with the local temperature would attribute this entire effect to the size of the local market and ignore the effect of other changes in local market conditions.

To overcome this threat to identification, I do not use the local temperature at each power plant as an instrument. Instead, I use temperatures at all 75 weather stations throughout the state of Texas as potential instruments. Due to the complex nature of the transmission grid, a temperature shock at any point in the network can affect the operation of the entire grid, including the incidence of transmission congestion and the formation of local markets. Because these other temperatures are geographically farther from the power plant, they are less likely to be correlated with market conditions in the plant's local market. Additionally, I can control for the local temperature closest to the power plant to isolate temperature variation that is orthogonal to any temperature-related characteristics of that plant's local market.[37]

To determine which of the 75 weather stations to use as instruments and what functional form these temperatures should take, I develop a prediction model of local market size. I next discuss this prediction model and its results, and I then present my empirical specification that incorporates the results of the prediction model.

5.1 Prediction model

Ambient air temperature is a key driver of local market formation, as described in Section 2 and depicted in Figure 6. I use machine learning tools to develop a prediction model of local market size as a function of temperatures throughout Texas, as well as other explanatory variables. The results of this model determine the first stage of my two-stage least square empirical strategy. This methodology is similar to that of Belloni et al. (2012), but I consider additional variable selection procedures beyond the LASSO model of their approach.

In addition to determining the first stage of my 2SLS regression, this prediction modeling exercise also mimics the prediction problems that firms face when submitting offer curves to the system

37. This identification strategy is similar to that of Schlenker and Walker (2016). They study the effect of air pollution due to taxiing airplanes at California airports. The endogenous variable is the taxiing time of these airplanes, and they instrument with taxiing time at large airports in the Eastern United States, an indicator of flight delays that propagate throughout the flight network.

operator. Because firms must submit their offers one to two hours before the market clears, I seek to develop a simple model using data that are available to firms when they submit offer curves. By doing so, my first stage isolates variation in local market size that is not only exogenous but also predictable by firms operating in this market.[38]

I predict hourly local market capacity for each power plant as a function of 75 temperatures throughout Texas, indicators for each hour of day within a month of the year, system-wide demand, and local market capacity lagged two hours. I consider five different model specifications, which I describe below, in order to explore the trade-off between estimating the effect of temperature on local market capacity as flexibly as possible for one weather station and including interactions between temperatures at multiple weather stations. In each case, I use stratified five-fold cross-validation to ensure the models are not overfitted.[39]

The first model simply accounts for the mean local market capacity at each power plant to serve as a baseline against which to compare the performance of the other models. That is, I separately estimate for each power plant the time-series model:

$$Size_t = \alpha + \xi_t \tag{7}$$

where α is the plant mean and ξ_t is the prediction error.

The next two model specifications use a subset selection procedure. For each specification, I estimate a model for each pair of power plant and weather station and then, for each power plant, select the model that minimizes the cross-validated mean squared error (MSE). For each pair of power plant and weather station, the model is:

$$Size_t = \alpha_{hm} + f(Temp_{st}) + \mathbf{X}_t\boldsymbol{\rho} + \xi_t \tag{8}$$

where α_{hm} represents a set of indicators for each hour of day (h) within a month of the year (m), $Temp_{st}$ is the time series of temperatures at weather station s, $f()$ is estimated nonparametrically using 2.5°C (4.5°F) temperature bins, $\mathbf{X}_t$ is an additional set of potential explanatory variables, and

38. With more than $12 billion transacted annually in ERCOT's real-time wholesale electricity market, firms benefit greatly from predicting markets and submitting optimal offer curves. The most sophisticated firms in this market have teams of employees to model future market outcomes. Consulting firms offer these services to firms that do not conduct forecasts in-house. I do not seek to duplicate these efforts. Instead, my prediction model mimics a forecast that could be conducted by even the less sophisticated firms in this market.

39. I stratify each fold by hour of day within a month of the year to ensure each fold sufficiently captures the seasonality and daily cycles present in electricity market data.

ξ_t is again the prediction error. In the first specification, I consider no variables beyond temperature. In the second specification, I include system-wide demand and lagged local market capacity in $\mathbf{X}_t$, up to a third-order term for each variable. I use LASSO to select which of these additional explanatory terms to keep in the model for each pair of power plant and weather station. For each power plant, I select the two models — one for each specification — that minimize the cross-validated MSE. This matches each power plant to the single weather station that is most predictive of the plant's local market capacity.

The final two model specifications include temperatures from all 75 weather stations in a single model for each power plant, and I use LASSO to select which temperatures to keep in the model. That is, I separately estimate for each power plant:

$$Size_t = \alpha_{hm} + \mathbf{Z}_t\boldsymbol{\zeta} + \xi_t \tag{9}$$

where α_{hm} is as above, Z_t is the set of all explanatory variables, and ξ_t is again the prediction error. In the first of these specifications, Z_t includes only temperatures, up to a third-order term for each of the 75 weather stations and an interaction for each pair of weather stations. In the second, I again add system-wide demand and lagged local market capacity; I include up to a third-order term for each variable, the interaction of the two, and the interactions of each one with each temperature measurement. Each specification has at least 3,000 possible explanatory variables, in addition to the indicators, and I use LASSO to select which terms to keep in the model for each power plant.[40] This procedure yields two additional prediction models — one of each specification — for each power plant.

Table 2 summarizes the results of these five models. I calculate each metric at the plant level, and the table displays the mean and standard deviation of all power plants. Column (1) corresponds to the baseline model of Equation (7), Columns (2) and (3) correspond to the subset selection models of Equation (8), and Columns (4) and (5) correspond to the LASSO models of Equation (9). The subset selection models perform better across all measures than the comparable LASSO models — that is, Column (2) performs better than Column (4), and (3) better than (5). This result suggests

40. When using LASSO to select terms to include in the model, I use the one standard error rule to select the value of the tuning parameter, typically denoted as λ, that scales the shrinkage penalty. That is, I consider many possible values of the tuning parameter and, for each value, calculate the mean and the standard error of the cross-validated MSE, where the mean and standard error are taken over the five folds of data. I consider all tuning parameter values that yield a mean MSE within one standard error of the minimum MSE. From this set of tuning parameters, I select the value that yields the models with the fewest number of variables. This common procedure produces the simplest model with a mean MSE that is statistically indistinguishable from the minimum MSE.

that the relationship between temperature and local market capacity is sufficiently nonparametric that a flexible function of a single weather station's temperature is a better predictor of local market capacity than is the full set of temperatures with higher-order terms and interactions. For this reason, I focus on the subset selection models that match each power plant to a single weather station.

Figure 7 plots the distribution of prediction errors from the baseline and both subset selection models. The lightest distribution shows the variation that exists in local market capacity, which is bimodal due to the mass of large negative values, corresponding to hours when a plant faces a local market that is much smaller than its average. The middle distribution, which plots the errors from the model in Column (2) of Table 2, shows that a single, nonparametric temperature and temporal indicators account for roughly one-third of the variation in local market size, with much more mass at the center of the distribution. The further addition of more explanatory variables increases this predictive power to nearly 45% of baseline variation, as shown in the darkest distribution and Column (3) of Table 2. While these simple models do not perfectly predict the size of the local market in which a power plant will operate, they account for a sufficient amount of variation in local market size to be useful in both the first stage of my 2SLS estimation and providing a simple heuristic that even the less sophisticated firms could use when operating in this market.

The main result that I take from this prediction model to my 2SLS regression specification is the matching of each power plant to the weather station that best predicts its local market capacity. Although both system-wide demand and lagged local market capacity add predictive power to the time series model for each power plant, I do not include them in my panel 2SLS regression for econometric reasons.[41] Thus, I use the matched weather stations from the subset selection model, summarized in Column (2) in Table 2, that includes temperature but no additional explanatory variables.

I instrument for a power plant's local market size with the temperature at its matched weather station while controlling for the local temperature near the power plant and, hence, the temperature-dependent characteristics of the power plant's local market. This approach, however, requires that these temperatures are only weakly correlated so the matched temperature has sufficient variation that is orthogonal to that of the local temperature. Figure 8 plots two histograms of the distance from power plants to weather stations, one for the closest weather station to each power plant, and

41. I include hour of sample fixed effects, which are collinear with system-wide demand. I also include unit fixed effects, so I omit lagged local market capacity to avoid dynamic panel concerns.

the other for the matched weather station. The majority of power plants have a weather station within 25 km (15.5 miles), so the temperature at the closest weather station is a good representation of the temperature at the plant and in the plant's local market. A matched weather station, however, can be more than 800 km (500 miles) away from the power plant, and the average distance from a power plant to its matched weather station is roughly 305 km (195 miles). These large distances suggest that, even when controlling for local temperature, the matched temperature has sufficient variation that is correlated with the plant's local market size but orthogonal to other local market conditions.

Figure 9 plots the distribution of matched temperatures, as both a kernel density and a histogram by temperature bins. These distributions depict an important feature of the Texas electricity market: high temperatures are common. The median temperature is 22.7°C (72.9°F), and the modal temperature bin is just above 25°C (77°F). At the upper end of the temperature distribution, temperatures greater than 30°C (86°F) occur in 21.5% of observations, and temperatures exceed 35°C (95°F) in 7.5% of observations. These high temperatures drive transmission congestion and local market formation, as described in Section 4, and the prevalence of these high temperatures underscores the importance of local markets in this setting.

5.2 Empirical specification

I estimate the causal effect of market size on markups by estimating the following two equations with a two-stage least squares (2SLS) regression:

$$Markup_{it} = \beta \widehat{Size}_{it} + g(LocalTemp_{it}) + \gamma_{ihm} + \delta_t + \varepsilon_{it} \tag{10}$$

$$Size_{it} = f(MatchTemp_{it}) + g(LocalTemp_{it}) + \eta_{ihm} + \theta_t + \omega_{it} \tag{11}$$

where the second stage in Equation (10) uses the fitted values of $Size_{it}$ — the market size facing unit i at hour t — estimated by the first stage in Equation (11). The parameter of interest is β, which gives the relationship between this fitted value of market size and $Markup_{it}$, the markup of price over marginal cost.

In the first stage I estimate a model of market size as a function of $MatchTemp_{it}$, the average temperature during hour t at unit i's matched weather station. I consider two different functional forms of temperature. In the more flexible specification, I nonparametrically estimate market size as a function of temperature using 2.5°C (4.5°F) temperature bins, as in the prediction models using subset selection. In the other specification, I estimate market size using a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). When reporting 2SLS results, I indicate which first-stage specification was used.

I flexibly control for the temperature during hour t at unit i's closest weather station, $LocalTemp_{it}$, in each stage. I again use 2.5°C (4.5°F) temperature bins. This function nonparametrically controls for conditions in the local market that are influenced by temperature, such as the level of demand due to indoor cooling and heating, as well as any potential effects of temperature on power plant operations, including reduced efficiency due to extreme temperatures.

I include in each stage a set of unit fixed effects, γ_{ihm} and η_{ihm}. Electricity markets exhibit cyclical behavior over the course of a day, with demand and wholesale price often peaking in the afternoon and reaching a low in early morning hours. Electricity markets also exhibit seasonality, with demand and wholesale price often at their highest during summer months, and this is particularly true of the Texas electricity market due to a large demand for air conditioning at high temperatures. To control for these regular patterns in the market and the manner in which firms may respond to them at each power plant, I allow for a different unit-specific effect for each hour of day (h) within a month of year (m). I also include in each stage a set of hour-of-sample fixed effects, δ_t and θ_t, to control for contemporaneous shocks common to all units.

After controlling for the local temperature and these fixed effects, the identifying variation is within-unit variation in deviations from market-wide averages. For example, isolated weather patterns that induce not only deviations in the mean temperature across the market but also deviations in the distribution of temperatures across the market. Or, in the case of market size, transmission congestion that results in not only heterogeneous local market sizes but also deviations from the typical grid topology.

The remaining terms, ε_{it} and ω_{it}, are idiosyncratic errors. These errors are likely to be highly correlated over time. For example, an unobserved shock that affects a unit's market size or markups in one hour often persists into the future. These errors are also likely to be highly correlated over space because all units within a local market face the same local market size. To account for this correlated error structure, I two-way cluster standard errors at the unit and date levels, which allows for arbitrary serial correlation within a unit and arbitrary cross-sectional correlations within the same date. This level of clustering provides the correct inference so long as one unit's error term is not correlated with the error term of a different unit on a different date.

In order to interpret my estimate of β in Equation (10) as the causal effect of market size on markups, the standard assumptions of 2SLS must hold:

$$\text{Cov}(MatchTemp_{it}, Size_{it} \mid LocalTemp_{it}, \eta_{ihm}, \theta_t) \neq 0 \tag{12}$$

$$\text{Cov}(MatchTemp_{it}, \varepsilon_{it} \mid LocalTemp_{it}, \gamma_{ihm}, \delta_t) = 0 \tag{13}$$

The relevance assumption in Equation (12) ensures that variation in the matched temperature captures some of the variation in market size; I have already shown this to be true in the prediction models and will address it again when discussing results. The exclusion restriction in Equation (13) ensures that, conditional on the local temperature and high-dimensional fixed effects, the effect of matched temperature on markups acts only through market size and through no other channels.

As with any instrumental variables approach, the primary threat to identification is the validity of the exclusion restriction. In this case, the exclusion restriction would be invalid if a power plant's matched temperature is correlated with the plant's markup in a way that is not explained by the local market size, the local temperature, the plant's typical daily or seasonal patterns, or a shock common to all plants in the full market. For example, the consumers in the plant's local market would have to adjust their demand for electricity in response not to the local temperature, but rather to the temperature at the geographically distant matched weather station, which is unlikely. These controls leave no clear channel, other than local market size, through which the matched temperature could be correlated with the markup decision at the power plant.

6 Results

6.1 First-stage results

I consider three different measures of market size: the production capacity of the local market, the number of units in the local market, and the number of firms in the local market. I first show how temperature affects market size by plotting the first-stage results of Equation (11) for each of these size measures. The results for capacity, units, and firms are depicted in Figure 10. The points and error bars show estimates and 95% confidence intervals for the binned specification;[42] the curve and shaded area show the estimate and 95% confidence interval for the natural cubic spline. In each

42. All interior bins are 2.5°C (4.5°F) wide. The first and last bins correspond to temperatures below -10°C (14°F) and above 40°C (104°F), respectively. Point estimates are placed at the midpoint of each bin and give the average response to a temperature within that bin.

specification, I estimate effects relative to a temperature of 16.25°C (61.25°F),[43] which roughly corresponds to the midpoint of the temperatures I observe. Because results are relative to this reference temperature, negative values on the figures do not imply local markets with negative size, but rather local markets that are smaller than a comparable market at the reference temperature.

Local market capacity, the number of units in the local market, and the number of firms in the local market are highly correlated, so their temperature-response functions share a similar shape in Figure 10, although the units and scale of the y-axis vary. For all three measures, the effect on market size is relatively flat and precisely estimated throughout the middle range of temperatures, and even at the lowest temperatures in most cases. Much of the temperature mass lies within this range, as plotted in Figure 9, so a unit's matched temperature has no marginal effect on its market size for many hours of my sample. When temperatures exceed 30–35°C (86–95°F), however, the average local market size faced by a power plant is significantly reduced. At the extreme, when temperatures exceed 40°C (104°F), local markets decrease in size by 6.5 GW, 15 units, or 4 firms, as compared to a moderate temperature. This represents a reduction of 9–12% from the mean local market size.

These first-stage results confirm the general intuition described in Section 2 and the aggregate data presented in Section 4. It is worth considering, however, exactly what is being identified in these regressions. Because of the empirical specification and the identifying variation that remains after controlling for local market conditions with local temperature and for unobservables with fixed effects, these results do not imply that local markets only form at extreme temperatures or that local market sizes vary by only 6.5 GW, 15 units, or 4 firms. Local markets do form at more moderate temperatures, as shown in Figure 6, but the size of these local markets are fully explained by a combination of the local temperature near the power plant and the fixed effects, so these observations provide no identifying variation in the first stage.

Instead, these first-stage regressions are identified off of variation that differs from the local temperature and hourly or unit-level averages. For example, when one unit's matched weather station experiences a temperature greater than 40°C (104°F), most weather stations throughout the state also experience high temperatures, so transmission congestion is likely to be pervasive during that hour. All units face relatively small local markets, but the first-stage regression identifies that

43. In this analysis, the fixed effects allow for each unit and each hour of sample to have a different intercept, and I estimate the average temperature-response function relative to those intercepts. One temperature must be selected as the reference temperature, or the point at which the intercept is calculated. This choice is arbitrary and innocuous. If I selected a different temperature to be the reference, the temperature-response function would simply shift up or down to intersect the x-axis at that temperature.

units matched to the most extreme temperatures face the smallest local markets. This is precisely the innovation of my research strategy. I exploit panel variation in market size, not simply time series variation, so I can causally estimate how firms in the smallest local markets markup their electricity relative to firms in larger local markets during the same hour.

6.2 Reduced-form results

I look at the markups submitted at three quantities of production: 85%, 75%, and 65% of nameplate capacity.[44] I show how matched temperature affects markups at each of these quantities through the reduced-form regression:

$$Markup_{it} = f(MatchTemp_{it}) + g(LocalTemp_{it}) + \upsilon_{ihm} + \phi_t + \psi_{it} \tag{14}$$

where the terms in Equation (14) are equivalent or comparable to those in Equations (10) and (11).

The results of these reduced-form regressions are plotted in Figure 11.[45] The effect on markups is relatively flat and not statistically different from zero at temperatures below 30°C (86°F), and this is true for all three quantities. When temperatures exceed 30–35°C (86–95°F), however, markups increase significantly. At the extreme, when temperatures exceed 40°C (104°F), markups at 85% capacity are as much as $220 per MWh higher than during comparable hours at a moderate temperature; this effect of extreme heat is substantial — more than six times as large as the average marginal cost for a fossil fuel-fired power plant and nearly twice as large as the average markup at this quantity. Similarly, markups at 75% and 65% of nameplate capacity are as much as roughly $140 and $120 per MWh higher, respectively, when temperatures exceed 40°C (104°F), relative to moderate temperatures; the effect at each of these quantities is 1.35 and 1.65 times as large the average markup at the respective quantity. The range of temperatures that induces these large effects on markups corresponds to temperatures that cause local markets to shrink, further suggesting a strong correlation between smaller market size and higher markups.

44. See Sections 2 and 4 for a description of offer curves and an explanation for focusing on these three points on the curve.

45. As in Figure (10), the points and error bars show estimates and 95% confidence intervals for the binned specification, and the curve and shaded area show the estimate and 95% confidence interval for the natural cubic spline. I again estimate effects relative to a temperature of 16.25°C (61.25°F), so effects are given not in absolute levels but relative to a comparable hour at the reference temperature.

6.3 Results for local market capacity

I estimate the causal effect of market size on markups using 2SLS as described in Equations (10) and (11). I first consider market size measured by local market capacity, and the results are shown in Table 3. This table reports results for markups at all three quantities; for each quantity level, the table reports three regressions: OLS, 2SLS using the temperature bin first-stage specification, and 2SLS using the natural cubic spline first-stage specification. Looking first at markups at 85% capacity, the OLS regression finds no effect of local market capacity on markups; the estimate is small and not statistically significant. The 2SLS estimates, however, are large and indicate markups are higher when markets are smaller. For every one less gigawatt of power plant capacity in the local market, a firm marks up prices by as much as an additional \$25 per MWh. Put differently, for a reduction in local market capacity equal to 10% of mean local market capacity, markups increase by up to \$135 per MWh. This effect is large — more than 3.5 times the average marginal cost of a fossil fuel-fired power plant and nearly 1.2 times the average markup I observe at 85% capacity.

A more than doubling of markups for a relatively small change in local market capacity may seem extreme, but it is again useful to consider the identifying variation for this estimation. These 2SLS regressions are identified off of variation in local market capacity caused primarily by hours of extreme heat. During these hours, the grid is likely stressed due to congested transmission lines and high levels of demand. When this situation occurs, most local markets are operating on the inelastic portions of their aggregate offer curves, so firms face inelastic residual demand and have strong incentives to exercise market power by marking up prices. If a local market in this scenario becomes even smaller, the residual demand facing every unit becomes substantially more inelastic, strengthening the incentives for every firm to exercise market power and yielding large increases in markups.

While these hours of extreme heat events represent only a minority of the hours when this market operates, they correspond to some of the most challenging hours for the operation of the electrical grid, and extreme levels of market power may exacerbate these challenges. Additionally, these times of extreme heat will become more common in the face of climate change, which I explore in Section 7. By employing an empirical strategy that exploits variation during these hours, this analysis contributes to better understanding one of the driving factors of market power during these critical times.

Returning to the results in Table 3, the effects of local market size on markups at 75% and 65% of capacity are similar to those at 85% of capacity, although effect sizes are smaller in absolute value at each subsequently smaller quantity. These results show that, as local market capacity shrinks, not only do markups increase but also offer curves become more steep.[46] Thus, when a local market is smaller, firms face fewer competitors, and those competitors submit offers that are less elastic. Both of these factors contribute to a firm facing less elastic residual demand, which increases the firm's incentive to further markup prices. These regressions estimate the combined effect of these two factors.

Across all three quantities, the OLS and 2SLS estimates are vastly different. The OLS results show that markups are uncorrelated with the size of the market ultimately faced by the firm, which is consistent with local market size being, on average, unpredictable by firms. As shown in Section 5, however, a portion of the variation in local market size is predicted by temperature. By instrumenting for market size with matched temperature in the 2SLS regressions, I isolate the variation in market size that a firm is able to predict when submitting offer curves for its power plants. As a result, the 2SLS regressions estimate a local average treatment effect (LATE). That is, they estimate the effect on markups of a predictable change in local market capacity that is induced by temperature, but not the effect of a change in local market capacity due to a different and potentially unpredictable cause.

One concerning set of results in Table 3 is the first-stage F-statistic for each 2SLS estimation. The largest F-statistic results from using the natural cubic spline first-stage specification for markups at 85% capacity, shown in Column (3). But this F-statistic is only 6.47, which falls short of the F-statistic "rule of thumb" of 10 (Staiger and Stock 1997), indicating the relevance assumption of Equation (12) may not be sufficiently strong and these 2SLS estimates may be biased due to a weak instruments problem. Using the Stock and Yogo (2005) test for 2SLS bias due to weak instruments, I reject that the bias of 2SLS relative to the bias of OLS is greater than 30% for the spline specifications, which is less than one standard error of my estimate. Additionally, the direction of the bias is toward the OLS estimate, so 2SLS estimates are biased toward zero. Thus, I interpret these 2SLS estimates as conservative estimates of the true LATE with only a small potential bias. For the remainder of this analysis, however, I focus on results for markups at 85%

46. I assume a unit's marginal cost is constant in quantity, so larger markup effects at larger quantities must be due to larger increases in the offer price — that is, an offer curve that becomes more steep in response to less local market capacity. I directly estimate effects on offer curves in Appendix D and show this to be true.

capacity using the natural cubic spline first-stage specification, which uniformly exhibit the largest first-stage F-statistic.[47]

6.4 Results for number of units or firms in local market

Tables 4 and 5 display the results of regressions using the number of units in the local market and the number of firms in the local market, respectively, as the measure of market size. As with the first-stage results in Figure 10, these regression results are qualitatively consistent with those for local market capacity because all three measures of market size are highly correlated, although the estimates differ to account for the scale of the market size measure being considered.

Focusing on effects at 85% capacity, a firm increases markups by an additional \$10 per MWh for every one fewer unit competing in a local market and by more than \$40 per MWh for every one fewer firm competing in the local market, as shown in Column (3) of each table. These estimates imply that a firm increases markups by \$155–180 per MWh if the size of the market shrinks by 10% of the mean market size. These increases in markups are even larger than for local market capacity — more than four times the average marginal cost of a fossil fuel-fired power plant and 140–160% of the average markup I observe at 85% capacity. The effects on markups at smaller quantities are similarly comparable to those for local market capacity.

6.5 Decomposition of effects

The results in Tables 3–5 give the average effect of market size on markups. As shown in the comparative statics of Section 3, however, different types of capacity and units may have vastly different effects on markups. For example, Equation (5) predicts potentially opposite effects on a plant's markups depending on the ownership of local market capacity — more competing capacity reduces markups, whereas more capacity owned by the plant's own firm may increase the incentive to mark up offer prices. Similarly, Equation (6) predicts different magnitudes of effects depending on whether capacity in the local market is dispatchable or not, with dispatchable capacity yielding larger effects in absolute value.[48]

47. I present results for markups at 75% and 65% capacity using the natural cubic spline first-stage specification and for markups at 85% capacity using the binned first-stage specification in Appendix D.

48. I define nondispatchable capacity as nuclear and renewable plants, and the remaining capacity as dispatchable. The nondispatchable plants almost always submit offer curves at negative prices and with approximately no slope at the quantities where the plants typically operate, whereas the dispatchable plants typically submit offer curves with prices that increase in quantity. See Figure 4 for example offer curves and Figure 5 for the mean offer curve by plant types.

To test these predictions, I regress markups on two market size variables. In the first case, these variables are local market capacity owned by competing firms and local market capacity owned by the plant's own firm; in the second, dispatchable capacity and nondispatchable capacity. I again use a 2SLS approach and estimate each size variable as a function of matched temperature in the first stage. Table 6 displays the results of regressions to test these predictions on markups at 85% capacity.[49] Column (1) duplicates the main regression result from Column (3) of Table 3 to serve as a reference.

Column (2) of Table 6 shows the effect on markups by ownership, which confirms the predictions of the model. Less competing capacity in a local market causes markups to increases at nearly double the rate of the average effect, although these point estimates are not statistically different from one another. Conversely, and as predicted, markups increase when there is more capacity in a plant's local market that is owned by the plant's firm, and the size of this effect is an order of magnitude larger than the effect of competing capacity.

Column (3) of Table 6 shows the effect on markups by unit type. Dispatchable capacity, which contributes more to the slope of the residual demand curve, has an effect on markups that is roughly equivalent to the average effect. Nondispatchable capacity, however, has no statistically distinguishable effect on markups. These results confirm the results of the model, that capacity with less elastic offer curves have a larger effect on markups because the offer curves have a larger effect on the elasticity of residual demand.

6.6 Robustness checks for local controls and temperature matching

An important feature of my identification strategy is matching each power plant to a weather station that is geographically far from the power plant, so I can also control for the temperature at the weather station closest to the power plant. This approach, however, has two potential concerns. First, local temperature may not be sufficient to fully control for conditions in the plant's local market, particularly when that local market is large. Second, if there is insufficient variation in matched temperature after controlling for local temperature and fixed effects, then estimates may be attenuated. Second, local temperature may not be sufficient to control for local market conditions when the power plant's local market is large. To explore these concerns, I use alternate local controls and matched weather stations.

49. Results for markups at other quantities, as well results using number of units or firms in the local market, are in Appendix D.

I first consider different local controls. The primary reason to include local temperature is to control for the level of demand in the local market. Higher levels of demand cause the local market to clear further out on the aggregate offer curve, which corresponds to less elastic supply and may affect markups independent of local market size. To directly control for where the local market clears, I include a higher-order function of local capacity factor — the ratio of production to capacity in the local market — as a control, both instead of and in addition to local temperature. To address the second concern, I also consider a model with no local controls. Table 7 displays the results of these regressions for the effect of local market capacity on markups at 85%.[50] The estimated effect changes little and no estimates are statistically different from one another.

Another method to address both concerns is to require a power plant's matched weather station to be a given distance away from the plant.[51] A minimum matching distance both increases the likelihood that a plant's matched temperature is outside its local market and decreases the correlation between matched and local temperature, to ensure more temperature variation remains when controlling for local temperature. Table 8 displays the results when requiring a plant's matched weather station to be 100, 200, or 300 km (62, 124, or 186 miles) away from the plant. The effect of local market capacity on markups at 85% capacity is not statistically different at any of these minimum matching distances.

6.7 Matched temperature or local temperature

An important innovation of this analysis is the use of a prediction model to match each power plant to the weather station that is most predictive of the plant's local market capacity, rather than simply using the local temperature closest to the power plant. Not only is this matching procedure necessary for causal identification — instrumenting with local temperature does not satisfy the exclusion restriction, as discussed previously — but it also exploits the variation that firms are more likely to consider when submitting offer curves. As a result, 2SLS regressions using matched temperature should yield less bias toward zero, as compared to estimates that use local temperature. To see the importance of this strategy and its effects on my estimates, I also perform 2SLS regressions using local temperature, rather than matched temperature, as the instrument for local market size.

50. See Appendix D for results using the other measures of local market size and other quantities.

51. I enforce no such requirement in the original match, even allowing a power plant to be matched to its closest weather station. Only three plants are originally matched to the closest weather station.

Table 9 shows the results of these regressions; Columns (1)–(3) duplicate previous results using matched temperature, and Columns (4)–(6) show the estimates from comparable regressions but with local temperature as the instrument. For all three measures of market size, the estimates when instrumenting with local temperature are roughly one-fourth of the estimates when instrumenting with matched temperature, and the estimates using local temperature are not statistically significant. These results show that my empirical strategy is necessary not only to find the causal effect of market size on markups but to find any empirical relationship between market size and the exercise of market power, as predicted by theory, that is not subject to large attenuation bias.

7 Welfare effects

The exercise of market power through increased markups has two main effects on welfare in this setting: large transfers from consumers to producers and a deadweight loss of consumer surplus. Increased markups yield higher prices paid by the purchasers of wholesale electricity, primarily retail providers that purchase electricity to distribute to their end-use consumers. The retail electricity prices paid by the consumers are contracted in advance and rarely vary with the wholesale price in the real-time market. Thus, in the short run, increased markups cause transfers from electricity retail providers to producers of electricity but no change in the quantity of electricity produced and consumed.

In the long-run, however, retailers pass their increased wholesale costs on to consumers in the form of higher retail prices. Thus, in equilibrium, average retail electricity prices are higher in this setting than they would be in a competitive market, even though wholesale prices are not passed on to end-use consumers in real time. Higher average retail prices have two effects. First, they yield transfers from consumers to retailers and then, ultimately, to the producers of electricity. Second, they produce a long-run demand response with less electricity consumed than in a competitive counterfactual. The transfers do not create an inefficiency in this market but may be of interest for equity concerns. The reduction in consumption, however, creates a deadweight loss of consumer surplus.

I calculate the size of the transfers and deadweight loss due to increases in markups that are induced by transmission congestion and smaller local markets. I then calculate the extent to which these welfare effects will increase as a result of climate change and the higher prevalence of transmission congestion.

7.1 Consumer surplus and transfers to producers

To calculate the welfare effects of this source of market power, I consider the counterfactual in which firms submit offer curves as if temperature-induced transmission congestion did not reduce the size of local markets. The first-stage results in Figure 10 show the extent to which a plant's local market size varies as a function of temperature at the plant's matched weather station. I focus on temperatures greater than 30°C (86°F), which correspond to the matched temperatures at which firms change their markups, as shown in the reduced-form results in Figure 11.

I use the first-stage estimates from the natural cubic spline specification to calculate how much larger each plant's local market capacity would be in the absence of temperature-induced transmission congestion — that is, the local market capacity at a moderate temperature,[52] as compared to the observed matched temperature. At these moderate temperatures, local markets are, on average, 1.1 GW larger. There is great heterogeneity, however, and when compared to the most extreme matched temperatures, local market capacity can be as much as 7.7 GW larger.

If firms did not consider local market size when submitting offer curves, markups would be smaller during these hot hours. Using the 2SLS estimates from Column (3) in Table 3, I calculate that markups would be $27.1 per MWh smaller on average, and some markups would be as much as $195.5 per MWh smaller. These smaller markups would yield lower LMPs at every power plant; I assume the reduction in LMPs at each power plant is equal to the reduction in its markup. Due to these lower markups, and their resulting lower prices, the total value of the electricity transacted in the real-time wholesale electricity market during my study period is $8.3 billion less in this counterfactual. This value represents only the lower prices paid by electricity retailers in the wholesale market, which they pass on to consumers, and not a change in the quantity of electricity produced or consumed. Thus, the total transfer from consumers to producers due to this exercise of market power is $2.1 billion annually.

These lower wholesale prices would yield lower average retail prices paid by end-use consumers. Over the four years of my study period, nearly 1.3 billion MWh of electricity was consumed. Thus, without this source of market power, the average retail price of electricity would be $6.5 per MWh lower. During these years, consumers in the state of Texas paid an average of $87.9 per MWh for electricity,[53] so roughly 7.4% of the average retail price is due to this source of market power. Recent

52. I use the local market size at the reference temperature of 16.25°C (61.25°F). A different moderate temperature would yield similar results because the local market capacity is not statistically different over the range 5–30°C (41–85°F).

53. Annual average retail electricity price is from the EIA's Electric Power Annual.

estimates of the long-run elasticity of demand for electricity range from -0.09 to -0.27 (Deryugina, MacKay, and Reif 2017).[54] Assuming a constant elasticity of demand for electricity, these elasticity estimates imply electricity consumption is 0.69–2.09% higher in the counterfactual without these markups. Thus, as compared to this counterfactual, consumers in this market lose \$7.1–21.5 million of surplus annually due to higher retail prices and a lower level of electricity consumption.

These calculations show that, in this setting, there is little welfare loss due to the exercise of market power that is induced by transmission congestion and smaller lower markets. These markups result in only \$7.1–21.5 million of deadweight loss annually, in a market that transacts nearly \$12 billion of electricity each year. This source of market power does, however, result in large transfers from consumers to producers, \$2.1 billion annually. These transfers are substantial and may be a cause for concern on equity grounds.

7.2 Climate change effects

Because transmission congestion is driven by ambient temperature, climate change is projected to have an important effect on the size of local markets and the resulting exercise of market power. To see this effect, I consider a counterfactual in which temperatures are 2.5°C (4.5°F) warmer than I observe during my study period, which corresponds to roughly 50 years in the future under a business-as-usual warming scenario.[55]

I first examine how climate change will alter the prevalence of transmission congestion and the number of local markets that form. I calculate the distribution of the number of local markets as a nonparametric function of mean temperature in the state. I then assign to each hour the distribution of local markets that corresponds to a temperature that is one bin warmer. Figure 12 shows the results of this calculation. The top panel displays histograms of the number of local markets that form in the observed data and in the counterfactual under 2.5°C (4.5°F) warming; the bottom panel shows the difference between the warming scenario and the observed data. Due to a warmer climate, the portion of time that the market is fully integrated will decrease by 2.7 percentage points. Instead, it is more likely the market will fragment into many markets. In particular, the

54. These elasticity estimates are for residential consumers. Commercial and industrial consumers are likely to exhibit larger elasticities, so the resulting deadweight loss is a conservative estimate.

55. 2.5°C (4.5°F) warming by 2065, as compared to 2011–2014, is roughly consistent with temperature projections under RCP 8.5 (Riahi et al. 2011). Because I use 2.5°C (4.5°F) temperature bins in my nonparametric temperature-response functions, this warming is equivalent to counterfactual temperatures that are one temperature bin warmer than what I observe in the data.

amount of time that the market splits into 7 or more local markets will increase by 3.3 percentage points.

This increased frequency of transmission congestion results in smaller local markets and, hence, increased market power through markups. Again using the first-stage estimates from the natural cubic spline specification, as shown in Figure 10, I calculate that power plants will face local markets that are 0.3 GW smaller on average, due to warming. These effects are not linear, however, and for power plants with matched weather stations that already experienced extreme temperatures, climate change will reduce the capacity of their local markets by as much as 1 GW in some hours. Using the 2SLS estimates from Column (3) in Table 3, I calculate that markups will increase by \$7.1 per MWh on average, and some markups will be as much as \$26.0 per MWh larger.

These increased markups will yield higher LMPs; I again assume the increase in each power plant's LMP is equal to the increase in the plant's markup. Due to these higher prices, an additional \$2.5 billion would be transacted annually in the real-time wholesale electricity market if the same amount of electricity was consumed. In the long run, however, these increased markups will result in average retail prices that are \$7.8 per MWh, or 8.8%, higher than prices during 2011–2014. These increased retail prices will yield less consumption; using the same range of elasticities as above, consumption will decrease by 0.76–2.26%.[56] This reduction in consumption corresponds to a consumer surplus loss of \$9.4–27.8 million annually.

Thus, climate change will yield a deadweight loss of \$9.7–28.8 million annually because of the effect of warmer temperatures on local market formation and the resulting exercise of market power by firms. This value is roughly equivalent to the deadweight loss due to the same source of market power during the study period. Climate change will also generate large transfers from consumers to producers, \$2.4–2.5 billion annually, as a result of increased markups, which is again comparable in size to what I calculate for the years 2011–2014. Thus, both deadweight loss and transfers from consumers to producers, which result from higher markups due to temperature-induced changes in local market size, will roughly double as a result of climate change. An important caveat is that this analysis considers the effect of warmer temperatures on the supply and transmission infrastructure present in 2011–2014, and much of this infrastructure will be updated or replaced within the next

56. I consider the effect of climate change only on the supply and transmission of electricity, not on the demand for electricity. See Auffhammer, Baylis, and Hausman (2017) for estimates of how climate change is expected to increase electricity consumption. At higher levels of electricity demand, the effects of increased markups are larger, so my calculations provide conservative estimates of the welfare effects of climate change in this setting.

50 years. These results highlight the importance of considering climate change when making these long-run planning and investment decisions.

8 Conclusion

This paper studies the effect of market size on market power in the Texas electricity market. I exploit a novel source of exogenous variation in market size, transmission congestion, to estimate how the size of the market in which a power plant operates affects its markups. I find that a 10% reduction in market size causes markups to more than double, compared to mean values. These results show that even a relatively small change in market size can greatly exacerbate the ability of a firm to exercise market power in this setting. These markups yield a small loss of consumer surplus, \$7.1–21.5 million annually, because demand for electricity, even in the long run, is relatively inelastic. These markups do, however, generate large transfers from consumers to producers, \$2.1 billion annually, which may be a cause for concern on equity grounds.

This work makes an important contribution to the literature by estimating the effect of market size on market power using weaker assumptions than prevailing approaches. Transmission congestion provides an exogenous source of variation that I use to estimate this effect without making structural assumptions on firm behavior or the nature of competition. It is, however, an hourly source of variation, meaning my estimates are short-run effects. In this setting, short-run effects are particularly important because supply must perfectly meet demand in real time. While these estimates may not be directly transferable to other contexts, the direction and magnitude of the effects give important information about how sophisticated firms compete in a dynamic market with large stakes.

This distinction between the short-run and long-run effects of market size on market power points to an important area of future research. In electricity markets, increased markups due to transmission congestion may lead to new investments in the long run. Firms may site new power plants in areas of the grid that often experience small local markets, which would decrease the market power of the incumbents, but this investment may not be socially optimal since market power causes relatively small deadweight loss in this setting. Understanding how this source of short-run market power affects long-run market outcomes, including trade-offs between short-run and long-run inefficiencies, is an important area of future research.

References

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: An Inverted-U Relationship." *The Quarterly Journal of Economics* 120 (2): 701–728. Accessed November 11, 2018.

Auffhammer, Maximilian. 2018. "Climate Adaptive Response Estimation: Short and Long Run Impacts of Climate Change on Residential Electricity and Natural Gas Consumption Using Big Data." NBER Working Paper 24397. Accessed November 12, 2018.

Auffhammer, Maximilian, Patrick Baylis, and Catherine H. Hausman. 2017. "Climate Change Is Projected to Have Severe Impacts on the Frequency and Intensity of Peak Electricity Demand Across the United States." *Proceedings of the National Academy of Sciences* 114 (8): 1886–1891. Accessed November 11, 2018.

Auffhammer, Maximilian, and Erin T. Mansur. 2014. "Measuring Climatic Impacts on Energy Consumption: A Review of the Empirical Literature." *Energy Economics* 46:522–530. Accessed November 11, 2018.

Bain, Joe S. 1951. "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940." *The Quarterly Journal of Economics* 65 (3): 293–324. Accessed November 11, 2018.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* 80 (6): 2369–2429. Accessed November 11, 2018.

Bergquist, Lauren Falcao. 2017. "Pass-Through, Competition, and Entry in Agricultural Markets: Experimental Evidence from Kenya." Working Paper.

Berry, Steven, and Peter Reiss. 2007. "Empirical Models of Entry and Market Structure." In *Handbook of Industrial Organization: Volume 3,* edited by Mark Armstrong and Robert H. Porter, 1845–1886. Amsterdam: Elsevier, North Holland. Accessed November 11, 2018.

Borenstein, Severin, and James Bushnell. 1999. "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry." *The Journal of Industrial Economics* 47 (3): 285–323. Accessed November 11, 2018.

Borenstein, Severin, James B. Bushnell, and Frank A. Wolak. 2002. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92 (5): 1376–1405. Accessed November 11, 2018.

Borenstein, Severin, James Bushnell, and Steven Stoft. 2000. "The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry." *The RAND Journal of Economics* 31 (2): 294–325. Accessed November 11, 2018.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization: Volume 2,* edited by Richard Schmalensee and Robert Willig, 1011–1057. Amsterdam: Elsevier, North Holland. Accessed November 11, 2018.

Bresnahan, Timothy F., and Peter C. Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy* 99 (5): 977–1009. Accessed November 11, 2018.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Global Non-Linear Effect of Temperature on Economic Production." *Nature* 527 (7577): 235–239. Accessed November 12, 2018.

Bushnell, James B., Erin T. Mansur, and Celeste Saravia. 2008. "Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets." *American Economic Review* 98 (1): 237–266. Accessed November 11, 2018.

Bushnell, James B., and Catherine Wolfram. 2005. "Ownership Change, Incentives and Plant Efficiency: The Divestiture of U.S. Electric Generation Plants." Center for the Study of Energy Markets Working Paper 140. Accessed November 12, 2018.

Cournot, Augustin. 1838. "Of the Influence of Taxation on Commodities Produced under a Monopoly." In *Researches into the Mathematical Principles of the Theory of Wealth,* translated by Nathaniel T. Bacon, 67–78. New York: The Macmillan Company, 1897. Accessed November 11, 2018.

Davis, Lucas, and Catherine Hausman. 2016. "Market Impacts of a Nuclear Power Plant Closure." *American Economic Journal: Applied Economics* 8 (2): 92–122. Accessed November 11, 2018.

De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper 23687. Accessed November 11, 2018.

———. 2018. "Global Market Power." NBER Working Paper 24768. Accessed November 11, 2018.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics* 4 (3): 66–95. Accessed November 12, 2018.

Deryugina, Tatyana, Alexander MacKay, and Julian Reif. 2017. "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation." NBER Working Paper 23483. Accessed November 12, 2018.

Deschênes, Olivier, and Michael Greenstone. 2007. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *American Economic Review* 97 (1): 354–385. Accessed November 12, 2018.

———. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–185. Accessed November 12, 2018.

Deutch, John M., Charles W. Forsberg, Andrew C. Kadak, Mujid S. Kazimi, Ernest J. Moniz, and John E. Parsons. 2009. "Update of the MIT 2003 Future of Nuclear: An Interdisciplinary MIT Study."

ERCOT (Electricity Reliability Council of Texas). 2018. "ERCOT Quick Facts." August 8, 2018. Accessed November 12, 2018.

Fabra, Natalia, and Mar Reguant. 2014. "Pass-Through of Emissions Costs in Electricity Markets." *American Economic Review* 104 (9): 2872–2899. Accessed November 11, 2018.

Gaynor, Martin, and Robert J. Town. 2011. "Competition in Health Care Markets." In *Handbook of Health Economics: Volume 2,* edited by Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, 499–637. Handbook of Health Economics. Amsterdam: Elsevier, North Holland. Accessed November 11, 2018.

Green, Richard J., and David M. Newbery. 1992. "Competition in the British Electricity Spot Market." *Journal of Political Economy* 100 (5): 929–953. Accessed November 14, 2018.

Harberger, Arnold C. 1954. "Monopoly and Resource Allocation." *The American Economic Review* 44 (2): 77–87. Accessed November 11, 2018.

Hortaçsu, Ali, Fernando Luco, Steven L Puller, and Dongni Zhu. 2017. "Does Strategic Ability Affect Efficiency? Evidence from Electricity Markets." NBER Working Paper 23526. Accessed November 11, 2018.

Hortaçsu, Ali, and Steven L. Puller. 2008. "Understanding Strategic Bidding in Multi-Unit Auctions: a Case Study of the Texas Electricity Spot Market." *The RAND Journal of Economics* 39 (1): 86–114. Accessed November 11, 2018.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. "Quantifying the Influence of Climate on Human Conflict." *Science* 341 (6151): 1235367. Accessed November 12, 2018.

Hsiang, Solomon M., Kyle C. Meng, and Mark A. Cane. 2011. "Civil Conflicts Are Associated with the Global Climate." *Nature* 476 (7361): 438–441. Accessed November 12, 2018.

Ito, Koichiro, and Mar Reguant. 2016. "Sequential Markets, Market Power, and Arbitrage." *American Economic Review* 106 (7): 1921–1957. Accessed November 11, 2018.

Joskow, Paul L., and Jean Tirole. 2000. "Transmission Rights and Market Power on Electric Power Networks." *The RAND Journal of Economics* 31 (3): 450–487. Accessed November 11, 2018.

McDermott, Grant R. 2018. "Hydro Power. Market Might." Working Paper.

Mercadal, Ignacia. 2018. "Dynamic Competition and Arbitrage in Electricity Markets: The Role of Financial Players." Working Paper.

Puller, Steven L. 2007. "Pricing and Firm Conduct in California's Deregulated Electricity Market." *The Review of Economics and Statistics* 89 (1): 75–87. Accessed November 11, 2018.

Reguant, Mar. 2014. "Complementary Bidding Mechanisms and Startup Costs in Electricity Markets." *The Review of Economic Studies* 81 (4): 1708–1742. Accessed November 11, 2018.

Riahi, Keywan, Shilpa Rao, Volker Krey, Cheolhung Cho, Vadim Chirkov, Guenther Fischer, Georg Kindermann, Nebojsa Nakicenovic, and Peter Rafaj. 2011. "RCP 8.5—A Scenario of Comparatively High Greenhouse Gas Emissions." *Climatic Change* 109 (1): 33. Accessed November 12, 2018.

Ryan, Nicholas. 2017. "The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market." NBER Working Paper 23106. Accessed November 11, 2018.

Sathaye, Jayant A., Larry L. Dale, Peter H. Larsen, Gary A. Fitts, Kevin Koy, Sarah M. Lewis, and André Frossard Pereira de Lucena. 2013. "Estimating impacts of warming temperatures on California's electricity system." *Global Environmental Change* 23 (2): 499–511. Accessed November 12, 2018.

Schlenker, Wolfram, W. Michael Hanemann, and Anthony C. Fisher. 2005. "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach." *American Economic Review* 95 (1): 395–406. Accessed November 12, 2018.

Schlenker, Wolfram, and Michael J. Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields Under Climate Change." *Proceedings of the National Academy of Sciences* 106 (37): 15594–15598. Accessed November 12, 2018.

Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83 (2): 768–809. Accessed November 12, 2018.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586. Accessed November 12, 2018.

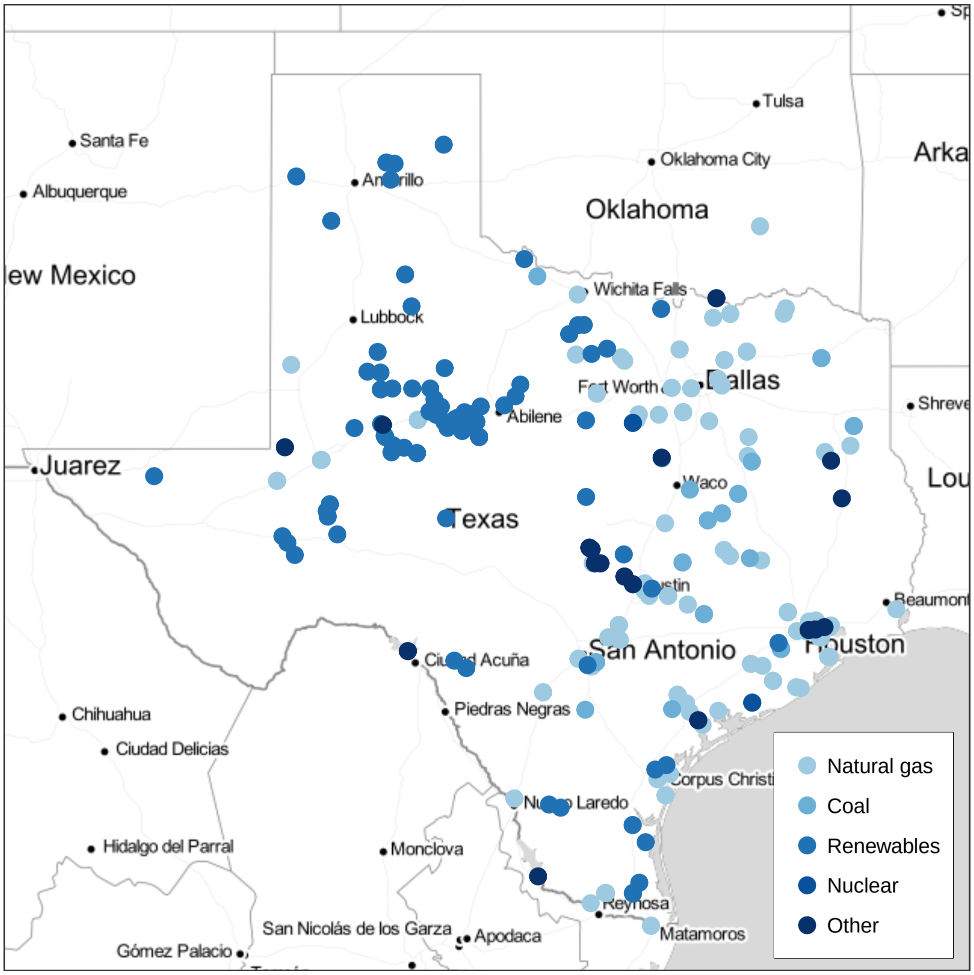
Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* edited by Donald W. K. Andrews and James H. Stock. Cambridge, UK: Cambridge University Press. Accessed November 12, 2018.

Wolfram, Catherine D. 1998. "Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales." *The RAND Journal of Economics* 29 (4): 703–725. Accessed November 11, 2018.

———. 1999. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review* 89 (4): 805–826. Accessed November 11, 2018.

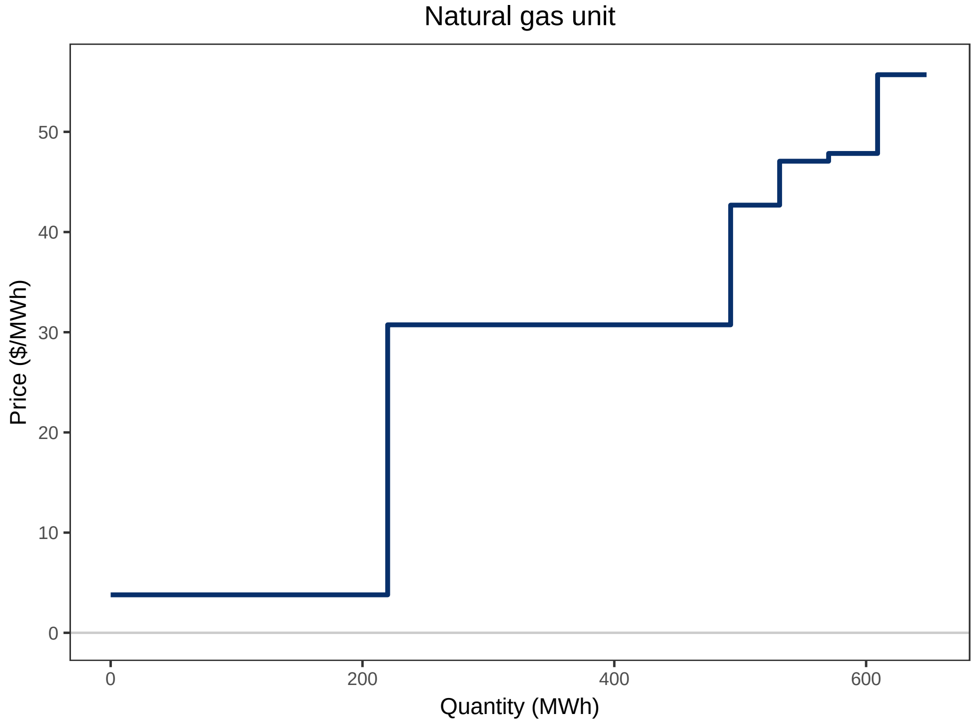
Figures

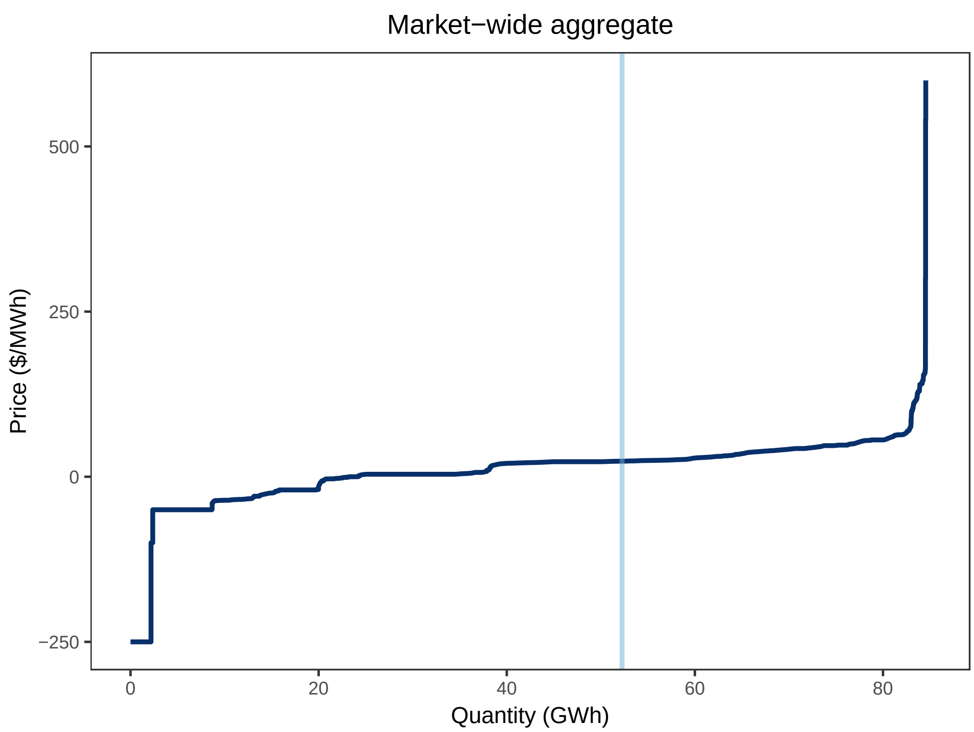
Figure 1: Units by fuel type



Notes: This figure maps all the power plants that supply power to the Texas electricity market. Each point represents one unit in my analysis. The fuel type of each unit is denoted by the color of the point. Plants are generally spread throughout the state, other than renewable plants, which cluster in areas with the greatest wind resources.

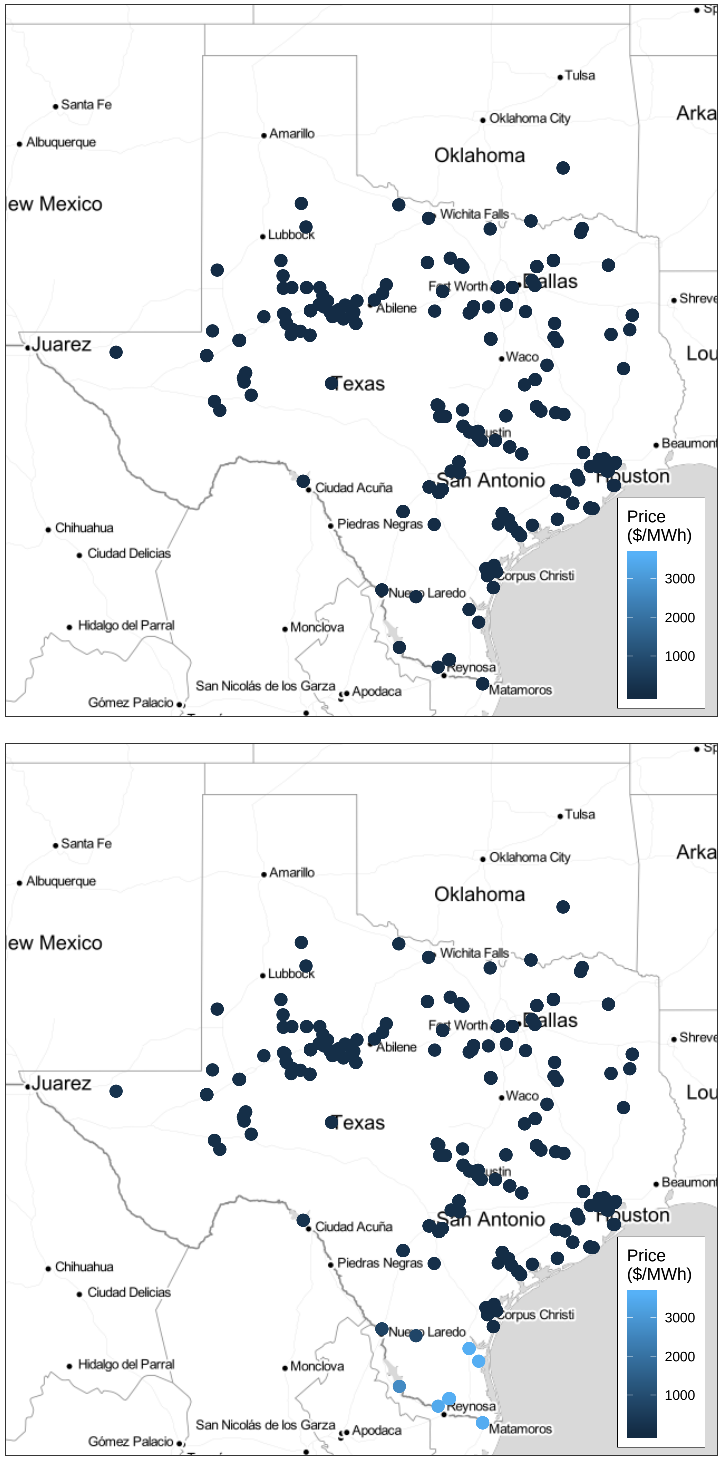
Figure 2: Supply offer curves





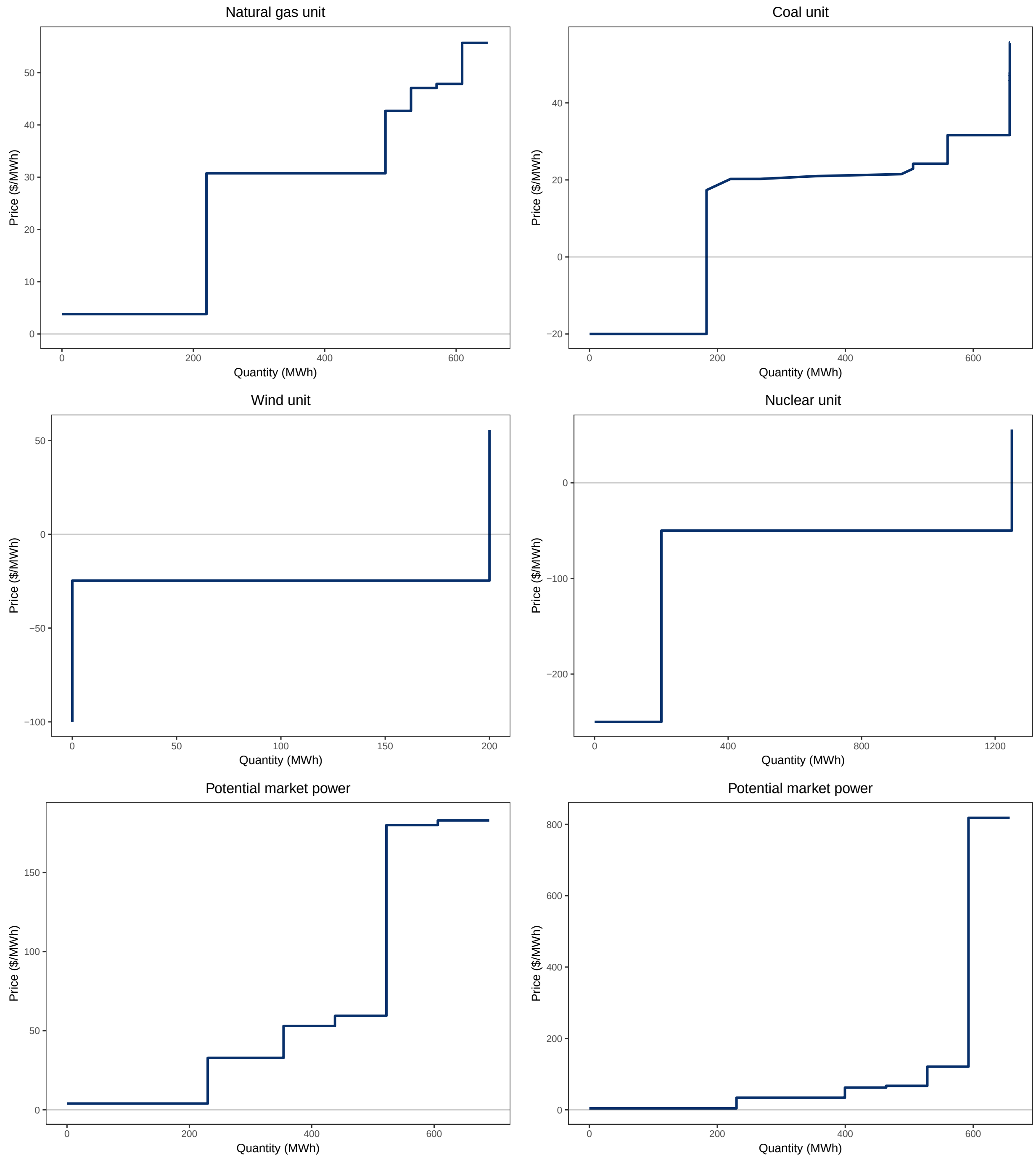
Notes: The top panel plots an example offer curve for a natural gas-fired unit. This curve specifies the amount the unit will produce at each price. The bottom panel plots a market-wide aggregate offer curve, which sums the quantity of all unit-level offer curves for each price. Note that the axes differ from the top panel. The vertical line plots the system-wide demand. In the absence of binding system constraints, the market will clear as a uniform-price auction at the intersection of the two curves.

Figure 3: Local markets and locational marginal prices



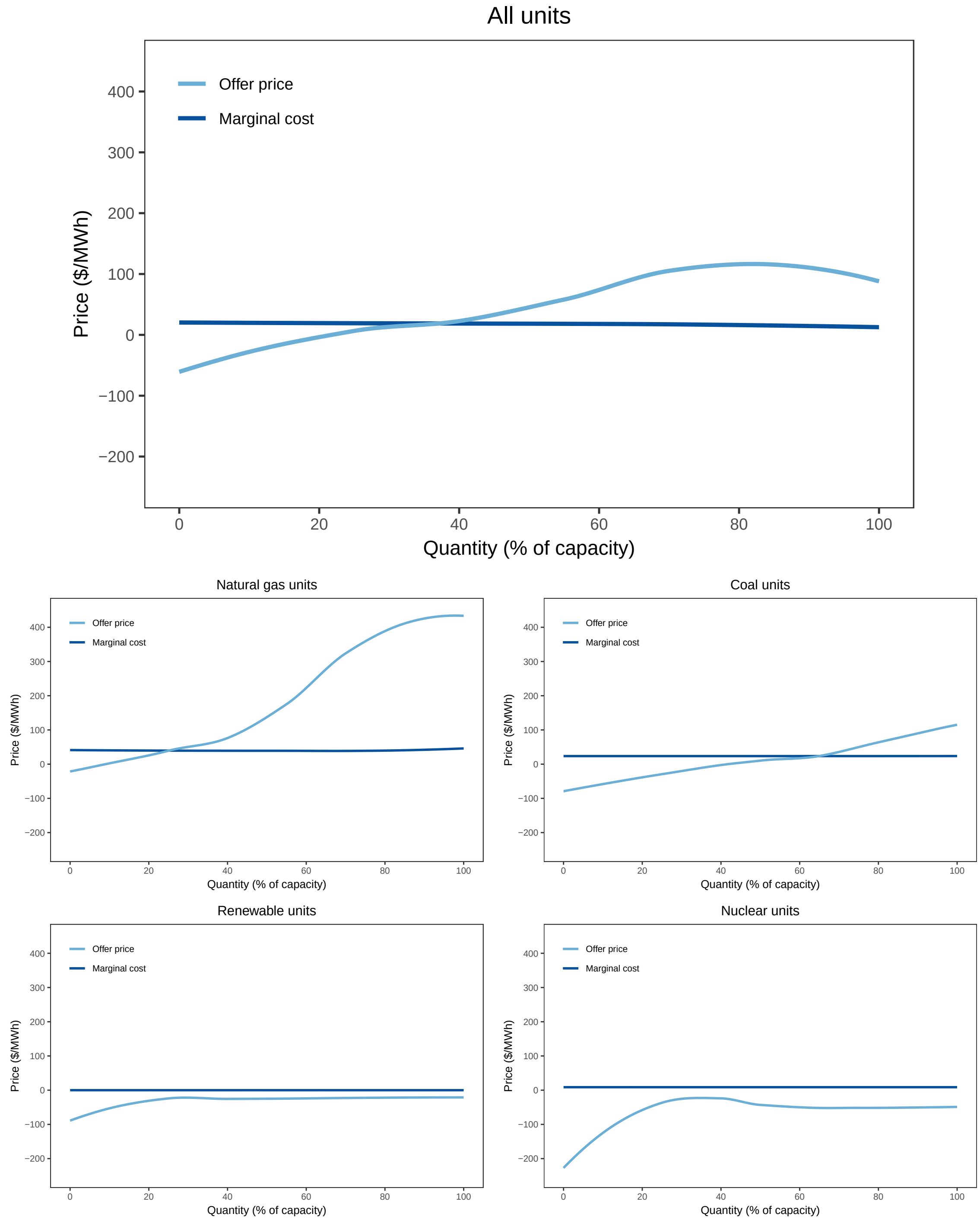
Notes: Both panels map the locational marginal price (LMP) at each unit in the Texas electricity market, but at two different times. The top panel depicts a fully integrated market with no binding constraints, which yields a uniform LMP of roughly $40 per MWh. The bottom panel depicts binding constraints that limit the flow of electricity into the southern portion of the grid and cause the full market to split into two local markets; LMPs in the larger local market remain at roughly $40 per MWh, while LMPs in the smaller local market exceed $3,500 per MWh.

Figure 4: Example supply offer curves



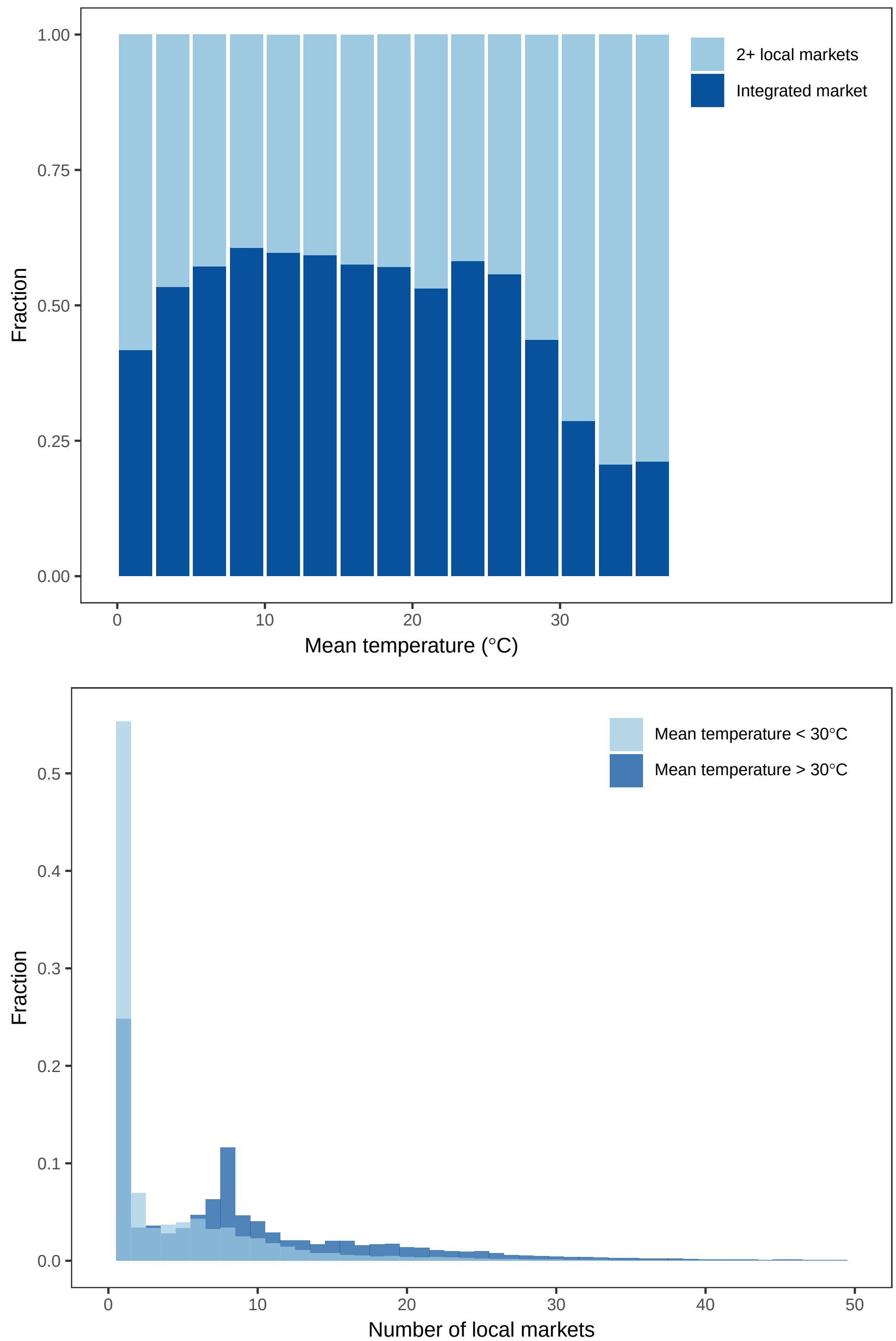
Notes: All six panels plot example offer curves. The top four panels show examples for a natural gas-fired unit, a coal-fired unit, a wind-powered unit, and a nuclear unit, as denoted above each panel. The natural gas-fired example is the same as in Figure 2. The bottom two panels depict examples of potential market power, in which the highest offer price greatly exceeds the unit's marginal cost. Note that the axes differ in all panels.

Figure 5: Average supply offer curves and marginal costs



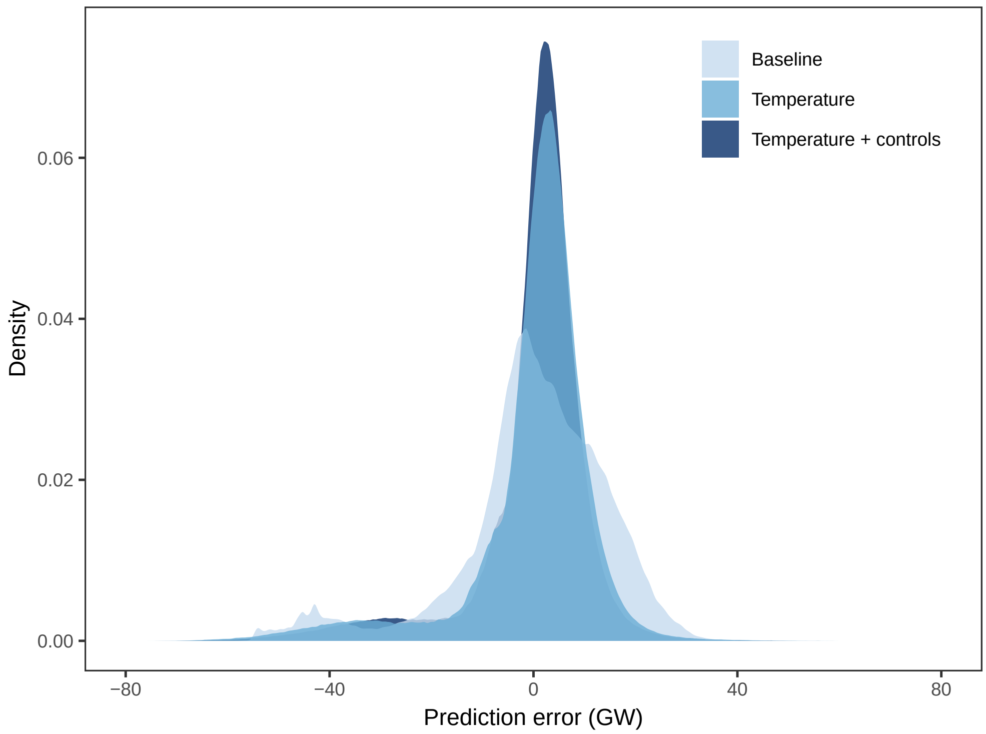
Notes: The top panel depicts the average unit-level offer curve and marginal cost. To construct the average offer curve, I first normalize each offer curve by the unit's nameplate capacity, which yields offer price as a function of the percent of capacity. I then perform a loess regression of offer price on percent of capacity to generate an average offer price at every quantity. Note that not all units offer their full capacity, so the sample changes slightly as quantity increases, particularly as quantity exceeds 90%, which produces the downward slope at large quantities. The bottom four panels depict the average offer curve and marginal cost for each of the four main fuel types, as denoted above each panel.

Figure 6: Number of local markets by temperature



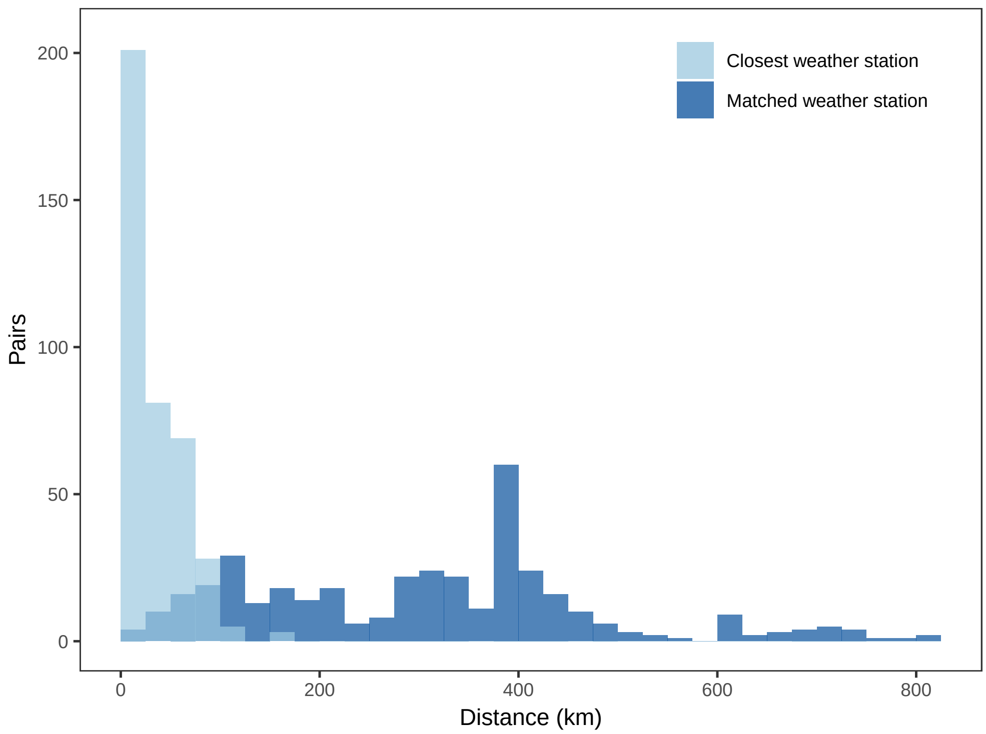
Notes: The top panel plots the fraction of time that the Texas electricity market is a fully integrated market and the fraction of time it is split into multiple local markets, as a function of the mean temperature in the state. Mean temperature is the simple mean of all 75 weather stations in the state. Results are shown by 2.5°C (4.5°F) temperature bins. The first and last bins also include all temperatures below 0°C (32°F) and above 37.5°C (99.5°F), respectively. The bottom panel plots histograms of the number of local markets when the mean temperature is below 30°C (86°F) and when the mean is above this temperature; I omit from this figure the presence of more than 50 local markets, which occurs rarely. At higher temperatures, transmission congestion is more prevalent and the market is more likely to split into many local markets.

Figure 7: Distribution of prediction errors



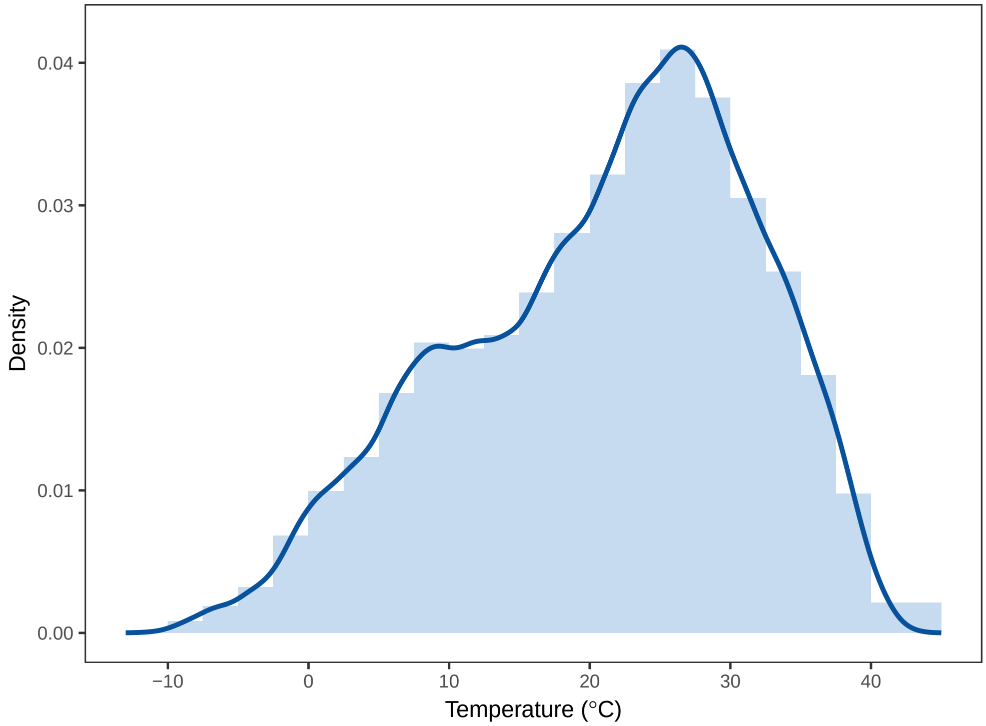
Notes: This figure plots the distributions of prediction errors from three prediction models. Baseline is estimated by Equation (7) and represents the raw variation in the data; this model corresponds to Column (1) of Table 2. Temperature is estimated by Equation (8) with no variables beyond temperature; this model corresponds to Column (2) of Table 2. Temperature + controls is also estimated by Equation (8) but adds system-wide demand and lagged local market capacity, up to a third-order term for each variable; this model corresponds to Column (3) of Table 2. Prediction error is substantially reduced in the Temperature model and further reduced in the Temperature + controls model.

Figure 8: Distance from power plants to weather stations



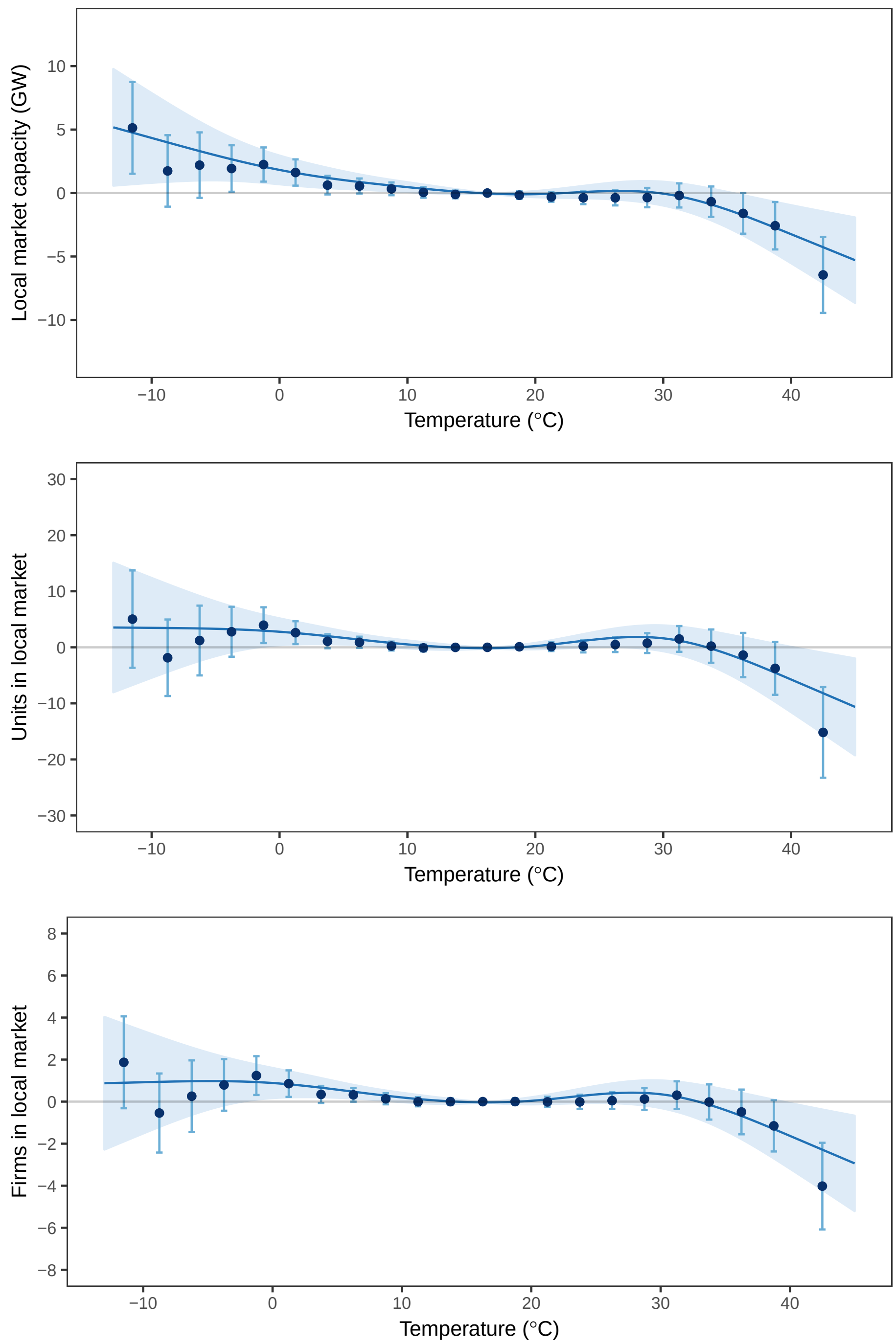
Notes: This figure plots histograms of distances from power plants to weather stations. One histogram corresponds to the weather station closest to the power plant. The other histogram corresponds to the weather station matched to the power plant. The majority of power plants have a weather station within 25 km (15.5 miles), but most are matched to weather stations much farther away.

Figure 9: Distribution of matched temperatures



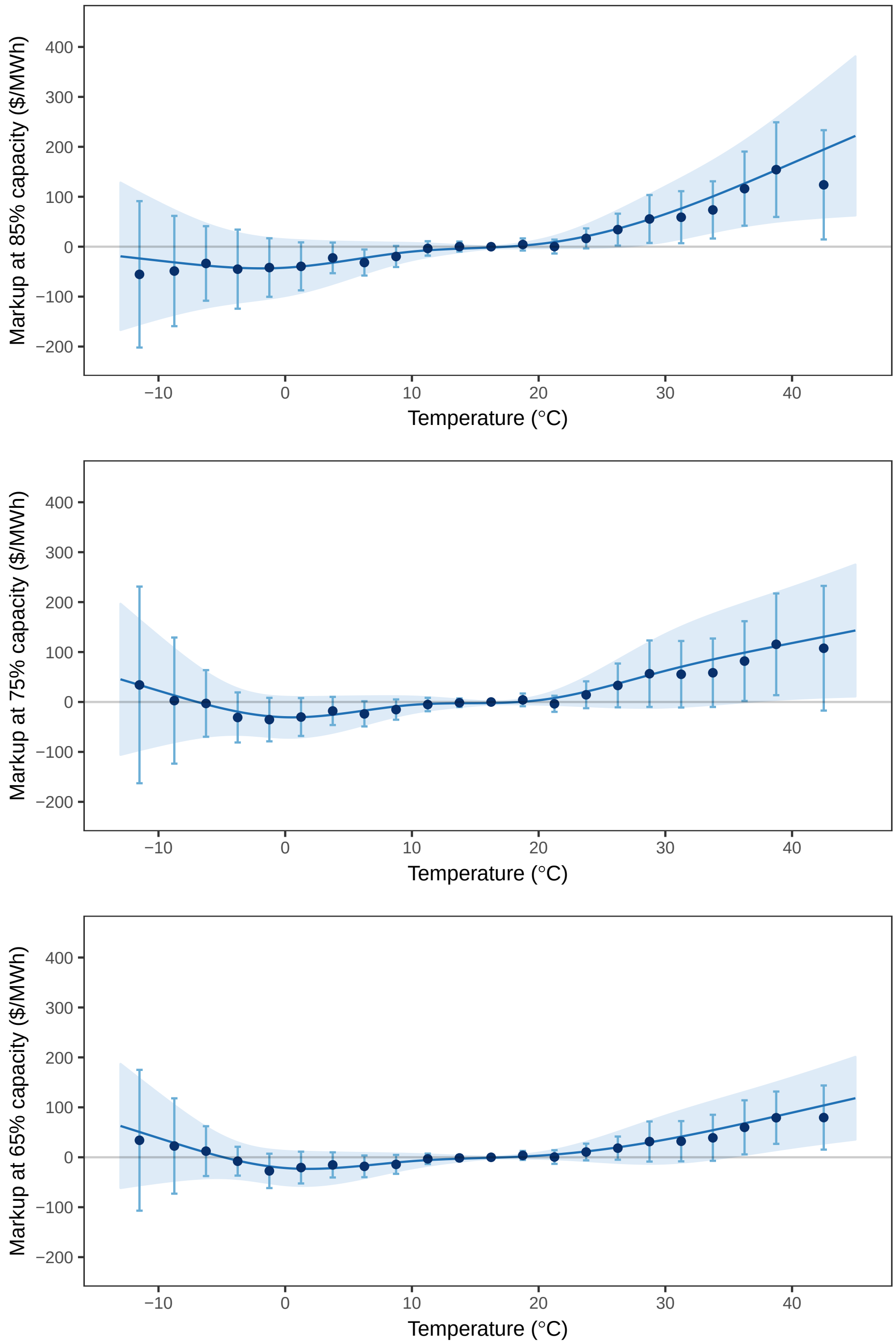
Notes: This figure plots the distribution of matched temperature as both a kernel density and as a histogram of temperature bins. Temperature bins are 2.5°C (4.5°F). The median temperature is 22.7°C (72.9°F), and the modal temperature bin is just above 25°C (77°F). At the upper end of the temperature distribution, temperatures greater than 30°C (86°F) occur in 21.5% of observations, and temperatures exceed 35°C (95°F) in 7.5% of observations.

Figure 10: Temperature-response functions for market size measures



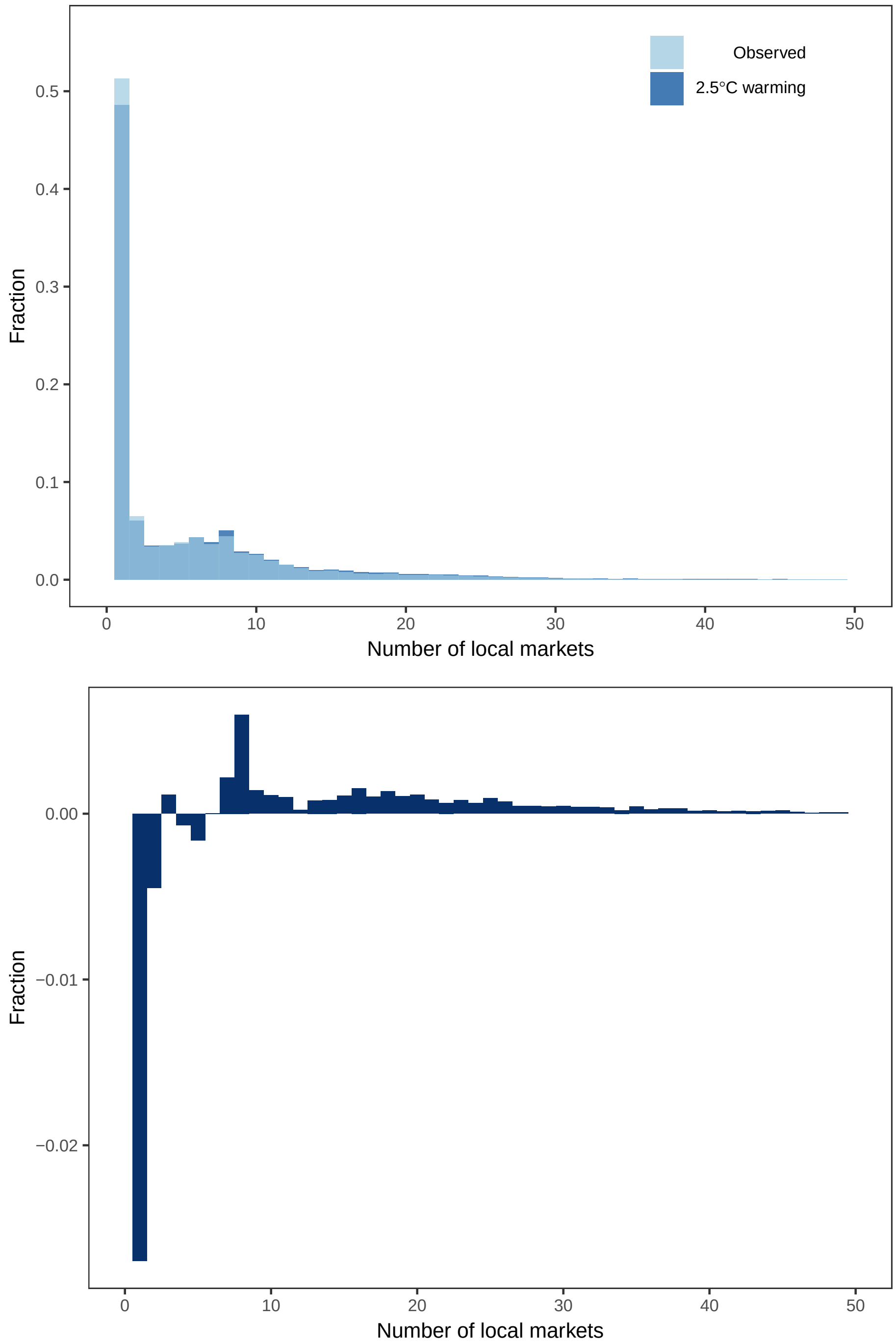
Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (11) for a different measure of local market size. From top to bottom, the measures are local market capacity, number of units in the local market, and number of firms in the local market, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure 11: Temperature-response functions for markups



Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (14) for markup at a different quantity of production. From top to bottom, the quantities are 85%, 75%, and 65%, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure 12: Number of local markets under climate change



Notes: The top panel plots histograms of the number of local markets that I observe in the data and that I project to occur under 2.5°C (4.5°F) warming. The bottom panel plots the differences between the two histograms. Negative values indicate a number of local markets that is less likely to occur with warming, and positive values indicate a number of local markets that is more likely. I omit from this figure the presence of more than 50 local markets, which occurs rarely. In the warming scenario, it is less likely for the market to be fully integrated and more likely for the market to split into seven or more local markets.

Tables

Table 1: Summary statistics by unit type and firm

	Capacity		Mean generation		
	GW	%	GWh	%	Units
Unit types					
Natural gas	58.9	61.6	15.8	41.8	232
Coal	19.9	20.8	13.8	36.7	32
Renewables	10.7	11.2	3.6	9.6	89
Nuclear	5.1	5.4	4.4	11.8	4
Other	1.0	1.0	0.1	0.2	33
Total	95.6	100	37.7	100	390
Firms					
Luminant	14.5	13.7	8.3	20.1	39
NRG Energy	13.9	13.1	6.3	15.1	44
Calpine	10.1	9.5	4.1	10.0	18
CPS Energy	5.8	5.4	2.2	5.3	19
NextEra Energy	5.3	5.0	2.4	5.8	21
GDF Suez	5.1	4.8	2.1	5.0	15
Lower Colorado River Authority	4.0	3.8	1.6	4.0	26
Exelon	3.9	3.7	0.6	1.5	16
Other (65 firms)	35.2	33.2	10.7	26.0	192
Total	106.1	100	41.3	100	390

Notes: The top panel of this table displays capacity, generation, and number of units by unit type. 99.0% of capacity and 99.8% of generation are accounted for by four fuel types: natural gas, coal, renewables, and nuclear. The bottom panel of this table displays capacity, generation, and number of units by firm. The eight largest firms account for 66.8% of capacity and 74.0% generation, with many small firms. By traditional metrics, this market appears competitive.

Table 2: Summary of prediction models

	Baseline	Subset selection		LASSO	
	(1)	(2)	(3)	(4)	(5)
Mean squared error	218.4	149.3	124.0	155.9	129.2
	[112.4]	[105.2]	[83.6]	[107.9]	[85.6]
Median absolute deviation of errors	7.4	4.3	3.8	4.5	4.0
	[0.8]	[1.2]	[1.1]	[1.2]	[1.0]
Interquartile range of errors	15.4	9.3	8.2	9.5	8.5
	[3.0]	[3.9]	[3.4]	[3.7]	[3.3]
Interdecile range of errors	35.9	24.3	21.7	25.6	22.9
	[12.8]	[12.8]	[11.3]	[13.0]	[11.5]
Single, non-parametric temperature		X	X		
Interacted and higher-order temperatures				X	X
Additive explanatory variables			X		
Interacted explanatory variables					X
Hour $\times$ month indicators		X	X	X	X

Notes: This table displays results from prediction models. Results are calculated at the unit level, and this table displays the mean and standard deviation of those results. Column (1) is the baseline model estimated by Equation (7). Columns (2) and (3) are subset selection models estimated by Equation (8). Columns (4) and (5) are LASSO models estimated by Equation (9). The subset selection models perform better across all measures than the comparable LASSO models.

Table 3: Local market capacity — Effect on markups

	Markup at 85% capacity (\$/MWh)			Markup at 75% capacity (\$/MWh)			Markup at 65% capacity (\$/MWh)		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Local market capacity (GW)	−1.09	−18.35**	−25.42**	−0.97	−11.98**	−15.24**	−0.72	−9.64***	−13.00***
	(1.02)	(7.63)	(11.22)	(0.97)	(5.35)	(6.46)	(0.79)	(3.05)	(5.03)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.93	6.47		2.75	5.88		2.68	5.28
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (10) with local market capacity as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the top panel of Figure 10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Number of units in local market — Effect on markups

	Markup at 85% capacity (\$/MWh)			Markup at 75% capacity (\$/MWh)			Markup at 65% capacity (\$/MWh)		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Units in local market	-0.17	-5.74^{**}	-10.27^{**}	-0.19	-3.54^{**}	-5.40^{**}	-0.15	-3.05^{***}	-5.30^{**}
	(0.36)	(2.79)	(5.06)	(0.31)	(1.76)	(2.58)	(0.25)	(0.97)	(2.21)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.11	4.98		2.04	4.65		1.97	4.29
Local temperature	X	X	X	X	X	X	X	X	X
Unit $\times$ hour $\times$ month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (10) with the number of units in the local market as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the middle panel of Figure 10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Number of firms in local market — Effect on markups

	Markup at 85% capacity (\$/MWh)			Markup at 75% capacity (\$/MWh)			Markup at 65% capacity (\$/MWh)		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firms in local market	−0.81	−23.82**	−41.10**	−0.82	−15.26**	−22.95**	−0.65	−12.94***	−21.64**
	(1.31)	(10.64)	(19.17)	(1.16)	(7.20)	(10.25)	(0.95)	(3.91)	(8.47)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.63	5.44		2.51	5.20		2.43	4.87
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (10) with the number of firms in the local market as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the bottom panel of Figure 10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Local market capacity by owner and unit type — Effect on markups

	Markup at 85% capacity ($/MWh)		
	(1)	(2)	(3)
Local market capacity (GW)	-25.42^{**}		
	(11.22)		
Capacity owned by competing firms (GW)		-50.10^{**}	
		(23.45)	
Capacity owned by own firm (GW)		456.92^{*}	
		(252.19)	
Dispatchable capacity (GW)			-27.59^{**}
			(12.15)
Non-dispatchable capacity (GW)			-2.00
			(73.96)
First-stage specification	Spline	Spline	Spline
First-stage F-statistic	6.47	6.48, 4.92	6.87, 4.57
Local temperature	X	X	X
Unit × hour × month FEs	X	X	X
Hour of sample FEs	X	X	X
Observations	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (10) with measures of local market capacity as the independent variables. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 3. Column (2) reports results with local market capacity split by ownership. Column (3) reports results with local market capacity split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for Columns (2) and (3) are in Appendix D. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Robustness to local controls — Effect of local market capacity on markups

	Markup at 85% capacity ($/MWh)			
	(1)	(2)	(3)	(4)
Local market capacity (GW)	-25.42^{**}	-20.69^{**}	-23.70^{**}	-29.80^{**}
	(11.22)	(8.30)	(9.68)	(14.10)
First-stage specification	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	9.33	9.29	5.86
Local temperature	X			X
Local capacity factor			X	X
Unit $\times$ hour $\times$ month FEs	X	X	X	X
Hour of sample FEs	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (10) with local market capacity as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 3. The remaining columns report results with different sets of local market controls, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(4) are in Appendix D. All regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Robustness to matching distance — Effect of local market capacity on markups

	Markup at 85% capacity ($/MWh)			
	(1)	(2)	(3)	(4)
Local market capacity (GW)	-25.42^{**}	-33.32^{**}	-30.32^{***}	-22.24^{**}
	(11.22)	(13.40)	(11.21)	(8.76)
First-stage specification	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	6.38	9.54	11.75
Minimum matching distance (km)	0	100	200	300
Local temperature	X	X	X	X
Unit $\times$ hour $\times$ month FEs	X	X	X	X
Hour of sample FEs	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (10) with local market capacity as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 3. The remaining columns report results with different matched weather stations. Each unit is matched to the weather station that best predicts the unit's local market capacity and meets the minimum matching distance, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(4) are in Appendix D. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Temperature selection — Effects on markups

	Markup at 85% capacity (\$/MWh)					
	Matched temperature			Local temperature		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	-25.42^{**}			-6.23		
	(11.22)			(5.26)		
Units in local market		-10.27^{**}			-2.63	
		(5.06)			(2.24)	
Firms in local market			-41.10^{**}			-8.25
			(19.17)			(7.57)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	4.98	5.44	20.09	19.04	19.29
Local temperature	X	X	X			
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (10) with different measures of local market size as the independent variable. The dependent variable is markup at 85% capacity. Columns (1), (2), and (3) duplicate Column (3) of Tables 3, 4, and 5, respectively. The remaining columns report results using local temperature rather than matched temperature. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (4)–(6) are in Appendix D. Regressions reported in columns (1)–(3) control flexibly for local temperature; all regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.