Measuring Market Power in Wholesale Electricity Markets: A Dynamic Competition Approach*

Stanley S. Reynolds†

May 8, 2018

Abstract

Restructured wholesale electricity markets in North America are subject to continuing concerns about the market power of sellers. Market power assessments have typically compared market outcomes to those predicted by a static competition model. However, electricity generation is subject to a variety of inter-temporal costs and constraints that are not adequately captured by static models. This paper develops a dynamic model of wholesale market competition that incorporates two important features of electricity generation technology: minimum generation rates and generator startup costs. Taking these features into account, this paper characterizes a dynamic competitive equilibrium and develops an approach for computing the equilibrium. This model's computed equilibrium prices provide an alternative competitive benchmark to assess market power. The model is applied to the 2014 wholesale electricity market of the Electric Reliability Council of Texas. Predicted peak hour prices from the dynamic competition model are above those from the static competitive model, but still less than actual peak prices on average. Average markups predicted by the dynamic model are about 20% less than markups predicted by the static model for the full sample of hours. When comparisons are restricted to hours of days with little or no transmission congestion, both the static and dynamic models predict much smaller markups over marginal cost. The variation in prices predicted by the dynamic model is more similar to observed variation in wholesale prices than is price variation from the static model.

*Thanks to Jeffrey Church and Joseph Cullen for helpful comments on a prior version.
†University of Arizona, P.O. Box 210108, Tucson, AZ 85721, reynolds@eller.arizona.edu

1 Introduction

Many economic analyses of the performance of wholesale electricity markets utilize static models of perfect competition. Such analyses are based on the merit order stack for generation units, which orders units from those with lowest marginal cost (MC) to those with highest MC and uses generation capacities to create a short run supply curve for the market. Mansur [2008] refers to this approach as the competitive benchmark analysis; I will refer to this as static competitive benchmark analysis, to distinguish it from the dynamic approach I use in this paper. This static approach has been the basis for assessments of the exercise of market power by generation suppliers in numerous studies [see, for example, Borenstein, Bushnell, Wolak [2002]; Joskow and Kahn [2002]; Bushnell et al. [2008], Wolak [2010]]. Long-run versions of this approach have been the foundation for assessments of regulatory policy changes, such as the introduction of real-time pricing for retail electricity customers [see Borenstein and Holland, 2005, Borenstein, 2005].

Borenstein et al. [2002] applied static competitive benchmark analysis to the California wholesale electricity market following restructuring in the 1990's. They use detailed information about generator characteristics, fuel costs, and emissions permit costs to estimate marginal generation costs. They combine estimated marginal costs with information about transmission line capacities, generator outages, electricity imports and exports, and market demand quantities to construct predicted hourly perfectly competitive prices for the wholesale market. Comparing actual wholesale prices to constructed competitive prices leads Borenstein et al. [2002] to conclude that producers exercised significant market power during peak summer demand periods of 1998, 1999, and 2000.[1]

Harvey and Hogan [2001] argue that tests of market power in wholesale electricity markets should recognize that generation units are subject to several types of technological constraints, such as minimum operating rates, ramping constraints, unit startup lags, minimum down-times, and unit startup costs. They criticize the static competitive benchmark

[1]In summer 2000 actual payments by California wholesale electricity customers Ire almost twice as high as predicted competitive payments - roughly $9 bwellion actual vs. $4.5 bwellion predicted. The difference is comprised mainly of production inefficiency and market power rents to firms.

approach used in Borenstein et al. [2002] and Joskow and Kahn [2002], arguing that the apparent large price-cost margins estimated in these papers could be due, at least in part, to unmodeled technological constraints. Mansur [2008] explores this issue in an analysis of markups in PJM, a regional transmission organization that operates in 13 states in the Eastern U.S. He uses production and cost data from the year prior to PJM restructuring, when firms operated as regulated utilities, to estimate a reduced form model of generation costs. These estimates (implicitly) take into account technological constraints that are ignored by the competitive benchmark approach. He applies estimates from the reduced form model to post-restructuring data in PJM to estimate markups of price over MC. He finds that this reduced form approach yields significantly smaller estimates of price-cost markups than does the competitive benchmark approach. Mansur's results suggest that static competitive benchmark analysis provides biased predictions of markups. Note however that his approach of using reduced-form estimates of costs is not viable in most cases because the required pre-competition data is not available.

In this paper I develop a dynamic model of a wholesale electricity market and use it to produce an alternative competition benchmark, which I term a dynamic competitive benchmark. The model incorporates important features of electricity generation technology: minimum operating rate constraints, unit startup lags, and unit startup costs. Supply-side dynamics introduced by these technology features may have important implications for competitive market outcomes. In addition, the significance of supply-side dynamics is likely increasing over time as penetration of intermittent renewable power increases and net load (load less renewable generation) becomes more variable; see Perez-Arriaga and Batlle [2012].

There are two main parts of the analysis. The first is a formulation and analysis of a dynamic competition model. Dynamics are important because decisions about startups and shut-downs of electricity generators are inherently forward-looking. A profit-maximizing generator operator will not start a generator unless she believes that wholesale prices during the next operating cycle will be high enough to recoup both production and startup costs. Incorporating these technology features into an economic analysis is challenging because it requires a dynamic formulation and must accommodate non-convexities in production. I build on results in Cullen and Reynolds [2017], who formulate and analyze a dynamic competition model in which demand fluctuates over time and generation suppliers have minimum operating rate constraints and unit startup costs. The analysis in Cullen and Reynolds [2017] allows for different types of fossil fuel generators (coal, combined cycle gas, etc.) but assumes all generators of a given type are identical. I extend their analysis to allow for unit startup lags and heterogeneity in generator efficiencies (i.e., heat rates) within a given type of generator. Generator heterogeneity is crucial for an empirically realistic short run model of wholesale electricity market supply.

The second part of the analysis is an application of the model to a particular wholesale market setting: the Electric Reliability Council of Texas (ERCOT) in 2014. supplier market power has been a concern in ERCOT, as it has been in many restructured wholesale electricity markets. The Public Utility Commission of Texas has market power mitigation rules aimed at detecting and deterring bidding consistent with exercise of market power; see Schubert et al. [2006]. Several features of ERCOT in 2014 facilitate application of the model. ERCOT operates a grid that has virtually no interconnections with the other major grids in the U.S., so that I can safely ignore energy imports and exports. In 2014 Texas had the largest amount of wind turbine capacity of all U.S. states; wind intermittency may increase the importance of supply side dynamics. Also, by 2014 ERCOT had completed most of the Competitive Renewable Energy Zone (CREZ) transmission capacity project linking ist Texas wind farms to load centers in the rest of Texas, so that transmission constraints were less significant in 2014 than in some prior years; see LaRiviere and Lu [2018].

I use data regarding generators, loads, and prices from ERCOT in 2014 to parameterize the model. I use the parameterized model to compute static and dynamic competitive benchmark prices, and compare these prices to actual ERCOT wholesale market prices. Several key results emerge. One is that markups of price over marginal cost implied by the static model are 20-40% greater than markups implied by the dynamic model. These results on the markup bias from the static competitive benchmark model are roughly consistent with the markup bias estimate in Mansur [2008]. Also, the dynamic model predicts a larger peak vs. off-peak price differential than the static model; this larger dynamic model price differential comes closer to matching the large peak vs. off-peak differential in actual prices. Many of the price spikes in the wholesale market occur during times of transmission congestion. When the sample is restricted to days with little to no congestion, the levels of peak and off-peak prices and the variation around these levels predicted by the dynamic competition model are similar to those of actual prices. By comparison, static competition benchmark prices fluctuate within a relatively narrow band that is well below the average level of actual wholesale prices.

There are several limitations of the approach used here. First, while the analysis captures some generation technology features that contribute to dynamics, there are additional unmodeled features that could affect market outcomes. Second, the analysis assumes a single integrated wholesale market with no transmission congestion. The ERCOT wholesale market experienced congestion in about 42% of hours in 2014; see Potomac-Economics [2015]. Third, the computations reported here abstract from uncertainty regarding future loads and wind generation, and rely on a perfect foresight assumption. Fourth, the analysis also abstracts from generator outages, and this will likely bias both static and dynamic model predictions toward lower and less variable prices.

Nonetheless, the results reported here suggest that a dynamic competition model is able to capture key features of wholesale electricity price movements and in addition, that omitting market dynamics may lead to a large upward bias in market power estimates. The concluding section of the paper discusses approaches for addressing some of the limitations of the analysis in this paper.

The next sub-section discusses related work. Sections 2 and 3 formulate and analyze the model. Section 4 provides data and information for ERCOT. Section 5 explains the computational approach. Section 6 describes results and Section 7 concludes and provides directions for future research.

1.1 Related Papers

A variety of features of electricity technology may be important for market performance: transmission and distribution system constraints, minimum and maximum output rate constraints for generation units, generation unit ramping constraints that limit changes in the rate of generation, time-lags for generation unit startup, and generation unit startup costs. There is a large power systems engineering literature that examines how a utility or electricity system should be operated and managed, taking these technological constraints into account. A common approach involves a so-called unit commitment model, which allows optimization methods to be applied to a complex system with multiple generation units and a connecting transmission network [see Bouffard et al. [2005] and Hobbs et al. [2002]]. Both unit commitment models and the analysis of the present paper use a dynamic optimization approach that takes generator operating constraints into account. The model in this paper abstracts from transmission constraints and some other details incorporated into unit commitment models, but allows for a more complex formulation of uncertainty and makes an explicit connection between optimization and competitive equilibrium.

Cho and Meyn [2011] formulate and analyze a theoretical model of a dynamic competitive market system in which demand evolves according to a stochastic process (continuous time Brownian motion) and generators are subject to ramping constraints. There is a single type of generation in their model that operates with constant marginal cost. Generation suppliers provide services both to an energy market and an operating reserves market. Cho and Meyn prove existence of a dynamic competitive equilibrium and show that, because of ramping constraints, equilibrium energy price frequently deviates from marginal generation cost. Their results suggest that using marginal generation cost as a basis for competitive prices (as in the static competitive benchmark analysis) may be problematic. However, the model of Cho and Meyn [2011] is very stylized and would not seem to provide a good basis for assessing deviations from competitive energy pricing in real-world markets.

Staffell and Green [2016] modify the basic merit order stack model for generator operations and investment to account for startup costs. They develop a heuristic approach aimed at approximating the solution to a full dynamic equilibrium model. They argue that a relatively simple heuristic may be valuable if it improves on the merit order stack model and, at the same time, does not require major computation.

Reguant [2014] estimates a dynamic structural model of generation costs using data from Spanish wholesale electricity auctions. She observes unit startups in this data and estimates generation cost parameters, including unit startup costs. She uses her estimates for counterfactual simulations, including a competitive simulation aimed at measuring exercise of market power. Markups estimated from a static competitive benchmark model with the Spanish data tend to be quite high during peak periods and low (sometimes negative) during off-peak periods. Reguant's estimates of markups from her dynamic model are smoother over time. In particular, the static competitive benchmark model overestimates price markups during peak periods due to its failure to consider unit startup costs.[2]

Cullen [2013] estimates a dynamic structural model of generation costs, using a different data set - ERCOT data on generation and prices. His model assumes that firms are price-takers, have rational expectations over future prices, and choose startup and shut-down decisions to maximize expected profit. Importantly, he uses estimated generation and startup costs to numerically simulate a dynamic competitive equilibrium, which he uses for counterfactual analysis of environmental policies. He does not consider market power, although his approach could be used for that purpose.[3]

This model formulation in this paper is closely related to that of Cullen and Reynolds [2017]; hereafter labeled as CR. As noted above, I build on the model formulation in CR, extending it to allow for heterogeneous efficiencies within generator types. This paper is focused on short run operating dynamics in wholesale electricity markets. By contrast, CR

[2]Reguant's approach to assessment of market power is quite different from the present paper. The dynamic competitive benchmark model she estimates is the solution to an individual firm's cost minimization problem, in response to actual bidding behavior of rivals. As such, it can be used to assess unilateral market power for individual firms. however, it does not provide a competitive equilibrium solution for the market that can be used to assess market-wide exercise of market power, as the present paper does.

[3]The approach used in the present paper offers advantages over the approach in Cullen [2013]. There is no guarantee that the computational approach used in Cullen converges to an equilibrium, in contrast to the approach used here. Also, his approach is computationally intensive, requiring iterations over dynamic optimizations for all firms in the market, in contrast to the solution of a single planner's problem here.

use a more stylized formulation of generators and extend the analysis to incorporate long run competitive equilibrium investment decisions.

2 A Model of Wholesale Electricity Competition

In the static benchmark competition model, suppliers make production (generation) decisions period by period. There are no dynamic linkages across periods in that model. However, if generation units have minimum production constraints and suppliers incur a startup cost each time a unit is started, then suppliers' decisions depend on current and expected future market conditions. A dynamic market model is required in order to capture the decision problem faced by suppliers and also to capture how dynamic constraints shape market outcomes.

In this section I formulate a perfect competition model that includes short-run market dynamics. The model specifies several different types of generation technologies (coal, natural gas combustion turbine, etc.), with an exogenous fixed amount of total generation capacity for each type of technology. Each supplier is assumed to operate a single generating unit that uses one of the generation technologies. The model is comprised of an unlimited number of time periods in which suppliers make operating decisions regarding their generation units; when to start a unit, when to shut down a unit, and generation rates for 'on' units. Time periods are indexed by $t \in \{1, 2, ..., \infty\}$, and the per-period discount factor is $\delta \in (0, 1)$. The application uses one-hour time periods. Wholesale market demand varies over time according to an exogenous stochastic process. Fuel prices, which influence marginal generation costs, also vary over time according to an exogenous stochastic process.

The model is similar to that of CR, with three main differences.[4] First, I allow for time lags for generator startups, whereas CR assume no lags in startups. Second, CR assume fixed fuel prices, in contrast to fuel prices that may vary over time in the present paper. Third, CR assume that all generation units within a given technology type are identical, whereas I allow for heterogeneity in the efficiency and marginal operating cost of units within each

[4]The model in CR includes a generator investment decision stage. Their application is focused on investment incentives.

type. This kind of heterogeneity is crucial for capturing empirically observed differences in heat rates for units within any technology type.

The model has a firm-specific binary state variable indicating each firm's current operating status; active or inactive. A firm must incur a lump sum cost in order to transition from inactive status to active. Firm-specific transitions are not continuous in firms' decision variables under this formulation. Results from Hopenhayn [1990] may not be applied because that paper assumes a continuity condition on firm-specific state transitions. I address the discontinuity in firm-specific state transitions by assuming that firms are 'small'; specifically that firm size is measure zero. While firm-specific state transitions are discontinuous, the transitions for the aggregate states that are relevant for the planner's problem are continuous, and this allows for a solution to the planner's problem that is equivalent to a competitive equilibrium allocation. To state things differently, binary states for firms coupled with lump sum transition costs pose an analytical difficulty in a model with large (positive measure) firms. In such a model, supply functions are not continuous in market prices and a competitive equilibrium need not exist. The small firms assumption side-steps this difficulty and also provides a way to link the planner's solution to a competitive equilibrium allocations.[5]

2.1 Market Demand

Demand varies across time periods according to the value of a demand shock (or, shift) variable, θ_t. The evolution of demand shocks over time is explained below. There is an *inverse* market demand function, $P(Q, \theta_t)$, that is continuous and weakly decreasing in total output Q. $P(0, \theta_t)$ is assumed to have a finite upper bound for all possible values of θ_t.

I define the gross benefit function B by:

$$B(Q, \theta_t) \equiv \int_0^Q P(u, \theta_t) du \tag{1}$$

For our electricity market application I focus on the wholesale market. Wholesale demand

[5] If firms are not small then non-convexities may require modifications to the equilibrium formulation. For instance, O'Neill et al. [2005] show that efficient market-clearing prices may be obtained by complementing energy prices with additional prices for generator startups.

is derived from downstream retail electricity demand. Many retail electricity customers face prices that are fixed over long periods of time, e.g., due to regulatory constraints. In such cases the wholesale inverse demand would not reflect marginal willingness to pay for energy, and its integral in (1) would not correspond to gross benefit. Regardless of the welfare interpretation of the function $B(\cdot)$, this function plays an important role in our dynamic competitive analysis.

2.2 Generation

There are J different types of generation technologies. Suppliers are atomistic, price-taking firms. Each supplier owns and operates one (arbitrarily small) unit of a particular type of generation capacity. I use the following notation for exogenous generation parameters:

s_j = startup cost per unit of type j capacity, $s_j > 0$

m_j = minimum generation rate per unit of type j capacity; $m_j \in (0, 1]$

f_{jt} = fuel price for type j unit in period t

k_j = total amount of type j capacity summed over all firms

Whenever a type-j generation unit is 'on' (i.e., active) it must produce positive output in the interval, $[m_j, 1]$. If a firm's current and near-term projected future prices are below its marginal generation cost then the firm may choose to shut down to avoid losing money. But a shut-down implies that the firm will have to incur a startup cost in order to begin producing later on when prices are higher. It is the combination of positive minimum generation rate and positive startup cost that yields a dynamic decision problem for generation suppliers.

I assume that startup cost and minimum generation rate per unit of capacity are the same for all generation units of any given type of generation technology. The marginal cost of a generation unit is the product of the fuel price for that type of unit and the heat rate (fuel required per unit of output) for the unit. Each active supplier with type j technology is assumed to have constant marginal cost for output in $[m_j, 1]$. I allow for heterogeneity in the efficiency of generators within each technology type, by allowing different units to have different heat rates. Let n_j be the number of different heat rates for technology j and order heat rates so that, $h_{1j} \leq h_{2j} \leq ... \leq h_{n_j j}$. Total capacity for units with heat rate h_{ij} is k_{ij} and these capacities summed over all type-j units add to k_j. A generation unit must be turned on in order to generate output. Let x_{ij} be the amount of capacity with heat rate h_{ij} that is turned on in a given period, where $x_{ij} \in [0, k_{ij}]$. Total output q_j for technology j is feasible if there exists a vector $(q_{ij}, ..., q_{n_j j})$ such that $\sum_{i=1}^{n_j} q_{ij} = q_j$ and $q_{ij} \in [m_j x_{ij}, x_{ij}]$ for $i = 1, ..., n_j$. The total cost of generation for feasible type-j output q_j when fuel cost is f_j may be written as,

$$C_j(q_j, f_j) = \min_{(q_{1j}, ..., q_{n_j j})} \sum_{i=1}^{n_j} f_j h_{ij} q_{ij}, \tag{2}$$

subject to, $\sum_{i=1}^{n_j} q_{ij} = q_j$ and $q_{ij} \in [m_j x_{ij}, x_{ij}]$ for $i = 1, ..., n_j$.

The notation that allows for heterogenous heat rates is cumbersome, but the underlying economic intuition is straightforward. Given a vector $(x_{1j}, ..., x_{n_j j})$ of on-capacities in a period, the total cost of generation for feasible output levels is continuous in output, with linear segments that have progressively steeper slopes as output increases. This implies that cost function C_j is increasing and convex in q_j, for feasible outputs.

2.3 Market, Feasibility, & Equilibrium

Demand shift variables and fuel prices are assumed to follow a stochastic process. Define a vector, ψ_t, to include period t demand shock, fuel prices, and exogenous variables Z_t - such as hour-of-day and lagged demand shocks and fuel prices - useful for predicting future demand shocks and fuel prices. That is, $\psi_t \equiv (\theta_t, f_{1t}, ..., f_{Jt}, Z_t)$. I assume that ψ_t follows a Markov process. Feasible values of this process are restricted to a convex set, Ψ; that is $\psi_t \in \Psi$ for all $t \geq 1$. A special case of the model is one in which there is no uncertainty, so that θ_t and f_t follow a deterministic path.

The market is comprised of a large number of small firms who operate as price takers. Each firm is identified with one unit of capacity of a particular type of generation. The production technology for a firm can be described quite simply. At the start of period t a firm's capacity is either 'off' or 'on'. If 'on' then the firm chooses a generation rate between the min and max rates for its generation type and decides whether or not to shut down for the next period. If the firm's unit is 'off' at the start of the period then the firm chooses whether or not to startup for next period.[6] A type-j firm incurs startup cost s_j in the period in which the startup decision is taken.

An implication of this formulation is that the production possibilities set for an individual firm is not convex. A firm's generation unit is either 'off' or 'on'; convex combinations of 'off' and 'on' for an individual firm are not permitted. This complicates showing existence of a competitive equilibrium. I address this difficulty by assuming that individual firms are of measure zero. This assumption yields a convex aggregate production possibilities set, even though production possibilities for an individual firm are not convex.

Firm-level heterogeneity also potentially complicates the analysis. This is made more manageable by our assumption that startup costs and minimum generation rates are identical for all generators of the same type. In a competitive equilibrium, high marginal cost units of a given type will never be started instead of lower marginal cost units of the same type. Likewise, a low marginal cost unit of a given type will never be shut down instead of a higher marginal cost unit of the same type in equilibrium. This implies that if a particular number (or mass) x_j of type-j firms are 'on' in a period in equilibrium, it must be that the type-j firms with the lowest heat rates are the ones that are 'on'.

The total mass of type-j technology firms is k_j. The following notation is used to describe the aggregate production technology. A vector $x = (x_1, x_2, ..., x_J)$ indicates the amount of each type of capacity that is 'on' at the beginning of the period; x_{jt} is equal to the mass of type-j firms whose capacity is 'on' at the start of period t. Define $X(k) \equiv [0, k_1] \times [0, k_2] \times ... \times [0, k_J]$ as the set of feasible vectors of 'on' capacities. The vector y is the amount of each type of generation that firms elect to continue to keep on into the next period; y_{jt} is the mass of type-j firms that keep their unit turned 'on' in t. The vector q is the amount of generation from the J types of generators, where, $q_j \in [m_j x_j, x_j]$ for $j \in \{1, ..., J\}$. startup decisions are

[6]Note that this formulation restricts the time for a unit startup to be one period (e.g., one hour). startup times for units such as coal-fired steam turbines or combined cycle natural gas units may be several hours. The model could be extended to permit multiple hours for startups, e.g. by expanding the state space to keep track of when the startup process started.

given by z, where $z_j \in [0, k_j - x_j]$ for $j \in \{1, ..., J\}$; z_{jt} is the mass of type-j firms that start up in t. The aggregate production technology has two parts: one that specifies constraints within each period and a second that describes generator transitions across periods. For the first part, define the constraint set:

$$PT(x) \equiv \{(q, y, z) : 0 \leq y_j \leq x_j, 0 \leq z_j \leq k_j - x_j, m_j x_j \leq q_j \leq x_j; j = 1, ..., J\} \tag{3}$$

$PT(x)$ specifies how aggregate vectors of outputs, shut-down decisions, and startup decisions for a period are constrained by x, the vector of 'on' capacities at the start of the period. Given the vector of capacites $k \in \mathbb{R}_+^J$, the set $PT(x)$ is compact for any vector $x \in X(k)$. Transitions over time satisfy:

$$x_{t+1} = y_t + z_t \tag{4}$$

The combination of compact constraint sets within each period ($PT(x)$) and linear transitions in (4) imply that the aggregate production technology is convex. This convexity property is key for establishing a connection between a planner's solution and competitive equilibrium. The definition of a competitive market equilibrium is below.

DEFINITION: An allocation $\{q_t, x_t\}$ together with a price process $\{p_t^*\}$ is a *market equilibrium* if:

(i) The allocation is feasible,

(ii) The allocation is consistent with profit maximizing policies for all firms, and

(iii) $p_t^* = P(\sum_{j=1}^J q_{jt}, \theta_t)$ for all $t \geq 1$.

Note that condition (iii) is a standard market clearing condition.

3 Dynamic Market Equilibrium

In this section I develop the solution to a planner's problem and rely on the claim that this is a market equilibrium. The equivalence of a competitive market equilibrium and a social optimum is, of course, a fairly standard type of result and parallels results for dynamic market equilibrium models in Lucas and Prescott [1971] Hopenhayn [1990]. The equivalence of market equilibrium allocations and solutions to the planner's problem is important because it provides a way to prove existence of market equilibrium and because it allows the use of the planner's problem as a vehicle for computation.

The planner has access to a vector k of total generation capacities for the J technologies and makes operating decisions in each period after observing (x, ψ), which serves as a state vector for the planner; the state $(x, \psi) \in X(k) \times \Psi$. Operating decisions are embodied in a vector, (q, y, z), where q specifies production rates, y specifies how much of available generation in the period is left 'on', and z specifies unit startups. The single period payoff, H, for the planner is total surplus for the period, which is equal to gross benefit less generation cost and startup cost.

$$H(q, z, \psi) = B(\sum_j q_j, \theta) - \sum_j C_j(q_j, f_j) - \sum_j s_j z_j \tag{5}$$

where $\psi = (\theta, f_1, ..., f_J, Z)$. H is concave and differentiable in (q, z); concavity follows from concavity of B in total output, convexity of generation cost for each generation quantity, and linearity of startup costs in startup quantities.

The planner makes operating decisions to maximize expected total surplus, where the single period return is H defined in (5). This can be described by an infinite horizon stochastic dynamic programming problem with the following Bellman equation,

$$V(x, \psi) = \max_{(q,y,z) \in PT(x)} \{H(q, z, \psi) + \delta E[V(y + z, \psi') \mid \psi]\} \tag{6}$$

where ψ' is the next period value of the exogenous stochastic process.

Proposition 3.1. *An allocation $\{a_t\} = \{q_t, x_t\}$ and price process $\{p_t\}$ constitute a market equilibrium iff the allocation solves the planner's problem of maximizing discounted expected total surplus.*

Proofs are omitted, as they are similar to those in Cullen and Reynolds [2017]. The *if* part of the proof of Proposition 3.1 is constructed by first showing that any welfare maximizing allocation, along with the associated market clearing price process, maximizes aggregate market profits of firms, taking the price process as exogenous. The second step is to show that there is an assignment of operating policies to individual firms such that aggregate market profit maximization implies maximization of individual firms' profits. The *only if* part of the proof uses concavity of the planner's single period return H and convexity of the set of feasible allocations, to show that no alternative feasible allocation yields higher payoff to the planner than a market equilibrium allocation.

Proposition 3.2. *A market equilibrium exists.*

The proof proceeds by showing that a solution to the planner's problem exists. An optimal policy for a planner's solution generates a feasible allocation and a price process which, by Proposition 3.1 constitute a market equilibrium. Note that we have existence of a dynamic competitive equilibrium here in spite of technology non-convexity at the firm level. Our assumption of small (measure zero) firms effectively smooths out what would otherwise be discontinuities in supply functions.

The planner's policy (i.e., optimal choice) function yields a competitive equilibrium allocation. In each period the planner's policy function yields choices for generation outputs q, unit shut-downs $x - y$, and unit startups z, as a function of the current state (x, ψ). The outputs and the demand state yield the competitive equilibrium price for the period, according to the inverse demand function, $P(\sum_j q_j, \theta)$. The planner's policy function, coupled with a current state (x, ψ), imply a distribution of prices in future periods. Firms have rational expectations over these future prices in the competitive equilibrium.

4 ERCOT & Data

ERCOT is the independent system operator (ISO) for an electric grid that covers most of the state of Texas. There are very few connections between the system managed by ERCOT and the rest of the U.S. electricity grid. Electricity generation and retailing are in ERCOT are largely deregulated. Transmission and distribution are regulated, and regulations specify that firms engaged in generation and retailing have open access to the grid to buy and sell power.

4.1 Generation

As of 2014 there were approximately 950 generators owned by 33 firms supplying electricity in ERCOT. The ownership of generation facilities is concentrated enough to raise concerns about supplier market power. Potomac-Economics [2015] estimate that there was a pivotal supplier in approximately 23% of all hours in 2014. Two large suppliers - NRG and Calpine - were participating in voluntary mitigation plans as a response to market power abuse claims that had been raised against them; see Potomac-Economics [2015].

The major generation technologies are coal, nuclear, natural gas, and wind turbines. There are small amount of hydroelectric power and utility scale solar PV; hydroelectric generation was quite low since its capacity factor was less than 5%. Table 1 provides summary information about different types of generation, including 3 types of natural gas generators: combined cycle (CC), gas turbines (GT), and steam turbines (ST). Wind turbine capacity increased by over a gigawatt during 2014.

Table 1: ERCOT Generation Capacity

Generation Type	Total Capacity (MW)	Avg Heat Rate (MMBtu/MWh)
Hydro	539	-
Nuclear	5,139	-
Coal	23,078	10.6
NG CC	36,785	7.5
NG GT	8,475	8.3
NG ST	11,076	11.3
Wind	11,050 - 12,472	-
Utility Solar	171	-

Data from EPA eGrid2014 report.

The marginal cost of fossil fuel generation is constructed using data on plant capacities, plant average heat rates, and fuel costs. I use monthly prices for natural gas delivered to Texas electric power producers.[7] Natural gas prices ranged from \$4.01/MMBtu to \$6.54/MMBtu during 2014. I use monthly prices for coal, drawn from EIA Electric power Monthly Reports.[8] Delivered coal prices in Texas varied from \$1.93/MMBtu to \$2.06/MMBtu in 2014. Prices for SO_2 and NOx permits had dropped to almost zero in 2014, so I do not include emissions rates or costs of emissions permits in the analysis. The marginal cost of fossil fuel generation is depicted in Figure 1. The marginal cost curve reflects the presence of substantial heterogeneity in heat rates, even within each fossil fuel generator type.

Dynamics enter the model through three features: startup costs, startup lags, and minimum output rates. If these features are removed then the model is completely static. Without startup costs, current actions to startup or shut-down generators have no implications for future profits. On the other hand, if generators are completely flexible in their output level, then they can avoid incurring startup costs by producing minuscule quantities.

Some economic studies of wholesale electricity markets have taken the position that startup costs are small enough that they can be safely ignored in an analysis of energy supply decisions.[9] This is likely true if one focuses on fuel and other energy costs associated with startups. However, the bulk of the opportunity cost of a generator startup is associated with additional maintenance and wear-and-tear on generators.[10] Kumar et al. [2012] estimate that capital and maintenance expenses comprise 80-98 percent of total startup costs, depending on generator and fuel type. The startup cost parameters used here and listed in Table 2 are based on Kumar et al. [2012], using their lower bound estimates for the capital and maintenance portion of startup costs. These startup cost parameters are broadly consistent with the structural estimates in Reguant [2014] and much lower than the structural estimates in Cullen [2013]. I assign the same startup cost (per unit of capacity) and minimum generation rate for NG combined cycle plants and NG steam turbine plants. Including both types as a single generation technology in the model reduces the number of state variables and simplifies computation. Note however, that generation units within the NG CC/ST

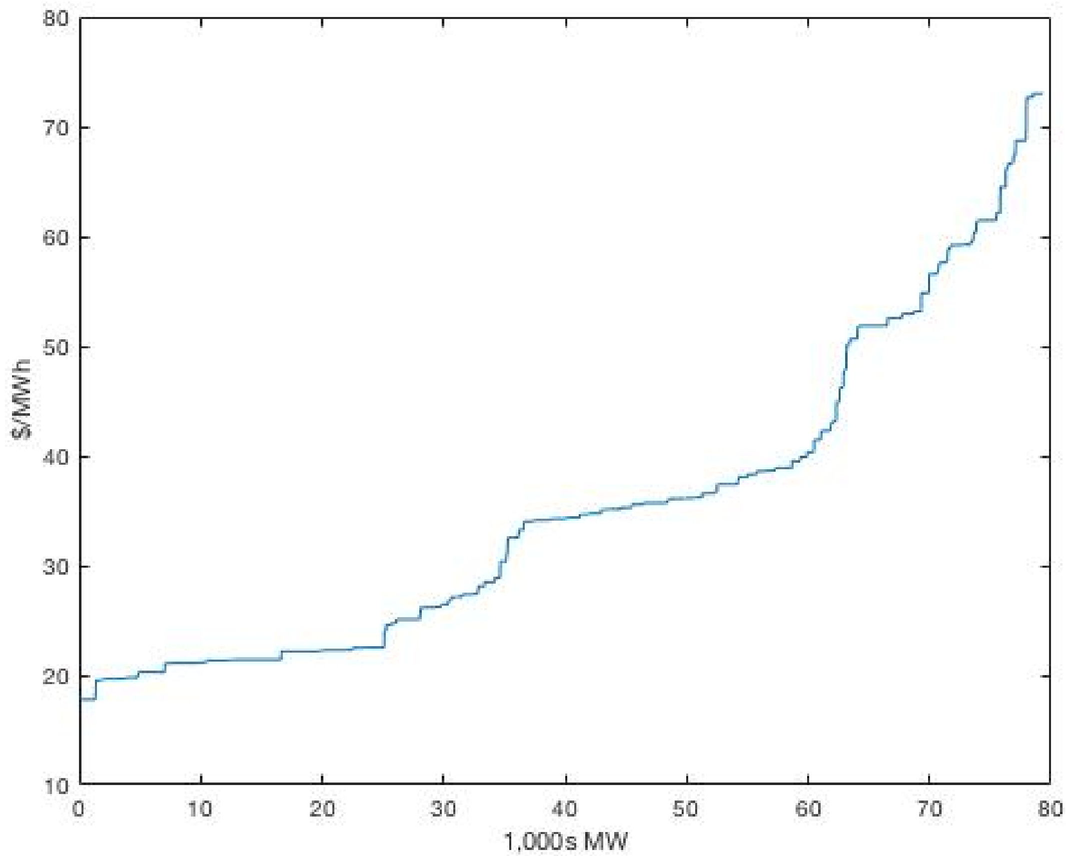
[7] https://www.eia.gov/dnav/ng/hist/n3045tx3m.htm

[8] Table 4.10.1 in 2014 and 2015 monthly reports.

[9] See for example, Borenstein et al. [2002], p. 1391.

[10] Perez-Arriaga and Batlle [2012] emphasize this point in their analysis of the effects of renewable intermittency on conventional power plant operations.

Figure 1: Marginal Cost of Fossil Fuel Generation (January 2014)



generation technology type are permitted to have different heat rates.

I calibrate minimum output rates for natural gas units using data from Cullen [2013]; see Table 2 below. Hentschel et al. [2016] reports minimum operating rates for new, state-of-the-art coal plants. I set the minimum operating rate for coal in Table 2 is somewhat higher than the rates in Hentschel et al. [2016], as the ERCOT plants are older and less flexible than new plants.

Table 2: Startup Cost and Minimum Generation

Generation Type	Startup Cost ($/MW)	Min Generation Rate (MWh/MW)
Coal	150	0.70
NG CC/ST	90	0.60
NG GT	65	0.45

There is a large amount of wind capacity in ERCOT. I use hourly wind generation data for 2014 from ERCOT. Wind generation is highly variable; swings of 7-8 thousand MWh's in a 24 period are not uncommon. Wind generation is sometimes curtailed during high-wind hours.

I treat nuclear and hydroelectric generation as must-run generation and fix output in each hour for each type of generation equal to capacity factor times total capacity. For hydro, this yields constant generation of 26 MW per hour. I do not have data for generation from utility scale solar in ERCOT. I use average hourly solar insolation data coupled with typical PV performance for the Dallas, Texas area to calculate hourly average solar PV capacity factors (see, https://pvwatts.nrel.gov/). I multiply these hourly capacity factors by total utility-scale solar PV capacity to yield hourly solar PV generation.

4.2 Markets, Prices and Transmission

The wholesale market in ERCOT operates with a nodal structure, with over 4,000 nodes (points of transmission system interconnection); ERCOT [2010]. Hourly locational marginal prices (LMPs) are determined for the nodes in the system for both a day-ahead market and a real-time market. Network nodes are organized into 4 regional competitive load zones. I use hourly real-time market settlement prices aggregated to the 4 competitive load zones. Summary statistics for hourly prices for 2014 are reported in Table 3.

The biggest transmission bottleneck involves transmission lines that move wind energy generated in West Texas to load centers in the rest of the state. By 2014 progress on the CREZ project had expanded capacity on these transmission lines, but there Ire still many hours with transmission congestion. When there is no congestion in the system there is a common price across the 4 load zones. I use price differences across the 4 load zones as an indicator of transmission congestion. I adopt an admittedly arbitrary indicator of no-system-congestion as a percentage difference between max and min hourly zonal price less than 10% or an absolute difference less than $5/MWh. This yields a congestion count of 42% of all hours in 2014. This is close to the figure of 44% of hours in 2014 with binding transmission constraints, reported in Potomac-Economics [2015].

Table 3: Hourly Price Data ($/MWh)

Houston Zone Average	37.76
North Zone Average	36.74
South Zone Average	38.80
West Zone Average	41.80
System-Wide Average	38.80
System-Wide Peak (2 - 8 pm) Average	49.33
System-Wide Peak Maximum	4,454
System-Wide Off-Peak Average	35.29
System-Wide Off-Peak Minimum	-1.21
Percent of Hours Congested	42%

The wholesale market price cap varied between $5,000/MWh and $7,000/MWh in 2014. The cap was not binding during the sample period.

4.3 Demand

Residential and commercial users purchased virtually all of their electricity at retail prices that were held fixed over long time periods. So in the short run, wholesale demand for electricity is almost completely price inelastic. Electricity load varies significantly both intraday and across seasons. For 2014 the minimum hourly load was 24,083 MW in April and the maximum hourly load was 66,464 MW in August. In addition, there were roughly 12,000 MW of wind capacity in 2014. The hourly variation in load coupled with hourly variation in wind generation yields a high degree of hourly variation in net load. Given that there was essentially no energy storage in the system, the variation in net load must be matched with variation in fossil fuel generation in order to maintain supply-demand balance.

5 Computation

I compute results for both a static competitive benchmark model and the dynamic competitive benchmark model. The models are calibrated using information and data described in Section 4. The dynamic model developed in Sections 2 and 3 allows for uncertainty about future demand levels, wind generation, and fuel prices. In this paper I simplify computation dramatically by adopting a perfect foresight assumption on future demand, wind generation, and fuel prices.[11] Under this assumption, owners of fossil fuel units have perfect foresight over future wholesale electricity prices and fuel prices. The corresponding condition for the social planner is perfect foresight over demand (load) levels, wind generation, and fuel prices. One point to emphasize is that the dynamic perfect foresight prices are the endogenous prices that emerge from the dynamic benchmark model; I am *not* assuming that agents have perfect foresight over actual ERCOT wholesale market prices. Also note that abstracting from uncertainty does not affect static competitive benchmark results, since that model is not forward-looking, but it does affect dynamic competitive benchmark results and likely yields

[11]Incorporating uncertainty in computations requires estimation of a stochastic process governing the evolution the ψ_t, and using the estimated process as a component of the planner's optimization problem. This approach is definitely feasible, but is not undertaken in this paper.

prices that are less variable compared to results from a model with uncertainty.

Computing the solution to the planner's problem was also simplified by replacing the infinite horizon problem with a series of finite horizon problems. Coupling perfect foresight with a finite horizon yields a planner's problem that is a collection of straightforward - though large-scale - multi-variate constrained optimization problems, instead of a complex stochastic dynamic programming problem.

The dynamic analysis assumes a fixed one-hour startup lag for fossil fuel units. This tends to be less than cold startup time for coal and NG CC/ST units. However, the specification of length of startup lags will be less significant for a perfect foresight model than for a model with uncertainty.

The theoretical formulation of generation does not provide for intermittent renewable generation, such as generation from wind turbines. In the computations I assume all wind generation is must-take energy and subtract this from load to yield a net load demand quantity. I do not account for possible curtailment of wind in the computations reported here.

The dynamic model assumes a downward sloping wholesale demand. In order to implement the model for the nearly completely inelastic demand of the ERCOT wholesale market, I specify a nearly-vertical linear inverse demand function for each hour with a fixed slope and horizontal intercept equal to actual net load for the hour. The slope parameter is set so that price elasticity is in the range of -0.001 to -0.002 over the domain of load variation.

Another simplification made for computations is to ignore generator outages. Borenstein et al. [2002] compute static competitive benchmark results for California by running a Monte Carlo simulation for each hour that incorporates multiple draws of (low probability) generator outages into the creation of merit order stacks. Ignoring these potential outages is likely to bias computation results for both the static and dynamic models toward lower and less variable prices.

Another feature of the Borenstein et al. [2002] analysis is incorporation of fast-response operating reserves. They add up-regulation quantities to load to arrive at a total demand quantity when computing equilibrium prices. The argument is that up-regulation would be provided mainly by partially loaded generators, and so the identity of the marginal generator would be determined by the quantity of load plus up-regulation. I did not find hourly up-regulation quantities for my analysis. Mago [2017] reports that average up-regulation is approximately 300 MW, so I added a fixed quantity of 300 MW to load for each hour to account for up-regulation for both static and dynamic computations.

For the dynamic model I use an annual discount rate of 8%, which corresponds to an hourly discount factor (δ) very close to one. For dynamic model computations, I divide the year into 3-day segments and solve an 80 hour (3.5 day) optimization problem for the planner for each segment. There are 3 decision variables for the planner for each of the 3 fossil fuel generation types for each one-hour period, so each constrained optimization problem has 720 choice variables. I discard the last 8 hours prior to the end of the horizon in reported results. Since each of my 3-day segments end at midnight when load is relatively low, the number of future periods incorporated past midnight has very little effect on startup or shutdown decisions, or on predicted equilibrium prices, within the 3-day segment. The Knitro optimization package was used for dynamic model computations.

6 Results

I report static and dynamic competitive benchmark results for hourly prices and compare these to actual prices. As a preview of these results, Figure 2 shows an hourly time series of actual and predicted prices for 8 days in January. This was a period of relatively low demand and one in which there was no transmission congestion, by my measure. Note that actual prices are much more volatile than static benchmark prices. Static prices fluctuate in a fairly narrow band between \$20 and \$34/MWh during this period, while actual prices fluctuate between \$5 and \$95/MWh. Dynamic benchmark prices exhibit variation similar to that of actual prices during this period, although it's also clear that these predicted price swings do not match up precisely with those of actual prices.

Figure 3 shows an hourly time series of generator startup capacities predicted by the dynamic model for the same 8 days in January. Startups occur in a relatively small number of hours that tend to line up with price peaks for the dynamic model. Total predicted startup capacity per day during this period ranged from a few hundred MW to about 10,000 MW.

Summary results for actual and simulated prices are reported in Table 4. Average peak and off-peak prices are reported, as well as standard deviations of peak and off-peak prices. For the full sample of all hours of the year, the static model under-predicts actual peak prices by about \$20/MWh and under-predicts off-peak prices by about \$10/MWh. Looking at the the static benchmark prices versus actual prices for the full sample would suggest that actual prices include markups over marginal cost of roughly 50% on average. The dynamic model predicts significantly higher peak prices than the static model, but overall it still would suggest that actual prices include substantial markups over marginal (opportunity) cost, of roughly 38% on average.

Table 4: Actual and Simulated Wholesale Prices (\$/MWh)*

	Actual	Static Model	Dynamic Model
Full Sample			
Peak Hours			
Average Price	49.33	28.23	38.05
Standard Deviation	48.86	5.03	23.20
Off-Peak Hours			
Average Price	35.29	25.27	24.81
Standard Deviation	74.39	4.78	16.79
Non-Congested Days			
Peak Hours			
Average Price	37.93	26.10	34.41
Standard Deviation	28.71	4.59	21.41
Off-Peak Hours			
Average Price	28.63	24.15	24.35
Standard Deviation	13.76	3.73	16.41

* Peak hours are defined as 2 pm - 8 pm.

The full sample includes many hours with transmission congestion, and some actual price spikes are undoubtedly related to congestion. Some of the highest hourly wholesale prices occurred off-peak, contributing to the high standard deviation of off-peak prices. Table 4 also

Figure 2: Hourly Prices for 8 Days in January 2014

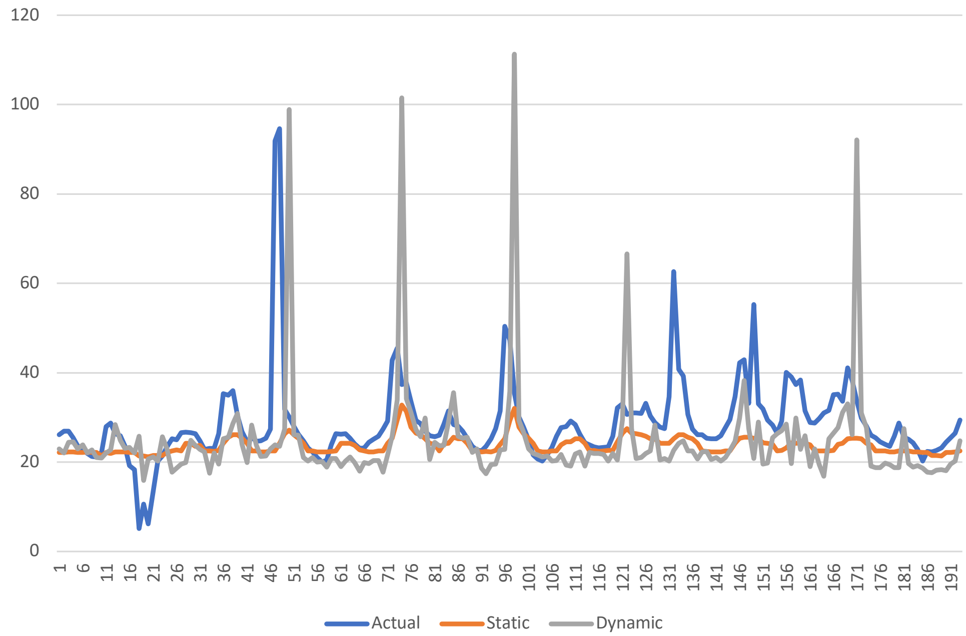
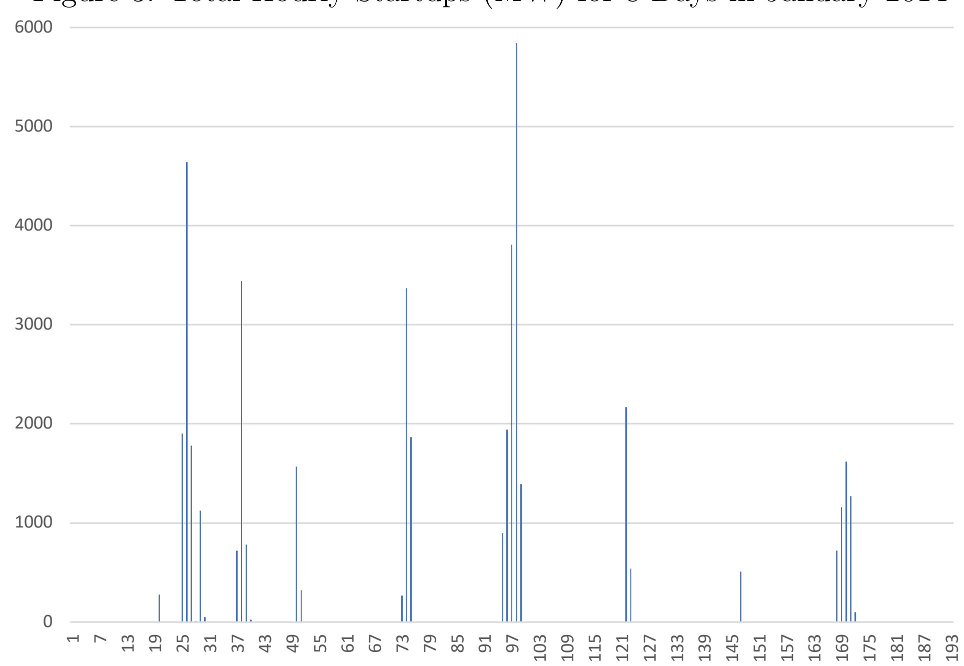


Figure 3: Total Hourly Startups (MW) for 8 Days in January 2014



reports average hourly peak and off-peak prices and standard deviations for non-congested days. Implied markups for both static and dynamic models are considerably smaller for this sample, compared to those for the full sample. The implied markup for the static model is roughly 25%, or one-half of the corresponding full sample markup. The implied markup for the dynamic model is roughly 15%, less than one-half of the corresponding full sample markup. Comparing the implied markups of the two models for the non-congested sample, we can see that the dynamic model markup is about 40% smaller than the static model markup.

The results for non-congested days reported in Table 4 suggest that the dynamic model yields a peak vs. off-peak price price differentiable and degrees of price variability that are roughly consistent with actual prices. By contrast, the static model predicts a smaller peak vs. off-peak price differential and much less price variability than we see in actual prices.

7 Conclusions and Future Research

This paper develops a model of short-run operating dynamics of a wholesale electricity market. The model provides a dynamic competitive benchmark for market power assessment. This benchmark is alternative to a commonly used static competitive benchmark. The dynamic model incorporates important features of electricity generation technology: minimum operating rate constraints, unit startup lags, and unit startup costs. Supply-side dynamics introduced by these technology features have important implications for competitive market outcomes. And the significance of supply-side dynamics is likely increasing over time as penetration of intermittent renewable power increases and net load becomes more variable.

There are two main contributions. The first is to extend the dynamic competition analysis of Cullen and Reynolds [2017] to allow for unit startup lags and heterogeneity in generator efficiencies (i.e., heat rates) within a given type of generator. Generator heterogeneity is crucial for an empirically realistic short run model of wholesale electricity market supply. The second contribution is to show how dynamics matter for market power assessment in a particular wholesale market setting: the Electric Reliability Council of Texas (ERCOT) in

2014. Markups of price over marginal cost implied by the static model are 20-40% greater than markups implied by the dynamic model. In particular, the dynamic model predicts a much larger peak vs. off-peak price differential than the static model; this larger dynamic model price differential comes closer to matching the large peak vs. off-peak differential in actual prices. Many of the price spikes in the wholesale market occur during times of transmission congestion. When the sample is restricted to days with little to no congestion, the levels of peak and off-peak prices and the variation around these levels predicted by the dynamic competition model are similar to those of actual prices. By comparison, static competition benchmark prices fluctuate within a relatively narrow band that is well below the average level of actual wholesale prices.

There are several limitations of the analysis in this paper. Some limitations are intrinsic to the approach used here. For example, the 'small firms' assumption is crucial for the perfect competition analysis, as it provides a way to deal with production non-convexities that arise because of startup costs and minimum generation constraints. In fact, electricity generation is from discrete, lumpy units of capacity, and this discreteness may have an impact on observed wholesale prices. Also, the analysis assumes a single integrated wholesale market with no transmission congestion. The ERCOT wholesale market experienced significant congestion in many hours in 2014. While it may be possible to include some transmission constraints in the analysis, extending this dynamic market analysis to incorporate an interconnected transmission grid is a formidable task.[12]

Other limitations of the analysis could be addressed in future versions of the paper. Some improvements could be made by incorporating better data. Acquiring hourly data for solar PV generation and for up-regulation quantities would provide more precise estimates of load served by fossil fuel generators. Fuel price data with a finer time scale would yield more accurate generator marginal cost estimates. Alterations in the model could address other limitations. Wind generation is taken to be exogenous in this analysis. The dynamic model could be extended to allow curtailment of wind output when prices become negative. The perfect

[12]A key transmission bottleneck involves lines that connect West Texas wind generation to Texas load centers. It may be possible to extend the dynamic analysis to include a constraint for this transmission link.

foresight assumption used for computations may be relaxed and replaced with a model with uncertainty about future electricity and fuel prices. This would require the use of numerical approximation methods for stochastic dynamic programming, rather than the constrained optimization approach used in this version. It should be possible to incorporate random generator outages, as done in the static competitive benchmark analysis of Borenstein et al. [2002]. Also, it should be possible to incorporate additional generation technology features and constraints, such as additional types of fossil fuel generators, longer startup lags, and (possibly) ramping constraints. The primary challenge associated with most of these changes is the increased difficulty of numerically solving the planner's optimization problem. Lastly, given the potential computation challenges for the dynamic model, it would be interesting to compare the static and dynamic competition benchmark results to competitive pricing results using the heuristic approach outlined in Staffell and Green [2016].

References

Borenstein, S. (2005). The long-run efficiency of real-time electricity pricing. *The Energy Journal*, 26(3):93–116.

Borenstein, S., Bushnell, J., and Wolak, F. (2002). Measuring market inefficiencies in California's restructured wholesale electricity market. *American Economic Review*, 92:1376–1405.

Borenstein, S. and Holland, S. (2005). On the efficiency of competitive electricity markets with time-invariant retail prices. *RAND Journal of Economics*, 36:469–493.

Bouffard, F., Galiana, F., and Conejo, A. (2005). Market-clearing with stochastic security - part I: Formulation. *IEEE Transactions on Power Systems*, 20(4):1818–1826.

Bushnell, J., Mansur, E., and Saravia, C. (2008). Vertical arrangements, market structure and competition: An analysis of restructured u.s. electricity markets. *American Economic Review*, 98(1).

Cho, I.-K. and Meyn, S. P. (2011). Efficiency and marginal cost pricing in dynamic competitive markets with friction. *Theoretical Economics*, 5:215–239.

Cullen, J. (2013). Dynamic response to environmental regulation in the electricity industry. Working Paper, Washington University.

Cullen, J. and Reynolds, S. (2017). Market dynamics and investment in the electricity sector. Working Paper 17-05, University of Arizona.

ERCOT (2010). Texas nodal market guide. Technical report.

Harvey, S. and Hogan, W. (2001). Identifying the exercise of market power in California. Working Paper, JFK School of Government, Harvard University.

Hentschel, J., Babic, U., and Spliethoff, H. (2016). A parametric approach for the valuation of power plant flexibility. *Energy Reports*, 2.

Hobbs, B., Rothkopf, M., O'Neill, R., and Chao, H., editors (2002). *The Next Generation of Electric Power Unit Commitment Models*. Kluwer Academic Publishers.

Hopenhayn, H. (1990). Industry equilibrium dynamics: A general competitive theory. Working Paper.

Joskow, P. and Kahn, E. (2002). A quantitative analysis of pricing behavior in Californiaś wholesale electricity market during summer 2000. *Energy Journal*.

Kumar, N., Besuner, P., Lefton, S., and Agan, D. (2012). Power plant cycling costs. Technical report. NREL Report SR-5500-55433.

LaRiviere, J. and Lu, X. (2018). Transmission constraints, intermittent renewables and welfare. Working Paper.

Lucas, R. and Prescott, E. (1971). Investment under uncertainty. *Econometrica*, 39.

Mago, N. (2017). Methodology for determining minimum ancillary service requirements. Technical report. ERCOT technical report.

Mansur, E. (2008). Measuring welfare in restructured electricity markets. *Review of Economics and Statistics*, 90(2):369–386.

O'Neill, R., Sotkiewicz, P., Hobbs, B., Rothkopf, M., and Stewart, W. (2005). Efficient market-clearing prices in markets with nonconvexities. *European Journal of Operational Research*, 164:269–285.

Perez-Arriaga, I. and Batlle, C. (2012). Impacts of intermittent renewables on electricity generation system operations. *Economics of Energy & Environmental Policy*, 1(2).

Potomac-Economics (2015). 2014 state of the market report for the ERCOT wholesale electricity markets. Technical report.

Reguant, M. (2014). Complementary bidding mechanisms and startup cost in electricity markets. *Review of Economic Studies*, 124:1708,1742.

Schubert, E., Hurlbut, D., Adib, P., and Oren, S. (2006). The Texas energy-only resource adequacy mechanism. *The Electricity Journal*, 19.

Staffell, I. and Green, R. (2016). Is there still merit in the merit order stack? The impact of dynamic constraints on optimal plant mix. *IEEE Transactions on Power Systems*, 31(1):43–53.

Wolak, F. (2010). Using restructured electricity supply industries to understand oligopoly industry outcomes. *Utilities Policy*, 18:227–246.