Market Dynamics and Investment in the Electricity Sector*

Joseph A. Cullen† Stanley S. Reynolds‡

February 15, 2017

Abstract

We model dynamic competition in which firms make initial capital investment decisions followed by repeated entry and exit choices as demand fluctuates. We show a correspondence between competitive equilibrium and the solution to a planner's problem, which establishes equilibrium existence and provides a platform for computation. We apply the framework to model electricity generator investment decisions, incorporating generator startup costs as the entry/exit friction. Market frictions are particularly important when evaluating renewable energy policies. The presence of startup costs reduces wind turbine profits, leading to as much as 60% lower uptake of wind investment for a given renewable subsidy level.

*We thank Ignacio Esponda, Gautam Gowrisankaran, Derek Lemoine, Glenn MacDonald, and Ignacia Mercadal as well as participants at the Berkeley Energy Institute seminar for helpful comments. Chuan Chen provided valuable research assistance.

†Olin Business School, Washington University in St Louis, St Louis, MO, jacullen@wustl.edu

‡Department of Economics, University of Arizona, Tucson, AZ, reynolds@eller.arizona.edu

1 Introduction

This paper analyzes competitive markets in which firms repeatedly enter and exit, or repeatedly open and close operations, in the face of fluctuating demand. Our analysis is motivated by the problem faced by electricity suppliers who make repeated costly decisions about starting up and shutting down fossil fuel generators in response to fluctuating wholesale prices. This problem is particularly relevant when environmental policies stimulate increased penetration of intermittent renewable power; see Perez-Arriaga and Batlle [2012]. Increased generation from intermittent renewables requires more frequent startups and shut downs of fossil fuel generators and leads to changes in market prices and profits for all types of generators. Moreover, such changes in prices and profits alter the incentives for investment in different types of generators.

Existing dynamic models are not well suited for analyzing this type of problem. Many dynamic model formulations allow for entry and exit, but restrict firms to at most one entry and one exit from the market; see for example, Hopenhayn [1992] for dynamic competition and Ericson and Pakes [1995] for dynamic oligopoly. Alternatively, some dynamic models allow firms to have a binary on-off state variable, but impose a continuity-in-actions assumption on state transition probabilities; see Hopenhayn [1990] and Adlakha et al. [2015].

In this paper we develop a dynamic competition model that allows for repeated entry/exit decisions, or repeated startup/shut-down decisions, without imposing the restrictions used in prior dynamic models. Our model allows for aggregate demand shocks, one-time capacity investment decisions from a menu of technologies, repeated entry/exit decisions, and production decisions for active firms. Firms incur a lump sum cost upon each entry and exit from the market, which introduces dynamic linkages across periods. This formulation presents theoretical and computational challenges, due to production non-convexities and the absence of an easily derived steady state equilibrium. We develop new techniques for solving such problems. The techniques rest on transforming the dynamic competitive problem into a social planner's problem. We characterize properties of competitive equilibrium and use the planner's problem as a computational platform for our application to electricity market

analysis.

In the theoretical portion of the paper, we prove equivalence between dynamic perfectly competitive allocations and the solution to a particular stochastic dynamic programming (DP) problem for a planner via a 'small firms' assumption. This equivalence is used to show existence of a competitive equilibrium. We also show that competitive equilibrium prices are unique. We allow heterogeneity in types to emerge endogenously in equilibrium via firms' initial capacity investment decisions. We further characterize properties of equilibrium allocations analytically using a special case of the model. Here we demonstrate how demand variability coupled with the magnitude of entry/exit costs influences the extent of heterogeneity. In addition, we examine a firm's option value of waiting to enter and waiting to exit. Our formulation introduces a novel channel for the option value of waiting; one that is linked to initial capacity investment.

We apply our theoretical model to analyze wholesale electricity market competition and examine the impact of renewable energy subsidies on market outcomes, focusing on wind turbines as the renewable technology. In this application, initial investments are in electricity generation units. Ex-ante heterogenous firm types correspond to different generator technologies (coal, gas combined cycle, peaker, etc.). Entry and exit correspond to generator startups and shut-downs. We show that incorporating generator startup costs into our model yields significantly different electricity prices and generator investments, compared to a static model.[1] These differences are magnified as renewable energy subsidies increase. The intermittency of renewable output implies that adding renewable capacity yields greater variability in net demand (demand less renewable output). Greater variability in net demand shifts the investment mix for conventional generators toward peaker units with relatively low startup costs and high operating costs. This shift is more pronounced for the model with startup costs than for a static model. The presence of startup costs in the model also reduces the profitability of renewable generators, leading to as much as 60 percent lower uptake of renewable technologies. Introducing startup costs into the model has large effects on equi-

[1] We use the term 'static model' to refer to a model with no entry or exit costs and no minimum output rate, and hence no dynamic linkages across periods within stage two.

librium outcomes even though total startup costs represent a very small percentage of firms' overall costs.

While our application is to electricity markets, the modeling approach is applicable to a variety of other market settings. Examples include airlines serving city-pair markets, mineral extraction firms that open and close mines, farmers that take plots of land in and out of agricultural production, and retailers that open and close stores as demand fluctuates. We use the term entry in our exposition but note that entry may also be interpreted as opening a plant or facility, or as starting up a piece of equipment. Similarly, exit may be interpreted as temporarily closing a plant, mothballing a facility, or shutting down a piece of equipment. In particular, note that exit does not imply that a firm dies, but rather that its asset is idle and remains available for re-activation later on.

Our analysis allows for heterogeneity in firm types via firms' initial decisions about which capacity technology to invest in. Different technology choices yield differences in operating costs, investment costs, and in entry and exit costs. For example, some types may have high investment costs and low operating costs with other types having low investment costs and higher operating costs. Similarly, some types may have greater flexibility than other types due to lower entry and/or exit costs. After firms' types are selected in an initial stage of the model, types are held fixed in a second stage that is comprised of an infinite sequence of production and trading periods. Our approach is similar to the dynamic competitive market analysis in Hopenhayn [1992] with two key differences. First, the entities in our model are persistent. This is important in settings where firms or consumers are long lived and can wait for the optimal entry or exit point. Persistence creates an option value not found in prior work. Second, unlike prior work, firms are ex-ante heterogeneous in stage two when entry and exit decisions are made. Combined with persistence, this permits us to model situations where differences between types are fixed and permanent, but at the same time allows the distribution of types which arises in equilibrium to be endogenous.

Our results build on several literatures: 1) dynamic competition modeling, 2) analyses of long-run electricity investment, and 3) an emerging literature on wholesale electricity market analysis using models with technology frictions. First, we contribute to the substantial

literature on competitive dynamic equilibrium. For example, Dixit [1989] models optimizing behavior of a price-taking competitive firm that can repeatedly enter, exit, and re-enter. Market prices evolve according to an exogenous stochastic process, and the firm incurs lump sum costs for entry and exit decisions. Our formulation mirrors the repeated entry/exit set-up of the Dixit model but also incorporates endogenous equilibrium prices and capacity investments. Several other papers analyze equilibrium models of dynamic competition with randomly varying market conditions; see Lucas and Prescott [1971], Jovanovic [1982] and Hopenhayn [1992].[2] Our model and results depart from these three papers in two economically important ways, however. First, we model one-time investment decisions of firms followed by a sequence of entry/exit decisions, where the latter decisions incorporate option value of waiting. This kind of option value is incorporated into the repeated entry/exit analysis of Dixit [1989], but is not incorporated into the entry and exit calculus of Lucas and Prescott [1971], Jovanovic [1982] or Hopenhayn [1992]. Second, we specify fixed, lump sum costs for transitions between active and inactive status; i.e., entry and exit costs. This is in contrast to continuity assumptions for firms' state transitions assumed in other papers.[3] Our assumption of fixed, lump sum entry and exit costs permits us to parameterize the model in applications without resorting to specifying distributions for entry and exit costs.

Second, our application contributes generally to the large literature on modeling electricity markets, but more specifically to the much smaller literature on investment in electricity markets. These models of investment in electricity markets have a short-run component - capturing market clearing prices and dispatch of generators under conditions of fluctuating demand - as well as a long-run component - capturing profit maximizing entry and investment in generation capacity. Borenstein [2005] and Borenstein and Holland [2005] use this type of model to analyze the long-run impact of the introduction of real-time pricing for

[2]Entry and exit are one-time decisions in these models. A firm does not have the opportunity to re-enter after an exit.

[3]In Lucas and Prescott [1971] this takes the form of convex costs of adjustment of continuous state variables. In Jovanovic [1982] and Hopenhayn [1990, 1992] this takes the form of continuously distributed costs of entry and exit, and/or probabilities of transitions between discrete states that are continuous in firms' actions. Continuity of state transition probabilities in players' actions is also a key assumption in Adlakha et al. [2015], which analyzes stationary equilibria of dynamic games with many players.

retail electricity. Bushnell [2010] uses this type of model to examine the impact of of increasing wind penetration on fossil fuel generation investment in four Western Electricity Coordinating Council regions. Gowrisankaran et al. [2016] analyze the effects of a renewable energy mandate for electricity generation. They focus on how increased solar photovoltaic penetration affects operating reserves, investment in conventional generators, and emissions. Blanford et al. [2014] analyze the effects of a clean energy standard on generation investment in the U.S., using the US-REGEN model. However, none of these papers incorporate dynamic generator technological operating constraints into a model with optimizing agents. Our paper incorporates these constraints into a wholesale market model with endogenous prices, investment, and rational, forward-looking generation operators.[4]

Finally, we also contribute to the emerging literature on the effects of market frictions in electricity markets. Cho and Meyn [2011] analyze a dynamic model of a wholesale electricity market and show that generator ramping constraints may lead to significant wholesale price volatility and sustained periods of prices that deviate from marginal generation cost. Reguant [2014] and Cullen [2013a] estimate dynamic structural models of electric generator operations that include startup decisions and startup costs. Both papers find economically significant estimated generator startup costs.[5] This paper is most closely related to Cullen [2013a], in which estimated generation and startup costs are used to simulate a dynamic competitive equilibrium, which he uses for counterfactual policy analysis. The counterfactual analysis of Cullen [2013a] applies to short-run competitive equilibrium with fixed amounts of generation capacity. This is in contrast to the long-run investment analysis of the present paper.

The rest of the paper proceeds as follows. In section 2 we characterize the model of competition. In section 3 we develop the connection between market equilibrium and the solution to the planner's problem. In section 4 we describe the data we use, explain how we set parameter values, and describe the computation approach. In section 5 we report results. Section 6 concludes.

[4]The US-REGEN model employed by Blanford et al. [2014] includes minimum generation rate constraints and allows for startup costs. However, the model uses ad-hoc decision rules for startup decisions, rather than optimization-based decision rules.

[5]Reguant [2014] shows how accounting for startup costs provides a natural correction for biased estimates of market power.

2 A Competitive Model with Repeated Entry and Exit

We formulate a dynamic competition model in which firms make decisions about investment, entry and exit, and production over time. Activity is divided into two stages. Firms make decisions about investment in stage one. Firms have a variety of technologies from which to choose for their initial investment. We assume that investment is irreversible, so that capacity investments made in stage one remain in place for all of stage two. The discount factor between stages one and two is $\tilde{\delta} \in (0, 1]$. Stage two is comprised of an infinite sequence of periods, where in each period firms make decisions about entry, exit, and re-entry, and active firms make production decisions. The per-period discount factor within stage two is, $\delta \in (0, 1)$.

2.1 Demand

Demand varies across time periods within stage two according to the value of a demand shock (or, shift) vector, $\boldsymbol{\theta}$, which is assumed to follow a Markov process. There is an *inverse* market demand function, $P(Q, \boldsymbol{\theta})$, that is continuous and decreasing in total output Q. Define Θ as the set of all possible $\boldsymbol{\theta}$-values. $P(0, \boldsymbol{\theta})$ is assumed to have a finite upper bound for all $\boldsymbol{\theta} \in \Theta$. In addition, we define the gross benefit function B by:[6]

$$B(Q, \boldsymbol{\theta}) \equiv \int_0^Q P(z, \boldsymbol{\theta}) dz. \tag{1}$$

2.2 Production

In stage one firms may invest in one of J different production technologies. We normalize the amount of production capacity that a firm may invest in to one for each technology. We

[6]For our electricity market application we focus on the wholesale market. Wholesale demand 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 $B(\cdot)$ in (1), this function plays an important role in our dynamic competitive analysis.

assume that all capacity units for any technology $j \in \{1, ..., J\}$ have indentical costs and characteristics. We use the following notation:

c_j = marginal cost of output for technology j

f_j = investment cost per unit of technology j capacity

g_j = entry cost per unit of technology j capacity

h_j = exit cost per unit of technology j capacity

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

Firms choose whether or not to invest in stage one; if a firm invests it also chooses which technology $j \in J$ to invest in. The total amount of investment in type-j capacity by all firms is k_j, and total cost of that investment is $f_j k_j$. Firms that invest in stage one compete over an infinite sequence of periods in stage two. Marginal cost of output is assumed to be constant for each technology type. A firm whose type-j capacity is inactive at the start of the current period may enter and begin production in the same period by incurring an entry cost of g_j.[7] A firm with an active unit is restricted to operate between the minimum and maximum output rate for their technology; type j units must produce at a rate in the interval $[m_j, 1]$ per active unit of capacity. A firm with an active unit at the start of a period may exit immediately. Firms may enter, exit, and re-enter repeatedly in stage two. There are five exogenous parameters - c_j, f_j, g_j, h_j and m_j - for each technology $j \in J$.

Entry and exit costs introduce non-convexities into production technology, which in turn complicates market analysis and may lead to non-existence of competitive equilibrium. In order to pursue our objective of a competitive market analysis that incorporates these technology features, we introduce a formulation in which non-convexities are permitted at the firm level, but for which the aggregate production technology is convex. We assume that individual firms are small relative to the size of the market; specifically, firms are assumed to be of measure zero in the formal model.

Later in the paper we apply this model to analyze competition in wholesale electricity

[7] The model may be extended to allow for lags in entry and/or exit. A one period lag in entry and/or exit is a straightforward extension. Lags of more than one period require an expansion of the state space for the planner's problem, which complicates computation of the model.

markets. Our formulation captures features of electricity generation technologies that are often abstracted away from in economic models of electricity markets; namely unit start-up costs and minimum generation rates.[8] In this context, entry costs, g_j , would be the costs of starting up an idle generator , m_j would be the technical constraint on the minimum operating level of a generator, c_j would the fuel costs of producing electricity, and f_j would be the cost of building a new generator. We show how incorporating these features into a model of wholesale electricity competition affects market outcomes and the impact of public policies. However, the model is not limited to this application. For example, in the context of the airline industry, the capacity investment decision might the be number of aircraft to purchase while the entry costs would be the cost of opening a new route with the stock of existing aircraft. In a labor context, the entry cost might be the fixed cost of hiring/firing workers while the investment cost would be the investment in capital used by workers (office space, machinery, etc) to create output. The framework is applicable to any situation where there are both long-run capacity investments and potential entrants which are persistent and heterogenous.

2.3 Market, Feasibility, & Equilibrium

The supply side of the market is comprised of a large number of small firms who operate as price takers. In stage one there is a mass $\overline{k}_j$ of type-j firms that may invest in production technology j. The mass of type-j technology firms that invests is $k_j \in [0, \overline{k}_j]$.

The production technology for a firm that invests can be described quite simply. In period t in stage two a firm was either active or inactive in the preceeding period; $\omega_{t-1} = 1$ indicates active status last period and $\omega_{t-1} = 0$ indicates inactive status last period. The firm chooses its status $\omega_t \in \{0, 1\}$ at the start of each period t. If a type j firm switches from inactive last period to active this period then the firm incurs entry cost g_j. A switch from active to inactive status involves an exit cost, h_j. Production for a type-j active firm is constrained to be between the min and max rates for its technology type; a type-j firm's output in t is

[8] On the other hand, we abstract from some other technology features, such as ramping constraints that limit the rate at which generator output may be adjusted over time.

$\gamma_t \in [m_j\omega_t, \omega_t]$.

The following notation is used to describe the aggregate production technology. A vector $\boldsymbol{x}$ indicates the amount of each type of capacity that is active in a period; x_{jt} is equal to the mass of type-j firms that have $\omega_t = 1$. The vector $\boldsymbol{q}$ is the amount of output from the J types of firms, where, $q_{jt} \in [m_j x_{jt}, x_{jt}]$ for $j \in \{1, ..., J\}$. The aggregate production technology is defined via the following constraints:

$$\boldsymbol{k} \in K \equiv [0, \overline{k}_1] \times ... \times [0, \overline{k}_J] \tag{2}$$

$$(\boldsymbol{x}_t, \boldsymbol{q}_t) \in A(\boldsymbol{k}) \equiv \{(\boldsymbol{x}, \boldsymbol{q}) : x_j \in [0, k_j]; q_j \in [m_j x_j, x_j]; j \in \{1, .., J\}\} \tag{3}$$

Capacities must satisfy the investment constraint, given by (2).[9] Constraint (3) specifies that active capacity in period t may not exceed total capacity for any technology, and total output in t must be between the minimum and maximum output achievable for a given amount of active capacity in t for each technology.

An allocation is defined by a a vector $\boldsymbol{k}$ of capacities and a sequence for $(\boldsymbol{x}_t, \boldsymbol{q}_t)$. In general, the values for this sequence of vectors will depend on realizations of the demand shock process, since as we will show, firms' decisions will depend on realizations of this process. Because of this, it is useful to describe an allocation as a stochastic process. It will be convenient to denote a history of realizations of demand shocks through time period t as, θ^t, and the set of all possible histories through t as, Θ^t.

DEFINITION: A *feasible allocation* is a stochastic process $\{\boldsymbol{k}, \boldsymbol{x}_t, \boldsymbol{q}_t\}_{t=0}^{\infty}$ that
(i) is measurable with respect to the set of possible histories of demand shocks, and
(ii) satisfies $\boldsymbol{k} \in K$ and for each realization of the process, $(\boldsymbol{x}_t, \boldsymbol{q}_t) \in A(\boldsymbol{k})$ for $t \geq 0$.

The set of feasible allocations is convex, since K is convex and $A(\boldsymbol{k})$ is convex for each $\boldsymbol{k} \in K$. Note that the set of feasible allocations for the market is convex, even though the production possibilities set for an individual firm is not convex. Measurability with respect to demand shock histories essentially means that there is an outcome for vector $(\boldsymbol{x}_t, \boldsymbol{q}_t)$

[9] This constraint is imposed for technical reasons. It insures that capacities are bounded for the planner's problem.

corresponding to each possible demand shock history, θ^t.

During stage two, in each period t firms are assumed to observe the current price p_t and the history of demand shocks through t, θ^t. Firms are assumed to have rational expectations regarding future prices. Equilibrium prices for a given period depend on the history of demand shocks through that period. A price process $\{p_t\}$ is a stochastic process that is measurable with respect to the set of all possible histories of demand shocks. Given a price process $\{p_t\}$, the value of a type-j firm in period t is given by:

$$v_{jt}(\omega_{t-1}, \theta^t) = \sup_{\{\gamma,\omega\}} \mathrm{E}_t \sum_{\tau=0}^{\infty} \delta^\tau [\omega_{t+\tau}\gamma_{t+\tau}(p_{t+\tau} - c_j) - \max\{\omega_{t+\tau} - \omega_{t+\tau-1}, 0\} g_j - \max\{\omega_{t+\tau-1} - \omega_{t+\tau}, 0\} h_j] \tag{4}$$

The suprenum in (4) is with respect to stochastic processes for decisions (γ_t, ω_t) that are measurable with respect to the set of possible demand shock histories. A policy for decisions regarding entry, exit, and output is *profit maximizing* for a firm with own initial state ω_{t-1} if it attains (4).

A type j firm will weigh the expected discounted payoff in stage two against the cost of investment. The value function for a type j firm in stage one is,

$$v_j^e = \max\{0, \tilde{\delta} E[v_{j0}(0, \boldsymbol{\theta}_0)] - f_j\} \tag{5}$$

where the expectation on the RHS of (5) is taken over initial stage two values of the demand shock. Note that a type j firm is indifferent between investing and not investing if the second term in brackets on the RHS of (5) is zero. A policy for a type j firm is profit maximizing *iff* it attains (4) in stage two and attains (5) in stage one.

DEFINITION: An allocation $\{\boldsymbol{k}, \boldsymbol{q}_t, \boldsymbol{x}_t\}$ 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) $p_t^* = P(\sum_{j=1}^{J} q_{jt}, \boldsymbol{\theta}_t)$ for all $t \geq 0$.

Condition (ii) states that all firms adopt policies that attain (4) and (5) when faced with price process $\{p_t^*\}$. Condition (iii) is a standard market clearing condition.

3 Competitive Market Equilibrium

In this section we show that a competitive equilibrium exists and characterize some properties of equilibrium. The key to our results is demonstration of equivalence between a competitive equilibrium allocation and the solution to a planner's problem. We then show that a solution to the planner's problem exists, and so the resulting allocation and price process constitute a market equilibrium. The equivalence of a competitive market equilibrium and a planner's solution is, of course, a fairly standard type of result and parallels results for dynamic market equilibrium models in Lucas and Prescott [1971], Jovanovic [1982], and Hopenhayn [1990, 1992]. However, our formulation differs from the papers cited above in an important way, and demonstrating this connection requires proof.[10]

In our model there is a firm-specific binary state variable indicating each firm's prior operating status; active or inactive. A firm must incur a lump sum cost in order to transition from one status to another. Firm-specific transitions are not continuous in firms' control variables under this formulation. Results from Hopenhayn [1990] may not be applied to our model because that paper assumes a continuity condition on firm-specific state transitions.[11] We address the dis-continuity in firm-specific state transitions that arises in our formulation by exploiting the 'small firms' (measure zero) assumption. While firm-specific state transitions are discontinuous, the transitions for the aggregate states that are relevant for the

[10]The model in Lucas and Prescott [1971] has a single type of representative firm, in contrast to the model in this paper which has several types corresponding to different technologies and firm-specific 'on/off' states. The model in Jovanovic [1982] has heterogeneous firms, but no aggregate shocks. The formulation in Hopenhayn [1990] allows for aggregate shocks and heterogeneous firms via a distribution of firm-specific states.

[11]Hopenhayn [1990] provides a relatively general framework for competitive dynamics and this continuity assumption is critical for his Theorem 1, which demonstrates existence of a competitive equilibrium. Models consistent with this continuous transitions assumption include: Lucas and Prescott [1971] which has each firm's next period capital as a continuous function of its current capital and investment, and Ericson and Pakes [1995] in which each firm's state is in a finite set with state transition probabilities that are continuous functions of its control variables.

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 put 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 for us to link the solution to a planner's problem to competitive equilibrium allocations.

3.1 The Planner's Problem

Before turning to equilibrium results we describe the planner's problem, which has a recursive structure. Let $\boldsymbol{x}'$ denote the vector of active capacities in the previous period. In any period $t \geq 0$ in stage two the planner has access to a vector $\boldsymbol{k}$ of total capacities for the J technologies and makes operating decisions in each period after observing $(\boldsymbol{x}', \boldsymbol{\theta}) \in X(\boldsymbol{k}) \times \Theta$, where $X(\boldsymbol{k}) \equiv \{\boldsymbol{x} : x_j \in [0, k_j]; j \in \{1, .., J\}\}$. $(\boldsymbol{x}', \boldsymbol{\theta})$ serves as a state vector in stage two for the planner.[12] Operating decisions are embodied in vector, $(\boldsymbol{q}, \boldsymbol{x})$, where $\boldsymbol{q}$ specifies production rates and $\boldsymbol{x}$ is the vector of active capacities for the current period. The values of $\boldsymbol{x}'$ and $\boldsymbol{x}$ together imply aggregate entry and exit decisions. The single period payoff, H, for the planner is total surplus for the period, which is equal to gross benefit less production cost and entry and exit costs.[13]

$$H(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}) = B(\sum_j q_j, \boldsymbol{\theta}) - \sum_j [c_j q_j - g_j \max\{x_j - x'_j, 0\} - h_j \max\{x'_j - x_j, 0\}] \tag{6}$$

H is bounded and is concave in $(\boldsymbol{q}, \boldsymbol{x})$ for each $(\boldsymbol{x}', \boldsymbol{\theta}) \in X(\boldsymbol{k}) \times \Theta$. Concavity follows from concavity of B in total output, linearity of production costs in output, and convexity of entry and exit costs in $\boldsymbol{x}$. That H is bounded follows from our assumption that $P(0, \boldsymbol{\theta})$ has a finite upper bound for all $\boldsymbol{\theta}$ and from the capacity constraints that bound outputs for each

[12] If $t = 0$ then we require that $\boldsymbol{x}' = \boldsymbol{0}$. Active capacity in $t = 0$ requires entry in $t = 0$.

[13] Our definition of total surplus implicitly assumes that there are not entries and exits for a single type of technology in the same period. In the proof of Proposition 3.1 we show that this property holds in the planner's solution.

technology.

The planner makes operating decisions in stage two to maximize expected discounted total surplus, where the single period return is H in equation (6). This can be described by a stationary stochastic dynamic programming problem with the following Bellman equation,

$$W(\boldsymbol{x}', \boldsymbol{k}, \boldsymbol{\theta}) = \max_{(\boldsymbol{x},\boldsymbol{q})\in A(\boldsymbol{k})}\{H(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{\theta}) + \delta E[W(\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{\theta}^+) \mid \boldsymbol{\theta}]\} \tag{7}$$

where $\boldsymbol{\theta}^+$ is the next period demand shock. The value function $W(\cdot)$ may be used to define the stage one capacity choice problem for the planner.

$$\tilde{W} = \max_{\boldsymbol{k}\in\boldsymbol{K}}\{-\sum_{j=1}^{J} f_j k_j + \tilde{\delta} E[W(\boldsymbol{0}, \boldsymbol{k}, \boldsymbol{\theta}_0)]\} \tag{8}$$

The expectation on the RHS of (8) is taken in stage one over initial stage two values of the demand shock.

3.2 Equilibrium Results

Proposition 3.1. *An allocation $\{\boldsymbol{a}_t\} = \{\boldsymbol{k}, \boldsymbol{q}_t, \boldsymbol{x}_t\}$ 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 in the appendix. 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.

Proposition 3.3. *Let $\{p_t\}$ and $\{p'_t\}$ be two equilibrium price processes. Then $\{p_t\} = \{p'_t\}$ almost everywhere.*

The proof of Proposition 3.3 is almost identical to that of Theorem 2 in Hopenhayn [1990], and therefore omitted. The idea of the proof is that if there were two distinct equilibrium price processes then there must be two distinct equilibrium allocations. Social welfare is equal at these two equilibrium allocations, since equilibrium allocations maximize the planner's objective. But if equilibrium price processes are distinct then the marginal social value of output differs across the two allocations for some histories of demand shocks. This would imply that a convex combination of the two allocations, which is feasible by convexity of the aggregate technology, would yield strictly higher social welfare.

Note that the allocations that generate the unique equilibrium price process need not be unique. We provide an example at the end of sub-section 3.3 with two equilibrium allocations with differing amounts of investment in technologies 1 and 2 that generate the same equilibrium price process. Moreover, even for a specific allocation of aggregate outputs, entry and exits that support the price process, there may exist alternative assignments of active/inactive status for individual firms of a particular type that are consistent with the same total amount of active capacity for that type.

Corollary 3.4. *If the investment constraints (2) are not binding then firms earn zero expected profit in equilibrium.*

The Corollary is a direct result of the defintion of payoffs in (4) and the definition of equilibrium. If $k_j < \bar{k}_j$ in equilibrium then there is a positive measure of type-j that choose to not invest. If type-j investing firms earn positive profits, then non-investing type-j firms are not maximizing profit since they could earn greater profit by investing.

3.3 A Special Case

We derive results below for a special case of our model in order to illustrate properties of equilibrium. We assume a linear demand function, two demand states, and two technologies. In addition we assume that there is no lag in long run investment (i.e., $\tilde{\delta} = 1$), zero exit costs, and minimum production rates equal to 100% ($m_1 = m_2 = 1$). While these assumptions are clearly restrictive, the qualitative nature of results for this special case are likely to carry over to more general specifications. In particular the assumption of a 100% minimum production rate is strong, but serves to simplify the model and to highlight the role of entry costs.

Inverse demand is given by, $P(Q, \theta) = \theta - Q$. We assume that the demand shock takes on either value θ_A or θ_B each period, with $\theta_A > \theta_B$. Demand shocks follow a Markov process with, $\text{Prob}[\theta_{t+1} = \theta_i \mid \theta_t = \theta_i] = \rho$, for $i \in \{A, B\}$. We assume that $1/2 < \rho < 1$ so that demand in successive periods is positively correlated. The long run probability of each demand state is 1/2 due to the symmetry of transition probabilities. We assume that the initial distribution of θ matches the long run probabilities.

We allow for two technologies, with $c_1 < c_2$ and $f_1 > f_2$. We define:

$$\Delta \equiv c_2 + 2(1-\delta)f_2 - c_1 - 2(1-\delta)f_1$$

and assume cost parameters are such that $\Delta > 0$. The assumption $\Delta > 0$ implies that technology 1 has a combined production and investment cost advantage over technology 2.

We first consider the case in which there is investment only in technology 1 in equilibrium. Proposition 3.5 below provides a sufficient condition for this case. Given just two demand states and a single technology, the structure of equilibrium is fairly simple. Output is equal to total capacity k_1 in all periods with high demand (state A). If the initial demand realization is low (state B) then entry and output are equal to x_1'; at the first transition to state A there is entry equal to $k_1 - x_1'$ and output rises to k_1. In any transition from state A to B there is exit equal to $k_1 - x_1''$ and output drops to x_1''. In subsequent transitions from B to A there is entry equal to $k_1 - x_1''$. The payoff to the planner may be expressed as a function of

(x_1', x_1'', k_1). In the appendix we derive the following solution:[14]

$$k_1^* = \theta_A - c_1 - 2(1-\delta)f_1 - (1-\delta\rho)g_1 \tag{9}$$

$$x_1''^* = \theta_B - c_1 + \delta(1-\rho)g_1 \tag{10}$$

$$x_1'^* = \theta_B - c_1 - (1-\delta)g_1 \tag{11}$$

$x_1'^*$ is an interim output level, occurring only during a sequence of periods following an initial draw of θ_B. Once state A occurs, output follows a Markov process that alternates between levels k_1 and $x_1''^*$, with ρ equal to the probability that the same output persists into the next period. The expressions for k_1^* and $x_1''^*$, coupled with the demand function, yield the following peak and off-peak prices for the steady state distribution of prices:

$$p_A = c_1 + 2(1-\delta)f_1 + (1-\delta\rho)g_1 \tag{12}$$

$$p_B = c_1 - \delta(1-\rho)g_1 \tag{13}$$

When the entry cost parameter g_1 is zero, equations (12) and (13) yield the familiar peak and off-peak prices from the peak-load pricing literature; Williamson [1966]. The off-peak price is equal to marginal operating cost (c_1) and the peak price is equal to the marginal operating cost plus the capacity rental rate ($(1-\delta)f_1$) divided by the frequency of the high demand state (1/2).

As the entry cost parameter g_1 increases the off-peak price falls and the peak price rises. In particular note that the off-peak price is less than marginal operating cost by the amount, $\delta(1-\rho)g_1$. This amount is equal to the discounted, expected savings in next period entry cost associated with keeping capacity active in the low demand state. Firms are willing to operate at a loss in low demand states in order to avoid the cost associated with exiting and then re-entering when demand reaches a high state. The peak price rises with higher values of g_1 since entrants must be compensated with higher peak prices to be willing to

[14] The results reported in section 3.3 are based on an interior solution with $x_1''^* < k_1^*$. A sufficient condition for an interior solution is that $g_1 < \theta_A - \theta_B - 2(1-\delta)f_1$.

incur higher entry costs for transitions between low and high demand states. The peak price p_A includes the term, $(1-\delta\rho)g_1$, which we define as an entry premium. When a firm enters in state A, the expected revenue associated with the entry premium for that entry episode is, $\sum_{\tau=0}^{\infty}\delta^{\tau}[\rho^{\tau}(1-\delta\rho)g_1] = g_1$. The entry premium component of the peak price yields exactly enough expected revenue to cover the entry cost.

The amount of capacity started up during transitions between low and high demand states $(k_1^* - x_1''^*)$ is increasing in the discount factor δ and demand persistence parameter ρ. While increases in these parameters reduce the peak price, they also raise the expected duration of a sequence of high demand states or yield less discounting of future payoffs for high demand states. The latter effects outweigh the price effects of greater demand persistence and higher discount factors. Note also that the amount of capacity started up during transitions between low and high demand states is decreasing in g_1.

Let's examine the entry and exit decisions of firms in more detail. A natural question that arises is whether there is an *option value of waiting* to enter. Dixit [1989] characterized a positive option value for waiting to enter for a firm facing an exogenous price process that was able to repeatedly enter and re-enter a market. On the other hand, Dixit and Pyndick [1994, ch. 8] show that in a dynamic competitive model with aggregate demand shocks, no firm-specific shocks, and endogenous prices, the option value of waiting to enter is zero. Entry of competing firms completely dissipates any option value of waiting. How do these forces play out in our model? Consider a firm that initially invests, enters in high states and exits in low states. Let π_A be the firm's value of profits beginning in a high state with with active status and π_B be its value of profits beginning in a low state with inactive status. These values satisfy the following conditions:

$$\pi_A = p_A - c_1 + \delta\rho\pi_A + \delta(1-\rho)\pi_B \tag{14}$$

$$\pi_B = \delta\rho\pi_B + \delta(1-\rho)(\pi_A - g_1) \tag{15}$$

where price p_A is defined in (12). The zero-profit investment condition implies that: $0.5(\pi_A - g_1) + 0.5\pi_B = f_1$. This may be verified by solving for π_B in (15) and substituting for π_B and

p_A in (14).

Consider the decision of whether to enter in the high demand state; let v^e be the value of entry and let v^w be the value of waiting.

$$v^e = \pi_A - g_1 = \frac{2(1-\delta\rho)(1-\delta)}{\psi} f_1$$

$$v^w = 0 + \delta(\rho(\pi_A - g_1) + (1-\rho)\pi_B) = \frac{2\delta(1-\rho)(1-\delta)}{\psi} f_1$$

where, $\psi \equiv (1-\delta\rho)^2 - \delta^2(1-\rho)^2 > 0$. Suppose that initial investment in capacity is costless; i.e., $f_1 = 0$. This yields the kind of competitive model considered by Dixit and Pyndick [1994, ch. 8]. Note that if $f_1 = 0$ then the value of entry and the value of waiting are both zero. As in Dixit and Pyndick, the option value of waiting is dissipated by entry of competing firms.

The situation is different if $f_1 > 0$. In this case, the value of waiting is positive, but less than the value of entry. There is an option value of waiting because the measure of firms that can potentially enter is limited by initial capacity investments. Because of this, profits are not completely dissipated by entry decisions. Similarly, one can show that if a firm is inactive and demand is low then the option value of waiting to enter is positive. So the feature of initial costly and irreversible capacity investments fundamentally changes the nature of option value of waiting for entry decisions. This investment feature limits the number of firms that compete in the market and make entry and exit decisions. This in turn limits the extent of profit dissipation due to entry decisions.[15]

Next we consider the case in which firms invest in both technologies in equilibrium. The structure of equilibrium mirrors the case of investment in a single technology. Output is equal to total capacity, $k_1 + k_2$, in all periods with high demand (state A). If the initial demand

[15] Our results on the option value of waiting to enter bear some similarity to those of Novy-Marx [2007] for the option value of waiting to invest. Novy-Marx [2007] formulates a competitive model with a fixed measure of firms and stochastic demand shocks. A firm may invest in new capacity at any time, with a strictly convex cost of investment, but when a firm invests its old capacity is scrapped. Thus firms face an opportunity cost of investment associated with replacement of capacity. He shows that with a particular initial distribution of heterogeneous firm capacity sizes, there is an option value of waiting to invest in equilibrium. The features of a fixed initial measure of firms and strictly convex investment costs prevent this option value of waiting from being dissipated by competition. These features in Novy-Marx's model play a role similar to the positive cost of investment in capacity in our model.

realization is low (state B) then entry and output are equal to x_1'; at the first transition to state A there is entry equal to $k_1 - x_1'$ for technology 1 firms and k_2 for technology 2 firms, and output rises to $k_1 + k_2$. In any transition from state A to B all technology 2 firms exit and output drops to k_1. In subsequent transitions from B to A all technology 2 firms enter. Technology 2 is the 'swing technology', utilized only during high demand states. In the appendix we derive the following solution for steady state equilibrium prices:

$$\tilde{p}_A = c_2 + (1 - \delta\rho)g_2 + 2(1 - \delta)f_2 \tag{16}$$

$$\tilde{p}_B = c_1 - \delta(1 - \rho)g_1 - \frac{1 - \rho\delta}{\delta(1 - \rho)}(\Delta - (1 - \rho\delta)(g_1 - g_2)) \tag{17}$$

In an equilibrium with technology 2 firms, steady state prices are similar to those for an equilibrium with only technology 1. The peak price $\tilde{p}_A$ has the same form as the peak price in (12), with technology 1 cost parameters replaced by technology 2 cost parameters. The off-peak price $\tilde{p}_B$ is equal to the off-peak price in (13) plus an additional term that is proportional to the expression, $\Delta - (1 - \rho\delta)(g_1 - g_2)$. This expression plays a key role in the following proposition.

Proposition 3.5. *If $\Delta > (1-\rho\delta)(g_1-g_2)$ then firms invest only in technology 1 in equilibrium and steady state equilibrium prices are defined in (12) and (13). If $\Delta < (1 - \rho\delta)(g_1 - g_2)$ then firms invest in both technologies in equilibrium and steady state equilibrium prices are defined in (16) and (17).*

Proposition 3.5 provides conditions under which an entry cost advantage for technology 2 is large enough to offset its production and investment cost disadvantage. Let $g_2 = \beta g_1$, where β is a fraction less than one; a low value for β indicates a large entry cost advantage for technology 2. Then the second inequality in Proposition 3.5 may be expressed as:

$$\rho < (1 - \frac{\Delta}{g_1(1 - \beta)})/\delta \tag{18}$$

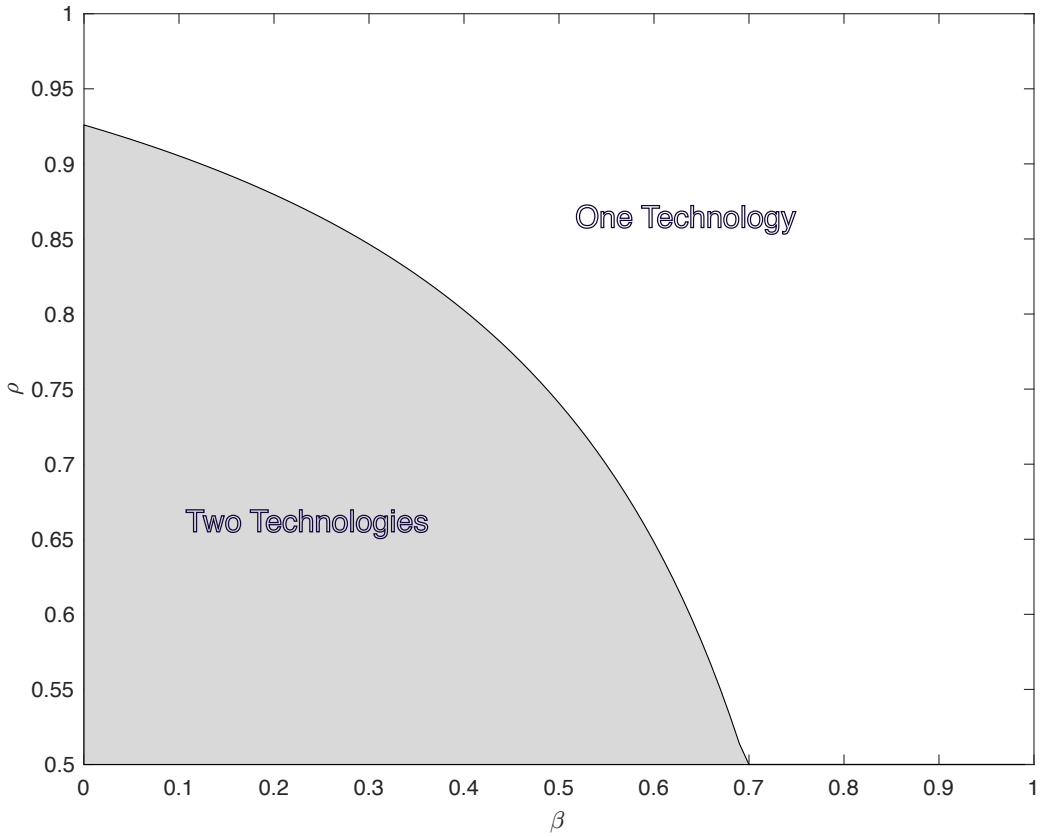
Figure 1 illustrates how entry cost advantage for technology 2 and demand persistence in-

fluence equilibrium investment. Parameter values for (β, ρ) that lie in the shaded area in Figure 1 satisfy inequality (18). Such parameter values are consistent with a large entry cost advantage for technology 2 and relatively low demand persistence; these parameters induce equilibrium investment in both technologies. Parameter values that lie above the curve in Figure 1 induce equilibrium investment only in technology 1. Note that a high value of ρ requires a large entry cost advantage for technology 2 in order for firms to invest in technology 2 in equilibrium. Put somewhat differently, as ρ falls entry and exit become more frequent, a small entry cost advantage is enough for technology 2 to offset its production and investment cost disadvantage.

$\Delta = (1 - \rho\delta)(g_1 - g_2)$ is a knife-edge case for which there are multiple equilibria. There is one equilibrium in which only technology 1 firms invest. Other equilibria involve investment in both technologies, with at least some of the entry that occurs in transitions between states B and A comprised of entry of technology 2 firms. Note that for this knife-edge case, steady state equilibrium prices defined in (12) - (13) and in (16) - (17) are the same, which is consistent with Proposition 3.3.

These results illustrate the importance of short-run dynamics and entry cost differences for equilibrium investment results. Under the $\Delta > 0$ assumption, the 'static' model with no entry/exit costs implies no investment in the higher marginal cost technology. When positive and heterogenous entry costs are taken into account, the dynamic model may yield investment in both technologies. Moreover, the incentive for investment in the high marginal cost technology is sensitive to the persistence of demand shocks.

Figure 1: Investment Incentives
(Parameters: $\delta = 0.9$, $g_1 = 6\Delta$)



4 Application to Investment in Electricity Markets

We apply the model developed in Section 3 to analyze the effects of environmental policies on electricity market outcomes. Specifically we investigate the short-run and long-run effects of subsidizing renewable energy on a typical electricity market by calibrating and numerically solving the model. We compare our results to a static model (without startup costs) to examine the extent to which ignoring market frictions would misrepresent market outcomes in this setting.

This application maps well into the fundamentals of the model. First, electricity markets are characterized by large, long-lived investments in a variety of well-understood generating technologies. Once built, generator characteristics are more or less fixed. Then it is up to the owners of the asset to decide how best to utilize it. Second, many generating technologies face costs associated with participating in the market. To supply electricity to the market, a generator must startup and being operating at a sufficiently high level. Starting up generators is costly for operators both in terms of fuel and mechanical wear-and-tear. This corresponds to the entry cost in our model. Finally, demand in electricity markets is quite variable across seasons and even within a day. Demand volatility necessitates that at least some generators startup and shut down frequently. Any increase in demand volatility will intensify the role of startup costs in determining market outcomes. Increasing intermittent renewable energy generation through subsidies will tend to increase the uncertainty and volatility of the demand facing conventional generators.

Not only does this application align well with the model, but investment in electricity generating capital is an important question in its own right. Energy is the lifeblood of the modern economy. Electricity markets are highly regulated and will likely to see increased government involvement that seeks to reshape the grid towards a cleaner production portfolio. Understanding how government intervention affects grid operation and investment is of first order importance.

To operationalize the model, we first need to construct operational and investment cost parameters for each technology type. Second, we need to characterize the distribution and

sequence of demand realizations.

4.1 Technology Parameters

For the simulations that follow, we calibrate the parameters for four technology types: coal, gas combined cycle (GCC), gas turbine (peaker), and wind. Coal, GCC, and peaker units are conventional fossil-fuel generating technologies. They have a high level of control over their production level, have significant marginal costs of production and need to be forward looking due to production adjustment costs. Wind is included as the clean renewable technology. Wind generators face very different incentives than the other generators. They have no fuel costs, but their production depends on the availability of wind. Since they have very little discretion over their production level, they are a 'passive' technology in the model; their production per unit of capacity will be calibrated using wind data.

The technology parameters can be split into two types. First, there are the characteristics of each technology that determine how an existing unit of capacity will compete in the daily operations of the market. These include, marginal costs, minimum operating rate, and the cost of startup. Second, there is the cost of investment that drives the decision to invest in a unit of capacity of that technology. They include data on construction costs, maintenance costs, and the expected lifetime of the generator. Both types of costs are calibrated using data from the Energy Information Administration (EIA), the Electric Reliability Council of Texas (ERCOT), and estimates from the previous literature. The full details of calibration approach can be found in appendix B.

The calibrated parameters are shown in table 1. The first row shows the range of marginal costs across the different technologies. If this were the extent of the differences between technologies, we might expect to see only one technology in the market. However, the technologies vary across other dimensions as well. For example, wind has negligible marginal costs, but the highest investment cost. In addition, it cannot control its production level which depends exclusively on the availability of wind. On the other hand, peaker plants have the highest marginal costs, but are very flexible with no startup costs. They also have the

lowest investment cost. It is these tradeoffs that will lead to a mix of technologies in the market.

4.2 Electricity Demand

We capture demand variation with using real-world electricity demand data and a simple demand function. Following the model, demand varies across time periods according to a demand shock which is assumed to follow a Markov process. We use a linear inverse market demand function for $P(Q, \theta)$ as shown in equation (19).

$$P(Q, \theta) = \theta - bQ. \tag{19}$$

The current demand state θ is the (continuous) demand shock in a given period. The parameter b is constant slope term across all demand shocks. In order to calibrate this demand function we need to specify the slope parameter b as well as the distribution of demand shocks, $F(\theta|Z)$ where Z is the agents' information set for predicting demand.

To estimate the distribution of demand shocks, we use data on demand from the Texas ERCOT market. In order to accommodate stochastic wind production into the model, we characterize evolution of demand using residual demand. Residual demand is defined as aggregate demand for electricity less wind generated electricity. The distribution of residual demand is the relevant demand metric for conventional generators to optimize against. This is consistent with the notion that wind production is exogenously driven by the availability of wind due to weather patterns.[16] Note that the amount of electricity generated by wind

[16] A limitation of this approach is that it does not allow for curtailment of wind generation. Curtailment occurs when a wind turbine discards free energy by allowing wind to slip over the turbine blades un-utilized.

Table 1: Model Parameters

		Coal	GCC	Peaker	Wind
Marginal Cost	($/MWh)	25	43	60	0
Min. Output	(%)	0.95	0.70	0	1
Startup Cost	($/MW)	150	80	0	0
Investment Cost	($/kW)	2250	1140	830	2340

will change with the amount of installed wind turbine capacity. However, we assume that the patterns of wind availability, or the wind capacity factor, will not change with the installed capacity. We use one year of hourly data from ERCOT on wind capacity and wind generated electricity to back out the distribution of wind capacity factors. We then combine this with electricity demand data from the same time period to estimate the distribution of residual demand given wind capacity. Using both aggregate demand and wind capacity factors from the same grid in the same time period allows us to richly capture the empirical correlation between demand shocks and wind shocks. It also simplifies the dynamic problem facing conventional generators; they form expectations over the future distribution of residual demand shocks rather than tracking separate processes for wind production and aggregate demand. Although we use data from the Texas grid, the calibration is intended to represent a 'typical' electricity market; the patterns for wind and electricity demand in ERCOT are broadly representative of wind and demand in other areas of the U.S. Further details on the estimation method used for characterizing the Markov process for residual demand can be found in appendix B.

To calibrate the slope parameter b, we look to the large body of literature on wholesale electricity demand. In general, short-run electricity demand is very price inelastic. This inelasticity coupled with demand variability imply that heterogeneity in the generation portfolio will be valuable. Elasticity estimates for market level wholesale electricity demand vary from zero to 0.35, with results concentrated in the low end of that range; Patrick and Wolak [2001], Johnsen [2001], Bigerna and Polinori [2014]. We choose a value for price elasticity of 0.05, and set the value of parameter b to yield this price elasticity at the average price and quantity observed in the data.

4.3 Computation

After calibrating the parameters of the model, we can solve the planner's problem, as specified in Section 3, to characterize competitive equilibrium investments in different technologies, generator outputs, emissions, and market prices. Stage 2 of the planner's problem is an

infinite horizon stochastic DP problem with a solution that is conditional on a vector of generation capacities. The state vector for that DP problem is $(\boldsymbol{x}, \theta, \boldsymbol{Z}) = (x_1, x_2, x_3, \theta, h)$, where x_1, x_2, and x_3 are continuous levels of 'on' capacities for the three fossil fuel generator types, θ is the current continuous demand shock variable, and h indicates the hour-of-day.[17] The current demand shock and hour of day constitute the information set for constructing beliefs about the possible demand shock in the next period. A computational complication of this application is that the hourly discount factor δ for stage two is very close to one, which can slow down DP computation. We address this by using DP methods whose solution time is independent of the discount factor.

The continuous state space combined with the large discount factor presents a formidable computational problem. We solve this continuous DP problem by discretizing the state space and applying policy function iteration as a solution method. Since the solution time using policy function iteration does not depend on the discount factor, we can solve the DP problem in a reasonable amount of time. We discretize capacities and transform the stochastic process for continuous demand shocks into a probability transition matrix for discrete demand shocks, conditional on hour of day. However, this is not without its trade offs. By evaluating the value function on a discrete grid of points for the continuous state variables, we restrict the planner's choice of the next period 'on' capacities to the points on the grid. We are limited in the number of discrete points by the curse of dimensionality in the DP problem.

Equilibrium capacity investments are found by nesting the stage two DP computation within a stage one optimization problem of selection of capacities by the planner. We perform a grid search over possible capacities to maximize the net benefit. Since the DP problem must be solved for each potential set of capacities, it is imperative that we solve the DP problem quickly. The solution to this optimization problem yields the long run competitive equilibrium described in the model. Additional details on the DP solution method and capacity search algorithm may be found in appendix C.

[17] Wind 'on' capacities do not enter the dynamic problem. Wind production is exogenous and is incorporated into the DP problem through net load.

5 Environmental Policy Analysis

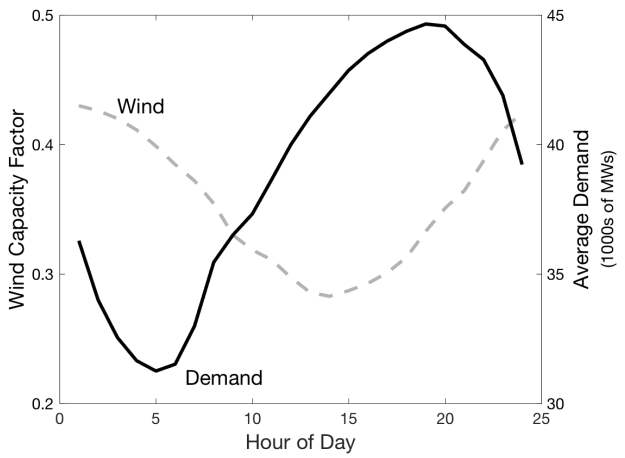
We use the competitive model developed in Section 3 coupled with model parameters specified in Section 4 to simulate the effect of renewable subsidies on investment in wind turbines and complementary technologies. Wind energy is the fastest growing source of renewable energy over the past 10 years. From 2005 to 2015, wind has accounted for 70% of new additions to renewable generating capacity and 32% of all new generating capacity. It has not been uncommon for annual wind capacity additions to be larger than either gas or coal capacity additions.[18] Renewable subsidies are a critical incentive for investment in wind; federal and state subsidies account for 40% to 60% of revenues for a wind farm; Cullen [2013b]. The expenditure on federal subsides alone for wind was 5.5 billion USD in 2010 and nearly 6 billion USD in 2013. While the short-run implications of increased wind power production have been studied extensively (Cullen [2013b], Kaffine et al. [2013], Fell and Kaffine [2014], Novan [2015]), the long-run effects of wind subsidies on investment have not. Critics of wind energy have argued that wind's uncertain and volatile patterns of production require expensive storage solutions or standby backup capacity. Indeed, land-based wind farms have highly variable production patterns that tend to supply energy when it is least needed. The first panel in figure 2 shows the hourly average wind production compared with electricity demand. As is clear from the figure, wind is most productive during periods of low demand. Looking at monthly average production and demand yields similar patterns. Wind power production is highest in the spring and summer when electricity demand is lowest (second panel). The off-cycle patterns in wind production thus tend to increase the variability of the residual demand facing conventional generators. Figure 3 shows that as wind capacity increases, production from wind turbines decreases the average residual demand, but at the same time significantly increases the volatility of residual demand. The introduction of 20,000 MWs of wind turbine capacity would decrease residual demand by 18%, but would increase in the variance of residual demand by 42%. Since conventional generators face production

[18] In 2008, 2009, 2012, and 2015 more new wind capacity was added in the United States then either coal or gas capacity additions. All capacity statistics come from data underlying figure MT-30 of the Annual Energy Outlook 2016; EIA [2016].

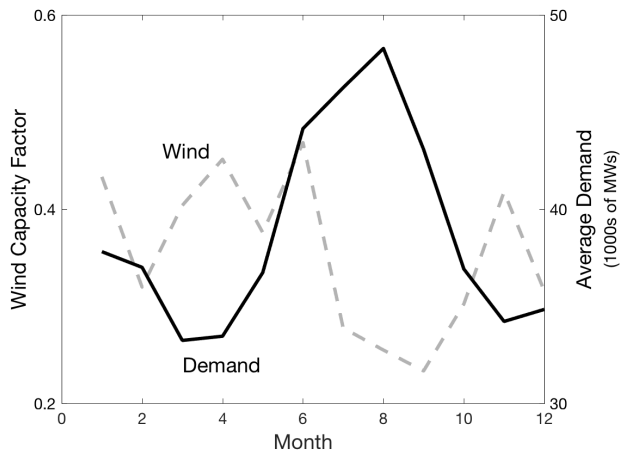
adjustment costs, this increased volatility will impose real costs on the system and will favor flexibility over efficiency. In the long run, this will change the portfolio of technologies in the market.

Figure 2: Wind Production and Electricity Demand Patterns

(a) Hourly Wind and Demand

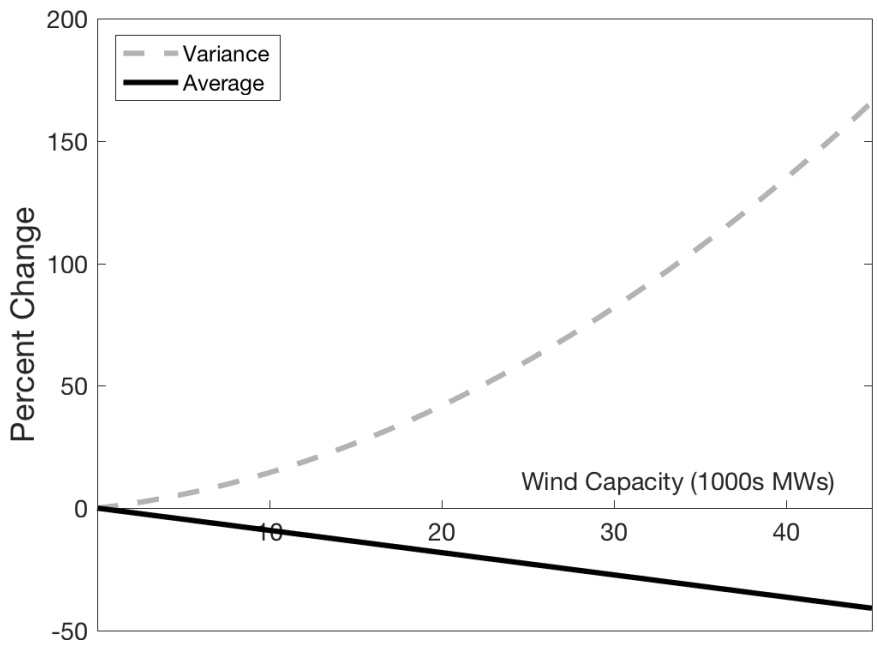


(b) Monthly Wind and Demand



For the results that follow, we first solve the model without a subsidy for renewable energy. In this baseline scenario, we compare the outcomes in the dynamic competitive equilibrium model with startup costs to the static competitive model calibrated with the same parameters, except for startup costs. We then introduce an investment subsidy for renewable technologies. The results will show that the presence of startup costs significantly changes the distribution of equilibrium prices, the mix of technologies invested in, and how the technologies are utilized. In addition, renewable subsidies are much less effective at spurring investment in wind capacity when startup costs for conventional generators are taken into account.

Figure 3: Wind Capacity and Residual Demand



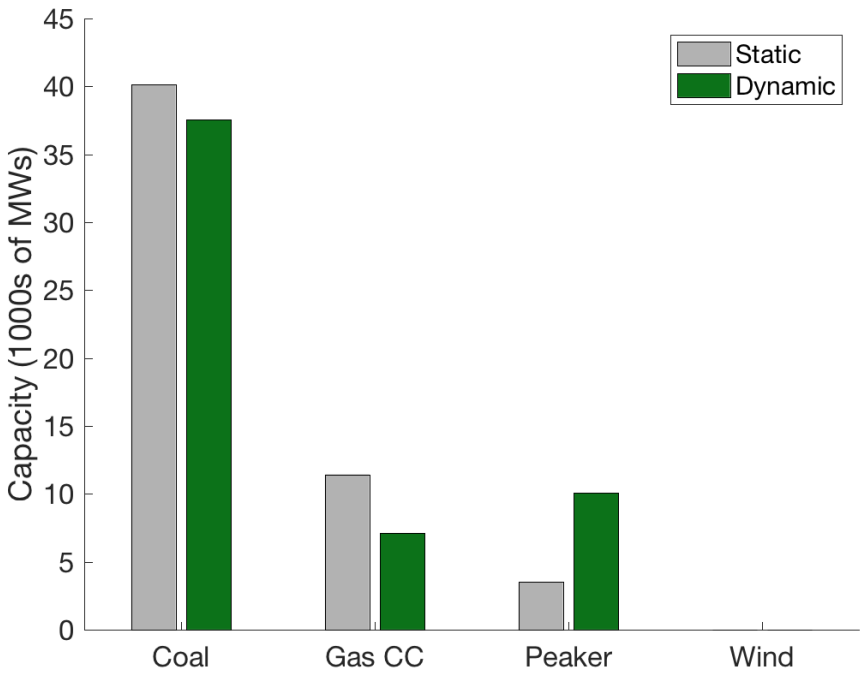
5.1 No Renewable Subsidy

First we examine market outcomes without a renewable subsidy. Figure 4 shows the portfolio of generators that would emerge in long-run equilibrium in the static model (no startup costs) and the dynamic model (with startup costs). Note that without a subsidy, we don't see any investment in wind in the static or the dynamic formulations of the model. Likewise in both models, coal is the primary technology used, with lesser amounts of gas capacity. This general pattern is driven by demand volatility and investment costs. Due to its high investment costs, coal is only cost-effective when operated near full capacity. Gas technologies on the other hand, may still be profitable when occasionally utilized. Even though more expensive to operate, they are cheaper to build. This is widely understood in the energy industry. Coal is employed to meet the base of electricity demand while other technologies are used to meet the peaks in demand.

Though there are many similarities in the investment profile, static and dynamic models

do differ significantly in the relative amounts of each technology used. Due to startup costs, we observe less investment in coal and GCC technologies in the dynamic model than in the static model We also see substantially more investment in flexible peaker capacity. This is entirely due to the startup costs facing coal and GCC generators. Without startup costs coal can operate like a peaker generator, flexibly responding to demand volatility. With startup costs this role is mainly relegated to generators with lower adjustment costs.

Figure 4: Capacity Investment

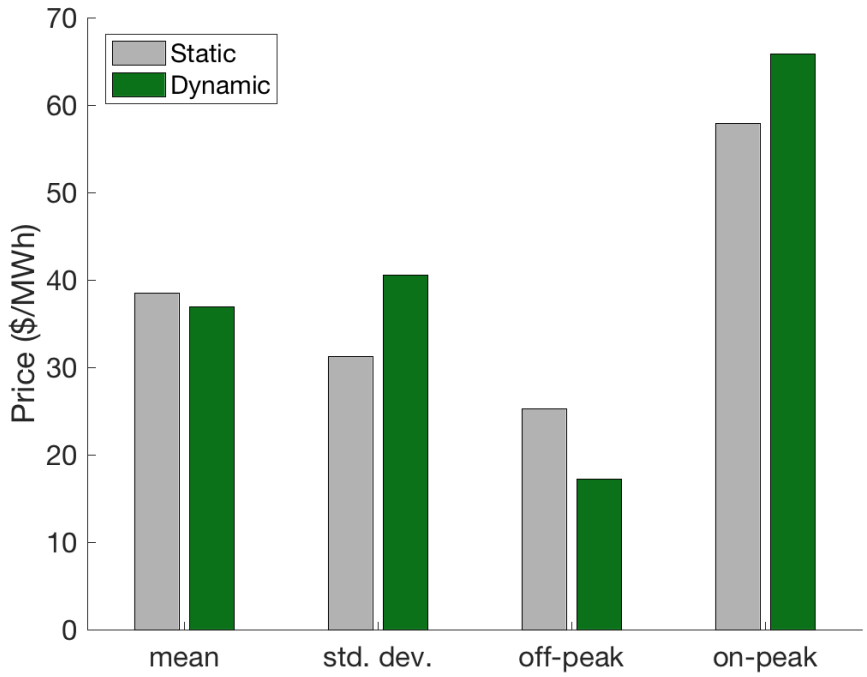


The implications of adjustment costs for equilibrium prices show up as predicted in the theory. Although average electricity prices are very similar across models, startup costs increase the variance of the distribution of prices as shown in figure 5. The increased variance is due to higher prices during high demand periods (on-peak) and lower prices during low demand periods (off-peak).[19] Higher prices occur when demand is high since firms must cover the costs of starting up their generators. Lower prices occur since forward looking firms are

[19]Here off-peak is defined as periods when demand is lower than average. On-peak are defined as periods when demand is higher than average.

reluctant to shutdown generators in low demand periods in order avoid later startup costs. In addition, there is an option value for waiting to shutdown an active generator that contributes to low off-peak prices.

Figure 5: Equilibrium Prices

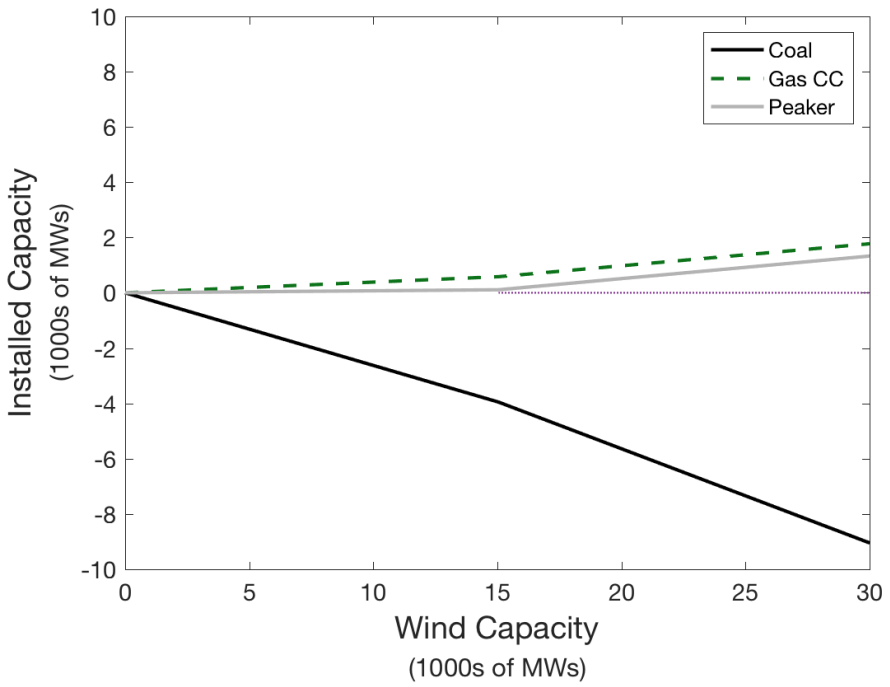


5.2 Wind Expansion

Next we explore how adding wind capacity affects the equilibrium price and investment in conventional technologies. For the moment we take these additions of wind capacity as exogenous in order to highlight the effect of wind capacity on equilibrium prices and investment in complementary technologies. We have already shown that adding wind capacity decreases average residual demand but increases the variance of residual demand (see figure 3). This change in the distribution of demand leads to increased equilibrium price variability and increases the returns for flexibility in the market. These forces affect the equilibrium investment in generating technologies. Figure 6 shows the changes in capacities of peaker,

GCC, and coal as wind capacity expands. Total capacity decreases due to wind energy production. However, the amount of flexible peaker capacity increases while the inflexible coal plants see the greatest reductions in capacity. Concurrently, the variance of equilibrium electricity prices increases significantly while the average price decreases slightly, as shown in figure 7.[20] In fact, even without startup costs, the variance of electricity prices increase as wind capacity expands. However, consistent with the theory, startup costs exacerbate the price volatility as shown in figure 8.

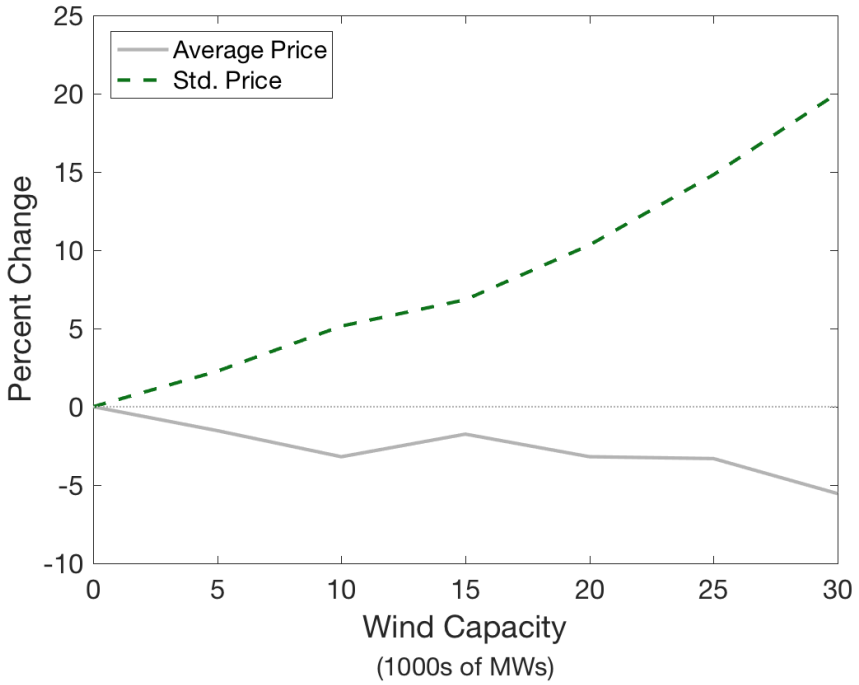
Figure 6: Wind Capacity and Conventional Capacity Change



Startup costs and their impact on market prices have a direct effect on the profitability of wind turbines. Lower off-peak prices due to startup costs decrease wind revenues since wind turbines are most productive in off-peak periods. As shown in figure 9, market frictions imply lower profits for even the first unit of wind capacity. Revenues for the first unit wind capacity are 5% lower when conventional generators face startup costs as opposed to a static

[20]The average price decreases since wind capacity is exogenously introduced into the market. This will change when the wind investment is endogenous and the costs of wind investment are reflected the market price.

Figure 7: Wind Capacity and Equilibrium Prices
(Dynamic)



dispatch model. This effect is stronger as more wind enters the market. Intermittent wind power production increases the volatility of residual demand for conventional generators and decreases average residual demand in off-peak periods, pushing down prices when wind is most productive. This suggests that adjustment frictions may lead to less investment in wind investment in equilibrium.

Figure 8: Wind Capacity and Price Volatility
(Dynamic vs Static)

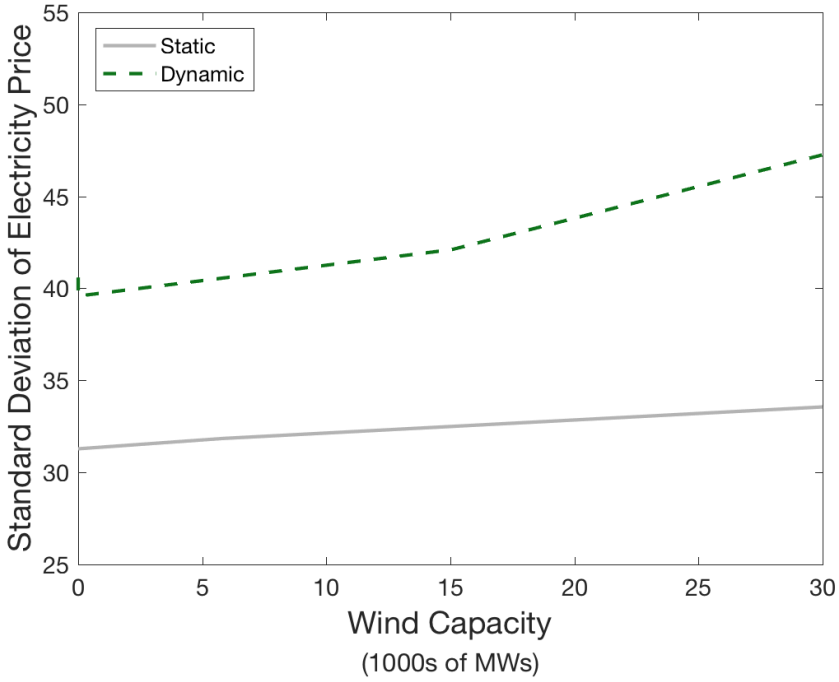
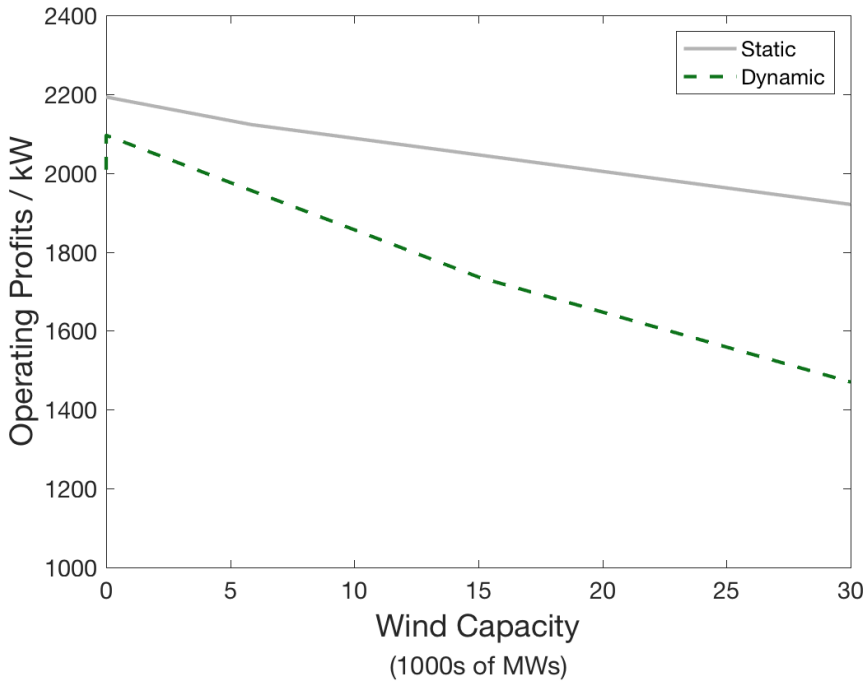


Figure 9: Wind Capacity and Wind Profitability
(Dynamic vs Static)



5.3 Renewable Subsidy

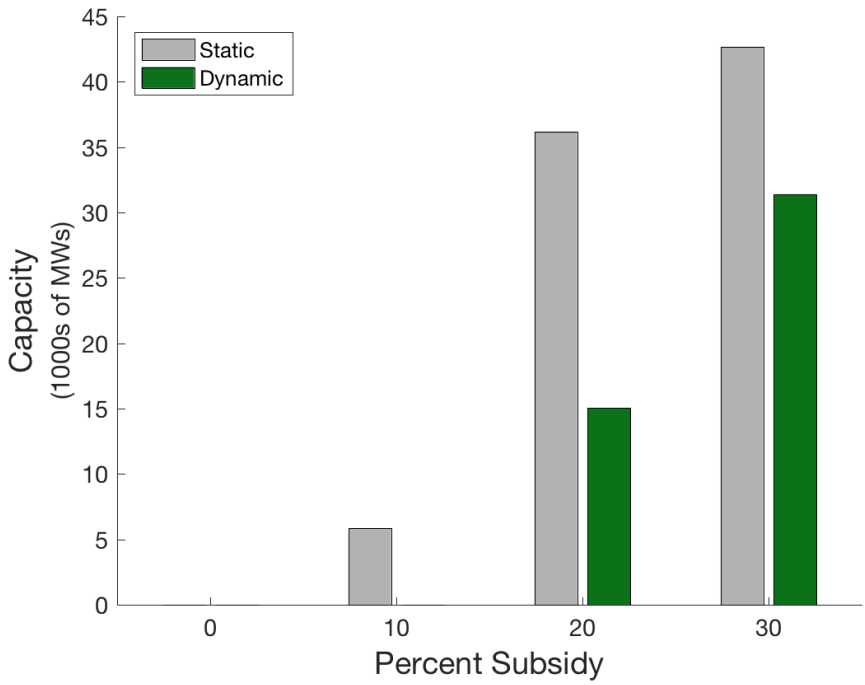
In this section we complete the analysis by allowing for endogenous investment of wind farms. Given the parameters of the model, no wind capacity is installed in the baseline model. To incentivize the investment of renewables, we introduce investment subsidies from 10% to 30% of the upfront cost of wind turbines. For high enough subsidies, the wind technology will be able to compete with conventional generators for a place in the market. The equilibrium investment portfolio reflects the interaction of two economic forces. First, in the presences of startup costs, the introduction of wind capacity reduces profitability and investment in inflexible technologies such as coal. Second, investing in wind capacity will be less attractive in a market in which conventional generators have startup costs. These forces combine to shape the equilibrium investment in conventional and renewable technologies. At a given level of renewable subsidy, these forces drive a wedge between the outcome of a frictionless market and one with startup costs.

Figure 10 shows the amount of wind capacity installed in long run equilibrium for various subsidy levels. Since wind power production increases the variation in residual demand, we would expect that startup costs will increasingly dampen incentives to invest in wind. First note that without an investment subsidy, wind energy is competed out of the market. At a 10% subsidy, wind generators start to enter the market without startup costs, but with startup costs, there is no investment in wind. At a 20% subsidy, we see wind investment in both models, but startup costs severely limit wind investment relative to the static model. The static model predicts more than twice as much wind investment. Once startup costs are accounted for, wind investment is significantly lower at every subsidy level. The startup costs of conventional generators greatly reduces the profitability of wind farms.

Next we compare the investment in each of the different technologies at each subsidy level as shown in figure 11. The technologies at extreme ends of flexiblity, coal and peaker, are most affected by the wind subsidy. Inflexible coal sees major reductions in capacity while peaker and GCC capacity grows. The added value of flexibility of GCC and peaker units implies consistently higher investment levels across all subsidy levels. In addition, the

difference between the static and dynamic investment levels is increasing for peaker plants as the demand for flexibility increases. On the other hand, the difference in GCC capacity investment between the static and dynamic frameworks is decreasing as more GCC capacity is demanded to substitute for even less flexible coal capacity. Interestingly, the amount of coal capacity in the dynamic setting may be higher or lower than in the static scenario due to differences in equilibrium wind capacity investment and the effect of that investment on coal profitability.

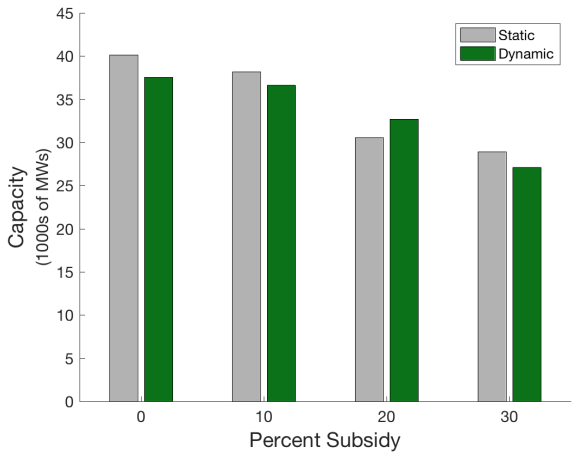
Figure 10: Equilibrium Wind Capacity Investment



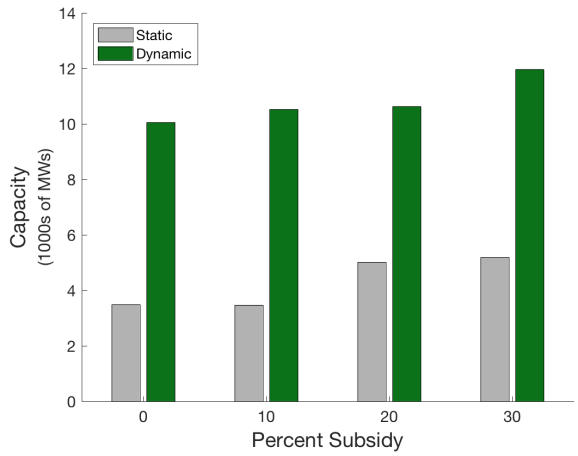
Although startup costs do shape the distribution of equilibrium electricity prices and capacity investment decisions, startup costs account for a small share of overall operating costs. As shown in figure 12, startup costs account for less than 0.1% of operating costs for coal plants and just under 0.2% of operating costs for gas combined cycle generators. Adding wind capacity does not affect the share of startup costs for gas cc generators, but does increase the role of startup costs for coal plants. However, even at high levels of wind capacity, startup costs account for less than 0.2% of operating costs for either technology.

Figure 11: Equilibrium Capacity Investment with Wind Subsidies

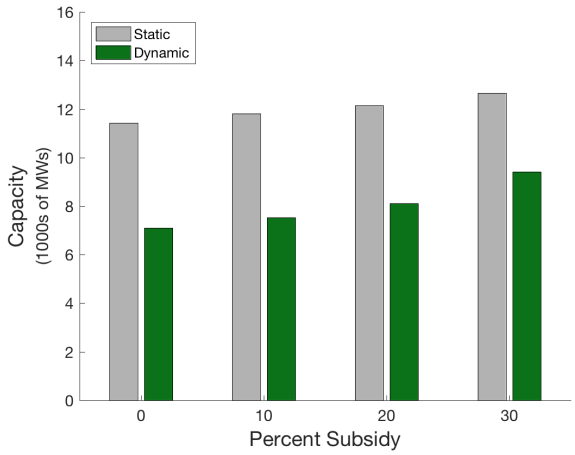
(a) Coal



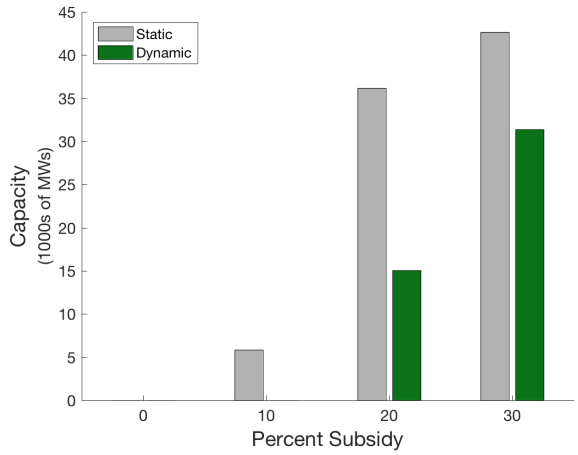
(b) Peaker



(c) Gas CC

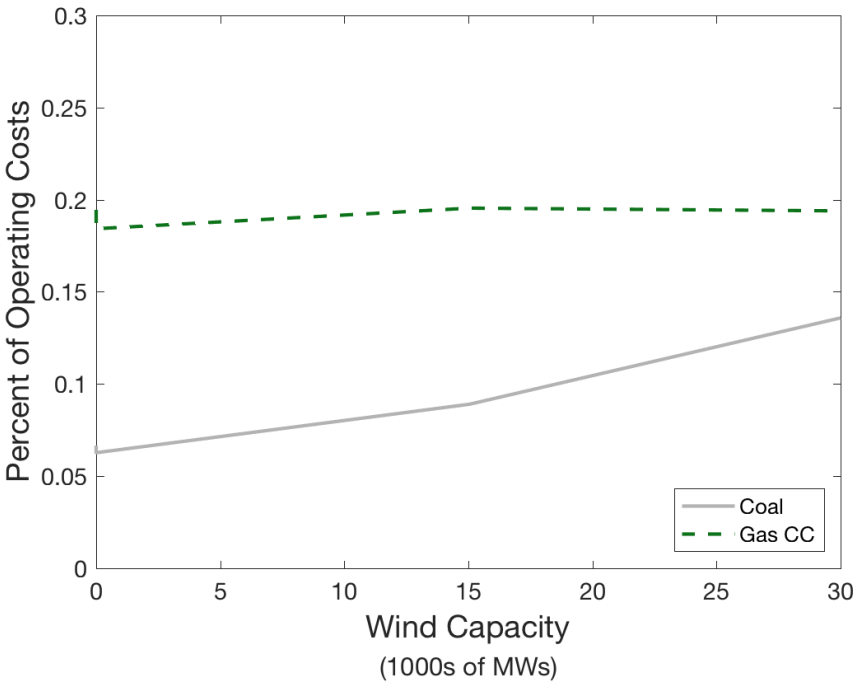


(d) Wind



This highlights the fact that even though market frictions may not directly contribute to costs in an accounting sense, they do shape behavior as firms seek to avoid paying those costs explicitly.

Figure 12: Wind Capacity and Startup Costs



All aspects of the results point to startup costs as a driving force for equilibrium investment in electricity sector. Ignoring these costs will lead to significantly different predictions for the mix of technologies on the grid. However, the results should not be interpreted as an exhaustive policy analysis of renewable subsidies. There are many aspects of electricity market that have been not been explored. For example, in order to highlight the economic forces in the model we maintain a parsimonious specification for technologies. We don't include technologies such as hydro, nuclear, or solar that could be important for the future portfolio of technologies. One could also consider adding wind power production flexibility into the model, either through discarding output (wind curtailment) or storing output (large-scale energy storage). In addition one could use the model to examine the investment implications of other policy interventions such as pricing carbon dioxide emissions. These extensions and

more are left for further research. Though not exhaustive, the application highlights how the framework can be used to model settings where persistent agents face entry/exit costs in a competitive market.

6 Conclusion

In this paper we develop a model of dynamic competition with persistent, ex-ante heterogeneous firms participating in a market with stochastic demand and entry/exit frictions. Firms make an initial capital investment followed by repeated entry, exit and re-entry decisions as demand fluctuates. The persistent nature of firms in our model introduces incentives for waiting to enter and for waiting to exit, even within a competitive framework. The inital investment phase allows for the endogenous formation of a pool of ex-ante heterogenous potential entrants. We extend prior work to show a correspondence between the model's dynamic competitive equilibrium and the solution to a social planner's problem. This allows us to establish equilibrium existence and also provides an attractive computational platform for numerical analysis.

We apply the framework to model long-run investment in electricity generating technologies in a way that has not been done before. We incorporate startup costs of conventional generators as an entry/exit friction in the market. The framework provides a tractable method of modeling competitive investment while accounting for the dynamics introduced by these startup costs. We find that incorporating generator startup costs into the model yields significantly different electricity prices and long-run generator investments, compared to outcomes for a 'static' model without startup costs. This is the case even though the incurred startup costs in equilibrium are relatively small. Startup costs are particularly important when evaluating the impact of policies to promote renewable generation; Perez-Arriaga and Batlle [2012]. Increased penetration of intermittent renewable generation yields more volatile net demand (demand less renewable output) served by conventional generators. A key finding is that the presence of startup costs reduces the profitability of wind turbines, leading to as much as 60% lower uptake of wind investment for a given renewable subsidy

level. These results underscore the importance of incorporating short-run market frictions when predicting long-run outcomes. The framework developed in this paper may be used for analysis of a range of counterfactual energy policies and is general enough to be applied to a variety of settings across economic fields.

Appendix

A Proofs of Propositions

Proof of Proposition 3.1

We introduce a variation of the model that was specified in Section 2. Let y_{jt} be the amount of exit of type-j capacity in period t; $\boldsymbol{y}_t$ is the vector of these amounts for all technologies. Let z_{jt} be the amount of type-j capacity that enters in period t; $\boldsymbol{z}_t$ is the vector of these entry amounts for all technologies. Using this notation, the aggregate production technology may be modified so that equation (3) is replaced by:

$$y_{jt} \in [0, x_{j,t-1}], \forall j, \forall t, \tag{20}$$

$$z_{jt} \in [0, k_j - x_{j,t-1}], \forall j, \forall t, \tag{21}$$

$$x_{j,t} = x_{j,t-1} - y_{jt} + z_{jt}, \forall j, \forall t. \tag{22}$$

The set of feasible allocations is convex since the constraint sets are compact and equation (22) is linear. Also, using this notation, the planner's single period total surplus function may be defined as:

$$\tilde{H}(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\theta}) = B(\sum_j q_j, \boldsymbol{\theta}) - \sum_j [c_j q_j + g_j z_j + h_j y_j] \tag{23}$$

$\tilde{H}$ is bounded and is concave and differentiable in $(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z})$. Note that we cannot have $y_{jt} > 0$ and $z_{jt} > 0$ in a socially optimal allocation for any $j \in \{1, 2, ..., J\}$. That is, an optimal allocation will not have simulataneous entry and exit for a single type of technology. This implies that maximal surplus for the planner's problem in the model of Section 2 and maximal surplus for the planner's problem in this variation of the model are identical. The version of the model with controls $\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z}$ has the advantage of differentiability of the planner's payoff, and this is the version used in the proofs.

($\Rightarrow$) Let feasible allocation $\{\boldsymbol{a}_t\} = \{\boldsymbol{k}, \boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t\}$ be a solution to the planner's problem. This allocation is a stochastic process that is measurable on the set of possible histories of demand shocks. This allocation induces a process $\{\boldsymbol{x}_t\}$ for active capacities and a price process, $\{p_t\}$, where $p_t = P(\sum_j q_{jt}, \theta_t)$.

Let $\{\boldsymbol{a}'_t\} \equiv \{\boldsymbol{k}', \boldsymbol{q}'_t, \boldsymbol{y}'_t, \boldsymbol{z}'_t\}$ be some alternative feasible allocation. Define a family of allocations by, $\{\boldsymbol{a}^\lambda_t\} = \{\lambda \boldsymbol{a}'_t + (1-\lambda)\boldsymbol{a}_t\}$, for $\lambda \in [0,1]$. $\{a^\lambda_t\}$ is a feasible allocation since the set of feasible allocations is convex.

Since $\{\boldsymbol{a}_t\}$ is a solution to the planner's problem, we have,

$$-\sum_j f_j k^\lambda_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}_t)] - [-\sum_j f_j k_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)]] \leq 0$$

for $\lambda \in [0,1]$. This implies that,

$$-\sum_j \frac{f_j(k^\lambda_j - k_j)}{\lambda} + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t [\frac{\tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}_t) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)}{\lambda}]] \leq 0$$

for $\lambda \in (0,1)$, and

$$\lim_{\lambda \downarrow 0}[-\sum_j \frac{f_j(k^\lambda_j - k_j)}{\lambda} + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t [\frac{\tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}_t) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)}{\lambda}]]] \leq 0. \tag{24}$$

Gross benefit B is bounded above, since $P(0, \boldsymbol{\theta})$ has a finite upper bound for all $\boldsymbol{\theta}$ and quantities are limited by capacities. Therefore, H is bounded above. Since H is bounded above by a δ-integrable function, we can pass the limit operator inside the expectation and t-summation operator. So,

$$\lim_{\lambda \downarrow 0} -\sum_j \frac{f_j(k^\lambda_j - k_j)}{\lambda} + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \lim_{\lambda \downarrow 0}\{\frac{\tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}_t) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)}{\lambda}\}] \leq 0. \tag{25}$$

Taking the limits as λ approaches zero in (25) yields:

$$-\sum_j f_j(k'_j - k_j) + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \sum_j [(\frac{\partial B(\sum_j q_{jt}, \boldsymbol{\theta}_t)}{\partial Q} - c_j)(q'_{jt} - q_{jt}) - g_j(z'_{jt} - z_{jt}) - h_j(y'_{jt} - y_{jt})]] \leq 0$$

Define the price process $\{p_t\}$ to equal the derivative of the gross benefit function with respect to total output for each period for each $\boldsymbol{\theta}_t$. Then,

$$\begin{aligned} &-\textstyle\sum_j f_j k'_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \sum_j [(p_t - c_j)q'_{jt} - g_j z'_{jt} - h_j y'_{jt}]] \\ &\leq -\textstyle\sum_j f_j k_k + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \sum_j [(p_t - c_j)q_{jt} - g_j z_{jt} - h_j y_{jt}]] \end{aligned} \tag{26}$$

The RHS of inequality (26) is aggregate market profit at the allocation $\{\boldsymbol{a}_t\}$ and price process $\{p_t\}$ induced by the solution to the planner's problem. Since $\{\boldsymbol{a}'_t\}$ is an arbitrary alternative feasible allocation, this inequality implies that, given $\{p_t\}$ and rational expectations on this process, there is no other feasible allocation that yields greater aggregate market profit than the allocation induced by the planner's solution.

The last part of the *if* proof is to connect aggregate market profit maximization to maximization of individual firms' profits. Given the allocation $\{\boldsymbol{a}_t\}$ we induce operating policies for firms as follows. For type-j firms, we assign the fraction $k_j/\overline{k}_j$ to invest and the remaining fraction to not invest. For period t and history θ^t there is a vector $(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{x}_{t-1})$ associated with $\{a_t\}$. For type-j firms, mass $x_{j,t-1}$ have $\omega_{t-1} = 1$ and mass $k_j - x_{j,t-1}$ have $\omega_{t-1} = 0$. If $x_{j,t-1} > 0$ then we assign the fraction $y_{jt}/x_{j,t-1}$ of active firms to exit in period t. Type-j firms are assigned output equal to $\gamma_t = q_{jt}/x_{jt}$. If $x_{jt,t-1} < k_j$ then we assign the fraction $z_{jt}/(k_j - x_{j,t-1})$ of inactive firms in $t-1$ to enter and the remaining fraction of firms to remain inactive. Any such assignment of operating policies yields a stochastic process for (γ_t, ω_t) that satisfies (5) for each possible demand shock history, for each firm. The assigned operating policy for a firm yields a (expected, discounted) profit for the firm, and total profits of all firms from period zero are equal to aggregate market profit on the RHS of (26). This equivalence holds since the operating policy assignment used implies that the aggregate production quantities and entry and exit capacities of firms add up to the q_{jt} quantities, z_{jt} entry capacities, and y_{jt} exit capacities on the RHS of (26).

Suppose that there is an alternative policy for a firm that yields greater profit than the policy assigned based on allocation $\{\boldsymbol{a}_t\}$. Furthermore, suppose that the measure of firms for which this is true is positive. Then an alternative market allocation can be constructed such that the firms that achieve greater profit with an alternative policy are assigned the alternative policy, and other firms retain the policy based on allocation $\{\boldsymbol{a}_t\}$. This alternative market allocation is feasible, since there are no externalities in production across firms. The aggregate profit for this alternative market allocation exceeds aggregate profit associated with $\{\boldsymbol{a}_t\}$ (the profit on the RHS of (26)). But this contradicts the result that aggregate profits are maximized using allocation $\{\boldsymbol{a}_t\}$; that is, there cannot be a positive measure of firms that achieve greater profit than the profit associated with the assigned policy from allocation $\{\boldsymbol{a}_t\}$. If there is an alternative policy for a firm that yields greater profit than the policy assigned based on allocation $\{\boldsymbol{a}_t\}$, and the measure of firms for which this is true is zero, then the policy assignment for these firms can be changed to the alternative policy without altering the market allocation or the price process.

We have shown that socially optimal allocation $\{\boldsymbol{a}_t\}$, along with associated process $\{p_t\}$, satisfies the 3 conditions for a market equilibrium:

(i) The allocation is feasible, since a socially optimal allocation must be feasible,

(ii) There is an assignment of policies to firms that maximize profit for each individual firm,

(iii) The price process satisfies the market clearing condition, since prices are set to clear the market for each period t and history θ^t.

($\Leftarrow$) Let $\{\boldsymbol{a}_t\} = \{\boldsymbol{k}, \boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t\}$ and $\{p_t\}$ be the allocation and price process, respectively, for a market equilibrium. Suppose there is an alternative feasible allocation $\{\boldsymbol{a}'_t\} \equiv \{\boldsymbol{k}', \boldsymbol{q}'_t, \boldsymbol{y}'_t, \boldsymbol{z}'_t, \boldsymbol{x}'_t\}$ that is not a market equilibrium and that yields greater total surplus than $\{\boldsymbol{a}_t\}$. Define $\{\boldsymbol{a}^\lambda_t\}$ as in the *if* part of the proof for $\lambda \in [0,1]$; we know that $\{\boldsymbol{a}^\lambda_t\}$ is a feasible allocation. By the concavity and differentiability properties of $\tilde{H}$ in $(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z})$, using Lemmas 1 and 2 from Hopenhayn [1990], the function $(\tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}))/\lambda$ is decreasing in λ, and increases to $\sum_j[(p_t - c_j)(q'_{jt} - q_{jt}) - g_j(z'_{jt} - z_{jt}) - h_j(y'_{jt} - y_{jt})]$ as $\lambda \downarrow 0$. In addition, concavity of $\tilde{H}$ implies that $(\tilde{H}(\boldsymbol{q}^\lambda_t, \boldsymbol{y}^\lambda_t, \boldsymbol{z}^\lambda_t, \boldsymbol{\theta}) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}))/\lambda$ is bounded below by

$(\tilde{H}(\boldsymbol{q}'_t, \boldsymbol{y}'_t, \boldsymbol{z}'_t, \boldsymbol{\theta}) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}))/\lambda$ for $\lambda \in (0, 1]$. These results are used in the string of inequalies below.

$$-\sum_j f_j k'_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \tilde{H}(\boldsymbol{q}'_t, \boldsymbol{y}'_t, \boldsymbol{z}'_t, \boldsymbol{\theta})] - [-\sum_j f_j k_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)]] > 0,$$

since $\{\boldsymbol{a}'_t\}$ yields greater total surplus than $\{\boldsymbol{a}_t\}$. This implies that,

$$-\sum_j \frac{f_j(k_j^\lambda - k_j)}{\lambda} + \tilde{\lambda}\mathrm{E}_0[\sum_t \delta^t[\frac{\tilde{H}(\boldsymbol{q}_t^\lambda, \boldsymbol{y}_t^\lambda, \boldsymbol{z}_t^\lambda, \boldsymbol{\theta}_t) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)}{\lambda}]] > 0$$

for $\lambda \in (0, 1]$, and

$$-\sum_j f_j(k'_j - k_j) + \lim_{\lambda \downarrow 0} \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t[\frac{\tilde{H}(\boldsymbol{q}_t^\lambda, \boldsymbol{y}_t^\lambda, \boldsymbol{z}_t^\lambda, \boldsymbol{\theta}_t) - \tilde{H}(\boldsymbol{q}_t, \boldsymbol{y}_t, \boldsymbol{z}_t, \boldsymbol{\theta}_t)}{\lambda}]] > 0.$$

This implies that,

$$\begin{aligned} &-\textstyle\sum_j f_j k'_j + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \sum_j [(p_t - c_j)q'_{jt} - g_j z'_{jt} - h_j y'_{jt}]] \\ &> -\textstyle\sum_j f_j k_k + \tilde{\delta}\mathrm{E}_0[\sum_t \delta^t \sum_j [(p_t - c_j)q_{jt} - g_j z_{jt} - h_j y_{jt}]]. \end{aligned}$$

In other words, we have shown that the supposition that there is some alternative feasible allocation $\{\boldsymbol{a}'_t\}$ that yields greater total surplus than a market equilibrium allocation $\{\boldsymbol{a}_t\}$ implies that total market profits at $\{\boldsymbol{a}'_t\}$ exceed those at $\{\boldsymbol{a}_t\}$. This is inconsistent with the property that individual firm profits are maximized at $\{\boldsymbol{a}_t\}$, since the strict inequality above implies that at least some firms could earn greater profits by choosing a different policy.

Proof of Proposition 3.2

The planner's problem has two parts: the stage one choice of the vector $\boldsymbol{k}$ of capacities and the stage two choice of a policy function for output, entry and exit in each period of stage two, conditional on $\boldsymbol{k}$. By the usual backward recursion logic, consider stage two first. Recall that function $\tilde{H}$ is bounded and is concave, and therefore continuous, in $(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z})$ for each $\boldsymbol{\theta}$.

Denote the constraint set for the planner's choice of $(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z})$ in a period by $G(\boldsymbol{x}', \boldsymbol{k})$:

$$G(\boldsymbol{x}', \boldsymbol{k}) \equiv \{(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z}) : y_j \in [0, x'_j]; z_j \in [0, k_j - x'_j]; q_j \in [m_j(x'_j - y_j + z_j), x'_j - y_j + z_j]; j \in \{1, ..., J\}\} \tag{27}$$

G is compact valued for each $(\boldsymbol{x}', \boldsymbol{k})$ and continuous in $\boldsymbol{x}'$ and $\boldsymbol{k}$. Define an operator T as follows:

$$(Tf)(\boldsymbol{x}', \boldsymbol{k}, \boldsymbol{\theta}) = \max_{(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z}) \in G(\boldsymbol{x}', \boldsymbol{k})} \{\tilde{H}(\boldsymbol{q}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\theta}) + \delta \mathrm{E}[f(\boldsymbol{x}' - \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{k}, \boldsymbol{\theta}^+) \mid \boldsymbol{\theta}]\} \tag{28}$$

Let C be the space of bounded functions that are continuous in $(\boldsymbol{x}', \boldsymbol{k})$ that map $X(\boldsymbol{k}) \times K \times \Theta$ into $\mathbb{R}$. Then by an argument similar to that in the proof of Theorem 4.6 in Stokey and Lucas [1989], the operator T maps elements in C into C and T has a unique fixed point, which is the stage two value function $W(\boldsymbol{x}', \boldsymbol{k}, \boldsymbol{\theta})$ that satisfies the Bellman equation (8) for the planner.

It can be shown that the stage two value function $W(\cdot)$ is concave in $\boldsymbol{k}$. There exists an optimal choice of $\boldsymbol{k} \in K$ in stage one for the planner, since the payoff function is concave and the constraint set K is compact. The optimal $\boldsymbol{k}$ and the optimal policy associated with (8) yield a feasible allocation and a price process, which by Proposition 3.1 constitute a market equilibrium.

Proof of Proposition 3.5

Part One: Assume that, $\Delta - (1 - \rho\delta)(g_1 - g_2) > 0$. We suppose initially that firms invest only in technology 1 in equilibrium, and then show that this must hold. Given just two demand states and a single technology, the structure of equilibrium is fairly simple. Output is equal to total capacity k_1 in all periods with high demand (state A). If the initial demand realization is low (state B) then entry and output are equal to x'_1; at the first transition to state A there is entry equal to $k_1 - x'_1$ and output rises to k_1. In any transition from state A to B there is exit equal to $k_1 - x''_1$ and output drops to x''_1. In subsequent transitions from B to A there is entry equal to $k_1 - x''_1$. The payoff $\tilde{W}$ to the planner may be expressed as a

function of (x_1', x_1'', k_1).

$$\begin{aligned}\tilde{W}(x_1', x_1'', k_1) = & \tfrac{1}{2(1-\delta)}[B(k_1, \theta_A) - c_1 k_1] + \tfrac{1}{2(1-\delta\rho)}[B(x_1', \theta_B) - c_1 x_1'] \\ & + \tfrac{\delta(1-\rho)}{2(1-\delta)(1-\delta\rho)}[B(x_1'', \theta_B) - c_1 x_1''] - \tfrac{\delta(1-\rho)}{2(1-\delta\rho)}[g_1(k_1 - x_1')] \\ & - \tfrac{\delta^2(1-\rho)^2}{2(1-\delta)(1-\delta\rho)}[g_1(k_1 - x_1'')] - \ -\tfrac{1}{2}g_1 x_1' - \tfrac{1}{2}g_1 k_1 - f_1 k_1\end{aligned}$$

The first term on the RHS of the payoff expression is gross benefit less production cost in the A states. The second term is gross benefit less production cost in B states that do not follow initial A state. The third term is gross benefit less production cost in B states that follow an initial A state. The fourth term is entry cost for a transition between an initial B state to the first A state. The fifth term is entry cost for all transitions from B states to A states after the first A state occurs. The sixth and seventh terms are entry costs for initial states B and A, respectively. The final term is investment cost. $\tilde{W}$ is quadratic and concave in its three arguments. The necessary conditions for an interior solution (with $x_1'' < k_1$) yield the solutions in (9) - (11). Peak and off-peak steady state prices are given by (12) and (13), respectively.

Suppose there is an equilibrium in which technology 2 firms invest, enter whenever demand is in state A, and exit in transitions from A to B. Let $\tilde{p}_A$ be the equilibrium price for state A; the price(s) in state B is irrelevant for such a firm's profits. Let π_2^A be the firm's value of profits beginning in state A with active capacity and let π_2^B be the firm's value of profits beginning in state B with inactive capacity. These values satisfy the following two conditions.

$$\pi_2^A = \tilde{p}_A - c_2 + \delta\rho\pi_2^A + \delta(1-\rho)\pi_2^B \tag{29}$$

$$\pi_2^B = \delta\rho\pi_2^B + \delta(1-\rho)(\pi_2^A - g_2) \tag{30}$$

The value of initial investment for this technology 2 firm is:

$$v_2^e = \frac{1}{2}[\pi_2^A - g_2] + \frac{1}{2}\pi_2^B - f_2 \tag{31}$$

Solving for π_2^A and π_2^B in (29) and (30) and substituting on the RHS of (31) yields:

$$v_2^e = \frac{1}{2(1-\delta)}\tilde{p}_A - \frac{1}{2(1-\delta)}[c_2 + 2(1-\delta)f_2 + (1-\rho\delta)g_2] \tag{32}$$

The first term on the RHS of (32) is the expected DPV of revenue and the second term is the expected DPV of total costs of investment, entry and operation. Note however that under the assumption that $\Delta - (1-\rho\delta)(g_1 - g_2) > 0$, if a technology 1 firm made the same investment, entry and production decisions as this technology 2 firm, its expected DPV of revenue would be the same and its expected DPV of total cost would be lower. By Corollary 3.4 firms earn zero expected profit in equilibrium, so technology 1 firms cannot earn strictly greater profit from investing in stage 1 than technology 2 firms. This contradicts the supposition that technology 2 firms invest and produce in state A in equilibrium. A similar argument establishes that technology 2 firms never invest in equilibrium if $\Delta - (1-\rho\delta)(g_1 - g_2) > 0$.

Part Two: Assume that, $\Delta - (1-\rho\delta)(g_1 - g_2) < 0$. If a technology 2 firm invests, enters whenever demand is in state A, and exits in transitions from A to B then the value of its profit is given by v_2^e in (32), where $\tilde{p}_A$ is the equilibrium price in state A. If a technology 1 firm invests, enters whenever demand is in state A, and exits in transitions from A to B then the value of its profit is strictly less than v_2^e, given the assumption that $\Delta - (1-\rho\delta)(g_1 - g_2) < 0$. Therefore, an equilibrium must involve technology 2 firms playing the role of 'swing' producers who enter in state A and exit in transitions from A to B. Furthermore, our assumption that $\Delta > 0$ insures that technology 1 firms are the firms that operate in all demand states, rather than technology 2 firms.

The structure of equilibrium is fairly simple. Output is equal to total capacity, $k_1 + k_2$, in all periods with high demand (state A). If the initial demand realization is low (state B) then entry and output are equal to x_1'; at the first transition to state A there is entry of technology 1 firms equal to $k_1 - x_1'$, entry of technology 2 firms equal to k_2, and output rises to $k_1 + k_2$. In any transition from state A to B all technology 2 firms exit. In subsequent transitions from B to A all technology 2 firms enter. The payoff $\hat{W}$ to the planner may be expressed as a function of (x_1', k_1, k_2).

$$\hat{W}(x_1', k_1, k_2) = \tfrac{1}{2(1-\delta)}[B(k_1+k_2,\theta_A) - c_1k_1 - c_2k_2] + \tfrac{1}{2(1-\delta\rho)}[B(x_1',\theta_B) - c_1x_1']$$
$$+\tfrac{\delta(1-\rho)}{2(1-\delta)(1-\delta\rho)}[B(k_1,\theta_B) - c_1k_1] - \tfrac{\delta(1-\rho)}{2(1-\delta\rho)}[g_1(k_1 - x_1') + g_2k_2]$$
$$-\tfrac{\delta^2(1-\rho)^2}{2(1-\delta)(1-\delta\rho)}[g_2k_2] - \ -\tfrac{1}{2}g_1x_1' - \tfrac{1}{2}[g_1k_1 + g_2k_2] - f_1k_1 - f_2k_2$$

$\hat{W}(\cdot)$ is quadratic and concave. The necessary conditions for optimality yield the following welfare-maximizing expressions for x_1', k_1, and k_2:

$$\hat{x}_1' = \theta_B - c_1 - (1-\delta)g_1 \tag{33}$$

$$\hat{k}_1 = \theta_B - c_1 + \delta(1-\rho)g_1 + \frac{1-\rho\delta}{\delta(1-\rho)}[\Delta - (1-\rho\delta)(g_1 - g_2)] \tag{34}$$

$$\hat{k}_2 = \theta_A - c_2 - (1-\delta\rho)g_2 - 2(1-\delta)f_2 - \hat{k}_1 \tag{35}$$

The expressions for $\hat{k}_1$ and $\hat{k}_2$ may be used to derive steady state equilibrium prices in (16) and (17).

B Parameter Calibration

The cost parameters for technologies are built up from data on cost and technology fundamentals. In this section, we describe how marginal costs, investment costs, and startup costs were calibrated. Rather than using the average characteristics of existing plants, we calibrate the paramters using the characteristics for the latest generation of newly constructed generators. Using state-of-the-art technology parameters corresponds best to the notion of long run equilibrium in the model. It reflects the best guess for what can expected in the future as old facilities are phased out and are replaced by newer construction.[21]

Fuel costs are projected future fuel costs in 2035, as estimated by EIA [2015]. Fuel costs

[21] It should be noted that these can be much lower than current average efficiencies for technologies. In 2014, average heat rates for existing coal plants was 10,485 BTU/kWh, approximately 19% less efficient than heat rates used here for newly constructed coal plants. Likewise, existing gas GT (11,378 BTU/kWh)and gas CC (7658 BTU/kWh)are 17% and 9% less efficient than newly constructed facilities used for the model.These parameters are based on estimates for new plants coming online in 2012 and 2013; see EIA [2010].

and heat rates are used to construct the marginal cost of production for each technology. The marginal cost of a generator is computed as the product of its heat rate and fuel cost plus any costs variable maintenance and emissions costs.

B.1 Marginal costs

The marginal cost of producing electricity from conventional generators is comprised of three parts: fuel costs, emission costs, and variable maintenance costs. To calculate fuel costs we need information on the the conversion efficiency, or heat rate, of the generator as well as the cost of fuel used for production. We use data from the Annual Energy Outlook published by Energy Information Administration on the characteristics of new gas and coal fired power plants; EIA [2010]. In particular, we use the heat rate, variable operating costs, and emission rate for new scrubbed coal, advanced gas combined cycle, and conventional combustion turbine.

The three cost components are shown in equation (36):

$$MC_j = \underbrace{HR_j * FC_j}_{\text{Fuel Cost}} + VOM_j + \underbrace{R_{CO_2 j} P_{CO_2} + R_{SO_2 j} P_{SO_2} + R_{NO_x j} P_{NO_x}}_{\text{Emission Cost}} \tag{36}$$

where

$HR =$ Heat Rate

$FC =$ Fuel Cost

$VOM =$ Variable Operating and Maintenance costs

$R =$ Emission Rates for CO_2, SO_2, and NO_x

$P =$ Abatement Costs for CO_2, SO_2, and NO_x

Of these three, fuel costs have always dominated the marginal cost calculation. The fuel component is the raw cost of fuel multiplied by the conversion efficiency, or heat rate, of

the generator. As shown in Table 2, a state-of-the-art coal generator with a heat rate of 8.8 MWh/mmBTU facing delivered coal prices of \$2.25/mmBTU incurs fuel costs of \$19.80 to generate a single MWh of electricity.

In addition to fuel costs, plants may emit potentially regulated pollutants such as SO_2, NO_x, or CO_2. Creating these pollutants as a byproduct of combustion creates a cost for the firm in terms of operation of pollution control technologies and/or purchasing pollution permits. However, in the absence of CO_2 regulation, these pollution costs have tended to be of secondary importance.[22] The additional costs of pollution for each technology are quite small relative to the fuel costs even with the relatively high pollution permit prices assumed. This is in part due to the fact that new power plants include emission control technologies such as SO_2 scrubbers for coal plants. These technologies increase the operational costs of the plant, but greatly reduce emissions. The cost of operating pollution control technologies are included in the Variable Operating and Maintenance (VOM) costs shown. VOM also includes any other costs that related to the amount of electricity produced. Costs from VOM account for roughly 5-20% of total marginal costs depending on the technology.

Total marginal costs vary greatly across technologies. Peaking gas turbine (GT) technologies have marginal costs that are roughly two and a half times greater than coal plants. Wind generators have essentially zero marginal costs. If this were the extent of the differences between technologies then we would only see the lowest cost technology in the market. However, the technologies vary in their investment costs as well.

B.2 Investment Costs

Investment costs are taken from the same EIA publication used to calibrate marginal costs. In particular, we use the "overnight construction" costs reported. These costs include only the costs of construction; they do not include financing costs or a project contingency factor.[23]

[22] However, even a modest price on carbon would elevate pollution to a primary driver of costs. For example, a \$20 / ton price on carbon would increase the marginal cost of a coal plant by approximately \$20. This nearly doubles the marginal cost of producing a MWh of electricity at a coal plant.

[23] According to the EIA, projection contingency factor is defined by the American Association of Cost Engineers as "the specific provision for unforeseeable elements of costs within a defined project scope; particularly important where previous experience has shown that unforeseeable events which will increase costs

Table 2: Marginal Cost Parameters

		Coal	GCC	GT	Wind
Heat Rate	(mmBTU/MWh)	8.80	7.05	9.75	–
Fuel Cost	($/mmBTU)	2.25	5.75	5.75	–
MC Cost of Fuel	($/MWh)	19.80	40.53	56.06	0
SO2 Rate	(lbs/mmBTU)	0.1	0.001	0.001	0
SO2 Cost	($/lb)	600	600	600	0
NOx Rate	(lbs/mmBTU)	0.06	0.007	0.03	0
NOx Cost	($/lb)	2000	2000	2000	0
Emission Cost	($/MWh)	0.79	0.05	0.30	0
VOM	($/MWh)	4.70	2.00	3.25	0
Total MC	($/MWh)	**25.29**	**42.59**	**59.61**	**0**

We must also calibrate the lifetime of the generator. This is no easy task as the decision to keep a plant open is a economic choice, not a fixed characteristic of the technology. All generators have the potential for very long lifetimes. The empirical distribution of plant ages in the US shows that there is coal and gas capacity in excess of 50 years old; EIA [2014]. Motivated by empirical distributions of plant lifetimes showing a much greater density of older coal plants, we assume coal plants have a relatively longer lifetime than gas or wind generators.

At the end of the lifetime of the generators, the firms could potentially decided to no longer participate in the market or not to reinvest in the same asset. However, since the distribution of demand is stationary and the technology parameters are unchanging, the decision problem 50 years in the future looks exactly the same as in the first period. Firms will always decide to reinvest in the same manner. Thus, we can model the decision to invest as a commitment to invest now <u>and</u> to reinvest in the same technology at the end of the life of the asset. The relevant investment costs for a firm in beginning period is the present discounted investment costs of the technology over an infinite horizon. This is shown in equation (37):

$$Inv = \sum_{i=0}^{\infty} (\beta^L)^i G \tag{37}$$

are likely to occur".

where

$$G = \text{Generator Construction Costs}$$

$$\beta = \text{Annual Discount Factor}$$

$$L = \text{Lifetime of the Generator in Years}$$

In practice, the reinvest part of the investment costs if small relative to initial costs the due to discounting over the long lifetime of the assets. Table 3 shows the infinite horizon investment costs for each technology. Note that wind and coal which have the lowest marginal costs, have the highest investment costs. Gas turbines, which have the highest marginal costs, are the least expensive technology to build. The tradeoff between low operating costs and high investment costs will lead to a mix of technologies even in the absence of startup costs.

The calibrated parameters for both investment costs and marginal costs are reasonable approximations of real costs. However, they are still only rough calibrations. To avoid the perception of precision, we round marginal costs and investment costs the nearest \$1 and \$10 for our analysis and the numbers reported in the paper.

B.3 Dynamic Parameters: Startup Costs and Minimum Output

Dynamics enter the model through two key parameters: startup costs and minimum output rates. Without either of these features, the model is completely static. Without startup costs, current actions have no implications for future profits. On the other hand, if generators are completely flexible in their output level, then they can avoid any startup costs by producing

Table 3: Investment Cost Parameters

		Coal	GCC	GT	Wind
Construction Cost	(\$/kW)	2078	897	653	1837
Lifetime	(years)	50	30	30	30
Infinite Horizon Cost	(\$/kW)	**2,251**	**1,142**	**831**	**2,339**

minuscule quantities.

Generator startup costs are a key feature of our application. 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.[24] 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.[25] Kumar et al. [2012] estimate that capital and maintenance expenses comprise 80-98 percent of total startup costs, depending on generator and fuel type. Our startup cost parameters are based on Kumar et al. [2012], using their lower bound estimates for the capital and maintenance portion of startup costs as shown in Table 4. Our startup cost parameters are broadly consistent with estimates from structural models in Cullen [2013a] and Reguant [2014].

Minimum output rates are also an important feature of the model. If generators can ramp output down to arbitrarily low levels, then they will never have to incur startup costs. Higher minimum output rates sharpen the effect of start ups, while lower minimum output rates weaken the effect. A generator with lower minimum output is more flexible than a generator with a high minimum output. We calibrate of minimum output rates for gas combined cycle using data from Cullen [2013a]. As noted in the body of the paper, we assume that combustion gas turbines are completely flexible. That is, they have no startup costs or equivalently that have no minimum output. Either assumption eliminates any dynamic implications of their production. In reality, gas turbines do incur small startup costs and do have non-zero minimum output constraints. However, we assume that the start costs are small enough as to be negligible. Coal technologies, on the other hand, are calibrated with a high minimum output rate. This is higher than the minimum output rate usually reported by these facilities. However, using a lower minimum output may overstate the flexibility of these generators. These generators are inflexible in ways not explicitly captured by the model. In addition to startup costs, coal generators must plan production far in advance

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

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

and move slowly to their targets. We capture this inflexibility by calibrating the minimum output at a higher level than would be indicated by the technological specifications of the plant.

Wind output is assumed to be at 100% of potential output. The potential output, or capacity factor, of a wind turbine in a given hour on a given day is determined by the wind speed in that period. Thus conditional on installed capacity, wind power generation in a given hour is fixed and exogenous. In a windy period, the capacity factor may be close to one. On average, the capacity factor of wind farms is around 0.35, though this varies greatly by season and time of day. We calibrate the capacity factor data from the Texas electricity grid (ERCOT) in 2014. ERCOT has substantial wind developments which are broadly representative of wind production patterns throughout the US. Using information on wind capacity and hourly wind production, we create an hourly wind capacity factor which can be then applied to the capacity chosen by the model to create the level of wind power production in the model. Implicit in this approach is that the capacity factor of wind investments is relatively constant over the relevant range of wind investments. This may overstate the production of wind capacity if wind farms must be sited in less productive locations as wind capacity increases. This would tend to overstate the attractiveness of investing in wind farms. Note that we calibrate demand using the same grid to allow us to carefully capture the correlation between wind availability and demand for electricity. Since we assume that all potential wind output is utilized, this rules out curtailment by wind generators. When wind generators curtail production, they are essentially throwing away free electricity by allowing wind to pass over the turbine blades unutilized. However, since wind farms have zero marginal costs curtailment would only occur if prices were to be below zero. In case of subsidizing wind with a production subsidy, prices would have to drop

Table 4: Dynamic Parameters

		Coal	GCC	GT	Wind
Startup Cost	($/MW)	150	80	0	0
Min. Output	(%)	0.95	0.70	0.70	1

substantially below zero in order to cancel out the subsidy received by production. Allowing for curtailment does have the potential to change the profitability of wind, especially at high levels of wind capacity. This investigation is left for future research.

B.4 Demand

Since we are using wind production data from ERCOT, we also use hourly demand realizations from ERCOT. This allows us to match hourly demand with hourly wind production to capture important correlations between electricity demand and production from renewables. While every electricity market is unique, the patterns observed in ERCOT are similar to the demand patterns in wholesale electricity markets in other parts of the U.S.

In order to accommodate stochastic wind variation without introducing additional state variables, we interpret the demand function as aggregate demand less wind generation. This is consistent with the notion that wind production exogenously driven by the availability of wind due to weather patterns. In this case, the conventional generators face the residual demand that is left after wind production.[26] Specifically, we estimate the demand process using hourly net load data from ERCOT in 2014 to estimate a process for demand shocks where net load is defined as realized demand less any production from renewable sources.[27]

The estimation process regresses hourly net load on hour-of-day dummy variables and interactions of hour dummies with lagged net load and squared and cubic lagged net load terms. The estimating equation is:

$$Q_t = \sum_{h=1}^{24} (\beta_{1h} D_{ht} + \beta_{2h} D_{ht} Q_{t-1} + \beta_{3h} D_{ht} Q_{t-1}^2 + \beta_{4h} D_{ht} Q_{t-1}^3) + \epsilon_t \tag{38}$$

where Q_t is net load in period t, D_{ht} is the hour h dummy for period t, the β terms are coefficients, and ϵ_t is the one-hour-ahead prediction error. The adjusted R^2 is very high at

[26] A limitation of this approach is that it does not allow for curtailment of wind generation. Curtailment occurs when a wind turbine discards free energy by allowing wind to slip over the turbine blades un-utilized.

[27] The wind generation data is conditional on the amount of wind capacity available during that time period. To allow for production to vary with the wind capacity choice, we convert wind production into a hourly wind capacity factor as described earlier. This adjustment allows for the sequence of hourly net loads to be conditional on the level of wind investment.

0.98. This reflects the fact that hour ahead net load can be predicted with a high degree of accuracy, though predicted net loads farther in the future are considerably less accurate.

We assume a linear inverse demand function for net load Q, given by:

$$P(Q, \theta_2) = \theta_1 - bQ \tag{39}$$

The current demand state is $\boldsymbol{\theta} = (\theta_1, \theta_2)$, where θ_1 is the (continuous) demand shock and θ_2 is hour-of-day. Parameter b is a common coefficient. As noted above, wholesale electricity demand is very price inelastic. Elasticity estimates for wholesale demand reported in Patrick and Wolak [2001], Johnsen [2001], and Bigerna and Polinori [2014] vary from zero to 0.35, with results concentrated in the low end of that range. We select a value for price elasticity of 0.05, and set the value of parameter b to yield this price elasticity at the mean wholesale price and quantity in the data. Values for the demand shift variable θ_1 are normalized using observed load data and the assumed form of inverse demand; specifically, a value for θ_1 is determined for each demand state by assuming that load equals the quantity demanded at the average wholesale price. This estimated net load regression (38) serves as a prediction equation for the planner's problem in our model.

C Computation

The inputs for each computation of our model are a set of parameters for generator operating and investment costs, stage two discount factor, and demand slope and a demand forecasting equation that predicts next hour demand shock conditional on current demand shock and hour-of-day. Computation involves an inner loop that takes a vector of generation capacities as exogenously fixed and finds the solution to the planner's stochastic DP problem; this is stage two of the model. The solution to this DP problem establishes a value for the planner for the given vector of generation capacities. The outer loop is an optimization routine that searches for the vector of capacities that yields maximum value for the planner; this is stage one of the model. The stage two DP computation is nested within this stage one optimization

problem. Note that this computation procedure is analogous to a procedure for estimating parameters of a dynamic optimization model, in which a DP problem is solved in an inner loop for a fixed set of model parameters and a likelihood function is maximized with respect to parameters in an outer loop.

The state vector for the stage two DP problem is (x_1, x_2, θ, h), where x_1, x_2, and θ are continuous state variables for 'on' coal capacity, 'on' GCC capacity, and demand shock, respectively, and h is hour-of-day. We discretize the state space, using 100 grid points for (x_1, x_2) and 40 grid points for θ. Since coal capacity exceeds GCC capacity given our model parameters, we use more grid points for coal than for GCC. Specifically, we use 20 grid points for coal (x_1) and 5 grid points for GCC (x_2). So the discretized state space for the inner loop DP problem has 96,000 states. The control variables for adjusting 'on' coal and GCC generation are restricted so as to maintain 'on' generation to the set of discrete grid points. In addition, we transform the stochastic process for continuous demand shocks into a probability transition matrix for discrete demand shocks, conditional on hour-of-day. Since the stage two discount factor is very close to one, we use policy function iteration as the solution method for the inner loop. The speed of convergence of this method is not sensitive to the discount factor.

Due to our discretization of the state space, the resulting stage one planner value may not be a smooth function of capacities. We therefore use an optimization method for stage one that can accommodate a non-smooth objective. Specifically, we use the particle swarm algorithm, a grid-based search method similar to genetic algorithms; Shi and Eberhart [1998].

If wind capacity is positive for stage two then we treat wind generation as a reduction in wholesale market demand. Wind capacity coupled with data on observed wind capacity factors yields a sequence of predicted hourly wind generation levels. The predicted wind generation levels are subtracted form observed hourly demand quantities (load data) to yield a sequence of net demand quantities (or, net loads). For each amount of wind capacity considered in stage two, we re-estimate the demand forecast equation to arrive at a forecast equation for net demand. This revised forecast equation is an input into the planner's stage two DP problem.

References

Adlakha, S., Johari, R., and Weintraub, G. (2015). Equilibria of dynamic games with many players. *Journal of Economic Theory*, 156:269–316.

Bigerna, S. and Polinori, P. (2014). Electricity demand in wholesale Italian market. *The Energy Journal*, 35(3).

Blanford, G., Merrick, J., and Young, D. (2014). A clean energy standard analysis with the US-REGEN model. *The Energy Journal*, 35(SI1).

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.

Bushnell, J. (2010). Building blocks: Investment in renewable and non-renewable technologies. In Moselle, B., Padilla, J., and Schmalensee, R., editors, *Harnessing Renewable Energy in Electric Power Systems*. RFF Press.

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

Cullen, J. (2013b). Measuring the environmental benefits of wind-generated electricity. *American Economic Journal: Economic Policy*, 5(4):107–133.

Dixit, A. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy*, 97(3):620–638.

EIA (2010). Annual Energy Outlook 2010. Technical report, U.S. Energy Information Administration.

EIA (2014). Annual Energy Outlook 2014. Technical report, U.S. Energy Information Administration.

EIA (2015). Annual Energy Outlook 2015. Technical report, U.S. Energy Information Administration.

EIA (2016). Annual Energy Outlook 2016. Technical report, U.S. Energy Information Administration.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62:53–82.

Fell, H. and Kaffine, D. T. (2014). A one-two punch: Joint effects of natural gas abundance and renewables on coal-fired power plants. Working Paper.

Gowrisankaran, G., Reynolds, S., and Samano, M. (2016). Intermittency and the value of renewable energy. *Journal of Political Economy*, 124:1187–1234.

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

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150.

Johnsen, T. (2001). Demand, generation and price in the Norwegian market for electric power. *Energy Economics*, 23:227–251.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50:24–52.

Kaffine, D. T., McBee, B. J., and Lieskovsky, J. (2013). Emissions savings from wind power generation in Texas. *The Energy Journal*, 34(1):155–175.

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

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

Novan, K. (2015). Valuing the wind: Renewable energy policies and air pollution avoided. *American Economic Journal: Economic Policy*, 8(1):291–326.

Novy-Marx, R. (2007). An equilibrium model of investment under uncertainty. *Review of Financial Studies*, 8(1):1461–1502.

Patrick, R. and Wolak, F. (2001). Estimating the customer-level demand for electricity under real-time market prices. NBER Working Paper No. 8213.

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

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

Shi, Y. and Eberhart, R. (1998). A modified particle swarm optimizer. *Proceedings of IEEE International Conference on Evolutionary Computation*, pages 69–73.

Stokey, N. and Lucas, R. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Williamson, O. (1966). Peak-load pricing and optimal capacity under indivisibility constraints. *American Economic Review*, 61:810–827.