Dynamic Response to Environmental Regulation in the Electricity Industry

Joseph A. Cullen

February 1, 2011

Abstract

Climate change, driven by rising carbon dioxide (CO_2) levels, has become an important economic and political issue. Governments around the world are implementing environmental regulations that tax or price carbon dioxide emissions or significantly increase renewable energy production. This paper seeks to understand the response of electricity producers to policy changes, given the current market structure. Electricity producers are the leading emitters of CO_2 and other pollutants. They make their output decisions in response to fluctuating prices for electricity given their costs of production which include substantial costs associated with starting up and shutting down generators. This paper, recovers the cost parameters of the industry with a dynamic price taking model. The parameters are used to solve for equilibrium prices and to simulate the supply of electricity, consumer surplus and firm profits under counterfactual environmental policies. Results evaluating a carbon tax policy show that total emissions from the industry do not change significantly when faced with tax rates at the levels currently under consideration by legislators. Even a very large carbon tax of twice that of expected levels, lowers emissions by only 7% in the short run.

1 Introduction

Climate change has become an important political and economic issue in recent years. Scientists point to rising carbon dioxide levels due to human activity as a major contributor to a warming environment. The costs associated with climate change are uncertain, but may be extreme. Governments around the world are implementing environmental regulations that tax or price carbon dioxide emissions or significantly increase renewable energy production. Regulations which reduce emissions in meaningful amounts will have major implications on a country's economy. Increased energy prices due to regulation will lead to different paths of consumption, production, and labor usage.

In this paper, I examine how environmental regulations to reduce carbon emissions may affect outcomes in the US electric industry. Electricity generation is the largest single source of CO_2 in the US accounting for 40% of annual CO_2 emissions.[1] Reducing emissions in the electricity sector will be an important component of any policy which aims to reduce aggregate emissions in the US.

Some policies have already been implemented to reduce emissions from electricity generation. Since 1992, solar, wind, and geothermal electricity generators have received generous production subsidies from the US federal government which has resulted in dramatic growth in renewable energy facilities. However, despite growth in new carbon free generators, CO_2 emissions from electricity production continue to rise in the aggregate. Legislators are now looking at market based regulations, such as cap and trade programs or carbon taxes, which directly price carbon emissions as a potential solution to rising CO_2 emissions. These different policies may have very different impacts on electricity production. For example, a carbon tax indirectly reduces pollution through a relative cost increase for high polluting generators and through reduced consumption of electricity due to overall higher energy prices. Renewable energy subsidization, on the other hand, will directly reduce fossil fuel electricity production, but may indirectly increase consumption by lowering equilibrium energy prices. In order to properly evaluate potential policies, it is important to accurately gauge the response of polluting industries. This research represents the first attempt to compute

[1] The contribution from other sectors excluding electricity use: transportation (33%), direct industrial emissions (17%), direct commercial emissions (4%), direct residential emissions (6%)(EIA 2008)

counterfactual equilibrium outcomes in the electric industry under alternative environmental regulations.

Carbon regulations interact with electricity generating decisions in a highly complex market. The supply and demand of electricity must be equated at every moment of every day. In addition, demand does not immediately respond to conditions in the wholesale market. As a result, wholesale electricity markets are characterized by dramatically higher prices during peak demand periods followed by low or even negative prices during off peak periods. A typical day will see average peak prices that are more than double that of off peak prices. The large variation in prices is partly due to the fact that generators cannot change output costlessly or instantaneously. For instance, industry reports on the cost of starting up a large coal plant range from $3,000 to $70,000. The fact that prices dramatically fluctuate over the course of day together with large startup costs imply a generator's decision is inherently a dynamic problem. Forward looking firms with costly output adjustment will anticipate price variations and plan output accordingly.

Certain types of environmental regulation have the potential to dramatically increase the level of electricity prices as well as exacerbate price spikes. For example, environmental policies which encourage the development of wind power will reshape the residual demand curve facing fossil fuel generators[2]. This residual demand curve will increase the need for conventional power plants to reduce or stop production during off peak periods while maintaining output levels during peak demand periods. Other environmental policies, such as carbon regulation, also have the potential to change the equilibrium production and pollution profiles. No studies to date have attempted to model the response of the electricity producers to environmental regulations within a dynamic framework.

In this paper, I develop a structural model to account for the dynamics in electricity production which arise due to generator startup costs. Startup costs are incurred whenever a generator turns on after a period of zero production. Using a detailed dataset from the Texas grid on generator output and energy prices, I estimate the startup costs for each generator using a

[2]Wind farms, which are on shore, have the highest output during times of off peak demand and have little output during high demand periods. Wind power thus reshapes the residual demand curve by increasing the difference in demand between on and off peak periods. Wind farms which are built offshore will have the opposite effect of residual demand since the usually blows off shore during peak demand periods when energy is most needed.

dynamic discrete choice model of generator operation.

Under the assumption generating capital is fixed and that firms are price takers, I can use the recovered parameters to simulate the electricity market under environmental policies. I develop a method to solve for this new dynamic equilibrium price path in way which ensures that firms' expectations for prices are consistent with the new equilibrium. Using equilibrium prices, I then simulate the supply of electricity, consumer surplus and firm profits under counterfactual environmental policies. This effectively simulates the response of firms to policies over a relatively short, two-year window, which is the approximate time required to build new generating capital.

I simulate the outcomes in the electricity market for two different policies currently under consideration: a carbon tax and an increase in renewable energy due to subsidies. For each counterfactual policy I solve for the dynamic equilibrium prices using a range demand elasticity estimates from the literature.

Results show that total emissions from the industry do not change significantly when faced with carbon tax rates at the levels currently under consideration by legislators. In fact, a very large carbon tax of twice that of expected price levels, lowers emissions by only 7% in the short run.

This model has several advantages over a reduced form approach to analyzing counterfactual outcomes. Since it explicitly solves each generators' dynamic problem, it is possible can simulate equilibrium outcomes that are very different from observed equilibrium outcomes. In contrast,reduced form approaches are not able to effectively deal with counterfactual equilibria which are too far out of sample. Second, the structural approach is more appropriate for simulating situations with increasingly volatile equilibrium prices. The reduced form approach cannot handle such situations since the firms' reactions are known only for the observed level of volatility in the market.

The remainder of the paper proceeds as follows. In section 1, I describe the operation of the Texas electricity market followed by a description the data in section 2. Section 3 introduces the model while section 4 details the estimation method. Section 6 contains the estimation results using a subset of the data. Section 7 simulates equilibrium outcomes under counterfactual environmental policies which is followed by a few brief concluding remarks.

2 Electricity Market

Before presenting the model, I first explain the basic structure of power systems and the institutional details of ERCOT.

2.1 Power System Basics

An electric system is composed of two main parts: generators and a transmission system. Electricity produced by generators flows over a transmission grid to end consumers of power. Electricity is an unusual commodity in several ways. First, demand for electricity is almost perfectly inelastic in the short-run; very few consumers of electricity are willing or able to adjust consumption in response to changing market conditions. Second, the quantity of electricity demanded at a given price varies cyclically over the course of a day and throughout the year. On a daily level, peak demand periods generally occur in the early evening hours while the lowest levels of demand are in the early morning. Peak demand can be twice that of off peak periods within the same day. On a yearly level, the demand for electricity is generally higher in the summer months than in the winter. Finally, electricity is unusual because it cannot be stored in meaningful quantities[3]. Electricity production and consumption on a grid must be balanced on a second-by-second basis. If more power is being consumed than is being produced then the reliability of the grid is threatened. Sufficient imbalances result in brownouts (dropping electrical frequency) or blackouts (complete loss of electrical service). Given that demand is inelastic and highly variable combined with the lack of energy storage puts high demands on generators to preserve the reliability of the grid by adjusting output to follow changing demand.

As generators follow demand, they face several output constraints. First, generators are capacity constrained. The maximum output capability of a generator is determined at the time of its construction and generally remains fixed over the life of the generator. Generators also face minimum output constraints. The minimum output constraint is the lowest level of sustained

[3]Chemical storage of electricity such as in lead-acid batteries are too costly to be used to store any meaningful amount of electricity in a system. Technologies do exist to turn electrical energy into potential mechanical energy which is storable such as compressed air or pumped hydro electrical storage. These technologies do make minor contributions on some grids,but such technologies have not been implemented on the electrical grid in my study.

output the firm can generate without shutting down. Operating below the minimum output level results in large inefficiencies and can damage generating equipment.

Generators also face costly adjustments to output. Adjustment costs include startup costs and ramping costs. Startup costs are incurred when bringing a generator online after a period of zero production. Bringing the generator online requires fuel to heat up equipment and bring the turbine up to speed was well as additional labor to supervise the process. In addition, startups are hard on equipment leading to increased maintenance costs in the long run. In fact, engineering studies estimate that wear and tear on generating equipment may account for the majority of the cost of startup (Chow, Ho, Du, Lee & Pearson 2002).

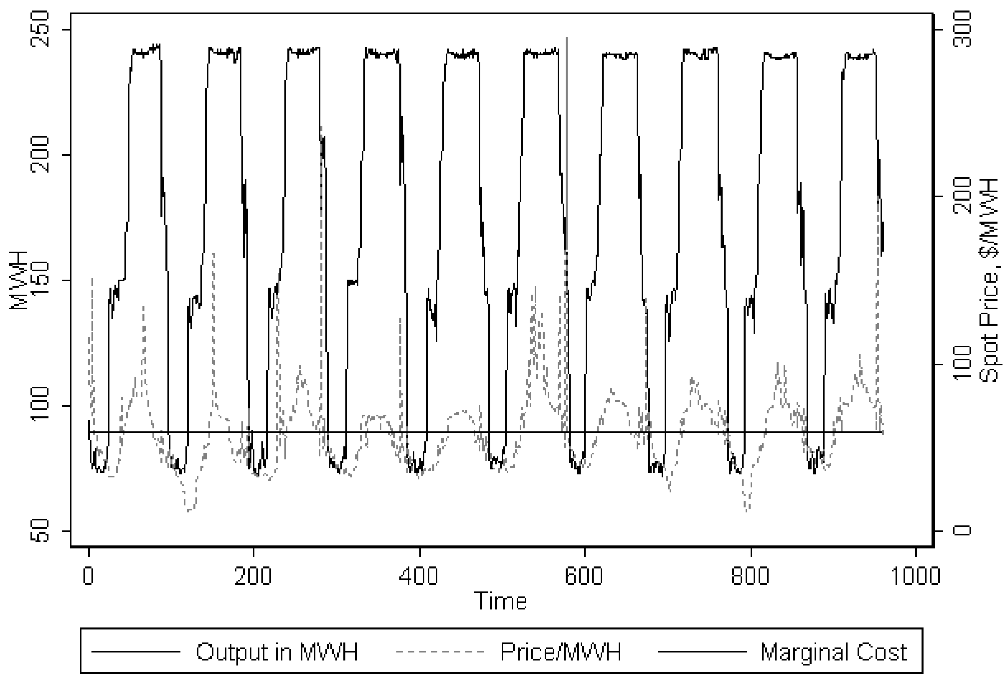
Ramping costs arise when firms change the level of output within their range of operation. These costs also increase wear and tear on machinery as well as decreased output efficiency. These costs increase with the severity of the adjustment; very large and quick output adjustments will be more costly than small gradual ones.

Both startup and ramping costs vary widely by generation technology and the age of the unit. For example, gas combustion turbines have lower startup costs than coal fired steam plants. Likewise, large plants will have higher startup costs than smaller generators even though they may use the same technology. Also, as generators age, they degrade in efficiency which will increase startup costs.

The costs associated with output changes are significant. Engineering estimates of startup costs range from a hundreds dollars to tens of thousands of dollars per start depending on the size and technology of the generator. Consequently, a generator with high startup costs may continue to run during low price periods to avoid startup costs. Likewise, a generator may not startup even though prices exceed its marginal cost of production if it believes that the profits will be sufficient to cover its startup costs.

Evidence of the importance of startup costs for firm behavior is illustrated in figure 1. This figure shows one generator's output over a 10 day period in July of 2006. The horizontal line shows the firm's constant marginal cost of production while the dashed line shows the spot price for energy. Notice that even though the spot price falls below the firm's marginal cost, the firm does not shut down. Rather, it reduces its output to some minimum level. As prices begin to rise, it again ramps up production. This is consistent with firm behavior in the presence of significant generator startup costs.

Figure 1: Operating Decision Example



Concrete information on startup costs is generally unavailable to researchers and policy makers. The information is considered proprietary and thus in not made publicly available. This paper provides a way to estimate startup costs given publicly available information.

2.2 ERCOT

This paper examines outcomes from the Texas grid which is managed by the Electricity Reliability Council of Texas (ERCOT). The ERCOT grid operates as a deregulated electricity market which serves most of the state of Texas. It operates almost independently of other power grids with very few connections to outside markets. Since the grid does not cross state lines it is also under less federal oversight than other grids in the US. Electricity generation and retailing are deregulated while the transmission and distribution of energy remains regulated to ensure that competitors in the generation and retailing markets have open access to buy and sell power. Unlike many regulated and even deregulated markets, companies in this market are vertically separated. There are no vertically integrated firms that control generating, transmitting, and retailing resources.

2.2.1 Generators

There are approximately 500 generators which supply electricity in ERCOT. Generators are split into four geographically distinct congestions zones. Each generators sells it energy to buyers either through bilateral contracts or through ERCOT's spot market called the Balancing Market. Approximately, 95% of energy produced is sold through bilateral contracts. The remaining 5% is allocated through the Balancing Market.

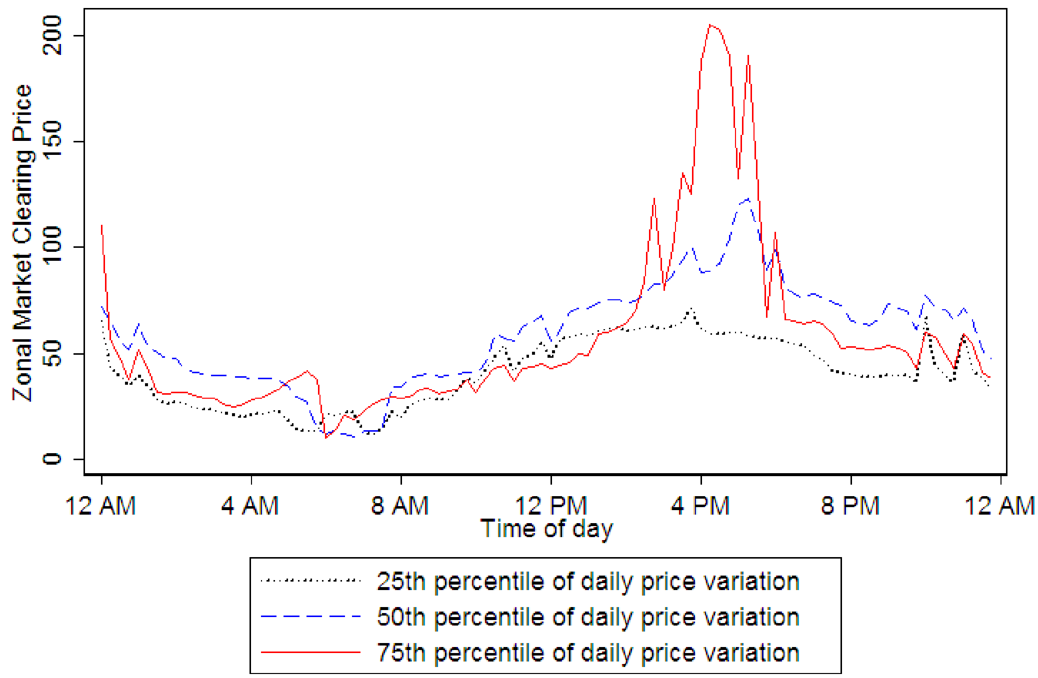
To ensure that there is sufficient supply, ERCOT requires generators and electricity retailers submit scheduled energy transactions a day ahead. These schedules are submitted through a Qualified Scheduling Entity (QSE) which generally submits schedules for a portfolio of generators and power purchasers. These schedules outline which generators are planning producing power and how that power will be transmitted to end users for each hour of the day. ERCOT allows QSEs to submit day-ahead schedules which leave them in long or short positions entering into the production period[4]. QSEs are also required to submit Balancing Market bidding functions for each hour of the day. The bidding functions show the willingness of generation portfolio to deviate from its scheduled output as a function of the price in the Balancing Market. The QSE must submit its willingness to both increase and decrease the portfolio output in response to price.

In real-time, ERCOT uses the Balancing Market to ensure adequate supply and to equate the marginal costs of production across generators. Every fifteen minutes ERCOT intersects the hourly bidding functions to arrive at a Market Clearing Price for Energy (MPCE) in each zone via a multi-unit uniform price auction[5]. If there is no congestion between zones then the prices are the same in each zone and the entire grid acts a single market. If congestion would occur between zones with a single MCPE, then ERCOT intersects the bidding functions separately by zone to achieve market clearing

[4]ERCOT also requires firms to have sufficient levels of ancillary power services

[5]See Hortacsu & Puller (2008) for a detailed explanation of the auction process.

Figure 2: Representative Daily Price Variation by Percentile



prices for each zone which do not exceed the transmission capability between zones. For example if more power is needed in the South zone, but the transmission lines are at capacity, ERCOT will raise the prices in the the South zone, while lowering or keeping constant the prices in the other zones. In any case, generators respond to MCPE based on their bidding functions. The Balancing Market also helps to ensure that the lowest cost producers are generating electricity. At a low MCPE, high marginal cost firms have incentives to reduce or shut down production and satisfy their contractual obligations through energy procured from the Balancing Market. In a static, price-taking setting the Balancing Market would ensure that only the lowest cost generators were production energy each period. With the introduction of dynamics in the generating process, this no longer holds.

The Balancing Energy prices can be quite volatile as shown in figure 2. This graph shows three examples of the daily path of Balancing prices in the Houston zone. The three lines show representative price paths with daily variation in the 25th, 50th, 75th percentiles of price variance for 2006. All three days exhibit higher prices during peak demand periods; the highest variance price path shown has peak prices that are twenty times that of off peak periods.

2.2.2 Transmission Congestion

Most of the time ERCOT operates as a single market with a single spot price for wholesale electricity. During peak periods when transmission congestion does arise, it is alleviated in two ways. First, congestion between zones is alleviated by having different prices for Balancing energy in each zone. For example, increasing the price Balancing energy in zones that are net importers of electricity while lowering the price in zones which are net exporters of energy will relieve demands placed on inter-zonal transmission lines. Thus, interzonal congestion can be relieved through pricing mechanisms in the balancing markets.

Congestion can also arise within zones. This type of congestion cannot be resolved with market prices since there is only one price for each zone. To deal with local congestion, ERCOT deploys generators out of bid order. That is, ERCOT deploys specific generators which are not willing to increase production at current prices by offering them prices higher than the prevailing market price. The costs of deploying these resources to alleviate local congestion is covered by an output tax levied on all generators in the zone. This amounts to a uniform increase in marginal costs across all generators. Thus, transmission congestion is either explicitly accounted for in the market price, if it occurs between zones or it arrives as a uniform output tax on all generators in a zone.

2.2.3 Demand

As in most electricity markets, demand in ERCOT does not respond directly to wholesale price signals[6]. Residential and commercial users purchase electricity at fixed prices which are constant for period of time ranging from one month to several years. As such they have no incentive to reduce con-

[6]Additionally some large industrial users negotiate lower energy prices by agreeing to have their supply of electricity temporarily interrupted in emergency situations when generating reserves on the grid reach critical levels. However, such contracts are confidential so are not available to support this hypothesis.

sumption during high price periods in the wholesale markets[7]. It is possible that industrial users could respond to price changes in the wholesale market through conditions in bilateral contracts with generators. However, I have not found any evidence that this is the case. Over a longer period of time, if average prices in the wholesale markets rise, this information will eventually be passed along to consumers in the form of higher rates. However, in the short run demand for electricity is inelastic.

3 Data

The data comes from the Texas electricity grid which is managed by ERCOT. The data cover the period from April 2005 to April 2007. During this period, there are approximately 80 different firms operating 180 power plants which supply electricity to the grid [8]. Each power plant hosts 1 to 10 generators. In total there are more than 500 generators which are connected to the grid supplying electricity to the wholesale market. Combined, these generators are capable of producing over 73,000 MW of electricity at full capacity. Generation technology includes coal, nuclear, natural gas, water, and wind power plants.

In the data, the output of each generator is observed every fifteen minutes over the two year period. I also observe the market clearing price for the balancing energy every 15 mins for each zone. For each generator and interval, it is also known if the generator was shutdown for maintenance or due to an involuntary mechanical failure. Other available generator level characteristics which include the maximum and minimum output capability for each generator, the age of the generator, its fuel type and its location.

[7] Some large industrial consumers do curtail electricity use when reserve capacity becomes short but they do not directly respond to fluctuations in the price of electricity in the wholesale market. These large industrial users negotiate lower energy prices by agreeing to have their supply of electricity temporarily interrupted in emergency situations when generating reserves on the grid reach critical levels. Industrial users with interruptible loads are called Loads Acting As Resources (LaaRs). In the event of an unexpected change in load, electricity delivery to the LaaR will be interrupted to maintain the frequency on the grid. Approximately half of responsive reserve services are supplied by LaaRs (MF7). Again, it is important to note that LaaRs respond to events that threaten the reliability of the grid, not to price changes in the wholesale market.

[8] There are additional generators which provide electricity on private networks, but which do not provide electricity to the grid controlled by ERCOT.

I supplement these data with information from the Environmental Protection Agency (EPA) and the Energy Information Administration (EIA) on the characteristics of power plants generators. Generator characteristics include a measure of the fuel required to produce 1 MWh of electricity, and the quantity of the SO_2, NO_x, and CO_2 emitted per MWh of output. average annual heat rate (MMBTU/MWH) across all generators at a power plant and the emissions rates for SO_2, NO_x, and CO_2.

To construct the marginal cost of electricity production for each generator, both fuel and pollution permit costs are needed. For fuel costs for coal plants, I use monthly information from EIA form 423 which gives the delivered quantity and cost of fuel for coal in Texas. I take the quantity weighted average coal price as the price for coal for all generators in the market for that month. For the cost of fuel for gas powered plants, I use daily spot prices for natural gas from transactions on the Intercontinental Exchange (ICE). For pollution permits, I use average permit prices from EPA permit auctions for both SO_2 and NO_x permits in 2006. Carbon dioxide is currently unregulated so there is no cost associated with CO_2 emissions. Marginal costs of production for each generator can then be calculated from the dost of fuel and the cost of pollution permits necessary to produce a unit of output.

The marginal cost of fuel for electricity production is the generator's heat rate times the average cost of delivered fuel. The marginal cost of emissions is the generators emissions rate times the cost of pollution permits. The total marginal cost of electricity is then simply the marginal cost of fuel plus the marginal cost of emissions.

The data does have some limitations. First, heat rate information is constructed by taking the annual electricity output of a plant and dividing by the heat content of the fuel used. If significant portion of a generators total fuel consumption is used during frequent startups then the efficiency of the generator will be understated and the corresponding marginal cost will be over stated.

Second, the prices used in the model are not necessarily the prices the firm received for its output since most energy in this market is sold via bilateral contracts with unobserved prices. However, spot prices do represent the opportunity cost of production for the firm. If the firm has no market power its contract position should not matter for its output decision. A firm can always shutdown production and fulfill its contract by buying power in the balancing market. Market analysis by ERCOT also suggests that forward contract prices for energy follow balancing price quite closely.

Third, some generators are paid to provide ancillary services for market such as regulation, capacity reserve, or out of merit order energy. These generators respond to price signals that I do not observe. For example, generators participating in responsive reserve service may start up and run at minimum capacity when price is below their marginal costs because their startup costs and minimum operating costs are covered by ERCOT. This implies that there will be unobserved states that generators optimize with respect to which are possibly serially correlated.

4 Model

I develop a dynamic model of firm output, that accounts for the impact of startup costs on firm behavior. In developing the model I make the following assumptions.

Assumption 1: Firms are price takers.

Assumption 2: The marginal cost of each generator is constant and known.

Assumption 3: There are no transmission costs or local constraints.

Assumption 4: A generator can costlessly adjust output within its operating range.

The first assumption allows the firm's decision problem to be modeled as a single agent dynamic problem since no firm's unilateral choice of output affects price. This price taking assumption also renders ownership of power plants irrelevant. This allows one to model each generator at each plant as a separate firm maximizing its own profit. Price taking is a strong assumption especially considering the active literature on the exercise of market power in electricity markets (Borenstein, Bushnell & Wolak 2002), (Mansur 2008), (Hortacsu & Puller 2008). There are several conditions specific to ERCOT that make this assumption more plausible. First, ownership rules limit a firms' ownership of generation facilities to 20% of the total generation capacity in any zone. Second, most of the energy is sold via bilateral contracts. Since most of the energy is not sold at the spot price, this reduces the incentives for a firm to withhold production to increase the energy price in the spot market (Wolak 2000), (Bushnell, Mansur & Saravia 2008). That said, price taking is an important and possibly restrictive simplifying assumption of the model.

The second assumption, that marginal costs are constant and known, is standard in the literature on electricity markets. In reality, the heat rate and thus the marginal cost of a generator is not constant within the operating range of the generator. In particular, as generators move away from full utilization of capacity efficiency tends to fall (Bharvirkar, Burtraw & Krupnick 2004). Increasing efficiency, or heat rate, over the output of a plant implies that marginal costs are increasing over some range of output. The degree to which a constant marginal cost assumption is reasonable depends in large part on the technology used. However, for most generators a constant marginal cost assumption is reasonable. Also, there are other marginal costs that are left out of the standard calculation. These include transmission costs, variable maintenance costs, or other variable input costs such as water for steam plants. However, these deviations from standard assumption are likely to be of second order importance.

The third assumption implies that firms are not constrained by local transmission bottlenecks when optimizing with respect to price. This assumption does allow for the primary paths of congestion, namely congestion between zones, to be represented by the model since this type of congestion is alleviated in ERCOT via price mechanisms. However, this assumption does rule out an congestion within a zone. Although this is ostensibly of second order importance, certain generators may be more sensitive to local congestion than others. In particular, a small number of generators may at times receive above market price payments to increase or start production to alleviate local congestion. I am not able to account for this directly in the model.

The fourth assumption allows me abstract away from the firm's choice of output level given that it is operating. With costless adjustment within its operating range, if a firm is operating it will produce at maximum capability if price is greater than marginal cost and will produce at minimum capacity if price is less than marginal cost. The power of this assumption is that the generator's decision collapses from a continuous choice of output level to a discrete choice of whether to operate or not.

This assumption is very plausible for some generators, but is less unconvincing for others. Figure 3 shows capacity utilization histograms for three representative generators. The first generator exhibits a production pattern that closely matches the assumption; the majority of production occurs at the generator's maximum or minimum output capability. The second generator also exhibits a bimodal distribution of capacity utilization, but the distribution is more diffuse and the upper mode is not at the generator's declared maximum capacity[9]. The third generator's production is not consistent with the assumption; much of the generation occurs far from the maximum or minimum output levels.

These figures suggest that although costless adjustment within a firm's operating range may be reasonable for some generators, it is clearly not a good assumption for others. A model which explicitly accounts for costly adjustment may more accurately model the behavior of certain firms. However, this greatly increases the complexity of the model. As such, the assumption of costless adjustment will be maintained through out this paper despite the deviation of some generators from the behavior implied by the assumption.[10]

Given these assumptions, I model each generator as a single firm with the following single agent dynamic problem. In each period, the firm observes the price in the market and the interval of the day. The firm can take one of two actions which are notate as:

$$a_{it} = \begin{cases} 1 & \text{if operate in } t \\ 0 & \text{if not operate in } t \end{cases} \tag{1}$$

$$\begin{aligned} \text{where} \quad & i \text{ indexes the generator} \\ & t \text{ indexes each fifteen minute time period} \end{aligned}$$

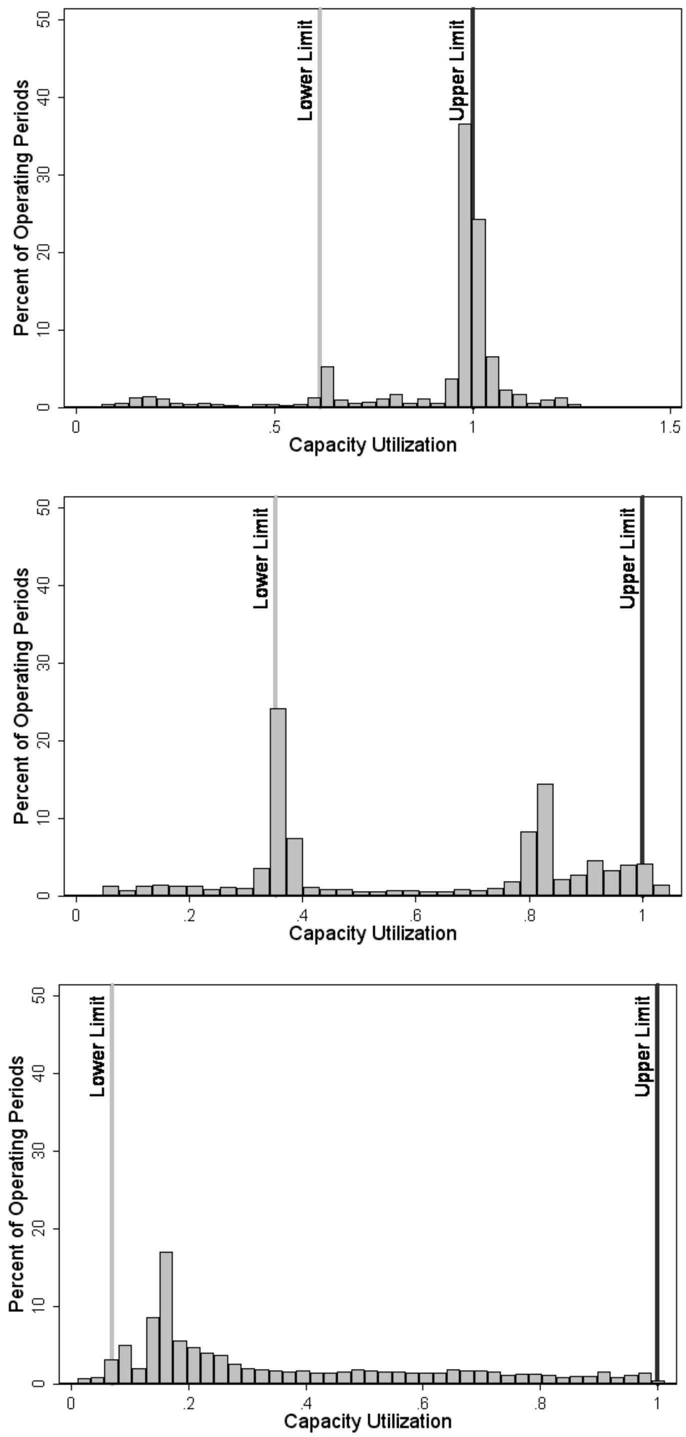
If the firm decides to operate, assumptions two and four imply that the firm's output will be one of two levels. If the price in the market is greater than the firm's marginal cost then the it will produce at maximum capacity. If the price is below marginal cost than the firm will produce at its minimum possible level.

$$\begin{aligned} q_{it} &= max \text{ if } P_t \geq c_i \text{ and } a_{it} = 1 \\ q_{it} &= min \text{ if } P_t < c_i \text{ and } a_{it} = 1 \end{aligned} \tag{2}$$

[9]It is interesting to note that the upper mode is centered around 80% of the generators capacity. In ERCOT, a generator may bid 20% of its capacity into responsive reserve markets. To the extent that the reserve markets are also competitive, the expected revenue that the generator receives for providing reserves should closely follow that expected revenue it gets from selling its power in the energy market. Thus, profits calculated as if the firm was producing at maximum level should provide a good approximation for actual profits.

[10]Modeling a continuous production choice with costly adjustment is the subject of the author's ongoing research.

Figure 3: Capacity Utilization Histograms



where c_i = constant marginal cost of generator i
P_t = price for electricity in the generator's zone

Each period when the firm is operating its profits are simply the price-cost differential earned on every unit produced minus any fixed costs associated with operating. The per period profit function for the generator is then:

$$\Pi(P_t, q_{it}, a_{it}) = \begin{cases} (P_t - c_i)q_{it} - OC_i & \text{if } a_{it} = 1 \text{ and } s_{it} = 1 \\ (P_t - c_i)q_{it} - OC_i - START_i & \text{if } a_{it} = 1 \text{ and } s_{it} = 0 \\ 0 & \text{if } a_{it} = 0 \end{cases} \tag{3}$$

where OC_i = non-variable operating cost for generator i
$START_i$ = cost of starting up generator i
$s_{it} = a_{it-1}$ = the operating state last period

I allow for a non-variable cost of operating each period with an additional startup cost that is incurred only if the firm was not operating last period. The structural parameters of the model are c_i, OC_i, and $START_i$. The constant marginal cost of production, c_i, is known for each generator. The structural parameters OC_i and $START_i$ are not known and will be the object of the estimation procedure. For notation simplicity the i subscript will be dropped for the remainder of the paper since each generator is modeled separately as a single agent.

In the dynamic model, the firm's expectations over future prices must be explicitly modeled. I assume that prices follow a conditional AR(1) Markov process described by the distribution $F(P_t|P_{t-1}, I_{t-1})$ where I_t is an indicator for each hour of the day. The price next period follows a distribution known to the generator and is conditional only on the current price and the time of day. Note that because of the price taking assumption the evolution of price does not depend on the action of the generator.

Although I observe output and price by 15 min intervals, I aggregate the data to an hourly level. I do this for two reasons. First, very few generators can turn on or off within a fifteen minute period. Thus, although one may observe a high price this period, a generator may not be technically able to respond to that price. Looking at the hourly prices averages out some of the noise introduced by temporary price spikes. Second, averaging over an hourly period more closely matches the scheduling decisions of firms which are typically done on the hourly level.

One might argue that the a simple Markov process is not sufficiently rich to accurately model the expectations of the firm. Indeed, firms have more information that simply the lagged price and time of day with which to form expectations for price in the next period. For example, firms may have expectations over future temperatures, load levels, and congestion. In addition they may use a long price history when predicting future prices. The extent to which our model of the evolution of price is adequate depends on the degree to which lagged price summarizes all of the other components of the expectations of price. While I would like to be as flexible as possible with respect to expectations, flexibility comes at a cost. Allowing for a richer specification for the formulation of price expectation increases the size of the state space which exponentially increases the computational burden for solving the dynamic programming problem. Simply adding several more state variables to allow for greater flexibility comes at a high cost.

To investigate further, I ran exploratory regressions of the price evolution process with our Markov process AR(1) and other more richly specified processes which used other explanatory variables such as temperature and transmission congestion as well as further lags of price and the other explanatory variables in a polynomial expansion using the same data which was used for our estimation procedure. The results showed that the simple Markovian model performed surprisingly well. With an adjusted R^2 of 0.72 it was able to account for much of the observed variation in the prices. Richer models of the price process did not explain the data significantly better with adjusted R^2 values ranging from .61 to .74 depending on the specification. A single lagged price conditional on the time of day, seems to captures most of the information that might be used to predict future prices. Although strictly speaking the price evolution may not be an AR(1) process, it has very favorable explanatory power when compared with other options in addition to providing a computable framework for estimation and simulation.

Given the specification of the transition and the profit function, the state space for the dynamic problem will then be (P_t, I_t, s_t) and the Bellman equation representing the dynamic problem can be written as:

$$V(P_t, I_t, s_t) = \max_{a_t \in \{0,1\}} \{\Pi(P_t, s_t, a_t) + \beta E[V(P_{t+1}, I_{t+1}, s_{t+1} | P_t, I_t, s_t)]\} \quad (4)$$

$$\text{where} \quad \begin{aligned} I_{t+1} &= I_t + 1 - 1(I_t = 96) * 96 \\ s_{t+1} &= a_t \end{aligned}$$

The expectation is taken with respect to P_{t+1} according to the distribution $F(P_{t+1}|P_t, I_t)$. The parameter β is a fixed discount factor.

The optimal policy for this dynamic problem is a cutoff rule in P_t for every pair of (I_t, s_t). That is, the firm should take same action whenever it encounters the same state (P_t, I_t, s_t). This creates a problem for using the solution to the dynamic problem to estimate structural parameters from data as the firm will invariably deviate from what appears to be the optimal policy. I address this by adding an additional state variable into the dynamic problem which is observed to the firm but unobserved to the econometrician. This is the approach taken by Rust (1987) and the long literature that follows from it. The unobserved state variable is interpreted as a choice specific shock to the fixed cost each period. I note the choice specific shock as $\epsilon_t(a_t) \in \{\epsilon_t(0), \epsilon_t(1)\}$. Like the Rust (1987) model, I assume that the shock is an iid process which simply introduces noise on the underlying decision process. Assuming that the process is iid, simplifies the joint distribution of the stochastic elements of the Bellman such that $H(P_t, \epsilon_t(a_t)|.) = G(\epsilon_t(a_t))F(P_t|P_{t-1}, I_{t-1})$. Because I observe profits, the scale of the error process is identified unlike in most discrete choice models. Accordingly I notate the choice specific shock to fixed costs as $\sigma\epsilon_t(a_t)$. For computational simplicity I make the distributional assumption that $\epsilon_t(0)$ and $\epsilon_t(1)$ are distributed as extreme value type I random variables. This allows for analytical integration over the unobserved shocks[11].

With the unobserved state variable the Bellman equation now becomes:

$$\begin{aligned} V_\theta(P_t, I_t, s_t, \epsilon_t(a_t)) = \quad & \max_{a_t}\{\Pi(P_t, S_t, a_t) + \sigma\epsilon_t(a_t) + \\ & \beta EV(P_t, I_t, a_t)\} \end{aligned} \tag{5}$$

where the function

$$\begin{aligned} EV_\theta(P_t, I_t, a_t) \equiv \quad & \int\int\int V(P_{t+1}, I_{t+1}, s_{t+1}, \epsilon_{t+1}(a_{t+1})|P_t, I_t, s_t, \epsilon_t(a_t)) \\ & dG(\epsilon_t(0))G(\epsilon_t(1))F(P_t|P_{t-1}, I_{t-1}) \end{aligned} \tag{6}$$

I denote the vector of unknown structural parameters as $\theta = (START, \sigma, OC)$.

[11] Using an iid assumption on the structural errors simplifies the computation of the model significantly. However, I may be concerned that certain factors affect the decision to run a generator will be highly serially correlated. For example, participation in ancillary services markets are likely to span several hours over the course of a day. Recent work in econometrics, such as Norets (2009) and Imai, Jain & Ching (2009), has provided some computationally attractive approaches to the estimation of discrete choice models with serially correlated unobserved state variables. I will allow for the possibility of serially correlated unobserved states in future work.

The function EV is the fixed point of a contraction mapping $EV_\theta = T_\theta EV_\theta$. Given my assumptions about the error process and the price transitions, the choice specific value expected value function is the solution to the following contraction mapping.

$$EV_\theta(P_t, I_t, a_t) = \tag{7}$$
$$\int_{P_{t+1}} \sigma ln \left(\sum_{a_{t+1} \in \{0,1\}} exp\left\{ \frac{1}{\sigma}(\Pi(P_{t+1}, a_t, a_{t+1}; \theta) + \beta EV_\theta(P_{t+1}, I_{t+1}, a_t)) \right\} \right) dF(P_{t+1}|P_t, I_t)$$

Since the value function does not have an analytical solution, I will need to solve the value function for discrete sets of values in the state space. I_t and s_t are already discrete, but P_t must be discretized or at least evaluated at a discrete set of points. The resulting state space could be quite large depending on how finely P_t is discretized. The dimension of I_t is 24 since there are 24 operating hours in each day. The operating state last period, s_t, is a binary outcome. The size of the state space is then $DP * 24 * 2$ where DP is the number of discrete prices used. For two hundred discrete prices, the total size of the state space would be 9,600 which is large, but not computationally burdensome. Solving the value function numerically amounts to finding the value of EV_θ for point in the state space through the contraction mapping that defines EV_θ. Once the value function has been calculated then the optimal policy function can be calculated.

The optimal policy function is viewed by the econometrician as the probability of operating at each state. Given my functional form assumptions about the fixed cost shock, the operating probability can be calculated from the choice specific value functions using the well-known logit formula with the addition of a scaling parameter for the fixed cost shock.

$$p(a_t|P_t, I_t, s_t) = \frac{\sigma exp\left\{\frac{1}{\sigma}(\Pi(P_t, s_t, a_t; \theta) + EV_\theta(P_t, I_t, s_t, a_t))\right\}}{\sum_{j \in \{0,1\}} \sigma exp\left\{\frac{1}{\sigma}(\Pi(P_t, s_t, j; \theta) + EV_\theta(P_t, I_t, s_t, j))\right\}} \tag{8}$$

Given a set of parameters, the probability of operation can then be used to construct a likelihood function.

$$L(\theta) = \Pi_{t=1}^{t=T} p(a_t|P_t, I_t, s_t; \theta) p(P_t|P_{t-1}, I_{t-1}) \tag{9}$$

where $p(P_t|P_{t-1}, I_{t-1})$ is derived from the conditional distribution $F(P_{t+1}|P_t, I_t)$ and is the probability of transitioning from one discrete price to another given the interval of the day. It should be noted that $p(a_t|P_t, I_t, s_t; \theta)$ implicitly depends on the transition probability matrix given by $p(P_t|P_{t-1}, I_{t-1})$ through the solution to the value function.

Since the transition probabilities do not depend on the vector of unknown parameters θ, they can be flexibly pre-estimated outside of the likelihood function. The simplified likelihood function can then be written as simply a function of the operating probability each period.

$$L(\theta) = \Pi_{t=1}^{t=T} p(a_t|P_t, I_t, s_t; \theta) \tag{10}$$

5 Estimation

Using the dynamic model, I estimate the vector of unknown structural cost parameters $\theta = (START, \sigma, OC)$ for each generator on the grid. I estimate the parameters via maximum likelihood using the likelihood function outlined in the previous section. While conceptually straightforward, solving for the parameters which maximize the likelihood function can be quite computationally intensive.

5.1 Nested Fixed Point

Rust (1987) suggests solving for the parameters which maximize the likelihood function derived from a single agent dynamic problem using a nested fixed point algorithm. The algorithm consists of set of nested loops. The inner loop solves the value function through the contraction mapping for a given vector of parameters θ. The outer loop uses the value function solution from the inner loop to evaluate the likelihood and searches over the parameter space for the set of parameters that maximizes the likelihood. For each guess of parameters by the outer loop the value function must be solved by the inner loop. The algorithm terminates when both loops reach a fixed point. A nested fixed point is achieved when the solution to the value function at a given set of parameters maximizes the likelihood.

There are two drawbacks to using this method. First, the value function must be numerically solved for each guess of the parameter vector. Depending on size of the parameter vector and the type of search used over the parameter space, this can involve solving the value function thousands of times. Solving the value function can be very computationally intensive especially for large state spaces. Second, solving the value function depends on discount factor implicit in the contraction mapping. The value function is usually solved by value function iteration where the solution time depends on the discount factor β. For any $\beta < 1$, the contraction is well defined and will converge from any initial guess of EV_θ. However, as β nears one the time to convergence increases exponentially. When modeling short time periods, such as hourly intervals as is done in this paper, β will be very close to one and solving the value function will be extremely computationally intensive.

An alternative to using value function iteration inside the nested fixed point algorithm is to use policy function iteration. The solution time for solving the value function by policy function iteration does not depend on the discount factor. However this method does requires inverting a potentially large probability matrix, which in some cases may be computationally infeasible. However, in my particular case, I can take advantage of the fact that the transition matrix is quite sparse. The sparseness of the matrix can reduce the computation time greatly which facilitates the direct computation of the value function.

5.2 Price Transition

A necessary input for the maximization of the likelihood is a set of price transition matrices which capture the firm's expectations about future prices at any state. Since the price transitions do not depend on the action of the firm in a price taking model, the transition matrix can be estimated outside of the likelihood function. Given that the conditional transition probabilities, $p(P_t|P_{t-1}, I_{t-1})$, depend both on the last periods price and interval of the day, there are 24 Markov price transition matrices , one for each hour of the day. The size of each time specific matrix depends entirely on how finely price is discretized. For example, if price were discretized into 100 bins, then each transition matrix has 10,000 elements. With 24 intervals in each day, this means that 240,000 conditional probabilities would need to be estimated. The large number of conditional probabilities render nonparametric estimation of the transition matrices infeasible even for very modest levels of price discretization. Consequently, I use a flexible parametric method to estimate the conditional probabilities. In particular, I use a 3rd order polynomial expansion of lagged price interacted with dummies for each hour of the day as shown in the equation,

$$P_t = \beta_0 + \beta_1 P_{t-1} + \beta_2 P_{t-1}^2 \beta_2 P_{t-1}^3 + D(P_{t-1})\alpha + D(P_{t-1}^2)\gamma + D(P_{t-1}^3)\delta + \epsilon_t \quad (11)$$

where D is a matrix of dummy variables for each interval of the day.

The parameter estimates from the above equation yield $E[P_t|P_{t-1}, I_{t-1}]$. To calculate all the conditional choice probabilities the distribution of $(P_t|P_{t-1}, I_{t-1})$ is needed rather than just the mean of its distribution. To create a conditional distribution for P_t, I use the empirical distribution of the errors recovered from the estimation procedure. I then calculate $p(P_t|P_{t-1}, I_{t-1})$ by integrating over the errors for each possible price.

Simple OLS estimation is used to estimate the parameters of equation 11. For maximum precision I estimate the parameters using the observed continuous prices and then calculate the conditional probabilities given the number of discrete prices.[12]

5.3 Structural Parameter Estimation

Once the transition matrix is defined, I use policy function iteration to solve the dynamic problem and estimate the parameters of interest for each generator on the Texas grid. The choice of the policy function iteration within a nested fixed point algorithm is motivated primarily by the very short intervals in the model. Since output and price are observed every hour the discount factor for each period is very close to one. If I assume an annual discount rate of 0.95, this translates into a discount factor of approximately 0.9999945 every hour. This renders value function iteration impractical. Since this is a single agent problem, where each generator's problem can be solved separately, the estimation process lends itself to cluster computing which can reduce estimation time by several orders of magnitude. Other computational methods, such as MPEC, were considered, but require the specialized software not present on the cluster computing resources available.

Although I have several years of production data available for each generator, I use a three month subset of the data, from May 2006 through July 2006, for estimation. I use only a subset of the data for two reasons. First, using more data increases the state space. Within the period, I assume that fuel costs are constant and that the the markov price process is uniform throughout. Extending the dataset would necessitate expanding the state

[12] Alternatively, I could discretize the price space before estimating the parameters with some loss of precision. As the number of discrete price increases, the results will converge the continuous price estimates. In practice I have found that even for 100 price bins the probabilities generated by the discrete estimation are very similar to the probabilities created using continuous prices.

space to account for seasonal changes, entry/exit of generators, and demand growth that would change the price transition probabilities. I would also need to explicitly model each firm's expectations for future fuel costs. The increased size of the state space would make computation infeasible. Second, by using these months of data, I am able to avoid maintenance periods for generators. When generators are offline due to scheduled maintenance, decisions are not motivated by price signals. In short, using a shorter period data prevents the model from becoming overly complicated.

5.4 Identification

The arguments for the identification of the structural parameters are fairly straightforward. First, the generator's start up cost is identified by the difference in the willingness to operate between two states with the same price and interval, but with a differing operating state last period. Consider the price/interval combination ($P_t = 50, I_t = 20$). The start up cost is identified by the difference in the firm's behavior at ($P_t = 50, I_t = 20, s_t = 1$) versus ($P_t = 50, I_t = 20, s_t = 0$). Startup costs imply that the probability of operation will be higher in the first case. In a world with no startup costs, the behavior of the firm would be identical when faced with either of those states. The scale of the variance, σ, of the fixed cost shock is identified by the willingness to operate in states outside of the cutoff rule implied by the deterministic model. More extreme or frequent deviations from the cutoff rule imply a higher σ. The fixed cost each period is simply the mean of the fixed cost shock.

The parameters for certain generators will not be identified. In order for startup costs to be identified, a generator needs to turn on/off voluntarily in response to price signals. Some baseload generators, such as nuclear plants or some coal generators, may only shutdown for scheduled maintenance or an equipment breakdown. For such generators, startup costs cannot be point identified although a lower bound on startup costs might be obtained. A lower bound would be identified by the lowest levels of observed prices under which the generator continues to operate. Ostensibly, there is some level of prices for which the generator would shutdown. How informative the bound is depends on how nearly the generator comes to shutting down at observed prices. In this paper, I do not attempt to bound startup costs on baseload generators, but rather use calibrated parameters for these few generators.

6 Results

At this time, I have focused my efforts on the smallest zone in the ERCOT grid which contains 37 fossil fuel generators. The composition of the fossil fuel generation facilities is summarized in table 1[13] As is the case for ERCOT as a whole, most generating capacity in this zone is gas fired and includes both combined cycle and simple cycle gas generators. There is one relatively new coal plant which is equipped with scrubbing equipment to remove SO_2 from the exhaust gases. Capacity is not highly concentrated in any one generator, but production is. The coal plant produces 34% of the electricity for the zone. The aggregate production from one combined cycle plant, Odessa-Ector, produces 44% of total output. These two generators have the lowest marginal costs in the zone and thus are the baseload producers.

Each generator in this zone is modeled as a single agent with one exception. To model each generator as a single agent, each needs to be able to react independently to prices. Combined cycle gas generators violate this rule since they run multiple turbines in sequence. In a combined cycle plant a simple combustion turbine first used to burn the natural gas. The exhaust of this turbine is used to heat water which powers a secondary steam turbine. Thus the operation of the steam turbine is closely linked to the operation of the combustion turbine. Some plants may have two or three combustion turbines which all feed a single steam turbine. Such plants can run in multiple configurations, such as with just one or two combustion generators feeding the steam turbine. Since the cost of starting up the steam turbine may be high, a plant may operate one gas turbine at minimum capacity to avoid the start up costs associated with restarting the steam turbine. If the gas turbine were modeled as a single agent, this would overstate the startup cost of this generator. To alleviate this problem I aggregate the output of all generators which are part of a combined cycle plant. In doing this I assume that the economically important startup costs are incurred when the entire plant starts production and abstract away for ramping costs associated with the output capacity of the plant.

Table 3 shows the estimates of startup costs all but two generators in the

[13] Wind generators are not included in the model since they lack the capability to increase production in response to price variations. They also do not reduce output during low price periods since their marginal cost of production is near zero. I also exclude one small hydroelectric plant for the analysis also because it cannot increase aggregate production.

Table 1: Generator Characteristics: West Zone

Name	**Fuel**	**Type**	**In-Service Year**	**Max (MW)**	**Min (MW)**	Capacity Share	Generation Share
Calenergy 1	Gas	CC	1988	76	40	1.6%	2.0%
Calenergy 2	Gas	CC	1988	76	40	1.6%	2.0%
Calenergy 3	Gas	CC	1988	60	4	1.2%	1.3%
Graham 1	Gas	ST	1960	229	46	4.8%	1.4%
Graham 2	Gas	ST	1969	377	26	7.8%	3.2%
Morgan Creek 5	Gas	ST	1959	127	15	2.6%	0.1%
Morgan Creek 6	Gas	ST	1966	450	90	9.4%	0.0%
Morgan Creek A	Gas	GT	1988	83	30	1.7%	0.2%
Morgan Creek B	Gas	GT	1988	85	30	1.8%	0.1%
Morgan Creek C	Gas	GT	1988	83	30	1.7%	0.1%
Morgan Creek D	Gas	GT	1988	85	30	1.8%	0.2%
Morgan Creek E	Gas	GT	1988	83	30	1.7%	0.1%
Morgan Creek F	Gas	GT	1988	84	30	1.7%	0.1%
Morris Sheppard 1	Water	HT	1941	12	3	0.2%	0.0%
Morris Sheppard 2	Water	HT	1941	12	3	0.2%	0.0%
Odessa-Ector C11	Gas	CC	2001	145	80	3.0%	7.2%
Odessa-Ector C12	Gas	CC	2001	145	80	3.0%	6.1%
Odessa-Ector C21	Gas	CC	2001	145	90	3.0%	6.3%
Odessa-Ector C22	Gas	CC	2001	145	90	3.0%	7.1%
Odessa-Ector ST1	Gas	CC	2001	215	115	4.5%	8.8%
Odessa-Ector ST2	Gas	CC	2001	215	115	4.5%	8.6%
Oklaunion 1	Coal	ST	1986	630	312	13.1%	34.3%
Permian Basin 5	Gas	ST	1959	116	7	2.4%	0.5%
Permian Basin 6	Gas	ST	1973	492	45	10.2%	6.1%
Permian Basin A	Gas	GT	1988	65	40	1.4%	0.2%
Permian Basin B	Gas	GT	1988	65	40	1.4%	0.3%
Permian Basin C	Gas	GT	1988	65	40	1.4%	0.2%
Permian Basin D	Gas	GT	1990	65	40	1.4%	0.2%
Permian Basin E	Gas	GT	1990	65	40	1.4%	0.1%
Sweetwater 1	Gas	CC	1989	31	25	0.6%	0.3%
Sweetwater 2	Gas	CC	1989	72	50	1.5%	0.8%
Sweetwater 3	Gas	CC	1989	68	50	1.4%	0.8%
Sweetwater 4	Gas	CC	1989	62	45	1.3%	0.7%
Wichita Falls 1	Gas	CC	1990	20	2	0.4%	0.1%
Wichita Falls 2	Gas	CC	1990	20	2	0.4%	0.2%
Wichita Falls 3	Gas	CC	1990	20	2	0.4%	0.2%
Wichita Falls 4	Gas	CC	1990	20	2	0.4%	0.1%

West zone[14]. The parameters were estimated using three months of data for each generator and using 100 discrete prices.

The first three columns of the table 3 show the estimated startup costs, fixed operating costs, and scale of the operating cost shock for each generator. Standard errors are shown in parenthesis below the estimates. The fourth column indicates what type of technology is used at the plant.

The estimates of startup costs are higher than expected. Engineering estimates of fuel and maintanence costs for startup range from \$500 to \$70,000 depending on the generator size and technology. The startup costs estimated here from production decisions start at \$15,000 small gas plants and are as high as \$125,000 for very large combined cycle plants. These are roughly an order of magnitude larger than engineering estimates.

The fuel and emission segments of these startup costs can be separated from other costs using EPA's Continuous Emissions Monitoring System (CEMS). The EPA tracks heat input and emissions output for generators on a continuous basis. Thus, it is possible to calculate average fuel usage and emissions releases over the period when a generator is starting up. The data reveal that the cost of fuel and emissions alone range from \$500 for small combustion gas turbines to \$55,000 for large gas steam turbines. The residual part of startup costs must be attributed to maintenance costs or other costs associated with changing output.

There are a number of explanations for large startup costs. First, the firms may not be responding to balancing energy prices in the way that the model predicts. Firms, for example, could decide to stick with their scheduled and contracted production while putting little weight on prices in the balancing energy market. This explanation would either imply that firms are not optimizing fully or that there are additional non-tangible costs associated with changing significantly from scheduled production. Second, the firms may be responding to prices in other ancillary markets, such as regulation or replacement reserve. However, while incentive to start or stop production in response the unobserved incentives of the ancillary services market may affect some generators, it is unlikely to be a systematic problem. Second, the exercise of market power would have a tendency to inflate startup costs as firm would withhold production in order to increase prices in the

[14]Startup costs could not be estimated for two of the generators. One generator never operated in the period and the other never shut down. In future work I may be able to estimate an upper or lower bound on the start up parameters for these generators.

market. However, a common, but puzzling contradiction in the data is that some generators tend to startup too early, earning apparently negative profits for several hours before becoming profitable. While the market power is an interesting extension, it greatly complicates counterfactual simulations and is beyond the scope of this paper.

Table 2: Plant Emission Rates

Plant Name	Fuel	Type	Heat Rate MMBtu/MWH	NO_x Rate lb/MWH	SO_2 Rate lb/MWH	CO_2 rate lb/MWH
Calenergy	Gas	CC	9.5	1.71	0.03	1114
Odessa-Ector	Gas	CC	7.1	0.57	0.03	832
Sweetwater	Gas	CC	9.8	0.96	0.01	1150
Wichita Falls	Gas	CC	9.3	0.47	0.03	1096
Graham	Gas	ST	11.4	2.46	0.02	1346
Morgan Creek	Gas	GT	15.2	2.16	0.16	1785
Permian Basin	Gas	GT	11.7	1.97	0.50	1376
Oklaunion	Coal	ST	10.7	3.45	1.72	2202

ST=Steam Turbine, GT=Gas Turbine, CC=Combined Cycle

7 Counterfactual

Given estimated parameters, the structural model can be used to simulate equilibrium outcomes under counterfactual environmental policies. In this section, I simulate unit level production, emissions, profits, and aggregate consumer surplus under several possible policy scenarios. The scenarios considered include increasing the share of renewable energy production and directly pricing carbon.

When simulating the counterfactual model, I use the estimated parameters from the previous section[15]. As in the estimation section, I limit the

[15] For the two generators which did not operate or did not shut down during sample period, it was not possible to estimate parameters. These generators include one large coal generator which operated continuously throughout the three month period and one older gas-steam plant. For the counterfactuals, I set the parameters for both of these generators equal to that of large gas-steam plant which was estimable. In future work I hope to be able to estimate bounds on the startup costs of these generators which will improve the accuracy of the simulations.

Table 3: West Zone Results

Unit	$START_i$	σ_i	FC_i	type
Calenergy	\$50,829	\$5,538	\$0	CC
	(28,824)	(2,972)	(0)	
Graham 1	\$88,979	\$15,691	-\$1,544	ST
	(8,707)	(1,572)	(311)	
Graham 2	\$35,636	\$6,679	-\$420	ST
	(2,707)	(426)	(114)	
Morgan Creek A	\$26,662	\$4,842	-\$145	GT
	(3,326)	(581)	(213)	
Morgan Creek B	\$40,103	\$7,207	-\$641	GT
	(6,288)	(1,083)	(330)	
Morgan Creek C	\$25,628	\$4,635	-\$252	GT
	(3260)	(562)	(243)	
Morgan Creek D	\$42,476	\$7,648	-\$619	GT
	(7,289)	(1,288)	(344)	
Morgan Creek E	\$26,550	\$4,725	-\$174	GT
	(3,378)	(574)	(227)	
Morgan Creek F	\$26,027	\$4,738	-\$236	GT
	(3,244)	(561)	(233)	
Morgan Creek 5	\$106,650	\$17,070	-\$3160	ST
	(64,980)	(10,350)	(1970)	
Morgan Creek 6	N/A	N/A	N/A	
Odessa	\$124,250	\$17,430	\$0	CC
	(34,860)	(4,990)	(0)	
Oklaunion	N/A	N/A	N/A	
Permian Basin A	\$33,214	\$6,731	-\$1,082	GT
	(5,609)	(1,073)	(341)	
Permian Basin B	\$36,912	\$7,378	\$1,069	GT
	(6,237)	(1,186)	(353)	
Permian Basin C	\$42,203	\$8,284	-\$1,243	GT
	(7,850)	(1,484)	(419)	
Permian Basin D	\$45,415	\$8,703	\$1,584	GT
	(9,167)	(1,734)	(530)	
Permian Basin E	\$58,804	\$10,625	-\$2,281	GT
	(12,476)	(2,177)	(713)	
Permian Basin 5	\$46,639	\$8,370	-\$1,407	ST
	(6,589)	(1,205)	(233)	
Permian Basin 6	\$122,540	\$20,520	\$0	ST
	(15,780)	(2,780)	(0)	
Sweetwater	\$19,291	\$3,852	\$0	CC
	(2000)	(369)	(0)	
Witchita	\$14,894	\$2,484	-\$210	CC
	(2,897)	(508)	(28)	

simulation to only the west zone of the Texas market. For the counterfactuals, I hold wind generation and electricity transfers between this zone and other zones of the grid constant. Wind output is held constant since wind farms have little control over the amount of power that can be produced in each period. While they can decrease, or curtail, production, it is not possible to increase production on demand. Federal tax incentives for wind power production make curtail production undesirable even when electricity prices are negative. Electricity transfers to other zones in Texas will eventually be incorporated into the counterfactuals when grid wide simulations are run.

For the counterfactual simulations, I need a model of demand for electricity. In the very short run, i.e. minute to minute, the demand for electricity is almost perfectly inelastic. This is because consumers of electricity generally face constant prices over some time period, ranging from one month to several years, which are invariant to changes in wholesale prices of electricity. Thus, consumers have no incentive, or even available information, to change consumption as wholesale prices change.

Even though consumers do not respond immediately respond to wholesale price changes, changes in the average wholesale price for electricity will, of course, eventually filter down to the prices consumers face. The literature on electricity demand reports very different demand elasticities depending on the time horizon. Dynamics exist on the demand side limit consumers' response to price changes in the medium run versus the long run. Just as owners of SUVs are temporarily "locked in" to a higher gas usage even as the prices of gasoline rise, likewise consumers of electricity must make costly adjustments to capital in order to full optimize with respect to prices. Purchasing more efficient appliances, upgrading heating/cooling systems, or insulating a home will allow consumers reduce consumption more in the long run than in the short run given higher electricity prices. Thus the response of demand to

In this paper, I do not estimate a demand side model. Rather, I build on previous studies which estimate price elasticities for electricity demand to capture a range of possible demand responses. I simulate market outcomes using three different assumptions about demand responsiveness: 1) short run inelastic demand, 2) medium run demand response, and 3) long run demand response.

Perfectly inelastic demand implies that consumers do not face the daily or seasonal variations of wholesale prices. However, in this model, it also implies that average wholesale price increases that will accompany many environmental policies, such as a price on carbon, are not passed along to consumers. Although inelastic demand is a realistic assumption for the very short run, it will not fully capture the new market equilibrium which will determine the profitability for different technologies going forward. However, this demand assumption will highlight the ability of the supply side to reduce emissions in response to environmental regulation in the counterfactual simulations. It is also the demand side model that, from a conceptual view, is most consistent with the short run supply side model which holds generating capital fixed.

For the medium and long run demand curves, I calibrate a simple demand function with parameters taken from the literature. In particular, I assume that each hourly period is characterized by a constant elasticity demand function, $D_t = K_t * p_c^{\alpha}$. Here, D_t is the observed hourly demand for electricity in time period t, p_c is the price consumers face for electricity, α is the demand elasticity parameter, and K_t is a positive constant. I assume that consumers face the average wholesale price for electricity over the simulation period. Given the constant price, the hourly changes in electricity demand are modeled as shifts in the constant elasticity demand curve. That is, for a given elasticity parameter, there is a K_t which rationalizes the observed quantity demanded for that hour. For each hourly demand curve over the simulation period, I can back out the K_t which rationalizes the observed quantities, given the elasticity parameter α and the observed average price in the wholesale market before any environmental regulation. These functions are then used to characterize the demand response when solving for counterfactual equilibrium prices for a given demand elasticity.

I rely on previous studies to inform the choice of the elasticity parameter, α. There is long literature which estimates the elasticity of demand for electricity which has produced a wide range of results. However many studies identify the medium run elasticity for electricity demand to be somewhere around 0.2 (Bohi 1981)(Espey & Espey 2004)(EIA 2008). I use this demand elasticity to simulate outcomes in a medium run situation where consumers observe higher priced electricity and respond accordingly but are not able to make capital adjustments to fully optimize to the new prices. Third, I assume demand varies with a long run supply elasticity. Pulling again from past literature, I use 0.7 as the long run elasticity of electricity demand. This represents a situation where consumers of electricity are fully able to respond to new equilibrium prices through capital adjustments. Using a long run demand elasticity is somewhat inconsistent with my supply model since I assume that the supply side is not able to adjust its capital; this implies that consumers can change capital much more quickly than electricity generators. However, just as inelastic demand gives a lower bound on short run emissions reductions, long run demand provides an upper bound on the emissions reductions that could be achieved by environmental policies holding electricity generating capital constant.

I solve for the counterfactual price taking equilibrium by ensuring three simple conditions are met. First, each firm must be acting optimally with respect to price. Second, the equilibrium prices must clear the market. Third, firm's expectations for prices must be consistent with equilibrium price vector.

The algorithm for solving is outlined as follows. Let P^0 be a Tx1 vector of observing equilibrium prices.

1. Estimate the price transitions, $p(P_t^0|P_{t-1}^0, I_{t-1})$.
2. Change structural parameters as determined by the policy change.
3. Solve the dynamic problem for each generator given the transition matrix $\Rightarrow p(a_{it}|P_t, I_t, s_{it})$.
4. Calculate expected supply, $E[s_{it}; P_t, \theta_i]$, for each generator at each possible price.
5. Choose a new vector prices, P^1, such that $\sum_{i=1}^{N} E[s_{it}; P_t, \theta_i] = D_t$ in each period.
6. Re-calculate D_t given the new average price, $E[P']$.
7. Re-estimate the price transitions $p(P_t'|P_{t-1}', I_{t-1})$.
8. Return to 3 and iterate until the market clearing price vector does not change between iterations.

I have not formally shown the convergence will occur or that the "found" equilibrium is unique. However, empirically the algorithm does converge to a solution which seems to be robust to initial conditions. The fact that the optimal policy functions are increasing in price given I_t, s_t and the transition probabilities may explain this consistency[16].

[16] The probability of operating in increasing in P_t because the value function is also increasing in P_t due to per period profits increasing in P_t

Since firms respond to an unobserved fixed cost shock, supply of a given generator can only be calculated in expectation. Given the optimal policy function implied by the transition matrix and a set of structural parameters θ_i, the expected supply function of a given generator in any period can be calculated as follows.

Let $\lambda_{it}(P_t, I_t, \lambda_{it-1}, \theta_i) =$ probability of operating generator i in time t

$$\lambda_{it}(P_t, I_t, \lambda_{it-1}, \theta_i) = \lambda_{it-1} p(a_{it}|P_t, I_t, 1) + (1 - \lambda_{it-1}) p(a_{it}|P_t, I_t, 0) \quad (12)$$

$$E[s_{it}; P_t, \theta_i] = \lambda_{it} Q_{it}(P_t, \theta_i) \forall t \in \{1, 2, ..., T\} \quad (13)$$

Where Q_{it} is determined as specified in equation 2.

The aggregate supply in any time period t is then simply

$$E[S_t; P_t, \theta] = \sum_{i=1}^{N} E[s_{it}; P_t, \theta_i] \quad (14)$$

Solving for the expected supply in each period t requires an initial condition, λ_{i0}, for each generator i in the market. For the initial conditions ,λ_{i0}, I simply use the actual operating state in the period before the beginning of the simulation.

I solve for the equilibrium prices under price on $C0_2$ of \$20/ton , \$50/ton carbon tax, and a \$200/ton. I solve each counterfactual under each of the three assumptions about the elasticity of demand, inelastic, short run elastic and long run elastic. This implies 9 counter factual results altogether. Table 4 shows the results.

Table 4: Counterfactual Results

	Inelastic			Elasticity -0.2			Elasticity -0.7		
	\$20	\$50	\$200	\$20	\$50	\$200	\$20	\$50	\$200
Δ CO_2 Emissions %	0%	-7%	-22%	-2%	-15%	-41%	-6%	-29%	-68%
Δ Avg Price	\$14	\$37	\$146	\$11	\$31	\$138	\$7	\$24	116\$
Δ Price %	26%	69%	272%	20%	57%	255%	13%	45%	215%
Δ Coal Prod. %	0%	-16%	-69%	-1%	-22%	-75%	-1%	-32%	-85%
Δ Gas Prod. %	0%	24%	105%	-9%	11%	58%	-19%	-8%	-10%
Δ Π Coal %	-20%	-46%	-98%	-9%	-68%	-100%	-41%	-79%	-100%
Δ Π Gas %	13%	34%	109%	-9%	-45%	34%	-57%	-63%	-26%
Δ Π Industry %	-7%	-14%	-16%	-9%	-59%	-46%	-47%	-73%	-70%
Δ Consumer Surplus	-\$33m	-\$84m	-\$328m	-\$17m	-\$54m	-\$296m	-\$1.8m	-\$34m	-\$244m
Δ Demand	0%	0%	0%	-4%	-9%	-13%	-9%	-13%	-55%

The first three columns of the table show the counterfactual results with an inelastic demand curve each period under each carbon tax. A \$20 carbon price is with in the range of prices that carbon permits have been selling for in the EU. A \$200 tax represents an extremely high price on carbon. With an inelastic demand curve, any reduction in emissions comes from production substitution from high carbon generators to lower carbon generators. Under a \$20 tax carbon emissions do not change. This is because the high carbon producers, coal generators, are still the lowest cost producers on the grid. At the same time, prices increase by 26% or \$14 a MWH. Average observed prices in this simulation before the tax were \$54 MWH.

Despite short run emissions remaining unchanged, the profitability of different technologies changes dramatically. The profitability of coal plants decreases by 20% while gas fired power plants profits increase by 13%. This underlies long run implications of carbon pricing; firms may make very different future investments even if current production decisions remain essentially unchanged.

Allowing demand to respond to a short run price elasticity of -0.2 produces a small reduction in emissions by 2%. This comes mostly from decreased consumption as opposed to fuel switching. Increasing the elasticity to -0.7, results in even lower overall consumption and a emissions reduction of 6%. The salient feature of these results is that a potentially politically feasible carbon tax of \$20 changes emissions from electricity generation only slightly even when demand is can completely adjust to the new, higher prices due to the tax.

With a much higher price on CO_2, emissions are reduced even with an inelastic demand curve. This reduction in emissions comes from a large substitution between gas and coal generation and a higher reliance on the most efficient gas generators. With a \$200 price tag on carbon emissions, aggregate emissions are reduced by 22% due to this supply side substitution, but the price consumers pay for electricity almost triples. Allowing for elastic demand greatly decrease emissions due to the fact that average electricity price increase by 200%. With a long run elasticity, CO_2 emissions are down 68% and coal production virtually disappears.

I also solve for a counterfactual which simulates the introduction of wind power installations such that 10% of power currently produced by conventional generators is produced by wind power. Since wind power installations already exist in west Texas, I can use the production profile of those wind farms to simulate the production of new wind farms. Even if new wind farms are less productive than existing wind farms, due to being placed on less desirable properties, scaling the production patterns of existing wind farms will provide good approximation of continued build out as long the diurnal and seasonal patterns of wind production are similar.

Wind farms in this region exhibit electricity production patterns that are heavily skewed toward off peak power production. In fact, it is quite common for on shore wind farms to have significantly higher levels of production at night and in the spring and fall, when demand for electricity is at its lowest levels. This pattern of production will have the tendency to exacerbate the variation in wholesale electricity prices by lowering off-peak prices. It may even increase on-peak prices as more generators may be forced to cycle on and off. My model is ideally situated to simulate production behavior with this increased price volatility.

Table 5 displays the outcomes resulting from wind power production offsetting 10% of power currently being produced by fossil fuel generators in the west zone.

Table 5: Counterfactual Results: 10 % Wind

	Inelastic	Elasticity -0.2	Elasticity -0.7
Δ CO_2 Emissions %	-10%	-6%	-4%
Δ Avg Price	-$13.20	-$7.93	-$4.82
Δ Price %	-22%	-12%	-8%
Δ Coal Prod. %	-12%	-8%	-6%
Δ Gas Prod. %	-7.5%	-4%	-2.5%
Δ Π Coal %	-14%	-8%	-4%
Δ Π Gas %	-4%	-2%	0.5%
Δ Π Industry %	-6%	-3%	-0.5%
Δ Consumer Surplus	$22.8m	$10.9m	$3.4m
Δ Demand	0%	2.5%	3.4%

The first column shows that emissions of CO_2 decrease by 10% when 10% of electricity is now produced by wind power. Both coal and natural gas plants reduce production, but coal reduce production more than gas reflecting the tendency of wind power to produce in off peak periods and to cut into base load electricity production. Since in this counterfactual wind power is being exogenously inserted into the market, prices also decline. Moving across the columns to look at thee elastic demand shows that emissions and price reductions are mitigated by the demand response to lower prices. Consumers demand more electricity due to lower prices resulting in CO_2 reductions of only 4% when demand fully adjusts. This highlights the potential feedback effects that can occur when subsidized renewable energy is inserted into the grid. Profitability of gas plants is not hit as hard as coal plants due to their ability to better respond to energy price changes. As compared to a price on carbon, the effects of wind power on emissions are more immediate, but do not incentivize investment in lower carbon emission technologies to the extent found for a price on carbon.

While all these counterfactual results specific to one zone in the ERCOT market, they illustrate the trends to be expected from a grid-wide analysis.

8 Conclusion

In this paper I build a dynamic model of electricity output in a price taking setting which accounts for the startup costs of generators. I abstract away for some institutions in the market such as transmission costs and continuous output adjustment costs to model the choice of the firm as a simple on-off decision. I use the model to estimate generator level startup costs using data from the Texas electricity grid. In doing so, I solve each firm's dynamic problem with a discount factor that is close to one. Estimates for generators' start up costs are higher than expected when compared with the range of values found in the engineering literature.

I also develop a method for computing the price taking equilibrium given estimates of the structural parameters of the model. The key condition for the equilibrium is that firms' expectations for prices must be consistent with the counterfactual equilibrium vector of prices. I exploit the monotonicity of a firm's optimal policy to solve for this equilibrium price vector. I use the estimated parameter values to simulate outcomes under several counterfactual environmental policies.

I find with a short run inelastic demand curve, $20/ton price on carbon has no effect on carbon emissions from generators even while wholesale prices increase by 26% on average. The negligible change in emissions due to the fact that very little substitution occurs between high marginal cost, low emissions gas generators and low marginal cost, high polluting coal generators. The lack of substitution is driven by the large initial marginal cost advantage enjoyed by coal plants; a moderate carbon tax still leaves coal plants as the low cost producer. A higher price on carbon is necessary to induce substitution towards gas generators. However, even a lower price on carbon drastically changes the investment incentives for coal versus gas generating technologies making coal power plants much less profitable compared with their gas fired counterparts.

Renewable energy production by wind power, on the other hand, has an immediate effect on carbon dioxide emissions. However, it lacks the same long run incentives for technology switching that a carbon price provides. Also, the long-run demand response to lower energy prices due to subsidized renewable energy investments mitigates some of the initial emissions benefits of wind.

These counterfactual experiments simulate the response of firms to environmental policies holding fixed the generation capital on the grid. This reflects probable outcomes over a relatively short two year window which is the approximate time required to build new generating capital. A larger decrease in emissions would be expected over a longer time period which would allow both generators and consumers of electricity to adjust their capital investments in response to new equilibrium prices. In fact, the output of this research is a natural input into an investment model which could look at the long run implications of carbon pricing and renewable energy development.

The results of this analysis paint a dismal picture for the short run emissions implications of carbon pricing. However, even moderate carbon pricing impacts the profitability of high carbon technologies profoundly. For policy makers, this implies that meaningful reductions in carbon dioxide emission may be able to be achieved without drastic environmental policies, but a long time horizon may be required to realize those benefits.

References

Bharvirkar, R., Burtraw, D. & Krupnick, A. (2004), No_x emissions trading and episodic air quality violations in maryland, Technical report.

Bohi, D. R. (1981), *Analyzing demand behavior: a study of energy elasticities*, John Hopkins University Press for Resouces for the future.

Borenstein, S., Bushnell, J. B. & Wolak, F. A. (2002), 'Measuring market inefficiencies in california's restructured wholesale electricity market', *American Economic Review* **92**(5), 1376–1405.

Bushnell, J. B., Mansur, E. T. & Saravia, C. (2008), 'Vertical arrangements, market structure, and competition: An analysis of restructured us electricity markets', *American Economic Review* **98**(1), 237–66.

Chow, J., Ho, K., Du, X., Lee, H. & Pearson, M. (2002), 'Experience from extensive two shift operation of 680mw coal/gas-fired units at castle peak power station', *Operation Maintenance and Materials Issues* **1**.

EIA (2008), Energy information adminstration: Annual energy outlook 2008, Technical report.

Espey, J. A. & Espey, M. (2004), 'Turning on the lights: A meta-analysis of residential electricity demand elasticities', *Journal of Agricultural & Applied Economics* **36**(1), 65–81.

Hortacsu, A. & Puller, S. L. (2008), 'Understanding strategic bidding in multi-unit auctions: a case study of the texas electricity spot market', *RAND Journal of Economics* **39**(1), 86–114.

Imai, S., Jain, N. & Ching, A. (2009), 'Bayesian estimation of dynamic discrete choice models', *Econometrica* **77**(6), 1865–1899.

Mansur, E. T. (2008), 'Measuring welfare in restructured electricity markets', *The Review of Economics and Statistics* **90**(2), 369–386.

Norets, A. (2009), 'Inference in dynamic discrete choice models with serially correlated unobserved state variables', *Econometrica* **77**(5), 1665–1682.

Rust, J. (1987), 'Optimal replacement of gmc bus engines: An empirical model of harold zurcher', *Econometrica* **55**(5), 999–1033.

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