Dynamic Response to Environmental Regulation in the Electricity Industry

Joseph A. Cullen *

September 3, 2009

Abstract

Climate change, driven by rising carbon dioxide (CO_2) levels, has become one of the most pressing economic and political issues. 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, taken as 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 startup costs. 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. Preliminary 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 ten times that of expected levels, lowers emissions by only 9% in the short run.

*I wish to thank Gautam Gowrisankaran for encouragement and guidance throughout this project. I also thank Allan Collard-Wexler, Alex Shcherbakov, Daniel Ackerburg, Gregory Crawford, Keisuke Hirano, Mo Xiao, Stan Reynolds, Price Fishback, and Jonah Gelbach, as well as the participants in University of Arizona Department of Economics seminar series, for their helpful comments on drafts of this paper.

1 Introduction

Climate change has become one of the most pressing political and economic issues 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 (Cullen 2008). In order to properly evaluate potential policies, it is important to accurately gauge the response of polluting industries. My research represents the first attempt to

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

compute 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 a 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. Moreover, environmental regulation has 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. 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.[2] With the introduction of a significant portion of wind generating capacity, the resulting 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. The equilibrium production profile, and its associated pollution profile, will depend crucially on the relative startup costs of generators. Low startup cost generators will find it profitable to shutdown during periods of low demand when prices are also low. Generators with high startup costs will either remain in a state of constant production, though possibly incurring loses during depressed price periods, or they may choose not to produce at all. Other environmental policies, such as carbon regulation, also have the potential to change the equilibrium production and

[2]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.

pollution profiles. No studies to date have attempted to model the response of the electricity producers to environmental regulations within a dynamic framework.

I develop a dynamic structural model to analyze output dynamics and simulate counterfactual equilibrium outcomes. I recover cost parameters for the industry using the optimal policy functions from the dynamic model. I do this under the assumption generating capital is fixed and that firms are price takers. The recovered parameters and assumptions about the elasticity of demand will govern the resulting price equilibria under different policy environments. 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 model 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 dynamic discrete choice model of generator operation. Given that firms make production decisions every 15 minutes, I use recent methods highlighted by Judd & Su (2008) to estimate the dynamic model with a discount factor that is close to one.

Using the estimated cost parameters, I simulate the outcomes in the electricity market under 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. Preliminary 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. A very large carbon tax of ten times that of expected price levels, lowers emissions by only 9% in the short run.

My model has several advantages over a reduced form approach to analyzing counterfactual outcomes. Since I explicitly solve each generators' dynamic problem I can simulate equilibrium outcomes that are very different from observed equilibrium outcomes. Reduced form approaches are not able to effectively counterfactual equilibria which are too far out of sample.

Second, my 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 preliminary 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

[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

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 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 maintanence costs in the long run. In fact, increased wear and tear on generating equipment may account for almost half 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. The costs increased 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 lose efficiency which will increase startup costs.

The costs associated with output changes are significant. Startup costs can range from a few hundred dollars to tens of thousands of dollars per

study.

start depending on the size and technology of the generator. Consequently, a generator with high startup costs may decide not to shutdown during low price periods if the cost of starting up later outweighs the potential financial loses of producing during low price periods. Likewise, a generator may not startup even though prices exceed its marginal cost of production if it believes that profits will not exceed its startup costs.

Adjustment costs make a generator's choice of output a dynamic problem. The generator must weigh the benefits of increasing/decreasing production this period with the costs adjusting to market conditions in the future[4]. The firm must solve the optimal path of output output given its expectations about future prices and levels of demand.

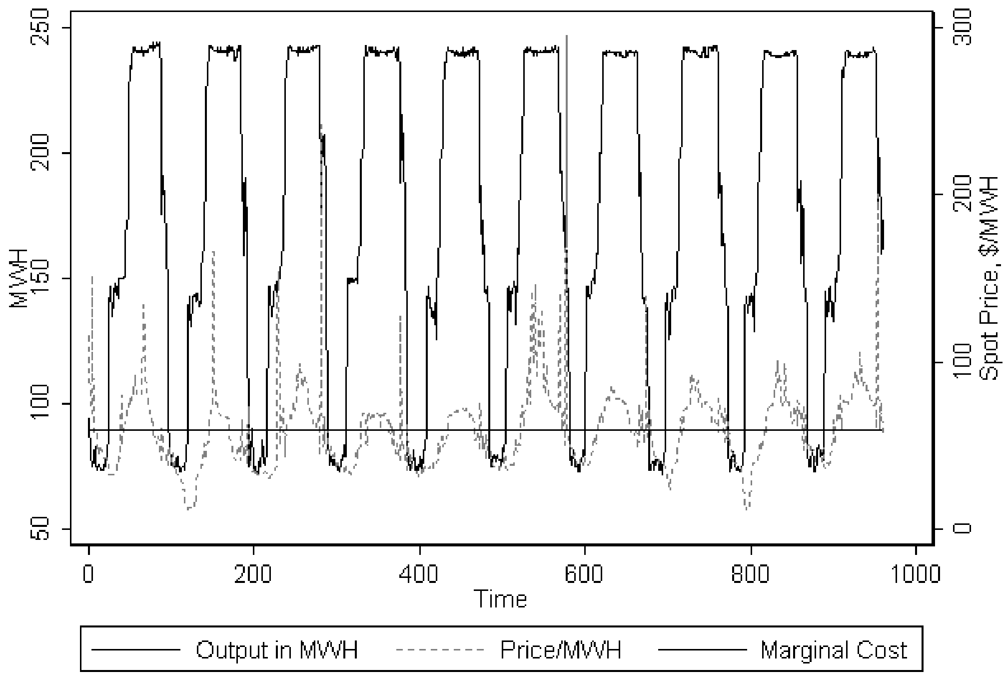
Evidence of the importance of startup costs for firm behavior is illustrated in figure 1. This figure shows one generator's output over a 10 day period in July of 2006. The horizontal line shows the firm's constant marginal cost of production while the dashed line shows the spot price for energy. Notice 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.

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 quasi-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

[4]For example, consider two generators facing demand which could be satisfied this period by just one generator. Suppose demand is expected to rise next period above the capacity of a single generator. Given adjustment costs, it may not be optimal shut down one generator this period, due to the costs incurred next period of restarting the generator. (Rather, both generators could be ramped down to a low level of production to avoid excess supply this period and startup costs next period.)

Figure 1: Operating Decision Example



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 to the ERCOT grid. 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.

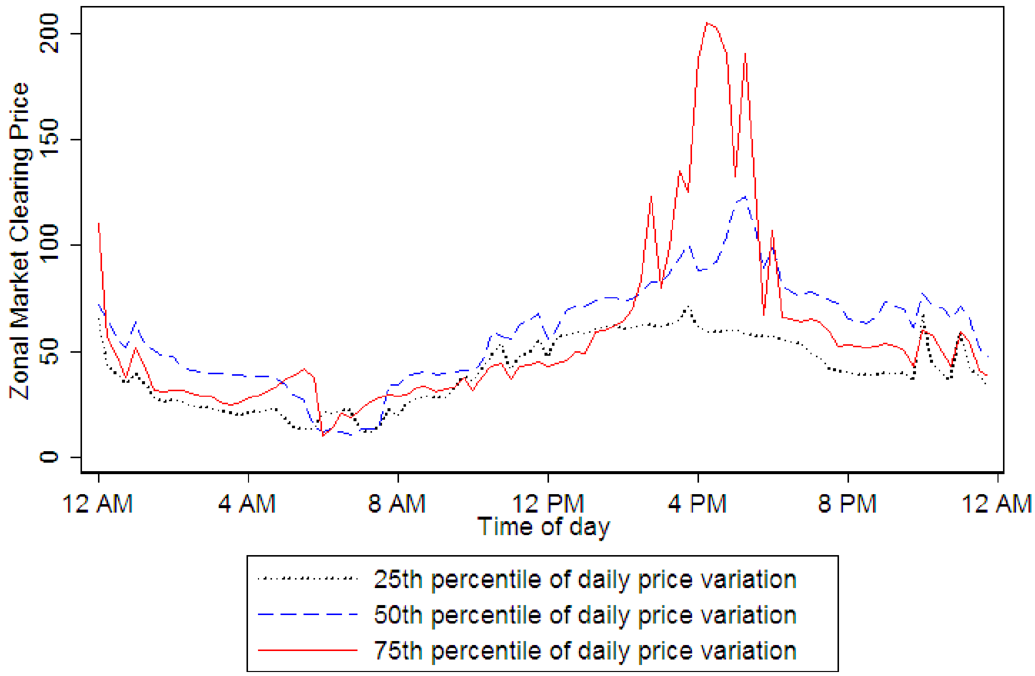
To ensure that there is sufficient supply, ERCOT requires generators and electricity retailers submit scheduled energy transactions a day ahead. These schedules are submitted through a Qualified Scheduling Entity (QSE) which generally submits schedules for a portfolio of generators and power purchasers. 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[5]. 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[6]. 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 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

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

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

Figure 2: Representative Daily Price Variation by Percentile



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. 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, inter-zonal congestion is accounted for by the price energy in the balancing.

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[7].

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 consumption a wholesale prices increase during peak demand periods. Some large industrial consumers do curtail electricity use when reserve capacity becomes short but they do not directly

[7] 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.

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. 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 describes the Texas grid which is managed by ERCOT over the period of 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]. With 2 to 4 generators at each power plant this makes for approximately 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, I observe the output of each generator every fifteen minutes from April 2005 through April 2007. I also observe the market clearing price for the balancing energy every 15 mins for each zone. For each generator and interval, I also know if the generator was shutdown for maintenance or due to an involuntary mechanical failure. I have 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.

[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. The plant characteristics include the 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. The heat rate is a measure of productive efficiency for the power plant defined as the amount of heat input necessary to produce one unit of electricity. The heat rate is usually measured in millions of British Thermal Units (BTUs) per megawatt hour (MWH) of electricity. For combined heat and power facilities, the heat rates are adjusted to reflect only the heat used to produce electricity. Emissions rates for SO_2 and NO_x are derived from continuous monitoring equipment installed in generators by the EPA. Emissions rates for CO_2 are derived from the heat rate and the typical CO_2 production for a given fuel type per unit of heat. All emissions rates are measured in pounds of pollutant per MWH of electricity.

To construct the marginal cost of electricity production for each generator, I need both fuel and pollution permit costs. For fuel costs, I use monthly information from EIA form 423 which gives the delivered quantity and cost of fuel for both coal and gas powered plants in Texas which use steam turbines. I take the quantity weighted average gas (coal) price as the price for gas (coal) for all generators in the market for that month. In doing this, I assume that the fuel cost for gas turbines is the same as for gas-fired, steam turbines. 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. 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, since emissions and heat rate information are plant level characteristics, they may be misleading when applied to generators if at a given plant the characteristics of generators are very different. However, generators at the same plant are usually of the same vintage and type so their characteristics are probably very similar for most plants. Second, 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. Third, 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. Fourth, 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. I am currently working on obtaining generator level information that would allow me to identity the generators which provide these services and to identify what portion of their output it produced in response to these signals.

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 me 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 exactly 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). However, over most of the range of the generator the heat rate is fairly constant. 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 guarantees the firms are not constrained in how they react to price. This assumption is very plausible for the vast majority of firms. The structure of the market in Texas allows for inter-zonal transmission constraints as reflected in the market clearing price in each zone. Intra-zonal congestion is alleviated by ERCOT deploying specific generators within the zone by offering a higher price than the market clearing price. Thus some generators may be responding to price signals other than what is observed. However, such deployments are specific to certain generators; the rest of the generators on the grid are not affected. Inter-zonal congestion posses no complications since it is alleviated through the market price for electricity. Intra-zonal or local congestion does pose problems, but only for a limited number of generators who receive above market price payments to increase or start production to alleviate local congestion. However, the set of generators who operate to relieve congestion is known and on going work will account for these.

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 plausibility of this assumption for generators in general is questionable although many generators, particularly those with more flexible technologies, do follow this pattern closely. In future work, I plan to relax this assumption and explicitly model the firm's choice of output level. 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 remains 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 exhibits a bimodal distribution of capacity utilization, but the distribution is much more diffuse and the upper mode is not at the generator's declared maximum capacity. 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.[9]

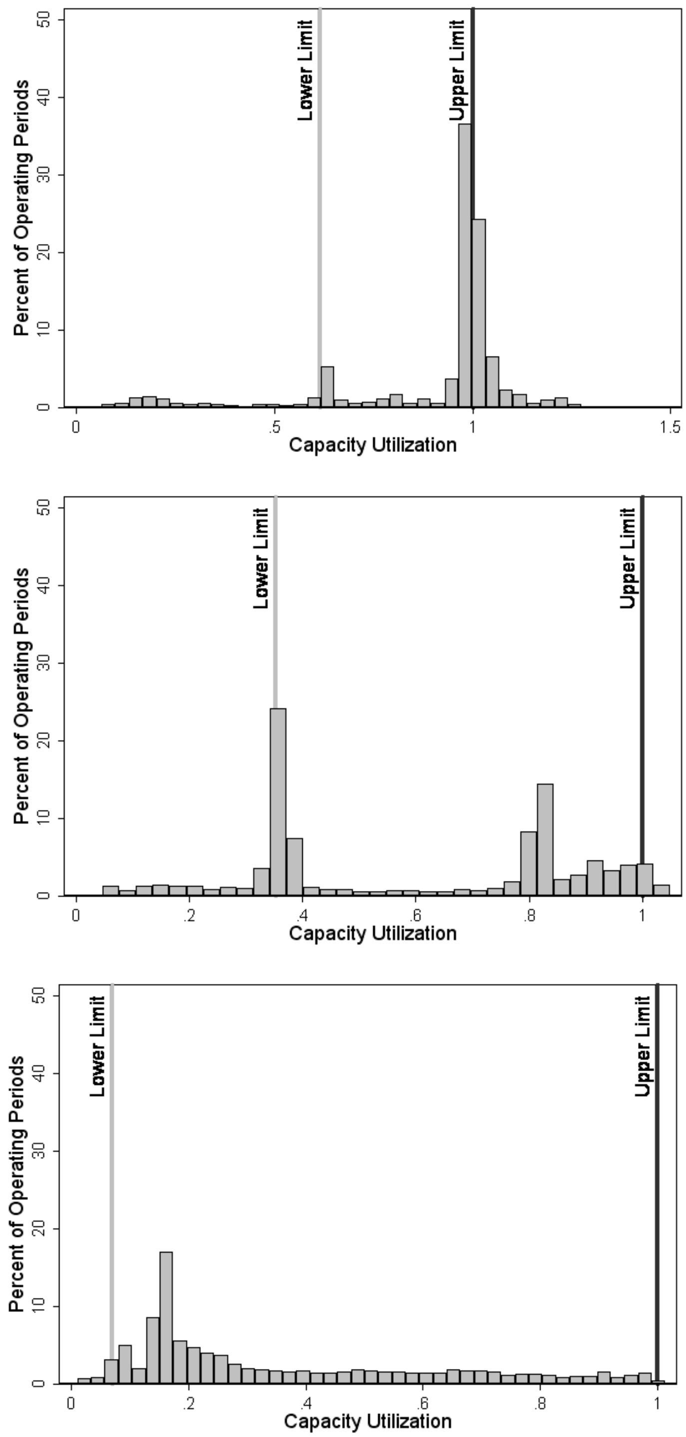
Given these assumptions, I model each generator as a single firm with the following single agent dynamic problem. In each period, the firm observes the price in the market and the interval of the day. The firm can take one of two actions which I notate as:

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

where i indexes the generator
t indexes each fifteen minute time period

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

Figure 3: Capacity Utilization Histograms



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.

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

$$\begin{array}{rl} \text{where} & c_i = \text{constant marginal cost of generator } i \\ & P_t = \text{price for electricity in the generator's zone} \end{array}$$

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:

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

$$\begin{array}{rl} \text{where} & OC_i = \text{non-variable operating cost for generator } i \\ & START_i = \text{cost of starting up generator } i \\ & s_{it} = a_{it-1} = \text{the operating state last period} \end{array}$$

I allow for a fixed cost of operation every 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 $START_i$. I have assumed that c_i, the constant marginal cost of production, 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 I will drop the i subscript 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 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 15 minute interval of the day. That is I assume that 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. 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 $F(P_{t+1}|P_t, I_t)$ and 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 as was done 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. Therefore the choice specific shock to fixed costs is $\sigma\epsilon_t(a_t)$. 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 to me to integrate over unobserved shocks using an analytical form.

With the unobserved state variable the Bellman equation now becomes:

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

where the function

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

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

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 I do not have an analytical solution for the value function or the resulting policy function, 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. The resulting state space could be quite large depending on how finely P_t is discretized. The dimension of I_t is 96 since there are 96 fifteen min intervals in each day. The operating state last period, s_t, is a binary outcome. The size of the state space is then $DP * 96 * 2$ where DP is the number of discrete prices used. For one hundred discrete prices the total size of the state space would be 20,000 which is large but not computationally infeasible. 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. Below I discuss two methods for maximizing the likelihood function. The first is an unconstrained optimization algorithm developed by Rust (1987) and the second is a constrained optimization method recently highlighted Judd & Su (2008).

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 twodraw backs 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 modelling short time periods, such as 15 min 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. Policy function iteration convergence does not depend on the discount factor. Although this method requires inverting the probability matrix which in many cases may be computationally infeasible, 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 MPEC

An alternative to the nested fixed point approach is to recast the *unconstrained* optimization problem into a *constrained* optimization problem. This direct optimization approach was suggested Luo, Pang & Ralph (1996) which they called Mathematical Programming with Equilibrium Constraints (MPEC). A similar constrained optimization technique was also used by Trick & Zin (1997) as an alternative way to solve the value function in a dynamic problem. Recently Judd & Su (2008) has highlighted the benefits of solving dynamic problems using MPEC when combined with the latest commercial solvers. These methods are very general and may not take advantage of the special features of an economic model. However they have the advantage of using very efficient algorithms and being relatively straightforward to implement using high level software. As a result, they can actually be computational faster than traditional techniques as shown in Judd & Su (2008). In my case, there is a significant advantage to the MPEC approach since the solution time is independent of the discount factor. This enables me to solve for the optimal parameters even with a β which is close to one.

The MPEC approach augments likelihood function to build a constrained optimization problem. Rather than solve for EV_θ for each parameter vector θ, the MPEC approach views EV_θ as a choice variable that must meet certain equilibrium constraints. The equilibrium constraints in my model are that EV_θ is the fixed point of the contraction mapping for each point in the state space. This new likelihood function with the contraction contraints is called an augmented likelihood function by Judd & Su (2008) to reflect the additional choice parameters in the likelihood function. The constrained maximization of augmented likelihood function is:

$$\begin{array}{rl} \max\limits_{\theta, EV_\theta} & \Pi_{t=1}^{t=T} p(a_t|P_t, I_t, s_t; \theta) \\ \text{subject to} & EV_\theta(P, I, a) = T_\theta EV_\theta(P, I, a) \quad \forall (P, I, a) \end{array}$$

The solution to the reformulated problem does not require solving the value function even once, but has as many non-linear constraints as the dimension of the state space.

5.3 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 96 Markov price transition matrices , one for each interval 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 96 intervals in each day, this means that 960,000 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 linear polynomial equation in P_{t-1} with dummies for each interval of the day as shown in the equation,

$$P_t = \beta_0 + \beta_1 P_{t-1} + \beta_2 P_{t-1}^2 + D\alpha + \epsilon_t \tag{11}$$

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

The parameter estimates from the above equation yield $E[P_t|P_{t-1}, I_{t-1}]$. However, to calculate all the conditional choice probabilities I need to know the distribution of $(P_t|P_{t-1}, I_{t-1})$ rather than just the mean of its distribution. To create a conditional distribution for P_t, I assume that the error term is normally distributed with mean zero and variance equal to the sample variance of the residual. I then calculate $p(P_t|P_{t-1}, I_{t-1})$ by integrating the normal distribution over each of the discretized price bins.

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.[10]

5.4 Structural Parameter Estimation

Once the transition matrix is defined, I use policy function iteration to solve the dynamic problem 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 I observe output and price every 15 minutes 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 0.999999 every fifteen minutes. Another reason this method is used has to do with the development of the algorithm.

The nested fixed point algorithm with policy iteration was chosen over the MPEC method for several reasons. First, the nested fixed point algorithm with policy iteration can be easily developed on a number of platforms. To implement the MPEC method efficiently one needs to take advantage of efficient algorithms used in commercial solvers such as SNOPT. These solvers are a essentially a black box which makes debugging errors very difficult. Second, the nested fixed point algorithm is easily parallelized to enable it to run on a computer cluster using readily available software. Software costs make implementation of the MPEC method on a cluster impractical which increases total computing time.

[10] 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.

For each generator, I select just one month of data to use to recover the structural parameters despite the fact that I have two years of data for most generators. I use only one month for two reasons. First, using more data increases the state space. Within a month, I assume that fuel costs are constant and that the price transition probabilities remain the same. Extending the dataset would necessitate expanding the state 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. Second, by using one month of data I am able to avoid maintenance periods for generators. When generators are offline due to scheduled maintenance or equipment failure operating decisions are not motivated by price signals. I avoid this by choosing a month of data for each generator which does not include any outages. In short, using one month of data prevents the model from becoming overly complicated.

5.5 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. Presumably there is some level of prices for which the generator would shutdown. How informative a 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

Due to the computationally intensive nature of this research, I have obtained only preliminary results at this time. I have focused my efforts on the smallest zone in the ERCOT grid which contain 23 generators. The composition of the fossil fuel generation facilities is summarized in table 1[11] Most generating capacity 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

[11] 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%

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 I abstract away for ramping costs with the output capacity of the plant.

Table 2 shows the some preliminary estimates of startup costs for 12 of the 23 generators in the West zone. The parameters were estimated using one month of data for each generators and using 100 discrete prices. G

The first three columns of the table 2 show the estimated startup costs, fixed costs, and scale of the fixed cost shock for each generator. Standard errors are shown in parenthesis below the estimates.

The third and fourth columns show the average profit per start and the average profit per 15 min interval when operating excluding any estimated costs[12]. For example, Calenergy earns on average \$1,493,934 each time it starts excluding startup costs. The next column shows that this same generator operated on average for 37 15min periods each time it started start. The next column shows the number of starts which occur in the month of data used to estimate the parameters.

The estimates of startup costs are much higher than expected. They are also not consistent with the calculated profits as the startup costs are often greater than the average profit per start. There are a number of factors which could explain this. First, the calculated marginal costs could be incorrect. Second, the firms may not be responding to balancing energy prices. 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 in effect implies that firms are not optimizing fully. Another reason the firms may not be responding to prices is that they be be participating in other ancillary markets in ERCOT such as regulation. 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. Third, the exercise of market power

[12]Profits each period are simply the price cost markup multiplied by the output of the generator in the period. The marginal costs are calculated as previously described. The output of the generator is not the observed output of the generator, but rather the output which is implied by the model(e.g. conditional on operating, the generator produces at minimum or maximum capacity depending on whether the price is above or below marginal cost)

would also have a tendency to inflate startup costs as firm would withhold production in order to increase prices in the market. However, a puzzling contradiction in the data is that generators tend to startup too early, earning negative profits for several hours before becoming profitable.

In addition, startup costs could not be estimated for two of the generators. No parameters were estimated for unit

Table 2: Preliminary Results

Unit	$START_i$	σ_i	OC_i	Average Π per start	Average Length	Average Π per period
Calenergy	\$109,665	\$10,079	\$0	\$1,493,934	726	\$2,600
Morgan Creek CT1	\$303,698	\$37,376	\$0	\$8,452	11.2	\$775
Morgan Creek CT2	\$24,089	\$2,979	\$28	\$14,713	15.75	\$1,018
Morgan Creek CT3	\$18,813	\$2,264	\$55	\$26,494	23	\$1,091
Morgan Creek CT4	\$23,184	\$2,919	\$0	\$16,242	17	\$923
Morgan Creek CT5	\$18,464	\$2,240	\$65	\$23,140	22.5	\$1,011
Morgan Creek CT6	\$17,844	\$2,236	\$83	\$21,092	20.5	\$946
Morgan Creek ST5	\$49,579	\$6,608	\$513	\$17,093	16.25	\$840
Permian Basin1	\$20,758	\$2,634	\$368	\$32,490	19	\$1,655
Permian Basin 2	\$22,387	\$2,952	\$409	\$28,376	17.1	\$1,569
Permian Basin 3	\$22,036	\$2,809	\$425	\$32,052	17.6	\$1,755
Wichita Falls	\$23,269	\$2,841	\$341	\$63,641	68	\$933

Note: average profits do not take into account estimated costs.

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, profits, and aggregate consumer surplus

Pending complete results, I have calibrated the counterfactual model with typical startup costs from the literature; these are displayed in table 3. The variance of the error term was chosen based on the preliminary estimates so far. The calibrated parameters quite coarse given that they are same for each technology without respect the individual characteristics of the each plant which can vary greatly within a generation technology. The calibration of parameters is only a temporary exercise to demonstrate the equilibrium solution. Once final parameter estimates have been obtained, they will be used to simulate the counterfactual.

Table 3: Calibrated Parameters: West Zone

Unit	$Start_i$	σ_i	OC_i
Calenergy	\$25,000	\$2,500	\$0
Graham 1	\$10,000	\$1,000	\$0
Graham 2	\$10,000	\$1,000	\$0
Morgan Creek A	\$10,000	\$1,000	\$0
Morgan Creek B	\$10,000	\$1,000	\$0
Morgan Creek C	\$10,000	\$1,000	\$0
Morgan Creek D	\$10,000	\$1,000	\$0
Morgan Creek E	\$10,000	\$1,000	\$0
Morgan Creek F	\$10,000	\$1,000	\$0
Morgan Creek 5	\$10,000	\$1,000	\$0
Morgan Creek 6	\$10,000	\$1,000	\$0
Odessa	\$25,000	\$2,500	\$0
Oklaunion	\$70,000	\$7,000	\$0
Permian Basin A	\$10,000	\$1,000	\$0
Permian Basin B	\$10,000	\$1,000	\$0
Permian Basin C	\$10,000	\$1,000	\$0
Permian Basin D	\$10,000	\$1,000	\$0
Permian Basin E	\$10,000	\$1,000	\$0
Permian Basin 5	\$10,000	\$1,000	\$0
Permian Basin 6	\$10,000	\$1,000	\$0
Sweetwater	\$25,000	\$2,500	\$0
Witchita	\$10,000	\$1,000	\$0

As proof of concept, I simulate the effect of the carbon tax on one zone of the Texas grid. In doing so, I hold wind generation and electricity transfers between this zone and other zones of the grid constant[13]. I also need to make some assumptions about the responsiveness of demand to price changes in

[13]When simulating the response of the entire grid, I will allow transfers between zone to change within the capacity of the transmission lines. However, I will need hold transfers outside the ERCOT grid constant since I do not model the decisions of operators outside of Texas

the wholesale market. In the very short run, i.e. minute to minute, the demand for electricity is perfectly inelastic. 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 responde to wholesale price changes, changes in the average wholesale price for electricity will eventually filter down to the prices consumers face. However, dynamics exist on the demand side which limit consumers' response to price changes in the medium 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.

In order to address these concerns, I simulate market outcomes using three different assumptions about demand responsiveness. First, I assume that demand is perfectly inelastic. This is equivalent to consumers not observing wholesale price increases such as before rate increases are implemented. While not a realistic assumption, highlights the ability of the supply side to reduce emissions in response to environmental regulation in the short run. Second, I assume that demand varies according to a short run supply elasticity. I assume a constant elasticity for demand which I obtain for other studies. The long literature which estimates the elasticity of demand for electricity has produced a wide range of results. However many studies identify the short 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 short 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 counter intuitive 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 several conditions are met. First and most obviously, 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. However, convergence is likely given the fact that the probability of operation, $p(a_t|P_t, I_t, s_t)$, is increasing in P_t given I_t, s_t, and the transition matrix[14].

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

[14] 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

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

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

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

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

The aggregate supply in any time period t is then simply

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

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.

As a proof of concept, I solve for the equilibrium prices under a carbon tax given my model and the calibrated structural parameters for the West zone in Texas. Using the calibrated parameters in table 3, I solve for the equilibrium price vector under a \$20/ton carbon tax and under a \$200/ton carbon tax. Each counterfactual is solved using 100 discrete prices and a discount factor of 0.9. I solve each counter factual under each of the three assumptions about the elasticity of demand, inelastic, short run elastic and long run elastic. This implies 6 counter factual results altogether. Table 4 shows the results.

The first two columns of the table show the counterfactual results with an inelastic demand curve each period under either a \$20 or a \$200 carbon tax. A \$20 carbon tax 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 are still the lowest cost producers on

Table 4: Counterfactual Results

	Inelastic		Elasticity -0.2		Elasticity -0.7	
	\$20	\$200	\$20	\$200	\$20	\$200
Δ CO_2 Emissions %	0%	-9%	-2%	-30%	-6%	-63%
Δ Avg Price	\$15.72	\$164.95	\$13.91	\$155.67	\$11.47	\$138.74
Δ Price %	20%	213%	18%	201%	15%	179%
Δ Avg Coal %	-2%	-66%	-2%	-88%	-3%	-100%
Δ Avg Gas %	3%	59%	0%	42%	-7%	-13%

the grid. Very little substitution between generators occurs; coal generated electricity decrease by about 2% while gas generation increases by 3%. At the same time, prices increase by 20% or \$15 a MWH. Average observed prices in this simulation before the tax were \$77 MWH. As I allow demand to respond to a short run price elasticity of -0.2, there is small reduction in emissions by 2% which is the result of decreased consumption. Increasing the elasticity to -0.7, results 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 win 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. Still even with a \$200 price tag on carbon emissions, aggregate emissions are reduced by only 9% due to this supply side substitution. 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 62% and coal production virtually disappears.

While these counterfactual results specific to one zone in the ERCOT market and use calibrated parameters, they still illustrate the trends to be expected from the final 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 take advantage of recent computational advances to solve each firm's dynamic problem with a discount factor that is close to one. Preliminary estimates for generators are consistent 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. As a proof of concept pending more complete estimation results, I use calibrated parameter values to simulate outcomes under counterfactual environmental policies.

I find with a short run inelastic demand curve, $20/ton price on carbon has almost no effect on carbon emissions from generators while wholesale increase by 20% in 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 $200/ton price on carbon is necessary to induce sufficient substitution towards gas generators to create 9% decrease in carbon emissions. 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.

There are several possible extensions to this research. The first extension involves modeling the generators continuous output choice with costly adjustment to output. To the extent that costly adjustment to output is an important factor in determining generator output, this model would be a more appropriate method of modeling the short run response of firms. The second more ambitious extension considers the long run effect of environmental regulation on the investment new generating capital and the retiring of old generating capital. Both extensions are the subject of the author's ongoing research.

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**.

Cullen, J. (2008), 'Measuring the environmental benefits of wind generated electricity', *Working paper* .

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.

Judd, K. & Su, C.-L. (2008), Constrained optimization approaches to estimation of structural models, Technical report.

Luo, Z.-Q., Pang, J.-S. & Ralph, D. (1996), *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK.

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

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

Trick, M. & Zin, S. (1997), A linear programming approach to solving stochastic dynamic programming, Technical report.

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.