

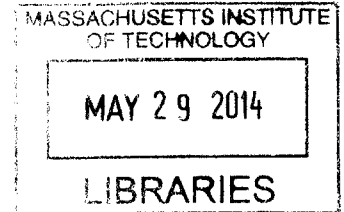
Unexpected Consequences of Demand Response:
Implications for Energy and Capacity Price Level and Volatility

by

Tal Z. Levy
B.A. in Economics
Brown University, 2006

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Technology and Policy

ARCHIVES



at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

© 2014 Massachusetts Institute of Technology. All Rights Reserved

Signature redacted

Signature of Author: _____

Engineering Systems Division
May 9, 2014

Signature redacted

Certified by: _____

Mort D. Webster
Visiting Associate Professor, Engineering Systems Division
Thesis Supervisor

Signature redacted

Accepted by: _____

Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

Unexpected Consequences of Demand Response: Implications for Energy and Capacity Price Level and Volatility

by
Tal Z. Levy

Submitted to the Engineering Systems Division on May 9, 2014
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Technology and Policy

Abstract

Historically, electricity consumption has been largely insensitive to short term spot market conditions, requiring the equating of supply and demand to occur almost exclusively through changes in production. Large scale entry of demand response, however, is rapidly changing this paradigm in the electricity market located in the mid-Atlantic region of the US, called PJM. Greater demand side participation in electricity markets is often considered a low cost alternative to generation and an important step towards decreasing the price volatility driven by inelastic demand. Recent experience in PJM, however, indicates that demand response in the form of a peaking product has the potential to increase energy price level and volatility.

Currently, emergency demand response comprises the vast majority of demand side participation in PJM. This is a peaking product dispatched infrequently and only during periods of scarcity when thermal capacity is exhausted. While emergency demand response serves as a cheaper form of peaking resource than gas turbines, it has recently contributed to increases in energy price volatility by setting price at the \$1,800/MWh price cap, substantially higher than the marginal cost of most thermal generation. Additionally, the entry of demand response into the PJM capacity market is one of primary drivers for capacity prices declining by over fifty percent.

This study investigates the large penetration of emergency demand response in PJM and the implications for the balance between energy and capacity prices and energy price volatility. A novel model is developed that dynamically simulates generation entry and exit over a long term horizon based on endogenously determined energy and capacity prices. The study finds that, while demand response leads to slight reductions in total generation cost, it shifts the bulk of capacity market revenues into the energy market and also vastly increases energy price volatility.

This transition towards an energy only market will send more accurate price signals to consumers as costs are moved out of the crudely assessed capacity charge and into the dynamic energy price. However, the greater volatility will also increase the risk faced by many market participants. The new market paradigm created by demand response will require regulators to balance the importance of sending accurate price signals to consumers against creating market conditions that decrease risk and foster investment.

Thesis Supervisor: Mort Webster

Title: Visiting Associate Professor, Engineering Systems Division

Acknowledgments

I would first like to thank my adviser and thesis supervisor, Mort Webster. Working with you has greatly expanded my understanding of power systems and modeling methodologies, which was one of the primary reasons I chose to attend MIT. Throughout my time at MIT, I have been impressed by your commitment to serving as a mentor and ensuring my research was structured to further my academic and professional goals.

I would also like to thank Ignacio Perez-Arriaga and Richard Tabors, who consistently made themselves available to discuss both my thesis results and other research. These discussions proved invaluable in helping guide and frame my work.

To my colleagues in the TPP class, I am grateful for having the opportunity to work with such an intellectually curious and dynamic group of people. The collective excitement about studying energy has been a truly great experience.

Finally, to my parents who have been a constant source of support for my 30 years. I'm thankful we've had the opportunity to live in the same city again.

Table of Contents

Abstract.....	3
Acknowledgments	4
Table of Contents.....	5
List of Figures.....	6
List of Tables	7
Chapter 1: Introduction.....	8
Chapter 2: Power Market Fundamentals with a Focus on the PJM ISO	15
2.1: US Electricity Restructuring and the Development of the PJM ISO.....	15
2.2: Generation Expansion.....	21
Chapter 3: Capacity Markets	25
3.1: Competing Justifications.....	25
3.2: Capacity Market Implementation	32
3.3: The PJM Capacity Market in Practice	35
3.4: Implications for Remuneration Shifting Between Capacity and Energy Markets.....	38
Chapter 4: Demand Response.....	42
4.1: Effects on Price	42
4.2: Value of Lost Load and Emergency Demand Response	44
4.3: Demand Response vs Price Sensitive Demand.....	47
Chapter 5: Methods and Model Description.....	51
Chapter 6: Results.....	62
6.1: Effects on Costs and Reliability.....	62
6.2: Generator Profits.....	67
6.3: Impacts on Energy Price	70
6.4: Energy Price Sensitivity.....	73
6.5: Monte Carlo Simulation.....	77
6.6: Reliability Option Analysis	79
Chapter 7: Discussion and Policy Implications.....	81
Bibliography	90
Appendix 1: Clustered Unit Commitment.....	95

List of Figures

Figure 1: Illustrative example of supply stacks with and without demand response	12
Figure 2: Schematic of operating reserve demand curve.	20
Figure 3: PJM 2013 load duration curve.	22
Figure 4: PJM capacity clearing prices in Base Residual Auction.....	37
Figure 5: Sources of demand response remuneration (Monitoring Analytics, 2013).	49
Figure 6: Modeled capacity demand curve.....	56
Figure 7: Modeled capacity clearing process.	60
Figure 8: Total generation cost (variable + fixed) and hours of unserved energy.....	63
Figure 9: Power cost to consumers by scenario expressed as a percentage of Case 1 (energy only market and no DR).	64
Figure 10: Generation profits by scenario	67
Figure 11: Power cost to consumers by year	68
Figure 12: Difference in profits between cases	69
Figure 13: Average annual price sensitivity to 1% changes in load.....	73
Figure 14: Profits of average CCGT in year 15 vs 1% changes in load.....	74
Figure 15: Modeled supply curve with and without demand response.	76
Figure 16: Mean centered energy cost to consumers from 500 iterations of the monte carlo.....	77
Figure 17: Mean centered average CCGT profits.	78
Figure 18: Distribution of energy prices from the Monte Carlo in summer months (June-August) for hours 8 – 23.....	82

List of Tables

Table 1: Components of PJM power cost (Monitoring Analytics, 2013).	10
Table 2: Penalty factor phase-in and effect on price cap (PJM, 2013b).....	21
Table 3: Variable and capital costs of the primary thermal generation technologies.....	22
Table 4: Sensitivity of capacity cost to the marginal cost of demand response.	67
Table 5: Frequency of high priced events, years 1-15 with and without the capacity market.	70
Table 6: Market response to large generation outage in July (Hours Ending 8-23).....	72
Table 7: Peak reduction due to ratio of peak to off-peak prices.....	83

Chapter 1: Introduction

There are several aspects of the power sector that make electricity a unique commodity. Electricity cannot be stored, and hence supply must instantaneously equal demand in order to maintain system operations. Other commodities, such as oil or natural gas can be stored allowing for short term deviations in production and consumption. If these imbalances occur on the electricity grid, however, they will result in rolling blackouts or complete grid failure. Additionally, short term electricity consumption is largely insensitive to price, forcing the equating of supply and demand to be almost entirely achieved through adjustments in production (Bushnell, Hobbs, & Wolak, 2009).

The paradigm of a completely inflexible demand side is rapidly changing in the electricity market in the mid-Atlantic region, called PJM. In the 2014/2015 and 2015/2016 planning years, demand response (DR) is contracted to provide 9.4% and 9.0% of total system capacity, although that value drops to 7.3% for the 2016/2017 planning year. Demand response has gained widespread support as an economical and environmentally friendly form of peaking capacity to help independent system operators (ISOs) manage peak load. The Federal Energy Regulatory Commission (FERC), through its oversight of US wholesale electricity markets, has consistently supported policies that favor demand response (Bushnell et al., 2009). Furthermore, within the academic literature, there is general acceptance that increased demand side participation will lower energy price levels and dampen the endemic volatility driven by inelastic demand (Albadi & El-Saadany, 2007; Borenstein, Jaske, & Rosenfeld, 2002; Centolella, 2010).

Recent events in PJM, however, have indicated that demand response has the potential to actually increase energy price volatility. After one such event, the ISO in a public response to

questions stated that demand response set price to the \$1,800/MWh price cap and this would have likely not occurred had thermal generation been the marginal resource. While demand response may be a cheaper form peaking technology than a gas turbine, it increases energy price volatility through a complex set of interactions between generator fixed and marginal costs and clearing prices in the energy and capacity markets.

Historically, power generation, transmission, and distribution were all provided by vertically integrated utilities that managed generation investment through their planning processes as opposed to market mechanisms. The vertically integrated utility determined the appropriate level of generation supply required to maintain reliability levels, and recovered its cost through tariffed rates. Reliability in this context refers to the ability of the generation supply to serve non-price sensitive load (security is the term used to refer to the grid's ability to maintain operation through exposure to unexpected shocks such as equipment failure). In other words, reliability measures the likelihood of avoiding non-price based demand rationing.

Electricity restructuring began in the US in the mid-1990s, and transitioned the power generation sector from a regulated monopoly to a competitive market. The management of the electricity grid and long term reliability planning was transferred from utilities to non-profit ISOs. PJM was the first US based ISO and initially served Pennsylvania, New Jersey, and Maryland. It has since grown to encompass most of the mid-Atlantic and some parts of the Midwest, and is currently the largest US ISO in terms of peak load.

Restructuring transferred the generation investment decision from vertically integrated utilities with regulated cost recovery to private actors that must recover costs from the various power market revenue streams. Early on during the process of electricity market restructuring

regulators realized that the revenue solely from the sale of electricity was inadequate to incentivize the level of installed capacity required to meet the NERC reliability standard of .1 day of unserved energy per year (Rodilla & Batlle, 2013).

In order to maintain reliability standards, ISOs developed capacity mechanisms as an additional form of remuneration to generators. The PJM capacity market functions as an auction in which generators and demand response bid to provide capacity. The auction clears based on these bids; resources' marginal costs and hence their impact on the energy market are not considered. The ISO procures adequate generation or demand response (both are considered capacity) such that, in expectation, load will only surpass the available capacity 2.4 hours per year.

In PJM, capacity charges have historically comprised a large portion of the total cost of wholesale power. The power cost refers to the total cost consumers pay for purchasing electricity off of the bulk power system, with the two largest components generally energy and capacity. Table 1 shows that capacity charges reached a high of 20% of total power costs in 2009, and steadily declined to 12% by 2012.

Table 1: Components of PJM power cost (Monitoring Analytics, 2013).

	2009	2010	2011	2012
energy	70%	72%	73%	73%
capacity	20%	18%	16%	12%
transmission	7%	6%	7%	10%
other	3%	4%	4%	5%

Gas turbines (GT) are the thermal generator technology used to meet peak demand and their fixed and marginal costs have a strong influence on the split between energy and capacity costs.

Gas turbines are peaking units because they have higher marginal costs, but lower fixed costs, than other types of generators. The marginal cost of a gas turbine is roughly \$70/MWh, but the annualized fixed costs are roughly \$72,000/MW (US Energy Information Administration, 2010). If a peaking unit gas turbine only runs for 10 hours per year, its average cost of production is \$7,270/MWh. In order for the gas turbine to recover its costs, capacity payments must make up the difference between the energy price, which is currently capped at \$1,800/MWh, and the average cost of \$7,270/MWh.

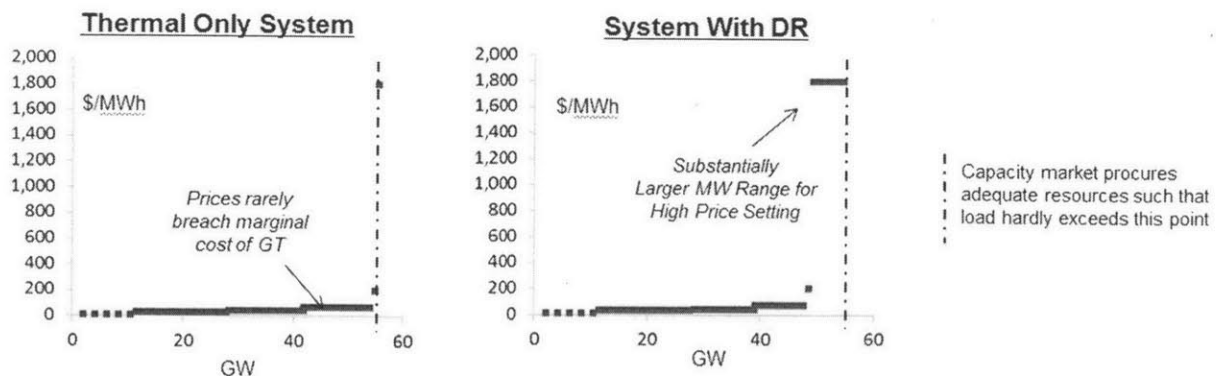
The fixed costs of demand response are likely trivial compared to generators. The cost of smart meters has been reduced substantially in recent years and demand response does not require the turbines, pollution controls, and natural gas and electrical interconnection infrastructure of a new generator. However, electricity is a key input into most modern economic activities and thus the marginal cost of curtailing electricity usage would be expected to be quite high. Estimates of the value of lost load (VOLL) tend to be in the \$2,000 - \$15,000/MWh range. A demand response participant with a VOLL of \$2,000/MWh and trivial fixed costs can provide peaking capacity far cheaper than a gas turbine that only runs for 10 hours per year, which has a cost of \$7,270/MWh. Despite having much lower total costs than the gas turbine, demand response will bid into the energy market at its marginal cost (which is likely bounded by the \$1,800/MWh price cap). Demand response, however, needs to recoup substantially less in the capacity markets and this is likely why capacity prices have been steadily declining in PJM.

This thesis examines how greater demand response penetration will alter the balance between energy and capacity prices, and the resulting implications for energy price volatility. While many types of demand response exist, this thesis focuses on the emergency demand response product that to date has constituted the vast majority of demand side participation in

PJM in capacity terms. This emergency product is dispatched during peak scarcity conditions and is likely bidding in at the price cap. Previous studies that concluded that increased demand side participation will lower energy price level and volatility tended to examine only the energy market in a steady state, assessing the impacts of increased demand side participation to a static generation mix. The balance between energy and capacity prices or the effects of generation entry and exit over time were not considered.

A dynamic perspective suggests that, over time, as demand response enters the market in greater quantities it will displace new entry by gas turbines, which are required to bid into the market close to their marginal cost of ~\$70/MWh. Demand response, in contrast, is allowed to bid into the market at the price cap, which is likely still below its marginal curtailment cost based on VOLL estimates. Thus, if a dynamic view were considered, over time we would expect demand response to replace gas turbines as the marginal resource during high load periods, replacing price setting at the \$70/MWh range with prices at the cap. Figure 1 demonstrates this dynamic.

Figure 1: Illustrative example of supply stacks with and without demand response



Assuming a competitive framework, this increase in energy market revenue will be matched by a decrease in capacity market revenue. While we would expect demand response to decrease total

power costs (energy + capacity), demand response could still raise energy prices by shifting revenue out of the capacity market and into the energy market.

This thesis extends previous studies by incorporating demand response into a generation expansion model that includes a capacity market based on the PJM mechanism. Based on a broad survey of the literature, this is the first study to quantitatively model how greater demand response may affect energy and capacity prices over a long term horizon in which the supply side can adjust to market prices. The generation expansion model is run for 3 cases. 1: only thermal generation entry (no DR entry) and no capacity market, 2: only thermal generation entry (no DR entry) combined with a capacity market, and 3: DR is allowed to compete with thermal generators for entry combined with a capacity market.

The results show that introducing a capacity mechanism to enforce NERC reliability standards in a traditional thermal system raises costs by about 1%, but shifts 24% of the energy market revenue into the capacity market. The entry of demand response shifts the vast majority of capacity payments back into the energy market reducing capacity to 4% of total power cost. This shift will provide more accurate price signals to consumers, as a substantial fraction of the power cost is moved away from a crudely assessed capacity charge and into the dynamic location marginal price (LMP). Additionally, this shift will encourage greater demand side participation from smaller customers that have the potential to adjust consumption in response to real time prices conveyed by a smart meter, but lack the sophistication to participate in the capacity markets.

The downside of this shift towards an energy only market is a vast increase in the risk to both consumers and generators. Capacity payments in PJM represent fixed payments set three

years into the future. Replacing these fixed payments with uncertain energy prices will in and of itself raise cash flow volatility. Additionally, changing the marginal cost of a peaking resource from that of a gas turbine to the price cap makes the energy price substantially more sensitive to changes in load and equipment failures such as generation outages. This sensitivity is large, with the potential for the dynamic range in both cost to load and generator profit to increase by three times over what is experienced in a traditional thermal system with a capacity market.

Furthermore, the results demonstrate that combining a well-functioning capacity mechanism with a primarily thermal system provides electricity markets with two services. First, it guarantees a certain level of reliability by incentivizing additional generation. Second, increasing the size of the supply stack both lowers and decreases the volatility of energy prices. The combination of known capacity payments with lower and dampened energy prices provides consumers a level of price protection and also ensures generators a more stable revenue stream. The entry of demand response will not affect the capacity market's ability to provide reliability, but it will vastly decrease its price depressing and dampening effect. In essence, demand response will lead to the reliability product and the financial hedging product no longer being intrinsically linked. This raises interesting public policy questions, because the market failures underlying the need for capacity mechanisms to support reliability are largely universally accepted. However, the need for capacity mechanisms to serve as financial hedges is substantially more controversial as there are not clear reasons for why electricity markets left to their own devices cannot provide the appropriate risk management products. Nevertheless, the power generation industry has repeatedly emphasized the importance of capacity payments in providing increased revenue predictability and lowering the cost of capital.

Chapter 2 provides background on US electricity industry restructuring, the price formation mechanisms currently prevailing in PJM, and the theory behind generation expansion in a liberalized market. Chapter 3 reviews the academic literature on capacity markets, and the tension between academic support based on the market failure caused by inelastic demand vs industry support premised on fixed guaranteed revenue streams. The chapter concludes with a discussion of the implications for costs shifting between the capacity and energy markets. Previous research on demand response is reviewed in chapter 4. Subsections of this chapter discuss how historical estimates of the VOLL are much higher than the price caps and the potential interactions of demand response products that participate in the capacity market vs elasticity in the demand curve. Chapter 5 describes the methods and model used to simulate generation entry and exit into the PJM market. The results are presented in Chapter 6, with subsections focusing on the impacts of demand response on costs and reliability, generator profits, energy prices, and volatility and risk. Chapter 7 discusses the policy implications and conclusions.

Chapter 2: Power Market Fundamentals with a Focus on the PJM ISO

2.1: US Electricity Restructuring and the Development of the PJM ISO

The electricity sector in many parts of the US has undergone profound change since the start of restructuring in the mid-1990s. During this period, many electricity markets transitioned from vertically integrated monopolies with average cost of service based prices to competitive wholesale power markets with transparent prices based on the clearing of the bids and offers of market participants.

The electricity industry can be divided into generation, transmission, and distribution. The power produced by generators is transported across high voltage transmission lines (generally defined as 69kV and above) to where it is needed. The voltage is then stepped down substantially and delivered over the distribution network to consumers. In the US, these three sectors were traditionally considered natural monopolies and were provided by utilities operating as regulated monopolies. Transmission and distribution represents a natural monopoly as competition would require building a redundant infrastructure. Currently these services are still provided almost entirely by regulated monopolies, although limited merchant transmission has been built. Generation was also considered a natural monopoly due to the economy of scales in power plants and the losses involved in transporting power long distances. By the 1970s and 1980s, however, decreases in generator size compared to total load and loss decreases due to improved transmission made these justifications less applicable. (Borenstein & Bushnell, 2000).

In 1978 Congress passed the Public Utility Regulatory Policy Act (PURPA), which required utilities to purchase power from “qualifying facilities” owned by independent power producers when the price was lower than the avoided cost of the utility producing the power itself. Qualified facilities were generally comprised of cogeneration or renewable resources, including refuse burning generation (PL Joskow, Bohi, & Gollop, 1989). The avoided cost calculations were left to the state regulatory agencies (PL Joskow et al., 1989), and in many cases states set prices that were likely much higher than the marginal savings (Borenstein & Bushnell, 2000). The combination of high cost PURPA contracts and cost overruns in nuclear plants led some states in the 1990s to have electricity prices much higher than the going-forward cost of new coal or gas fired plants. Other states that followed more cautious investment approaches were able to hold their electricity prices to substantially lower levels. This price

differential was one of the determining factors in which states decided to implement deregulation, with the higher cost states hoping that market forces could lead to more prudent investments (Borenstein & Bushnell, 2000).

FERC orders 888 and 889, issued in 1996, laid the groundwork for the development of Independent System Operators (ISOs) by requiring

transmission owners to provide access to their networks at cost-based prices, to end discriminatory practices against unaffiliated generators and marketers, to expand their transmission networks if they did not have the capacity to accommodate requests for transmission service, and to provide nondiscriminatory access to information required by third parties to make effective use of their networks. (P Joskow, 2006)

In 1997 PJM became America's first "fully functional independent system operator" (PJM, PJM History). As an ISO, PJM is responsible for managing the operations of the bulk power system. ISOs are non-profit organizations that do not own any generation or transmission assets, but they are responsible for both managing the day to day market operations and long term planning processes (PJM, Who is PJM).

The PJM Energy market is designed to match supply bids from generators with demand bids from load while maintaining grid security. PJM, in a fashion similar to most US ISOs, uses a two settlement system. In the Day Ahead market, prices are calculated for each hour of the next operating day using least cost security constraint unit commitment. In essence, a large optimization is minimizing the as bid cost of generation required to serve load, subject to operating constraints. These operating constraints are primarily the physical parameters of the generators (minimum and maximum output levels, ramping rates, minimum offline and online times, and startup time) and the line and interface capacities of the transmission system. The Day Ahead Market is required in order to give generators adequate notice to adjust their output to the required levels, and serves as the primary form of remuneration for generators. The real time

market serves more as a balancing market to instantaneously match load and generation throughout the ISO. The real time market relies on security constrained economic dispatch and LMPs are calculated at 5 minute intervals.

A key feature of PJM and other ISOs is the use of a nodal system with the LMP calculated for hundreds or thousands of nodes across the footprint. The LMP, outside of scarcity conditions, represents the marginal cost of consuming an additional unit of power at that location. Absent congestion and loss, the LMP would be the same across the entire ISO. Losses tend to be a small part of the total LMP, but congestion in the transmission lines can lead to substantial price differentials. One of the primary arguments for transitioning to a more market based system was that the LMP provides both accurate short term consumption and production signals, and under several assumptions, accurate long term signals as well.

Early on in the history of restructuring, the California energy crises highlighted electricity market's susceptibility to market manipulation. During peak load periods, almost every generator is required to serve demand. Most power generation companies during these peak periods are considered 'pivotal' in that their output is required for the supply curve to cross the demand curve. Thus, even smaller power generation companies can have a reasonable expectation of market power, either through reducing quantity or raising their bid price. In other markets, storage and demand elasticity limit the ability of suppliers to exert market power, but these factors are not present in electricity (Borenstein & Bushnell, 2000).

Most US ISOs, including PJM, have taken two approaches to managing market power. The first is the establishment of market monitoring and market power mitigation protocols. PJM monitors the bids submitted by generators, and if they diverge too far from accepted normal

levels, PJM will literally adjust their bid to what the ISO considers appropriate based on the plant's marginal cost. Power generation is fairly unique in that the marginal cost is well defined by the generator's heat rate and fuel price. Thus, generators cannot alter their bids substantially without drawing the attention and risking enforcement action from the market monitor.

The second approach to mitigating market power is the imposition of price caps (Borenstein & Bushnell, 2000). Price caps are a somewhat crude tool that limits the ability of generators to send prices above exorbitant levels. However, given that the marginal cost of a thermal generator is normally under a few hundred dollars per MWh compared to US price caps that are now generally above \$1,000/MWh, the market power mitigation measures generally activate well before the price cap binds. This is not to say that price caps can be abolished. In some situations demand can exceed supply and the market will not clear at any price. In these situations, a price cap is required to stop prices from going to infinity and simply being undefined (Cramton, Ockenfels, & Stoft, 2013).

PJM is currently in the process of phasing in new shortage pricing rules based on a demand curve for operating reserves. Historically, if the system was experiencing shortages of operating reserves it would take out of market emergency actions such as voltage reductions, manual load dumps, and dispatch of emergency demand resources. These actions reduce load in rather crude increments and had the unintended consequence of substantially depressing the energy price during emergency conditions. PJM also noted that due to quirks in the non-simultaneous nature of the dispatch of the energy and reserve markets, operating reserve shortages could occur while the reserve markets were clearing at \$0. Implementing a demand curve for operating reserves and simultaneously clearing the energy and reserve markets solves most of these issues (Federal Energy Regulatory Commission, 2012).

An operating reserve demand curve is built on the premise that decreasing the operating reserve margin increases the likelihood that the system operator will be forced to shed load.

Figure 2: Schematic of operating reserve demand curve.

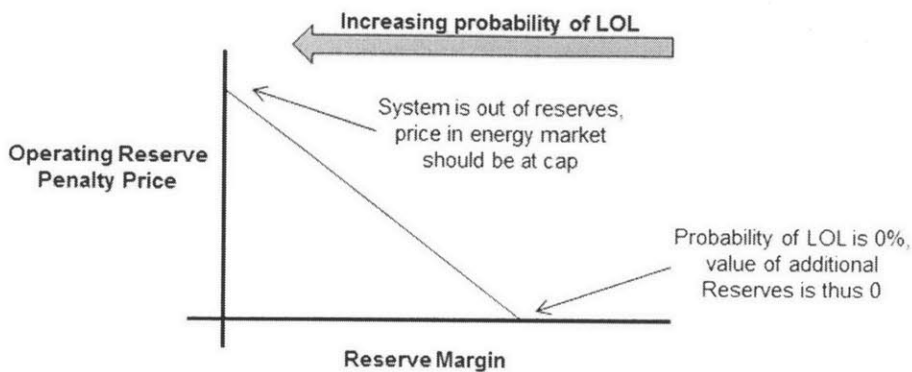


Figure 2 shows that at an adequate operating reserve margin the probability of a loss of load occurring is essentially zero. The Y axis is referred to as the operating reserve penalty price. The energy price is the result of the bid price of the marginal generator plus the penalty price. Hence, when adequate reserves are available (which is normally the case), the system energy price is set by the marginal generator. However, if a shortage of reserves starts to occur, the energy price should be increased by a reserve penalty factor to account for the fact that consuming an incremental MW of power decreases available reserves, making a loss of load (LOL) event more likely. Once reserves are completely exhausted, an increase in 1MW of consumption will be causing someone else to not consumer that MW of power, and hence the energy price should be at the cap. In PJM, the maximum offer price for generators is \$1,000/MWh. The price cap is currently at \$1,800/MWh. Thus, the maximum penalty that operating reserves can add to the energy price is \$800/MWh. Note that PJM is phasing in this shortage pricing protocol by

increasing the price cap and hence maximum reserve shortage penalty. Table 2 summarizes the phase in schedule.

Table 2: Penalty factor phase-in and effect on price cap (PJM, 2013b).

Dates	Generator Offer Cap (\$/MWh)	Price Cap (\$/MWh)	Max Operating Reserve Penalty (\$/MWh)
Oct 2012 - May 2013	1,000	1,500	500
Jun 2013 - May 2014	1,000	1,800	800
Jun 2014 - May 2015	1,000	2,100	1,100
Jun 2015 onward	1,000	2,700	1,700

2.2: Generation Expansion

Generators can be roughly classified into base-load, mid-load, and peakers. The key differentiating feature is the relationship between fixed and marginal costs. Baseload generators have high fixed costs but low variable costs while peaking plants have low fixed costs but high variable costs. In order for base load plants to be economically viable, they must run for a substantial portion of the year so that the high fixed costs are amortized over a large number of production hours. The reverse is true for peakers in that they shouldn't be run for too many hours per year, or else the high operating costs will overwhelm the savings from the low fixed costs (Ventosa, Linares, & Pérez-arriaga, 2013).

Figure 3 shows the PJM load duration curve for calendar year 2013, with illustrative boxes showing roughly the portions of the load curve served by each type of generation. In this case the few peak hours are not served by a generator and are hence referred to as unserved energy. Given that demand changes substantially, both between the hours of the day and seasonally, the optimal generation mix is some combination of baseload, mid-load, peaking units, and unserved energy (Ventosa et al., 2013).

Figure 3: PJM 2013 load duration curve.

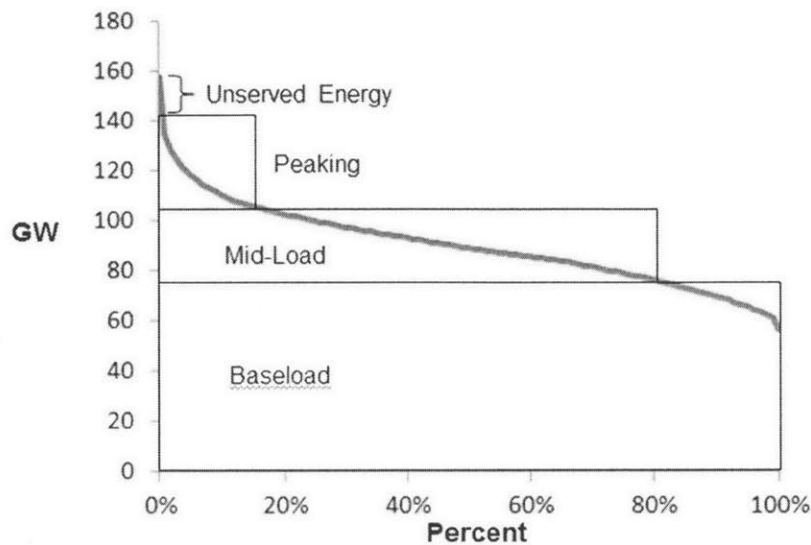


Table 3 shows the variable and fixed costs for the primary thermal generation types. The fixed costs assume a 15 year payback period.

Table 3: Variable and capital costs of the primary thermal generation technologies.

gen_type	Variable Cost (\$/MWh)	Capital Cost (\$/MW-yr)
Nuclear	10.0	444,683
Coal	17.5	219,270
CCGT	38.7	94,390
GT	69.0	71,913

Comparing coal to a gas turbine (GT), one can see that coal's fixed costs are over 3 times that of a GT, but its variable cost is also over three times lower. Computing an average cost of production for the two technologies if they both run for the entire year yields an average cost of coal of \$42.5/MWh compared to \$77.2/MWh for a GT plant. However, if these two technologies

only run for 100 hours per year the GT becomes the substantially cheaper technology with an average cost of \$788/MWh vs \$2,210/MWh for coal.

An additional takeaway from this exercise is that it becomes extremely expensive to serve peak load. The average cost of a GT that only runs for 50 hours per year is \$1,440/MWh and this number rises to \$7,270/MWh for only 10 hours of operation. As the number of hours of operation decrease, the fixed costs of a GT become by far the dominant component, and the average cost of production becomes extremely high. While some end users such as hospitals may be willing to pay for electricity at almost any price, the cost of curtailing load during peak events is often less than cost of providing that energy from a GT. Hence, it is generally most efficient to have at least some hours of unserved energy (Ventosa et al., 2013). Note that the North American Reliability Corporation (NERC) reliability standard, which US ISOs are bound by tariff to maintain, is an average of 2.4 hours of unserved energy per year. A gas turbine running for only 2.4 hours per year would have an average production cost of \$30,033/MWh. Since price caps are generally below \$2,000/MWh in the US, it is impossible to maintain the NERC reliability standard without a capacity mechanism.

Demand response can be considered a new generation resource that sits at the extreme end of the fixed cost vs marginal cost spectrum. The fixed costs for demand response are very low. Instead of requiring a gas turbine and the other physical infrastructure needed to support a peaking plant, demand response primarily needs a smart meter capable of metering real time load. However, the industrial and commercial customers that currently dominate the demand response space use electricity as a key input in their economic activity. Hence curtailing consumption often has a very high marginal cost. The phenomenon of extremely high average

production costs to serve peak load from gas turbines explains why demand response may be a cheap peaking resource even if it has marginal costs much higher than a GT.

Assuming a single node system (essentially ignoring transmission constraints) the system price is set by the bid price of the marginal generator. Adding in the assumption of perfect competition leads to the outcome of generators bidding at or very close to their marginal cost. Given the costs in the table above, if a coal plant is marginal the system price will be \$17.5/MWh, if a CCGT is marginal the system price will be \$38.7/MWh, etc. If a generator is only remunerated through energy payments, it must recover its fixed costs through inframarginal rents. In this simple stylized system of only thermal generation, GTs will always be the peaking thermal resource. If peak load does not exceed generation capacity, the system price will never rise above the marginal cost of a GT and these units will be unable to recover their fixed costs. Hence, in the absence of a capacity market, some hours of scarcity are required in which demand exceeds supply and prices are set to an administratively high level in order for peaking units to recover their fixed costs. The current PJM price cap of \$1,800/MWh implies that roughly 40 hours at the price cap are required for a peaking unit to recover its costs. In practice, this is far more hours than actually occurs and maintaining adequate revenue sufficiency has been a constant concern throughout the history of US restructured electricity markets.

Prior to electricity industry restructuring, revenue sufficiency was ostensibly guaranteed by the regulator. Utilities were tasked with determining the combination of baseload, mid-load, and peaking generation that would serve expected load levels at least cost. Barring regulatory risk, a reasonably performing utility would recover its costs from its rate base; the issue of market prices incentivizing adequate generation was largely irrelevant.

In an idealized system with several strong assumptions, Perez-Arriaga and Meseguer (1997) show that a marginal cost based energy price would lead the competitive market to produce the same generation mix as a central planner minimizing costs (Pérez-Arriaga & Meseguer, 1997). A corollary to this is that the “energy market price is all that is needed to remunerate the generators if the regulator is not seeking any specific objectives” (Rodilla & Batlle, 2013). The necessary assumptions are:

- 1: Short term spot price always reflects demand side marginal utility.
- 2: Risk neutrality of market participants.
- 3: Production cost function is convex.
- 4: Economies of scale and lumpy investments do not exist.
- 5: Perfect competition.

In practice, most if not all of these assumptions fail to fully hold. Of particular importance is that short term spot prices during scarcity events are often administratively set to a price cap. In the scarcity hours required to remunerate peaking generators, prices by definition do not reflect demand side marginal utility. Additionally, regulators are often imposing “specific objectives,” including reliability requirements. In the US, most ISOs are tasked with maintaining the 1 day in 10 year standard of loss of firm load (North American Reliability Council, 2012). This standard is a holdover from the days of vertical integration and there is little reason that a market based approach would produce the same reliability standard. Hence, if the regulator aims to combine a market based system for wholesale generation while maintaining NERC reliability standards, an additional regulatory mechanism is needed.

Chapter 3: Capacity Markets

3.1: Competing Justifications

Capacity mechanisms are a feature fairly unique to electricity markets. In other industries, the total production capacity is determined by private actors investing based on expectations of future demand and prices. If production capacity becomes scarce, prices will rise incentivizing new entry. Nevertheless, many electricity markets have determined that the minimum acceptable level of installed capacity should be determined by regulators, as opposed to market forces. To maintain these production levels, deregulated markets have introduced various forms of capacity mechanisms to raise compensation to generators in order to incentivize investment above what an energy only market would create.

The need for capacity mechanisms to maintain reliability has gained widespread acceptance in the academic literature based on the identification of several market failures. In an idealized system, capacity markets would not be necessary. In times of scarcity, prices will send a short term signal to consume less and generate more. Even if all generation supply is dispatched, the market can still clear with the demand curve setting price. Consumer's whose VOLL is below the electricity price will stop purchasing power until supply equals demand. Electricity demand, however, is generally on a tariffed rate and is not receiving the real time price of power. This makes electricity consumption extremely inelastic to short term price movements. Hence, during periods of scarcity, the real time price does not ration demand and rolling blackouts occur. Given the lack of price sensitive demand, the market will not clear at any price when demand exceeds supply and prices must be set administratively to the price cap (Cramton et al., 2013).

This need to administratively set prices when demand exceeds supply makes it impossible for electricity markets to optimize the frequency of rolling blackouts. The frequency of rolling blackouts depends on the level of installed capacity. However, peaking units depend on

the small number of very high priced hours during scarcity conditions to recover their fixed costs. Given that the price of energy during these scarcity conditions is set administratively, it is simply impossible for the market to optimize the duration of blackouts (Cramton et al., 2013).

In the absence of market mechanisms for determining reliability levels, US electricity markets utilize the NERC reliability standard of 0.1 day of unserved energy per year (North American Reliability Council, 2012). If a gas turbine only runs for the 2.4 hours per year (the equivalent of 0.1 day per year), its average cost of production is approximately \$30,000/MWh. This is much higher than most VOLL estimates, but many of the US ISO's including PJM are still bound by their tariff to meet the NERC standard. In the US, however, price caps are far too low for energy revenues alone to incentivize the level of investment required to meet this stringent reliability standard.

Raising price caps from their current levels in the \$1,000/MWh - \$2,000/MWh range to the \$2,000/MWh - \$15,000/MWh range more commensurate with VOLL estimates would raise generator energy market remuneration and also likely improve overall efficiency. Nevertheless, raising price caps to levels more consistent with the VOLL will not be a perfect solution. The VOLL is generally difficult to compute (Paul Joskow & Tirole, 2007), and at best it seems to be computed as an average value for various customer classes and locations. Within a customer class consumers will have a diversity of VOLLs, and the marginal value should be the one setting price when supply is exhausted.

Joskow (2008) and Joskow and Tirole (2007) identify additional market failures that further support the need for capacity mechanisms. First, scarcity conditions increase the likelihood of network collapse. When the electricity grid fails the electricity market ceases to

exist and generators and demand are unable to transact even though transactions exist that would increase surplus to both parties. An electricity price that does not include this risk of grid failure will not lead to an efficient level of investment. Second, they note that ISO operators often intervene with out-of-market actions in emergency situations. These actions may be needed to maintain grid safety but they can have depressing effects on price, which will again not provide the proper investment signals to achieve efficient levels of reliability.

While the market failure justification for capacity markets outlined above has gained widespread acceptance, a second more controversial justification centers on the need to provide stable revenue streams to generators. Energy market revenues tend to be highly volatile with extremely hot summer or cold winter temperatures driving substantial portions of generator remuneration. Power plants are capital intensive assets, however, and debt obligations must be serviced regardless of the weather realized. Capacity markets, by smoothing the generator's revenue stream, decrease the financial risk a generator faces and in doing so lowers its cost of capital.

The power generation industry has shown strong support for capacity mechanisms due their risk reduction impacts. For instance, the president of NRG's East Region, in describing how the PJM capacity market supported the upgrade of an existing facility in PJM stated that "each of these elements was critical to providing a level of price stability necessary to make this multi-year investment" (Davis, 2013). Similarly, AEP in comments to FERC in support of a PJM tariff change argues that "PJM's proposal will appropriately restore the original design feature of PJM's capacity market, ... the purpose of which was to reduce price volatility and increase the stability of the capacity revenue stream over time" (Federal Energy Regulatory Commission, 2014). NSTAR, with the perspective of both a utility and a generation owner writes that capacity

markets assure “investors that arrangements exist for providing long-term capacity revenues, thus providing an attractive environment for investment in needed generation and other resources, while relieving electric customers of at least some of the associated risks” (Holodak, 2013). AEP vice president and CFO Brian Tierney took an even stronger stance in an investor call arguing in response to prices hitting the \$1,800/MWh cap in PJM in January 2014 that “do you think the people being paid those prices are being paid for energy? They're not ... We need to start getting some of those capacity rents back in the capacity auction” (Testa, 2014).

In the US, available evidence points towards power plants being unable to gain required financing without some sort of a guaranteed revenue stream. A study by the American Public Power Association was able to compile data on 60% of the new power plants built in 2011 and found that 98% of the studied capacity was built either with a long term power purchase agreement with a utility or a simply by a vertically integrated utility. Only 2% of the capacity studied was built primarily for spot market sales (American Public Power Association, 2012).

Despite the power industry’s emphasis on the need for predictable revenue streams to lower capital costs, the academic literature tends to disagree on this topic. Paul Simshauser, who interestingly has both academic and industry appointments writes that “resource adequacy in the utilities industry is a well understood concept. Capital adequacy is not. ... Ignoring the importance of capital markets risks overlooking one of the most fundamental drivers of investment and price in the utilities industry” (Simshauser, 2009). Given that utilities are one of the most capital intensive industries in the world (Simshauser, 2009), electricity market design issues that impact cost of capital should merit consideration. Some of the seminal papers by Joskow, Tirole, and Cramton’s more recent articles that justify capacity markets through the

identification of market failures do not address how capacity markets affect risk and financing costs.

Neuhoff and De Vries (2004) mathematically show that if generators are only remunerated through the spot market, the combination of risk averse investors and consumers will lead to inefficient generation investment decisions. Specifically, “the uncovered price risk increases financing costs, reduces equilibrium investment levels, distorts technology choice towards less capital-intensive generation and reduces consumer utility” (Neuhoff & Vries, 2004). Allowing unlimited forward contracting between power plants and consumers eliminates this problem, but they cite analysis in Woo et al. (2003) showing that trading of long term contracts in restructured power markets tends to be thin.

Hobbs et al. (2007), in their evaluation of several capacity market demand curves proposed by PJM, conclude capacity mechanisms that lower revenue volatility to power producers will decrease capital costs and overall energy prices. This conclusion is fairly similar to the one reached by Neuhoff and De Vries (2004), in that risk averse agents prefer more stable revenue streams. In the Hobbs et al. (2007) study long term capacity payments are essentially substituting for forward contracting. The model they developed found that capacity mechanisms that create more stable revenue for generators could both increase reliability while decreasing capacity and scarcity costs by close to 50% compared to capacity mechanisms with higher volatility levels. This finding is important in that it emphasizes that capacity market design should potentially focus not only on procuring a certain capacity product, but also on providing revenue stability to generation.

Cramton and Stoft (2006), however, argue that a lack of forward contracting is not intrinsically an impediment to investment in new generation. They contend that “long-term investments are rarely financed on the basis of long-term product sales; and long-term investments are often made in industries that sell almost all of their product only a month or two in advance” (Cramton & Stoft, 2006). For instance, hotels are built before the rooms are booked, car manufacturers build plants without long term contracts for cars, and chip manufacturers build fabrication plants before chips are even designed, much less sold. According to Cramton and Stoft (2006), the expectation of making profitable sales, as opposed to near guaranteed profitable sales through long term contracts, is what drives investment in many industries. The key factor that separates investments in power generation from other industries such as hotels and automobiles is higher levels of regulatory risk. Forward contracts or long term capacity markets can counteract this regulatory risk, facilitating investment (Cramton & Stoft, 2006).

Similarly, Oren (2003) argues that regulatory intervention is one of the most important risk factors facing generation. He concludes, however, that loan guarantees by the regulator would be a more efficient mechanism than capacity payments to spur investment because it would be a signal of the “regulator’s commitment to uphold free market principles.” While Oren (2003) and Cramton and Stoft (2006) both conclude that the power generation industry faces large regulatory risks, they do not cite the volatility of the energy markets in and of themselves as impediments to investment. This implies that the primary purpose of capacity mechanisms is to maintain reliability levels and not lower risk posed by volatile energy prices.

The conclusions reached by Cramton and Stoft (2006) and Oren (2003), however, are asserted without considering how revenue volatility may affect the returns a potential investor would demand. Cramton and Stoft’s (2006) identification of industries that invest in new

capacity in the absence of long term contracts raises interesting questions, but more information on the financing structures and level of revenue volatility would be required to fully assess whether the case of hotels and automobile investment is directly comparable to generation expansion. For instance, a new peaking unit may hardly operate during a mild year and if neither a capacity mechanism nor long term contract existed its revenue would be close to zero. A hotel likely does not face the same revenue volatility. Neither of the two articles meets head on the issue that risk averse investors will accept lower profits in return for more stable revenue streams and that capacity mechanisms have a strong stabilizing influence.

3.2: Capacity Market Implementation

Capacity mechanisms take several forms and can be split between price based or quantity based mechanisms. If a price mechanism is employed, the system operator provides an additional administratively determined payment to generators. This requires the regulator to determine the size of the capacity payment necessary to achieve reliability targets. If the capacity payment is too large, a surplus of generation will be built; too small a payment will lead to inadequate capacity to maintain reliability targets. These types of mechanisms have been more popular in Europe and Central and South America than the US. The California ISO (CAISO) is the only ISO in the US that employs a price based mechanism that administratively sets the capacity payments to the generators.

Quantity mechanisms set a target level of capacity required to achieve reliability goals and use some form of a market based instrument to set the price for the reliability product. The ISOs in the eastern portion of the US were some of the first system operators to implement this approach. The benefit of quantity based approaches is that the regulator is not constantly

guessing at the level of payment required to incentivize adequate generation. Instead, the regulator sets the quantity of capacity to procure, and the price is set through a competitive framework in which generators bid to provide the capacity product. Thus, the eastern US ISOs refer to their capacity mechanisms as capacity markets.

Several important factors are required in order for quantity based capacity markets to work effectively. First, the capacity product must be well defined. In essence, the product is a commitment for generators to be available to produce power when the ISO needs the generation resource to maintain system reliability. Paying a generator for its installed capacity provides little benefit to system reliability if the generator is outaged during peak periods. Electricity system operators and regulators were surprisingly slow in officially recognizing the need for formal definitions of capacity and strong enforcement mechanisms to incentivize generator availability during peak periods. For instance, many of the capacity mechanisms in the restructured markets of the eastern US initially simply based payments on installed capacity (ICAP). Recognizing installed capacity was not the same as available capacity, ISOs moved towards payments based on unforced capacity (UCAP), essentially the capacity of the plant discounted for the plant's rate of forced outages (Rodilla & Batlle, 2013).

Additionally, in a quantity mechanism, the price is set through some sort of market like mechanism where generators bid to provide the capacity product. Thus, effective market power mitigation measures must be in place in order to ensure that the price is set in a competitive manner. This is a major issue in quantity based capacity mechanisms as generally most, if not all, of the installed capacity is required to meet expected peak load (Batlle, Vázquez, Rivier, & Pérez-Arriaga, 2007). Thus, the majority of larger market participants can expect to have price influence.

These market mitigation measures often include setting bid caps and floors at what the ISO determines to be reasonable max and min values. While bid limits prevent prices from skyrocketing (or plummeting), in many instances they do not ensure a competitive framework. For instance, the Boston zone in ISO-NE (NEMA) for a long time had a surplus of capacity with the market clearing at the price floor. However a few large generation retirements led to a capacity shortage in the zone. Given the difficulties of building new generation in the Boston area, the Salem Power Redevelopment Project was the only new resource available to replace the retiring units, and not surprisingly the capacity clearing price for the Boston zone jumped from the price floor to the price cap. Whether Salem Power was the marginal bid is impossible to ascertain from publicly available data, but the impending capacity deficiency in the Boston zone and the lack of any other major developments to replace this capacity was information clearly available to Salem Power (Energy Tariff Experts 2013, Public Power Weekly 2013, Maloney 2014). Hence, the developers must have known that they could exert substantial price influence.

Vazquez et al. (2002) develop a specific quantity based mechanism called the reliability option. They call for a centralized capacity market in which load serving entities must purchase adequate reliability options from generators in order to serve load at the desired reliability level. Reliability options include the key features of quantity based mechanisms previously discussed, but they also incorporate a call option with a specified strike price. For instance, if the strike price is \$300/MWh, load is purchasing a complete hedge against energy prices rising above \$300/MWh. Thus, in purchasing reliability options, load is both purchasing a guarantee of a certain level of reliability and a price hedge. Clearly, including the price hedge will raise the cost of the reliability option compared to capacity products not including this hedge. Generators

are selling off their right to collect energy revenues in excess of the strike price and need to be compensated for this sale.

If a regulator chooses to include a call option in the reliability product, they are implicitly asserting that the free market will not provide an adequate level of hedging. If the regulator trusted the utilities and generation to optimize their hedges there would simply be no need for requiring all of load to purchase a call option with the same strike price. PJM chose not to incorporate this call option into their reliability product, but many other ISOs including ISO-NE and the Spanish Electricity system chose to use reliability options. As will be shown later, the combination of a primarily thermal system with a capacity market enforcing the NERC 1 day in 10 years reliability standard leads reliability options to only have a small premium over capacity only products. However, the additional volatility induced by demand response will vastly increase the value of the call option embedded in reliability options.

3.3: The PJM Capacity Market in Practice

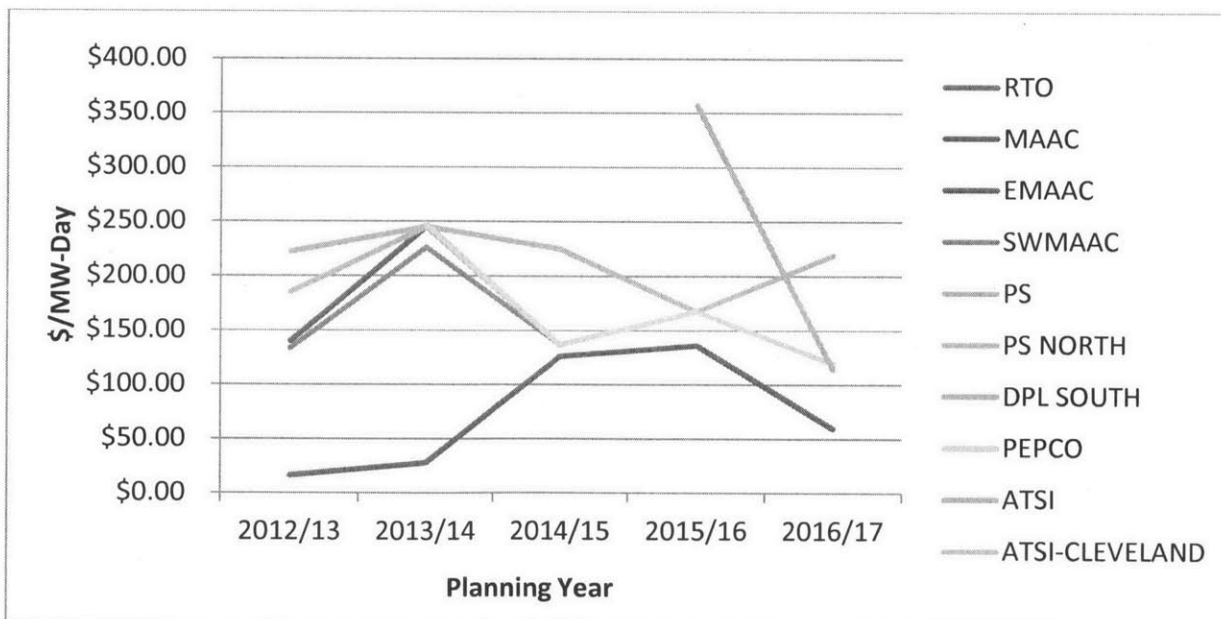
Capacity markets in PJM had an ignominious start. The auctions were marked by extreme price volatility and a review by the PJM market monitor concluded that the even when including capacity market revenue, generator remuneration was still too low to justify new entry (Rodilla & Batlle, 2013). The price volatility was largely due to a vertical demand curve combined with a short term daily capacity product. The capital costs of existing generation are sunk costs, so existing units will likely bid into the capacity market at very low prices. Most days have surplus capacity since adequate generation must be available to meet the peak demand day. The combination of a surplus of capacity and a vertical demand curve lead the capacity auction to clear at very low prices on most days. However, for the few days that capacity shortages

occurred, few new units can offer to provide the short term capacity product and the vertical demand curve led to extremely high clearing prices. The lack of market power mitigation rules likely also increased the price volatility. Hence, “instead of the market providing the stable price signal sought by investors, the end result was that another (even more) volatile short-term market was created” (Rodilla & Batlle, 2013). The initial PJM capacity market also lacked a locational element, leading generation to receive the same capacity payment regardless of whether they were in a region of capacity surplus or shortage (Sener & Kimball, 2007). Another notable concern was the weak generator performance mechanisms, raising the possibility that generators would not be available when their capacity was most needed (Brattle Group, 2011).

The reliability pricing mechanism (RPM) that PJM implemented in 2007 addressed many of these issues. Capacity is primarily purchased in the Base Residual Auction three years in advance, allowing new resources to more effectively compete in offering capacity. Elasticity was introduced into the demand curve allowing prices to rise if there is a capacity shortage, but also preventing them from falling to near zero if a capacity surplus exists. This reflects the understanding that the desired reserve margin is not a magic number, and the value of capacity slightly in excess of the reserve margin is not zero. The introduction of locational differentiation that accounts for transmission helps ensure that capacity is paid to enter into the correct locations. Importantly, the RPM allowed for several demand response resources to participate in the capacity market on an equal footing to generation. More robust performance requirements and market power mitigation procedures were also introduced (Sener & Kimball, 2007). Interestingly, capacity bids can be mitigated both up and down, reflecting a concern that prices could be either too high or too low.

A key input into the demand curve is the net cost of new entry (CONE) of a new gas turbine. Net CONE is the estimated cost of building a new gas turbine (called gross CONE) with the expected infra-marginal energy market rents netted out. The price of the demand curve at the desired level of capacity is set at net CONE. For the 2016/17 planning year, PJM calculated net CONE at a low of \$277/MW-Day for the more constrained zones PEPCO and SWMAAC to a high of \$363/MW-Day in ATSI. Figure 4 shows the capacity clearing prices for the various PJM zones in the Base Residual Auction. In general, the figure shows steep declines in the capacity clearing prices for the 2016/17 year, with the prevailing prices well below net CONE. Many factors have contributed to declining capacity prices, but large scale demand response is often considered one of the primary drivers.

Figure 4: PJM capacity clearing prices in Base Residual Auction.



3.4: Implications for Remuneration Shifting Between Capacity and Energy Markets

Capacity markets fundamentally shift part of the scarcity signal out of the energy market and into capacity payments. Capacity markets broadly constitute a method for increasing generator remuneration in order to increase production capacity. However, this increase in supply inevitably decreases prices in the energy market. The PJM market monitor notes that:

The exogenous requirement to construct capacity in excess of what is constructed in response to energy market signals has an impact on energy markets. The reliability requirement results in maintaining a level of capacity in excess of the level that would result from the operation of an energy market alone. The result of that additional capacity is to reduce the level and volatility of energy market prices and to reduce the duration of high energy market prices. (Monitoring Analytics, 2013)

Furthermore, the large differential in prices between the price cap and the marginal cost of thermal generation creates the opportunity for capacity markets to have a large depressing effect. The marginal cost of a new GT, assuming \$5/MMBtu gas, is roughly \$70/MWh. The price cap in PJM is currently \$1800/MWh. By ensuring that enough thermal capacity exists to serve load in almost all hours, capacity markets can create an order of magnitude change in the market clearing energy price experienced during high load periods.

The academic literature is surprisingly quiet on the impacts of capacity markets on the price signal. Besser et al. (2002) view the reduction in price volatility in a positive light arguing that capacity mechanism “facilitates the implementation of competitive generation markets while protecting consumers with respect to price and reliability risks” (Besser, Farr, & Tierney, 2002). Paul Centolella, however, views this price suppression quite negatively writing that “capacity markets ... dilute the price signals in the energy and ancillary service markets” contributing to moving electricity markets “away from the ideal of an efficient market, a market in which consumers simply see prices and can make individual decisions about whether or not to

purchase” (Centolella, 2010). The qualitative articles by Besser et al. (2002) and Centolella (2010) are some of the few academic studies to consider the impact of capacity markets on the energy price and the implication for the signal sent to consumers. Furthermore, the two articles clearly differ on whether the impact is positive or negative.

Shifting revenue from the energy market into the capacity market has several important implications for both consumers and generation. Generation companies strongly prefer these fixed revenue streams because it decreases their cash flow risk. In PJM a large fraction of a generator’s net revenue (revenue in excess of variable operating cost), comes from capacity payments. Between 2009 and 2012 the PJM market monitor estimates that a new GT would have received 61% of its net revenue from the capacity markets and 35% from the energy markets (the missing few percent are due to compensation for providing reserves and voltage support). These values for a new combined cycle plant are 35% from the capacity market and 64% from the energy market (Monitoring Analytics, 2013). Hence shifting to an energy only paradigm would represent moving a very large fraction of the industry’s revenues from fixed payments guaranteed 3 years into the future to highly volatile energy market payments.

The energy markets are also less susceptible to exercises of market power than the capacity markets. Unlike quantity based capacity mechanisms, the energy market often does not need output from every available generator to clear the market. This increases competition between suppliers because generators face the risk that too high of a bid simply won’t clear the energy market. Additionally, the PJM tariff requires generators to bid into the energy market at prices close to their marginal cost. The marginal cost bidding requirement is relatively straightforward to monitor given the transparency of plant heat rates and fuel prices. Monitoring whether a capacity bid falls within a reasonable range is substantially more difficult because

factors such as debt obligation, forward contracts, and maintenance requirements must be considered. A shift in remuneration towards the energy market would likely increase competition and decrease the risk of large exercises of market power in the capacity markets.

Under the current electric paradigm, most residential and smaller commercial and industrial customers are largely indifferent to whether payments are being made via the capacity mechanism or energy market. This is due to these customers paying a tariffed rate that does not vary with the spot market price. The capacity payment can simply be averaged over the expected consumption in the same manner as the energy price. The clear downside to this approach is that consumers see prices that are divorced from the short term supply conditions. In essence, this leads consumers to pay far too much for power during most hours in order to subsidize consumption during peak hours. Historically, hourly electricity meters were far too expensive for use in anything but large industrial and commercial applications. This gave utilities little choice but to charge most of their consumers flat rates. In this paradigm, the wholesale market only needed to send accurate price signals to supply, as load mostly was not exposed to the wholesale price.

However, recent innovations, both in solid state meters and cheap networking technology, have made it quite feasible to install smart meters at the residential level. These technological advances make it possible to relay wholesale market prices to the vast majority of consumers, raising the importance of sending accurate price signals to load. Given the tight coupling between energy and capacity prices, this amplifies the implications of shifting power cost between the capacity and energy markets.

The LMP sends a much more accurate signal to consumers than capacity charges. In restructured US electricity markets the LMP is highly granular, available at hundreds or thousands (depends on market) of locations at hourly and 5 minute intervals. The LMP is the result of a unit commitment algorithm clearing offers provided by generators with demand bids by load serving entities and large consumers. The LMP accounts for both system topology and generation outages. Hence, if a transmission line fails or a generator trips the LMP sends an immediate price signal that reflects those new conditions. Capacity prices, on the other hand, do not get more granular than zonal in location and monthly in time. Capacity charges in most US ISO's are based on a customer's peak coincident load usage. This provides a strong incentive for customers not to consume during the system peak, but this might not actually be the peak scarcity of the system. If a generator or line outage causes the system to be most stressed outside of the peak load period, the capacity charge may be discentivizing consumption at the wrong, or at least not the optimum time. Without the continuous clearing component of the LMP, the capacity charge cannot send an accurate signal to load that reflects current market conditions.

This timing problem will likely become more acute with increases in renewable penetration. Non-dispatchable renewables have the potential to shift the thermal generation peak generation hour away from the peak load hour and towards some combination of high load and low renewable output. Errors in wind and solar forecast also have the potential to create shortages in available thermal generation outside of peak load periods. In these cases the LMP will react immediately sending a price signal to consumers to decrease consumption, but the capacity signal with its current emphasis on peak coincident load will not react to these changes in market conditions. Another area for concern is that increases in renewable power often lead to

new congestion patterns. These changes in power flow are reflected immediately in LMPs, but will likely take years to flow through to administratively determined capacity zones.

Furthermore, capacity charges in their current US implementation are essentially a retroactive fee. The magnitude of the fee assessed to the market is determined in a competitive framework in which generators bid to provide capacity. The apportionment of that fee between consumers in the market, however, is based on a measure of peak coincident load (PJM, 2010). This means that on a hot day a consumer cannot simply check the market clearing price of power, but instead also must try to predict whether the current period is the peak load interval. While better assessment of capacity charges could improve the price signal over the rather crude current mechanism, a dynamically clearing market will inevitably provide a more accurate price signal than a retroactively assessed fee.

Chapter 4: Demand Response

4.1: Effects on Price

Within the literature, it is generally assumed that increases in demand response and price elasticity will lower price volatility. For instance, Borenstein et al. (2002) write “price-responsive demand holds the key to mitigating price volatility in wholesale electricity spot markets.” Similarly, Albadi and El-Saadany (2007) find that an important contribution of demand response “is the reduction of price volatility in the spot market.” Centolella 2010, arguing in favor of price responsive demand (as opposed to demand response provided by a curtailment provider) writes “it may reduce business risks for market participants by mitigating extreme price volatility.” Note that demand response generally refers to a product primarily

remunerated through the capacity market and paid to reduce consumption from baseline levels during scarcity conditions. Price sensitive demand refers to elasticity in the demand curve in which short term consumption changes based on the price level, but the consumer is not receiving payments for reductions compared to a baseline level.

Quantitative studies generally confirm that increases in price sensitive demand and demand response will lower overall system costs and depress peak prices. Caves (2001) conducts a simulation based on aggregate supply and demand curves and shows that small increases in demand elasticity have large effects on the peak prices realized (Caves, Eakin, & Faruqui, 2000). Kirschen et al (2000) found that incorporating cross temporal demand elasticity (essentially load shifting) into demand bids substantially lowers power price during peak periods (Kirschen & Strbac, 2000). Sing et al (2011) develop a day ahead dispatch algorithm that incorporates price responsive demand shifting (PRDS) to show the potential for PRDS to flatten the demand curve, reduce transmission congestion, and consequently flatten the distribution of LMPs (Singh, Padhy, & Sharma, 2011). Rassenti et al. (2002) find that demand side bidding can significantly reduce the exercise of generator market power and vastly reduce the occurrence of price spikes (Rassenti, Smith, & Wilson, 2003). A recent NREL study based on the Colorado electricity system found that various forms of demand response could economically provide ancillary services (Hummon et al., 2013). It is important to note that all of these studies were conducted assuming a static thermal generation mix and the potential for demand response to perturb the balance between energy and capacity prices was not considered. Furthermore, many of these studies tend to assume highly flexible demand side participation, which can move consumption from peak to non-peak periods at little or no cost. This is quite different to the emergency demand response in PJM that is highly compensated to reduce load during peak conditions.

A few more recent articles have examined the nuances of price setting in electricity markets to show that increases in demand curve elasticity and demand response could in some situations increase energy price level and volatility. Wu 2013 using a network constraint unit commitment model finds that demand response “may cause abrupt and unwarranted price increases with the objective of maximizing system social welfare, when it displaces less flexible generating units and/or causes additional transmission congestions as observed in numerical case studies” (Wu, 2013). It is important to note that dispatching demand response is maximizing social welfare (and hence minimizing production cost). However, the high marginal costs of demand response are setting the modeled LMP to high levels, whereas the higher startup cost of thermal generation in the US is generally remunerated out of market and hence do not enter the market clearing price. Centolella (2010) asserts that introducing an operating reserve demand curve based on estimates of VOLL and co-optimized with the energy markets would “place greater reliance on the price signals in the energy and ancillary service markets. And, revenue will shift from the capacity market to energy and ancillary service markets, such that total costs are reduced.” (Centolella, 2010).

The literature is not in disagreement that introducing more price sensitive demand will decrease production costs. However, Wu (2013) highlights that while demand response can in some situations have lower production costs than generators due to startup costs, its higher marginal cost leads to higher and more volatile prices. Centolella (2010) identifies interesting methods by which costs can be shifted between energy and capacity markets, and notes the implication for the changing price signal sent to consumers.

4.2: Value of Lost Load and Emergency Demand Response

An important underlying input to demand response studies is the price level and form of the introduced demand elasticity. Given that markets have traditionally not existed for short term electricity reliability, there are not clear market prices for what consumers are willing to pay for electricity to avoid short term supply interruptions. Historically, the value of lost load (VOLL) has been estimated through econometric studies that attempt to estimate the loss of production or harm from a short term loss of power. These types of studies have consistently estimated VOLL well above the ISO imposed price caps (P Joskow & Tirole, 2004). Most estimates of system average VOLL tend to be in the range of \$2,000/MWh to \$15,000/MWh, although some range as high as \$50,000/MWh (de Nooij, Licshout, & Koopmans, 2009; PL Joskow, 2006; Leahy & Tol, 2011). Nooij, et al. (2009) conduct a detailed study in the Netherlands examining the VOLL by municipality, sector, and time of use. In general they find that manufacturing has the lowest levels of VOLL at \$2,300/MWh while construction and government has the highest estimated VOLL at \$42,000/MWh – \$45,000/MWh. Households came in somewhere in between with VOLL of \$21,760/MWh (de Nooij et al., 2009). Leahy and Tol (2011) perform a study of Ireland and find, similar to Nooij et al (2009), that manufacturing has the lowest VOLL, although they estimate a value of \$5,300 euro/MWh. The fact that these VOLL estimates are well above the price caps in most deregulated markets implies there is limited demand willing to decrease short term usage at foreseeable price levels. Indeed, demand elasticity in the range of \$2,000 – \$15,000/MWh will do little to decrease electricity price volatility when electricity prices rarely breach the \$500/MWh and are often capped in the range \$1,000/MWh - \$2,000/MWh.¹

Other studies examining residential customer response to time of use pricing schemes or assessing the technical potential for water heaters and pumps to shift consumption have indicated

¹ Nooij et al. and Leahy and Tol (2011) presented their findings in Euros. These values were converted to Dollars using the 2013 average exchange rate of 0.7533 Euros per Dollar.

demand elasticity may exist at lower price levels. Faruqui and Palmer (2011) reviews the results from 109 studies on residential consumer response to dynamic pricing and find that switching to dynamic pricing leads to reductions in peak load. The 109 studies include various types of dynamic pricing, ranging from more predictable time of use to completely market determined spot prices. Despite the diversity of pricing structures, Faruqui and Palmer (2011) finds that 60% of the variation in peak load reduction can be explained by the ratio of peak to offpeak prices (Faruqui & Palmer, 2011). Furthermore, given the prevailing residential price of power in the US, the results indicate that demand elasticity exists for residential consumers at prices below the cap. Joskow (2012), however, notes that most of the studies rely on volunteer participants, so the levels of demand elasticity are likely over-estimated because consumers who would benefit most from dynamic pricing are most likely to volunteer (P. L. Joskow, 2012).

While these studies indicate the potential for the development of demand elasticity at levels below the price cap, the vast growth of demand response in PJM has been as an emergency resource that is primarily remunerated through the capacity market. In the June 2012 – May 2013 planning year, emergency demand response was only dispatched on two days (PJM, 2013a). Given that capacity prices in many of the constrained zones cleared for roughly \$130/MW-Day, and emergency demand response can only be dispatched for 6 hours per day, this results in emergency DR receiving a capacity payment of ~\$4,000/MWh for every MW of reduction. Furthermore, since the maximum strike price of demand response is \$1,800/MWh and demand response is guaranteed to receive this payment regardless of LMP, emergency demand response received about \$5,800/MWh for every MW curtailed. Hence, the prices being paid to demand response in PJM during the 2012/2013 operating year are within the range of econometric estimates of VOLL. Additionally, demand response as a percentage of total capacity

has declined from ~9% for 2014/2015 and 2015/2016 to 7.3% for the 2016/2017 planning year. This is likely a combination of the increasing frequency with which PJM has been dispatching demand response and substantially lower capacity clearing prices. Since emergency demand response is guaranteed to recover its strike price (currently \$1,800/MWh) and is receiving a capacity payment on top of this, the reduction in cleared demand response indicates that these resources have a marginal cost above \$1,800/MWh, in line with VOLL estimates in the thousands of dollars per MWh range.

4.3: Demand Response vs Price Sensitive Demand

Demand response often takes a variety of definitions in the literature. Some studies define demand response very generally as altering consumption from normal levels based on any signal, be it an ISO ordered curtailment or a price signal (Cappers, Goldman, & Kathan, 2010). However, Bushnell et al. (2009) draw an important distinction between traditional demand response and true dynamic pricing, or what this thesis will call price sensitive demand. Demand response refers to customers being compensated (historically primarily through the capacity market) to reduce demand below pre-determined baseline levels (Bushnell et al., 2009). Price sensitive demand refers to electricity consumers altering their short term electricity consumption based on the commodity price. If the price of energy rises, price sensitive demand may consume less to avoid costly purchases, but it is not actively being compensated by the ISO for its reduction.

Bushnell et al. (2009) argue that the highly favorable regulatory treatment of demand response risks crowding out more flexible and economically efficient price responsive demand. Demand response's favorable treatment is largely a symptom of FERC's support for more

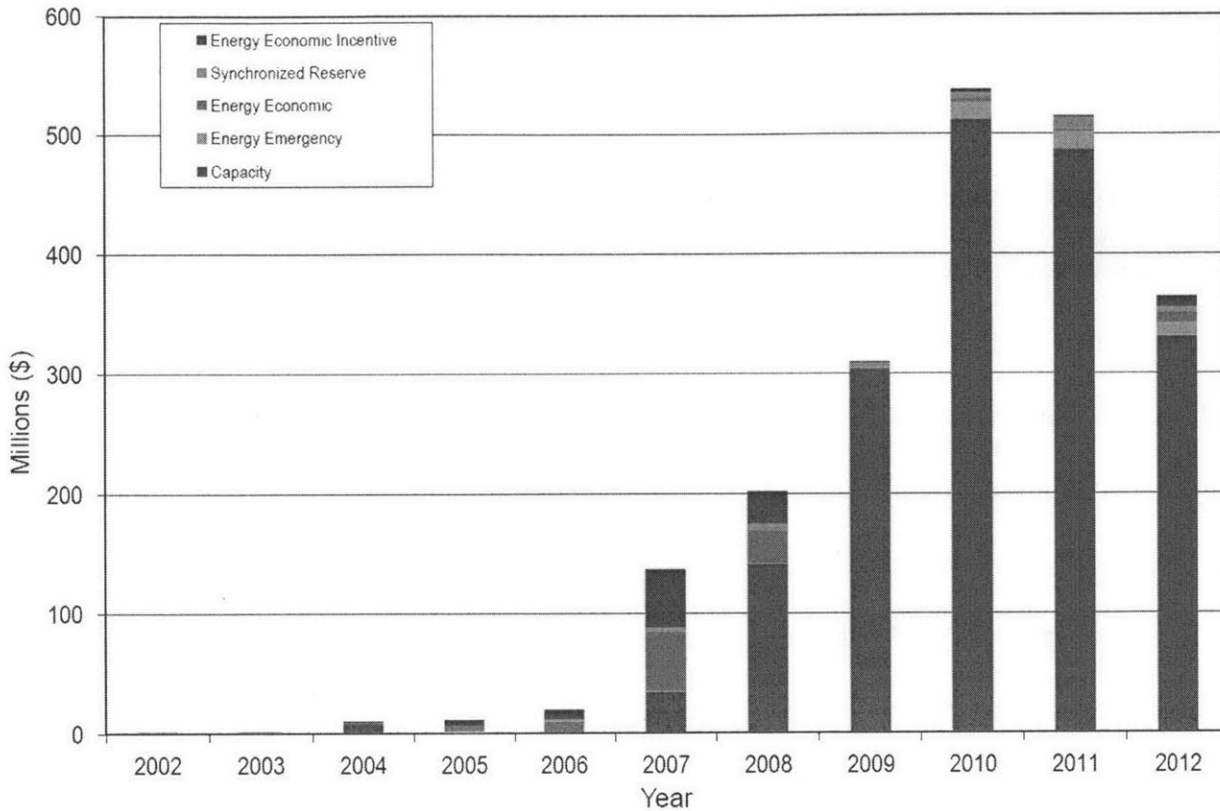
flexible demand while “state public utility commissions (PUCs) and legislatures have been at best indifferent and at worst openly hostile to the expansion of dynamic electricity tariffs for consumers” (Bushnell et al., 2009). FERC has used its jurisdiction over the wholesale power markets to treat demand response favorably, but jurisdiction over retail rates received by end use consumers lies at the state level.

Demand response has several drawbacks over true price responsive demand. According to Bushnell et al. (2009) the largest drawback is that it creates an asymmetric relationship between the treatment of load and generation. Demand response is compensated based on reductions from an administratively set expected level. Generation, on the other hand, is compensated based on what it produces, in other words generation has a baseline of zero. Furthermore, paying for reductions from baseline levels creates a moral hazard where end use customers have an incentive to inflate their consumption prior to the start of the demand response program. Other studies have also noted the incentive for demand response providers to distort their baseline usage (Chao & DePillis, 2013). Furthermore, Bushnell et al. (2009) note that in the CAISO control area shortages can be highly localized, which could be addressed by price sensitive demand receiving the granular LMP signal, but not by the more blunt instrument of demand response, which, tends to be dispatched zonally.

Bushnell et al. (2009) highlight two mechanisms by which demand response can crowd out true price responsive demand. The first is quite simply that if demand programs are focusing on demand response, there might be less of a policy emphasis on developing price responsive demand. Second, consumers might find participation in demand response programs more lucrative than price response programs given the potential for payouts from the capacity market and gaming of the baseline from which reductions are measured from.

In addition to the two mechanisms identified by Bushnell et al. (2009), capacity markets in and of themselves may crowd out price sensitive demand. Demand response has traditionally been remunerated primarily through the capacity markets (Bushnell et al., 2009). Figure 5 shows the extent to which capacity payments dominate demand response remuneration in PJM. In order for a resource to participate in the PJM capacity markets, it must provide a well specified product that curtails when the ISO instructs it to do so. Failure to comply with ISO orders will result in fines. Hence, demand response is primarily remunerated for providing an emergency product that responds to a curtailment order from the ISO during a few peak load events per year. This is quite different from an end use customer seeing a high energy price and deciding they will turn down the air-conditioning this instance, but not committing to reductions at future instances. Large industrial consumers may prefer to provide the specific capacity product and receive capacity payments. Smaller customers, however, may prefer or be simply unable to provide the capacity product the ISO requires, which includes the risk of the associated non-performance penalties. Historically, combining a capacity market with a thermal system has led to large capacity payments and dampened price volatility. This combination favors larger customers who can access the capacity markets over small consumers who may be able to access the real time energy price through a smart meter, but not the more complex set of requirements required for capacity market participation. On average, large numbers of smaller consumers responding to energy price could be an economical form of peak load reduction, but the combination of capacity mechanisms with a primarily thermal system does not favor this form of demand side participation.

Figure 5: Sources of demand response remuneration (Monitoring Analytics, 2013).



As demand response becomes a larger part of the installed system capacity, it will be increasingly be the marginal resource. Very simply, increasing the quantity of demand response will lead to greater price setting by demand response. Given the dramatically higher marginal cost of demand response compared to a GT, this has the potential to significantly increase the energy clearing price during peak load events. While Bushnell et al. (2009) identify several avenues by which demand response may crowd out price responsive demand, the potential for demand response to drastically increase energy price volatility could provide a strong incentive for the development of price responsive demand. Demand response, by transferring power costs away from the capacity market and towards the energy market will allow more equal competition between price responsive demand (remunerated through energy market savings) and demand response (remunerated through capacity payments and energy market savings).

Chapter 5: Methods and Model Description

Power systems modeling has historically been divided into separate categories depending on the time horizon of interest. Very detailed models of the transmission grid are required to maintain system security in real time operations, which assess market operations on time scales of between seconds and minutes. Unit commitment refers to the decision of which generators to turn on and off, and is generally made for time horizons between 1 day and 1 week. These decisions need to be made well in advance because many generation technologies require substantial time to turn on and off and adjust their output level. These generation constraints are then combined with transmission constraints in an attempt to ensure that the pre-determined dispatch schedule is feasible. Most European markets use a zonal system where transmission limitations between zones are modeled, but constraints within zones are ignored. The US restructured markets use a nodal system where most of the transmission line limits are included in the unit commitment model as constraints. Operations planning generally occurs over a time period of weeks to 1-2 years. Here generation and transmission line outages are scheduled to avoid significant overlapping outages. Some grid level modeling may occur in operations planning, but historically a full unit commitment analysis is not performed at this step. Planning and investment refers to the entry and exit decisions of new generation, and also to the decision of whether to build new generation. Historically, given the long time periods over which generation expansion occurs, modeling the generation operation and transmission constraints was not feasible. Thus, generation expansion models tended to assume very simplified operations in order to model such a long time frame.

Mixed integer linear programming (MIP) has been the primary method employed in electricity modeling. Linear programming allows an objective function to be maximized or

minimized subject to a set of constraints. Binary variables are required in order capture the non-convex generator start up and shut down decision variables. Including these non-convex constraints and control variables, however, has a large computational cost as the solver can no longer take advantage of the convex nature of the problem space.

Research in recent years has focused on making unit commitment modeling more efficient, and incorporating unit commitment into generation expansion planning. Modeling unit commitment is an innately difficult problem because of the large number of binary variables involved. For every time step (typically an hour), binary variables for the generator status (on/off), start-up, and shut-down decisions are required. In a large system with hundreds of generators, the number of binary variables quickly explodes as the number modeled of hours increases. Combining unit commitment with generation adds another set of binary variables to indicate whether the generator has been built. Thus, trying to combine generation expansion with a standard unit commitment formulation quickly leads to an intractable problem (Palmintier & Webster, 2014).

One approach to solving this problem is developing a more efficient unit commitment algorithm. Palmintier and Webster (2014) develop a method to efficiently model clusters of similar generators. In many systems, there are large numbers of similar generators, and these can be combined into groups to vastly decrease the number of binary variables. Instead of modeling each generator individually with multiple binary variables per generator, generators are combined into clusters with the on/off decisions transformed from binary to integer variables. Palmintier and Webster (2014) incorporate this integer block unit commitment method into a generation expansion model and achieve improvements in solution time on the order of 5,000x. Their results further showed that simple generation expansion planning models that ignored

generation operation constraints would arrive at materially different generation mixes and expected system costs.

The integer block unit commitment algorithm is one strategy for tractably including unit commitment within a generation expansion model. However, modeling generation expansion as a mixed integer formulation has a few drawbacks specific to this study. Primarily, it is unclear how to formulate a complex capacity market along the lines of the one implemented in PJM. In linear programming, one maximizes or minimizes a given scalar objective function. Hence, mixed integer capacity expansion models take the perspective of the central regulator minimizing costs. It has been shown that under several strict assumptions, the least cost central planning solution is the same as the competitive market solution. Nevertheless, a mixed integer program is not an appropriate framework in which to model generators that submit bids into a capacity market and have the bids cleared against a sloping demand curve. Fundamentally, including any sort of revenue sufficiency requirement on generators leads to a set of non-linear equations in that a generator's revenue is determined by the multiplication of market price with cleared quantity. Because both of these variables are endogenous to the model, including the product of the two variables in the model makes it non-linear and extremely difficult to solve.

The standard methodological approach to modeling competitive behavior in electricity markets is the mixed complementarity (MCP) approach. This involves first formulating the non-linear set of equations, and then taking the first-order or Karush Kuhn Tucker (KKT) conditions. A capacity market based on generator bids could theoretically be constructed because a mixed complementarity approach allows modeling multiple profit maximizing generators. The primary downside to this approach is that the KKT conditions require taking derivatives, and hence these methods cannot accommodate binary variables. Additionally, multiple equilibriums may exist,

and it can be difficult to determine whether the single solution found by the solver is representative of the other solutions.

Dynamic models represent a third approach and have received less focus in the electricity modeling literature. A notable application of this approach was Hobbs et al. (2007), which assesses the merits of the various demand curves that PJM proposed for its capacity auction. The fundamental difference between dynamic models and the optimization methods previously discussed is that dynamic models step through time in discrete intervals. Linear programming and mixed complementarity methods, on the other hand, examine the entire model space simultaneously. This latter approach can be thought of as representing the central planner in a MIP, or the individual agents in an MCP having perfect foresight of the input parameters. In a stochastic optimization, the exact realized value of an input may not be known when the initial decision is made, but the potential range of values that can be realized in every year is known. In a dynamic model, the agent(s) is making a decision at each time step based on a subset of the information available at the current time step. The decision logic in each stage can certainly include forecasts of conditions in future stages, but the actual realized outcome in time $t+n$ will not be fed back into decision for time t .

Focusing on the decisions in each time step, as opposed to the entire model horizon, can break a large intractable optimization into many small tractable problems. Additionally, in a dynamic model, decision rules as opposed to mathematical equations written within the confines of a MIP or MCP are employed. This vastly expands the flexibility of decisions that can be considered, and issues such as the need to express all equations in linear form for a MIP or no binary variables for an MCP disappear.

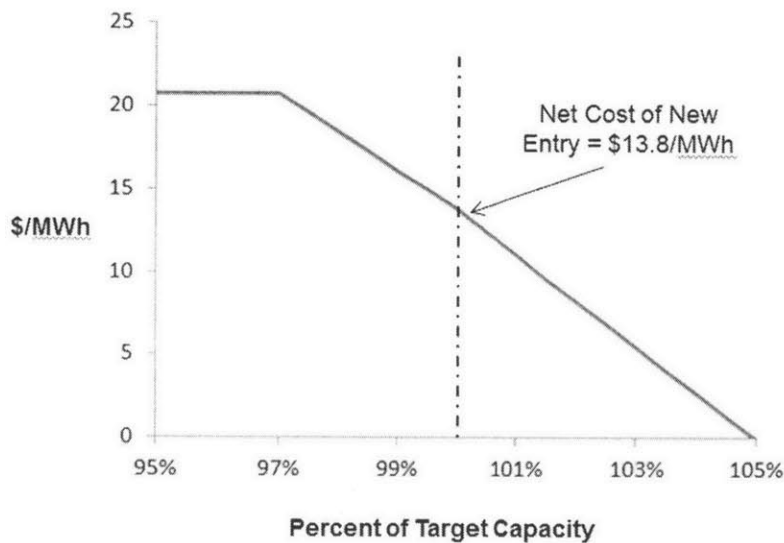
The fact that dynamic models allow decisions to be made without complete information can be considered either a strength or a weakness of the modeling technique. Given the predominance of rational decision making with complete information, there is generally very little opposition to constructing models with this assumption built in. However, real world investment obviously doesn't follow this perfectly rational and perfectly informed decision making process. Investments in manufacturing capacity, including electricity, are often marked by boom and bust cycles. The market overreacts to periods of scarcity leading to a surplus of supply. Low prices then discourage entry, and the long lead time required to build these capital intensive assets can lead to extended periods of tight supply. MIP and MCP models are poorly suited to capturing these boom and bust cycles because the entire model is being solved simultaneously. Dynamic models, however, are much better suited to capturing these real world investment dynamics because they step through time iteratively. The key is to provide the decision making agent in the model an amount of information that would be reasonably available to them at that time.

Most dynamic models do require additional decision rules in order to ensure that the model is representing reasonable conditions and not over reacting to the currently available information. For example, the generation expansion dynamic model in Eager et al. (2012) caps the potential annual revenues of a generator received from scarcity rents to the annualized cost of a GT. This ensures that generation entry is not based on scarcity rents higher than what would occur in a competitive market in which a GT would enter to capture these rents (Eager, Hobbs, & Bialek, 2012).

For this thesis, a dynamic model was developed to simulate the entry and exit decisions of merchant generators under a market paradigm based on PJM. Energy prices are derived from a

unit commitment algorithm that dispatches the least cost generation mix required to serve load while observing the physical operating limitations of the generators. The clustering algorithm developed by Palmntier and Webster (2014) was employed to speed the unit commitment. The detailed mathematical formulation for clustered unit commitment is given in Appendix 1. Unit commitment decisions are solved for one week at a time for every other week in a year, as opposed to every week, in order to further improve model execution speed. The model includes a capacity market premised on the PJM system and clears capacity supply offers against the capacity demand curve shown below.

Figure 6: Modeled capacity demand curve.



The pseudo code for the generation expansion model is as follows.

Sets:

t: years

i: generators (updates for every *t* based on entry and exit)

tech: technologies to enter the market (no entry, GT, CCGT, DR)

Important Variables:

EnergyRevenue: Revenue earned from sale of energy

VarCosts: all variable costs (fuel + variable O&M)

FixedCosts: Fixed costs that must be met for generator not to exit the market. For existing generators capital costs are considered sunk and only fixed O&M is considered. For new entrants, capital costs are also included.

InfraRent: revenues of generator in excess of variable operating costs (infra-marginal rents)

ExpectedInfraRent: the expected future revenues of generator in excess of variable operating Costs

CapacityPrice: Clearing price of capacity market

BoolCapMarket: Whether the model run includes a capacity market

MostProfitableTech: Expected profit of the most profitable new entrant technology

Code

```
##time period loop  
for t in 1:15
```

Create time series of energy prices and generator output for year t based on current i using clustered unit commitment. Note the set i can include generators under construction and not in service yet, only in service generators enter the unit commitment.

```
##calculate average inframarginal rents by technology type
```

```
 $InfraRent_{i,t} = EnergyRevenue_{i,t} - VarCosts_{i,t}$ 
```

```
 $InfraRent_{tech,t} = mean(InfraRent_{i,t}, grouped\ by\ tech)$ 
```

```
##(set Capacity Price to arbitrarily large number)
```

```
 $CapacityPrice_t = 9999$ 
```

```
##(set Profit to arbitrarily small number)
```

```
 $MostProfitableTech = -9999$ 
```

```
##forecasting loop
```

```
foreach tech in (no entry, GT, CCGT, DR)
```

```
##expand generator set with new entry
```

```
 $i\_new = i + tech$ 
```

create time series of energy prices and generator output for year $t+4$ based on i_new and year $t+4$ load.

```
##calculate infra-marginal rents for every generator in  $i\_new$ 
```

```
 $InfraRent_{i\_new,t+4} = EnergyRevenue_{i,t+4} - VarCosts_{i,t+4}$ 
```

```
if existing plant
```

```
 $ExpectedInfraRent_{i\_new,t} = (InfraRent_{i,t} + InfraRent_{i\_new,t+4})/2$ 
```

```
##if plant is new entrant  $InfraRent_{i,t}$ , does not exist, and thus must use
```

```
 $InfraRent_{tech,t}$ 
```

if new entrant

$$ExpectedInfraRent_{i_new,t} = (InfraRent_{tech,t} + InfraRent_{i_new,t+4})/2$$

$$CapBid_{i_new} = FixedCosts_{i_new} - ExpectedInfraRent_{i_new,t}$$

Capacity supply curve = rank order ascending of $CapBid_{i_new}$

Capacity Demand curve bid = prices ISO willing to pay based on current PJM RPM

see Figure 6 for a graphical example of the capacity market clearing

If tech = no entry

$CapacityPriceNew$ = demand bid that the minimizes vertical distance
Between supply and demand curves (i.e. comes closest to equating supply
and demand)

If tech != no entry

$CapacityPriceNew$ = supply bid that the minimizes vertical distance
Between supply and demand curves (ie comes closest to equating supply
and demand)

if $CapacityPrice_t > CapacityPriceNew$ & $BoolCapMarket=TRUE$

$$CapacityPrice_t = CapacityPriceNew$$

the generator set for the next period is updated, only generators that
cleared the capacity market (and hence make profit) enter

$$i_{t+1} = i_new \text{ s.t. } (CapBid_{i_new} \leq CapacityPrice_t)$$

$MostProfitableTechNew = mean(ExpectedInfraRent_{i_new,t} - FixedCosts_{i_new})$ s.t. i is
a new entrant

if $MostProfitableTechNew > MostProfitableTech$ & $BoolCapMarket=FALSE$

$$MostProfitableTech = MostProfitableTechNew$$

the generator set for the next period is updated, only profitable new
generators enter

$$i_{t+1} = i_new \text{ s.t. } (ExpectedInfraRent_{i_new,t} - FixedCosts_{i_new} >= 0)$$

End foreach tech loop

End Time Period Loop

The model starts in year 1 by running the unit commitment algorithm using the existing generation mix to create an hourly time series of energy price and generator output. The model then enters the forecasting module to determine entry and exit for that year. The possible new

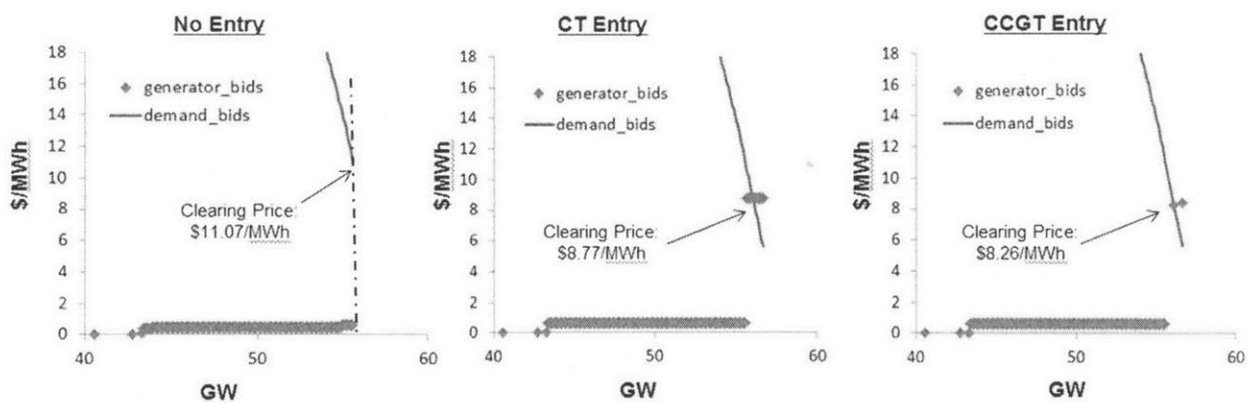
entry scenarios are 1: no new entry, 2: GT entry, 3: CC entry, and 4: Demand Response entry (if the model run is considering DR entry). The forecasting step generates energy prices for year $t+4$ for each of these scenarios. The expected infra-marginal rent of each generator is then calculated from the simple average of the infra-marginal rents from the year t and the forecasted rents from $t+4$. Note for new entrants, a value does not exist for year t , and instead the average value for that technology is used. Hence, the energy prices determining entry are based on the current year and a year forecasted in the future.

Each generator then bids into the capacity market at the price required to cover the difference between their fixed costs and infra-marginal rents from the energy market. In other words, generators bid into the capacity market at the price required to stay operational. Note that capacity bids are floored at zero. If the generator has already entered the market, its capital costs are considered sunk and hence the fixed costs it needs to cover are very low. New entrants, however, include capital costs in their bids and thus bid at prices much higher than the existing units of the same technology. The capacity market is cleared using a sloping demand curve developed to be similar to the one used by PJM's Reliability Pricing Model. The technology that minimizes the capacity clearing price is the technology that enters the market. Given that in the base runs of the model, generation outages are not considered, a planning reserve requirement was not modeled in the capacity market.

Figure 7 graphically depicts the capacity market clearing process for a model run in which demand response was not allowed to enter. In the no entry case, the existing generation supply does not provide adequate capacity and the supply curve from the generator offers does not cross the demand curve for capacity. In this case the clearing price is set by the demand curve at \$11.07/MWh. The demand curve sets price when the supply and demand curves fail to

cross in order to prevent a shortage of capacity from causing low market clearing prices. In the case with new GT entry, the supply and demand curves cross and one of the new GTs now sets the capacity price at \$8.77/MWh. The first new CCGT, however, needs to recover slightly less money in the capacity market in order to cover its fixed costs and clears the markets at \$8.26/MWh. Note that the second new CCGT does not enter in this case because \$8.26/MWh price is not quite enough to allow the unit to have the expectation of recovering its fixed costs. In essence, the most profitable generation technology will enter. Only one generation technology is allowed to enter in a given year to decrease the size of the search space. Future improvements of the model could include a second round of entry in which the model checks if additional generators could profitably enter the market. However, in the current setup, the size of new generators is tuned to PJM's expected load growth so additional profitable entry after the first entry round is not possible.

Figure 7: Modeled capacity clearing process.



If any existing generators were not covering their fixed costs, they would exit at this point. If the system requires additional capacity, however, no units will retire due to a lack of profitability because the capital costs included in the new unit's capacity market bid will

inevitably clear the market above the level required for an existing unit just to cover its fixed O&M costs. Note that the model can also simulate an energy-only market. In this case the most profitable type of generation technology enters the market, with the caveat that expected profits must be positive in order for the generator to enter.

Once entry is determined for $t=1$, the model steps ahead to $t=2$ and the process is repeated. Hence, for each year the model outputs an hourly time series of energy price and generator output, the capacity market price, and generation entry and exit.

The modeled power system (generation mix and demand) is loosely based on PJM, but with the total capacity reduced by 1/3 of the PJM system uniformly across all technologies in order to increase model execution speed. Load and wind generation levels are the actual PJM 2013 values, scaled by 1/3. Load grows at the PJM forecast rate of 1.2% rate. Coal retirements are based on the 8% - 11% of total capacity range predicted by The Brattle Group (Aydin, Graves, & Celebi, 2013), with the low end of the range chosen because of higher than expected energy prices and coal utilization in PJM during the winter 2012-2013 (Testa, 2014). A natural gas price of \$5.10/MMBtu is used, based on the forward curve of Henry Hub and an adder to account for transport costs. The price cap is set at \$1,500/MWh, the level that prevailed in PJM between June 2012 – May 2013. The installed thermal generation mix is based on the current thermal capacity in PJM, but these values have been reduced to account for the load scaling and the removal of the reserve factor. Generator costs (both variable and fixed) and physical operating characteristics are based on data published by the EIA (US Energy Information Administration, 2010).

The modeled energy market is simplified in three important ways compared to the actual PJM system in order to reduce the execution time of the model. First, the modeled energy market does not consider transmission constraints. This is often referred to in the literature as modeling the system as a single node, or a copper sheet. Second all inputs into the energy market are deterministic and as such ancillary service markets such as regulations and spinning and non-spinning reserves are not modeled. This also means that the PJM reserve shortage pricing mechanism is not incorporated. Third, following from the deterministic inputs, only a DA market is modeled as a RT market is not needed to instantaneously balance supply and demand. By only running a DA market, this model assumes that demand response is a resource available for dispatch and price setting in the DA. Currently, emergency demand response can only be dispatched in the RT market. The market monitor has recommended making emergency demand response eligible for DA participation, so this rule may change in the near future. Additionally, the ability of virtual energy traders to arbitrage price differentials in the DA and RT markets means that price setting in the RT market will be transitioned into the DA market.

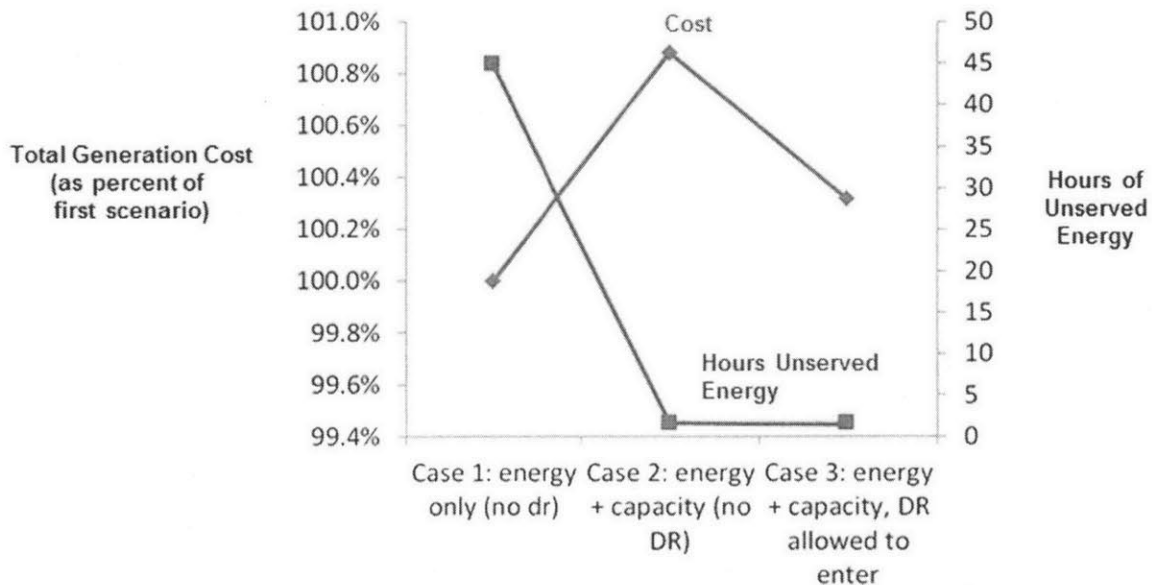
Chapter 6: Results

6.1: Effects on Costs and Reliability

The model was initially run for 15 years ($t=1$ to $t=15$) for 3 cases. These were 1: no demand response entry and an energy only market, 2: no demand response, but the capacity market was implemented to maintain the NERC reliability standard, and 3: demand response entry and the capacity market. Demand response for these cases was modeled with a marginal cost of \$2,000/MWh and no fixed costs.

Figure 8 summarizes the total generation cost and hours of unserved energy by scenario. Total generation cost refers to the sum of all costs incurred by generators. This includes variable costs, fixed O&M, and fixed capital costs. Note the capital costs of generators in the exogenously imposed starting mix are calculated as if the generators came online at $t=1$. The capacity market reduces the average number of hours of unserved energy from 42 per year to 1.6, which is much closer to the NERC reliability standard of 2.4 hours per year. Total generation costs to achieve this reliability requirement increased by 0.9%. The demand response entry in Case 3 decreased the incremental cost required to achieve the NERC reliability standard to 0.3%. This is expected, as demand response is a cheaper form of peaking capacity than a gas turbine and as such should lower the total generation cost required to meet the reliability level.

Figure 8: Total generation cost (variable + fixed) and hours of unserved energy.

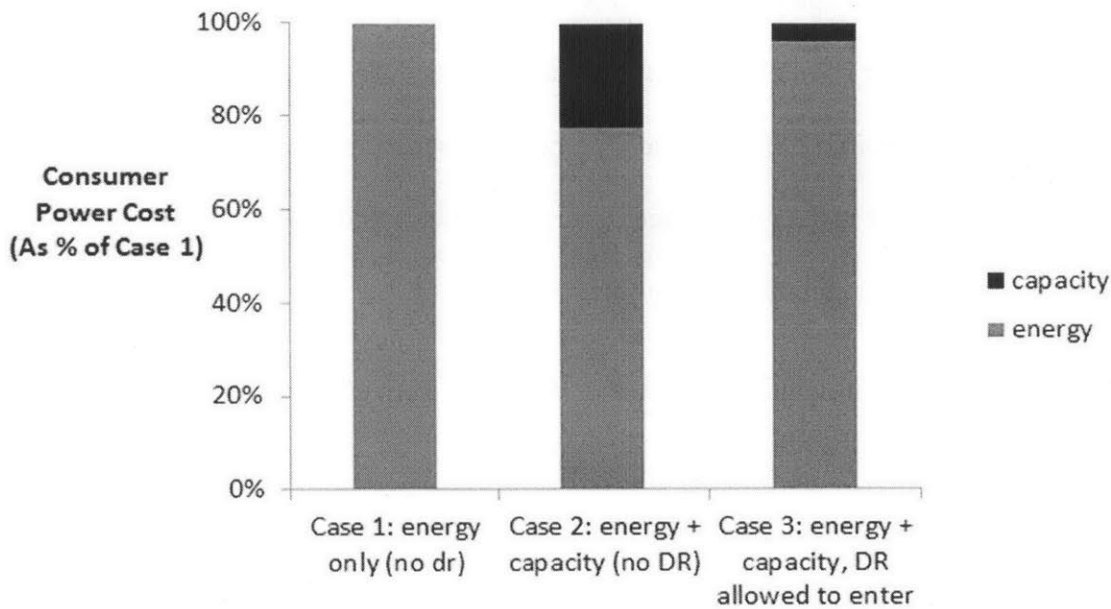


Introducing the capacity market caused a large redistribution of consumer power cost between the energy and capacity markets. Consumer power cost refers to the total amount that

consumers must pay for their power and includes both energy and capacity. In other words, this is the amount of money that consumers are paying to generators for the power they purchase. Figure 9 summarizes the split in energy and capacity payments by scenario. Clearly, in the energy only market, 100% of consumer power cost is in the energy market. Introducing the capacity market into the thermal only system (no DR) shifts 24% of the consumer power cost from energy into capacity. Hence, while the capacity mechanism increased total generation costs by only 0.9%, it shifted almost a quarter of the consumer power cost away from the energy market and into the capacity market. The modeled value of capacity comprising 24% of the consumer power cost is consistent with the 22% that occurred in PJM in 2009.

Occasional criticism of capacity markets has focused on the very high capacity prices required to achieve reliability standards. However, this critique appears to assume that the entire capacity charge being billed to consumers is the cost of the additional reliability. These results, however, show that total generation cost (the cost that society must bare) has only increased by roughly 1%. Capacity's large share of the consumer power cost in case 2 is much more a function of costs shifting between the energy and capacity market, rather than an increase in total generation cost.

Figure 9: Power cost to consumers by scenario expressed as a percentage of Case 1 (energy only market and no DR).



Introducing demand response in Case 3 substantially alters the balance between energy and capacity payments. With demand response, capacity payments drop from 24% of consumer power cost to 4%. Importantly, this massive shift in cost from the capacity market and into the energy market occurs with demand response rising to only 10% of installed capacity. In the 2015/2016 planning year, PJM is expecting demand response to comprise 9.3% of capacity, so the level of demand response entry predicted by the model is not unreasonable.

Demand response alters the balance between energy and capacity costs because it has a very different cost structure compared to a gas turbine. With gas in the roughly \$5/MMBtu range, the marginal cost of a GT is approximately \$70/MWh. However, the fixed cost component becomes very large when amortized over a small number of hours. If a GT only runs for 10 hours in a year, its annualized average cost of production is about \$7,200/MWh. The difference in the average cost per MWh and the prices in the energy market must be recovered in the capacity markets. Furthermore, the capacity market ensures that the energy price is rarely

greater than the marginal cost of a gas turbine (\$70/MWh). Demand response drastically lowers capacity as a share of consumer power cost because it needs to recover far less in the capacity market. Demand response bids into the market at the price cap, setting price at \$1,500/MWh when it is dispatched. This makes the gap between energy payments and its cost of \$2,000/MWh far smaller than the gap experienced by a gas turbine.

Given the short history of demand response in PJM, it is difficult to determine exactly the correct marginal cost to use for demand response. Estimates of VOLL are in the range of \$2,000/MWh- \$15,000/MWh. Since VOLL are averages estimated for subgroups, demand response should have a VOLL near the bottom of this range because these consumers are volunteering to reduce load in exchange for payment. Presumably customers with the lowest cost of load curtailments would be the most likely to participate in such a program. Additionally, the cost of curtailment is above the price cap as demand response would set the PJM capacity clearing prices close to zero if it were willing to participate only based on energy market revenues.

Table 4 shows the sensitivity of capacity costs on the marginal cost of demand response. Since the marginal cost of demand response is above the price cap, demand response is in affect losing money for every hour of curtailment. This loss needs to be recovered in the capacity markets, and hence the larger the marginal cost of demand response the greater the capacity costs. Nevertheless, increasing the marginal cost of demand response from \$2,000/MWh to \$3,000/MWh had a very small effect on the shape of the supply stack, with the amount of demand response dropping from 10.1% of installed capacity to 9.6%.

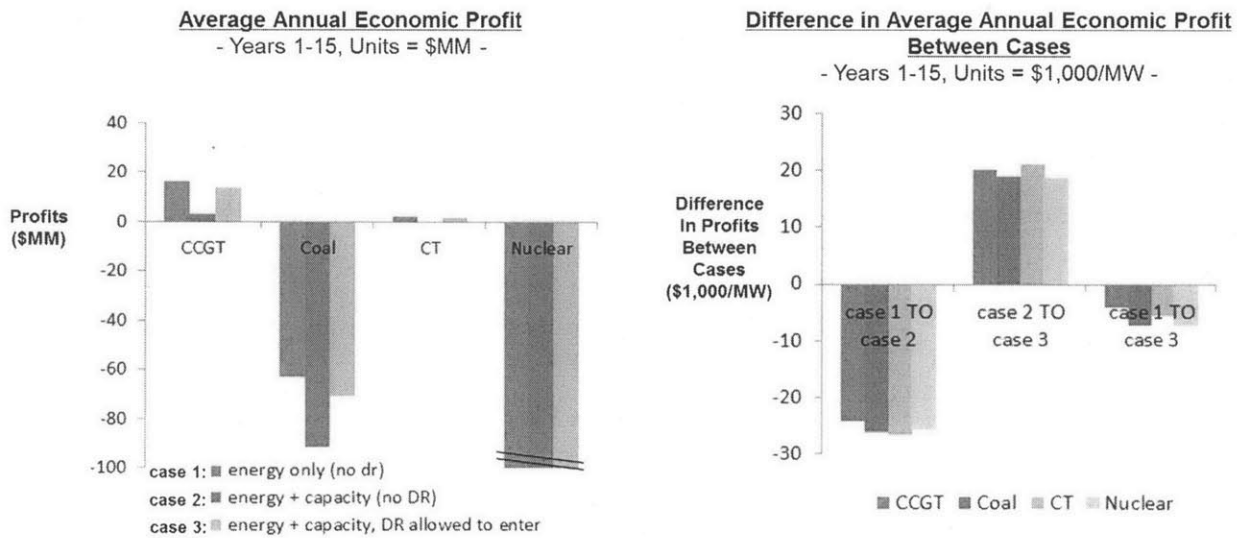
Table 4: Sensitivity of capacity cost to the marginal cost of demand response.

Marginal Cost (\$/MWh)	Capacity Costs (percent of consumer power cost)
2,000	4.0%
2,500	6.4%
3,000	9.7%

6.2: Generator Profits

Generation profits can be calculated as the difference between consumer cost (the amount consumers pays) and total generation costs.

Figure 10: Generation profits by scenario

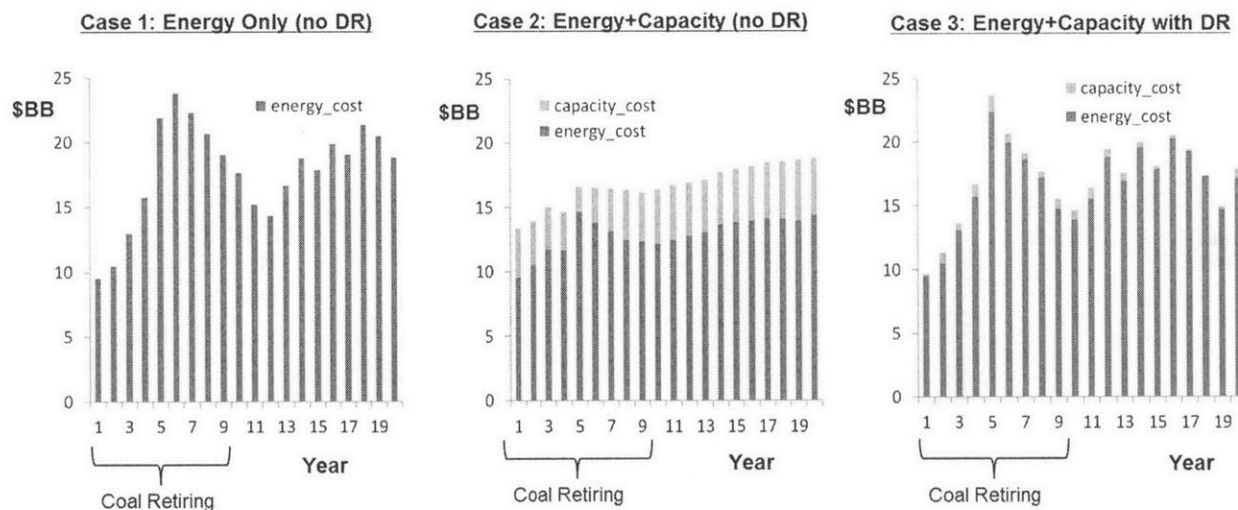


Interestingly, generator profits change across cases, with the combination of a capacity market and thermal only system (Case 2) leading to the smallest economic profit for generators. Note that economic profits tend to be small, with CCGTs earning \$16MM per year in the most profitable case on an asset worth roughly \$650MM. The losses incurred by coal and nuclear

plants are expected as these plants are not economical to build and the profit calculation includes capital costs. The graph on the right of Figure 10 shows that at a per MW basis, the capacity market and the entry of demand response have roughly the same effect on the profits of each generator type. The model finding that introducing a capacity mechanism does not favor any particular generation technology is in line with capacity market theory (Cramton et al., 2013).

In order to more fully understand why the cases produced different levels of generator profit, the time horizon was increased from 15 years to 20. Figure 11 shows that costs to consumers in Case 1 and Case 3 are substantially more sensitive to the coal retirements than in Case 2.

Figure 11: Power cost to consumers by year



The model structure, in which energy prices are based on the average of the current year and a forecast future year, prevents investment decisions from fully incorporating future retirements. This will clearly have a large effect on Case 1 (energy only, no capacity market) because energy prices are the only determinant of new plant entry. The capacity market in Case 2, however, is

forward looking and will procure adequate capacity to maintain the reliability standards. Thus, Case 2's consumer power cost at $t=1$ is much higher than Case 1 because the market is already paying for the new generation required to meet the coal retirements. Furthermore, replacing hours of unserved energy (in which prices are set at the cap) with price setting by CC and GT thermal units (\$40 - \$70/MWh) vastly reduces the sensitivity of energy prices to the generation outages. In case 3, however, demand response tends to enter instead of GT plants. This shifts the majority of capacity costs out of the capacity market and into the energy market, making the market more sensitive to the coal retirements. Additionally, the vastly different cost structure of demand response compared to thermal generation likely pushes the market further out of equilibrium. Note that hours of unserved energy in Case 3 are the same as Case 2, so the entry of demand response does not impact the capacity market's ability to maintain reliability.

Figure 12: Difference in profits between cases

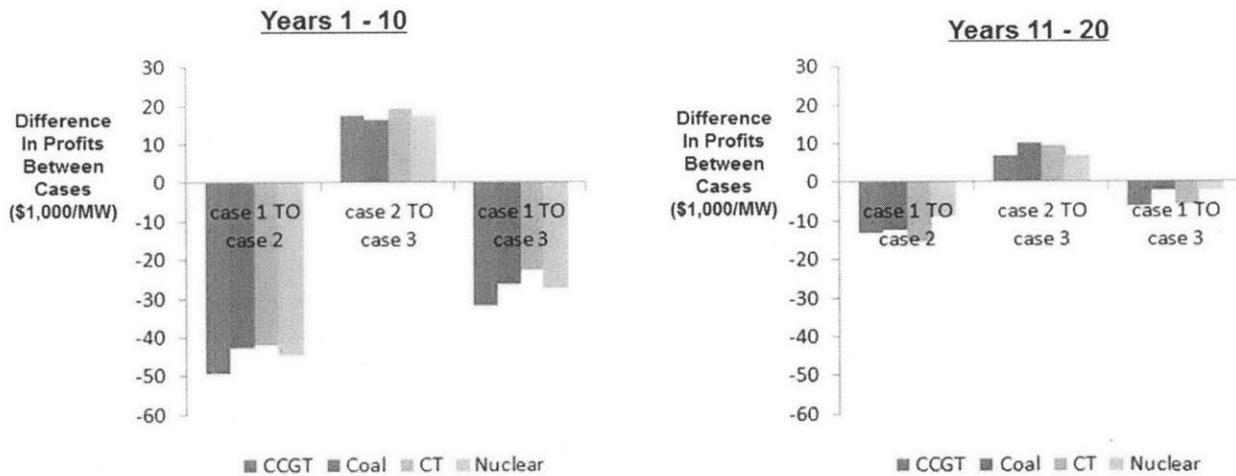


Figure 12 shows that the difference in generator profits between cases for years 1-10 (the time period with the coal retirements) are substantially larger than the difference in profits for

years 11-20. Thus, the model results are consistent with the expectation that over time, the economic profits of generators should trend towards zero in the presence of perfect competition. However, the combination of coal retirements and the partial dependence of investment decisions on historical energy prices push Cases 1 and 3 out of equilibrium for the first 10 years, creating temporary positive economic profit to generators.

6.3: Impacts on Energy Price

Altering the balance in consumer power cost between energy and capacity has a large effect on the energy price. Table 5 summarizes average energy price level and the frequency of high priced events. Adding a capacity mechanism to a primarily thermal system substantially lowers the frequency of high priced energy events, and this leads to a large reduction in the average energy price. Thus, the model finding is line with the PJM market monitor’s assertion that capacity markets depress energy prices and decrease the frequency of high priced events. However, the entry of demand response shifts the pricing regime substantially back towards prices experienced in an energy only market.

Table 5: Frequency of high priced events, years 1-15 with and without the capacity market.

	Case 1: Energy Only (No DR)	Case 2: Energy + Capacity (No DR)	Case 3: Energy+ Capacity with DR
Average Energy Price (\$/MWh)	52.57	41.33	50.2
Percentage Hours Price > \$200/MWh	1.20%	0.16%	1.00%
Percentage Hours Price at Cap	0.80%	0.06%	0.65%

Transitioning a larger percentage of the cost to consumers away from capacity and towards the more granular energy market will send a more accurate price signal to consumers.

Capacity charges are administratively recovered fees based on coincident peak load usage. However, the system peak load may not be the period in which the system is reaching maximum scarcity. For instance, the combination of generator outages, transmission outages, and high but not peak load may actually place higher stress on the system than the peak load hour with all equipment operational. The energy price, which clears at least hourly, captures these instantaneous market conditions. Capacity charges, though, are a fee assessed based on their own schedule irrespective of spot market conditions.

To explore the potential for capacity charges to send an incorrect price signal, a large loss of generation was modeled in July for $t=15$ for Case 2 (capacity market and thermal entry only, no DR) and case 3 (capacity market and thermal and DR entry). Peak load of 55.06GW was reached in June, so the capacity charge in its current PJM construct would not be assessed during the outage month. However, with 3.3GW of generation outaged, peak scarcity (measured as non-outaged generation - load) occurred on the third highest load day (July 10th).

First, consider the system with a capacity market, but without demand response. Assuming that a price sensitive consumer correctly predicts that 6/25 is the peak load day of the year, and that they apply the capacity charge to the 10 highest priced hours (HE 12 - 21), the price signal sent by the capacity payment alone is \$6,169/MWh. This is more than 4 times the price cap. Even if the consumer amortizes their capacity charge over 3 peak days, the capacity charge is sending a marginal cost signal of \$2,467/MWh. Only one of those days is within the outage period when true scarcity is occurring. Of these 30 hours, 23 have price set by a GT with at least 1.5GW of additional capacity. Based on the underlying spot market conditions, there is no need to decrease consumption in 23 of these 30 hours. Nevertheless the capacity charge's fee based nature prevents it from reacting to spot market conditions, leading it to send a strong signal

(larger than the energy price cap) to stop consuming in hours with spare capacity. The imposition of a capacity market in a thermal system without demand response shifts a large amount of remuneration into the capacity market. These large costs must be recovered, and the current fairly crude mechanism of basing them on coincident peak load creates the opportunity for large disincentives to consume in potentially the wrong hours.

Introducing demand response does not eliminate the issue of the capacity costs being allocated to hours with spare capacity, but the importance of this misallocation is severely reduced because demand response has reduced the capacity charges by roughly 85%. Furthermore, with a much larger amount of the scarcity costs running through the energy markets, the energy price during the outage rises substantially more than in the case without demand response.

Table 6: Market response to large generation outage in July (Hours Ending 8-23)

	Without Outage		With Outage		Outage Driven Increase in Cost to Load (\$MM)
	Average Energy Price (\$/MWh)	Hours at Price Cap	Average Energy Price (\$/MWh)	Hours at Price Cap	
Case 2: No Demand Response	58	0	177	18	1,382
Case3: With Demand Response	141	13	342	46	2,409

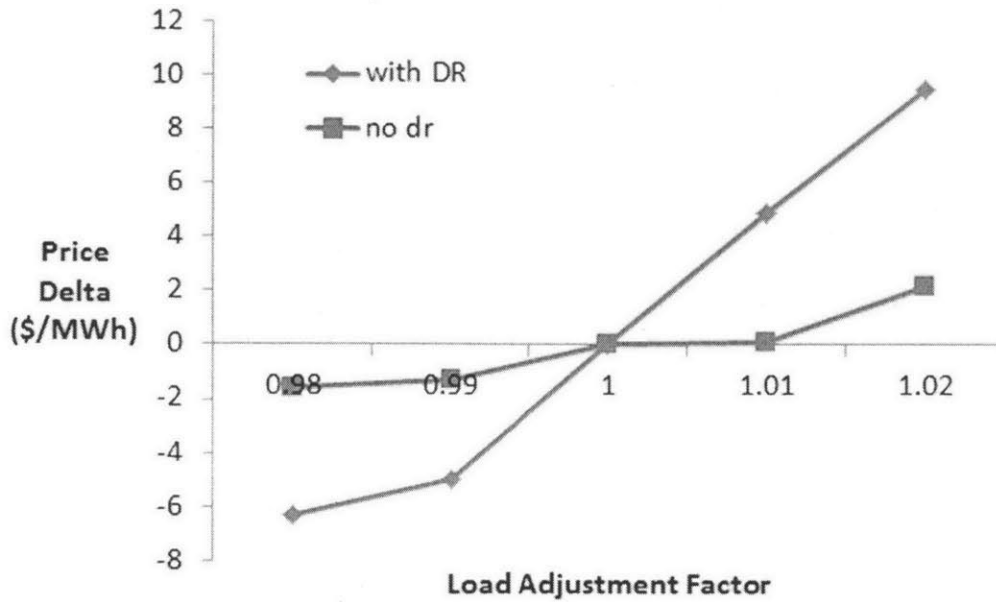
Table 6 shows modeled prices in July, for Cases 2 and 3, with and without the generation outage. The first item of note is that even without the generation outage, average energy prices in July during peak usage hours (HE 8-23) are substantially higher when demand response is allowed to

enter. The capacity costs under the current PJM construct would be assessed from the peak load day in June, so already demand response is sending a much stronger signal to reduce consumption during peak summer periods. However, if the generation outage is imposed, the absolute increase in energy price level is far higher in the demand response case (\$201/MWh vs \$119/MWh for the no demand response case). Demand response, by shifting remuneration from the capacity to the energy markets, allows the energy market to send a much stronger price signal in response to changing system conditions. Capacity is extremely tight during the outage and hence prices should rise to incentivize decreased consumption. On the other hand, consumer risk has clearly been elevated as the outage increases the energy cost to load by an additional \$1BB compared to the no demand response case. This example demonstrates that the shift towards an energy only market increases the accuracy of the price signal, but also makes prices far more sensitive to deviations from expectations and hence raises the risk for market participants.

6.4: Energy Price Sensitivity

To further examine changes in price sensitivity, the generator mix online in year 15 was used to serve load adjusted between -2% and 2% in 1% intervals.

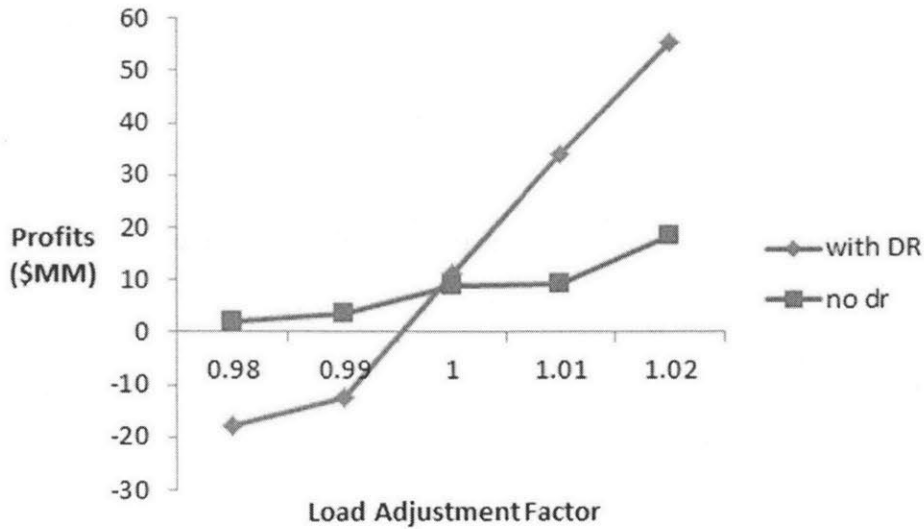
Figure 13: Average annual price sensitivity to 1% changes in load.



The traditional thermal system shows reasonable price sensitivity to load. However, the sensitivity is dwarfed by the price changes experienced in the system with demand response. A 1% decrease or increase in load changes average prices over the year by a massive -\$5.0/MWh and \$4.9/MWh, respectively. This about a 10% change in average price due to a change in load of only 1%!

Examining the average profits of the CCGT fleet shows similar increases in sensitivity. Note that profits are calculated as total revenue from energy and capacity markets minus variable costs, fixed operations and maintenance costs, and the annualized capital cost.

Figure 14: Profits of average CCGT in year 15 vs 1% changes in load.

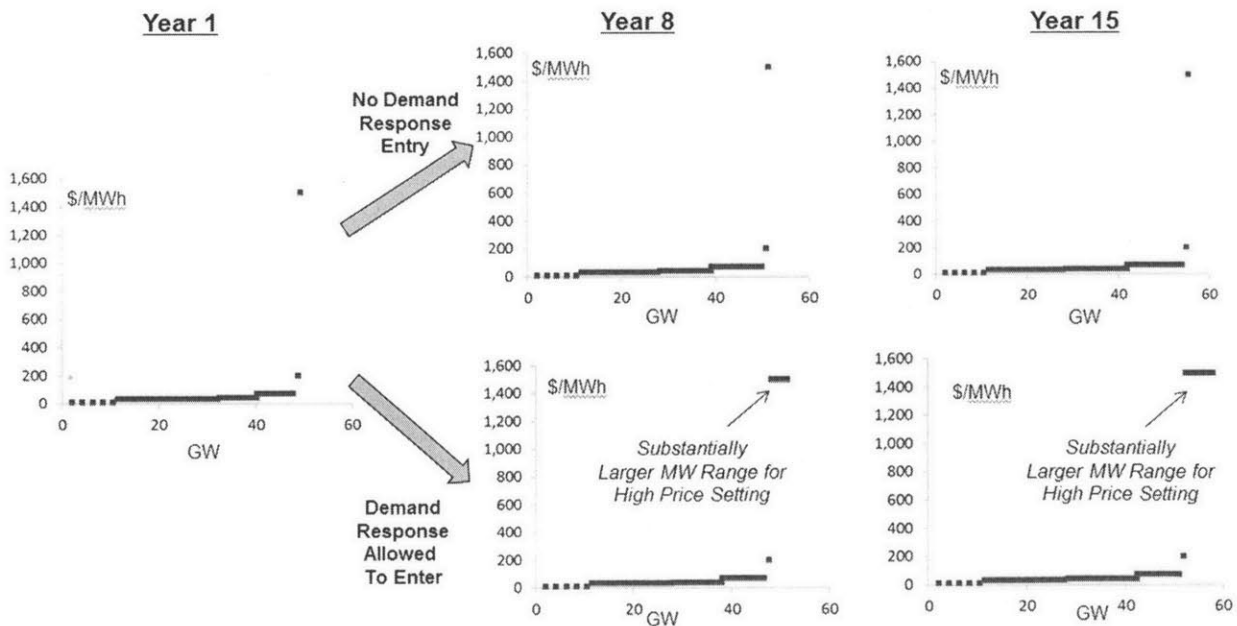


Initially, these results appear counterintuitive. Holding all else equal, increasing demand flexibility either in the form of demand response on the supply side or demand curve elasticity should lower both the energy price level and volatility. Indeed, numerous qualitative studies posited this, and numerous quantitative studies confirmed this. This study, however, is the first to combine capacity markets, generation expansion, and demand response in the form that has entered in large quantity in the PJM market. Thus, the results uncover the effect of demand response over a longer period of time in which generators exit and enter in response to market forces.

Figure 15 shows how allowing demand response to enter drastically alters the shape of the supply stack for peak load hours. In year 1, both scenarios start with the same exogenously imposed supply stack. In the no demand response scenario, new entry is comprised of both GTs and CCGTs. Importantly, a new GT has a marginal cost of about \$70/MWh; additionally, the capacity mechanism insures that peak load very rarely surpasses installed generation. Hence the

amount of supply offered above the marginal price of the GT does not increase and prices are very rarely set at the cap of \$1,500/MWh. However, when demand response is allowed to enter, it often proves to be the cheaper peaking resource, crowding out most of the new GT entry. Given that the demand response implemented in the model is based on the PJM emergency DR product, it leads to a substantial increase in the size of the stack at the price cap. This means that during the peak load hours, prices are set at the cap much more frequently, vastly increasing the remuneration flowing through the energy market. Demand response is a cheaper peaking resource than a gas turbine, but due to the interaction of its fixed and marginal cost structure, in conjunction with the imposition of a capacity mechanism, it leads to substantially higher energy prices and lower capacity prices.

Figure 15: Modeled supply curve with and without demand response.

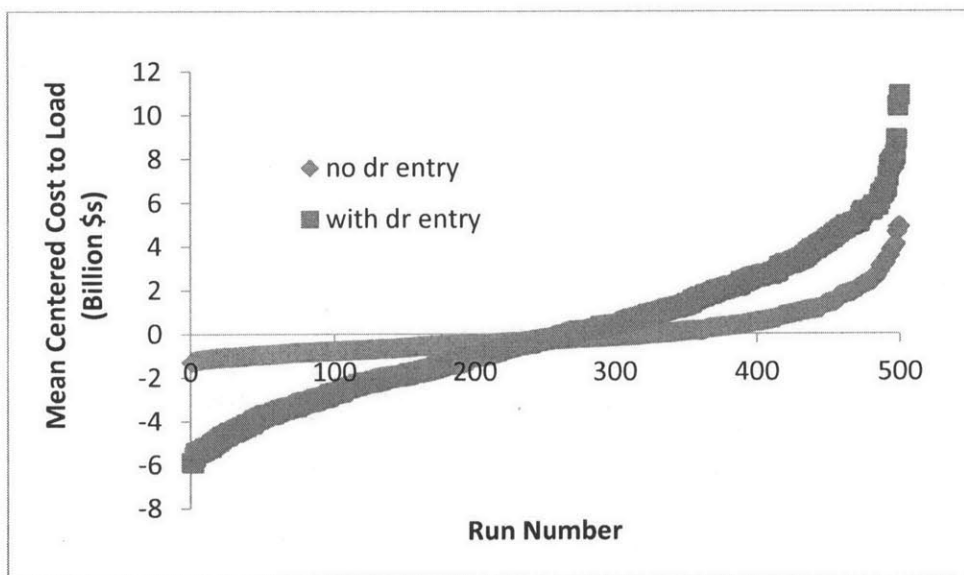


6.5: Monte Carlo Simulation

In order to assess the effect on overall volatility, a monte carlo was run sampling from different combinations of load levels and generator outages. For every monte carlo scenario, the load level was adjusted by a uniform sample between -2% and 2%, and generators were outaged at a weekly level based on PJM reported average forced outage rates by technology type. The generator supply stack was the modeled stack that would be in effect for year 16, after running the model for years 1-15. The increase in volatility due to demand response entry can be examined by comparing the output created by the supply stacks from Case 2 and Case 3.

Figure 16 shows energy cost to load to load as a difference from the mean value for 500 iterations of the monte carlo. The potential range of energy costs consumers incur in any given year is substantially larger under the demand response entry scenario.

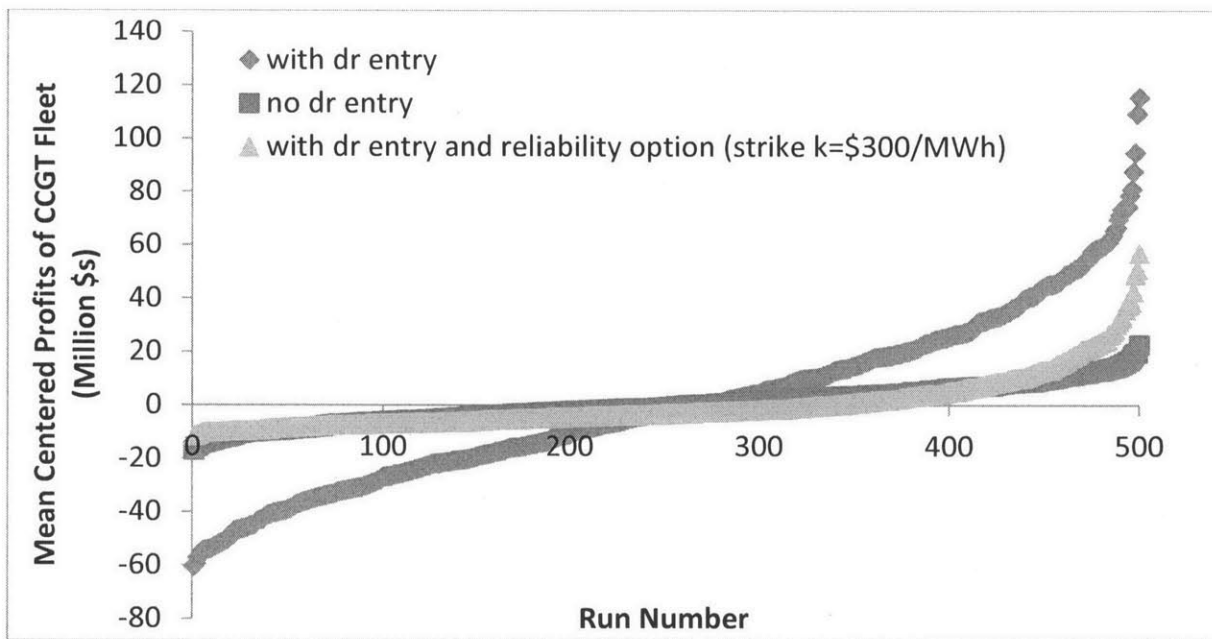
Figure 16: Mean centered energy cost to consumers from 500 iterations of the monte carlo



The introduction of demand response increases the standard deviation from \$1.1BB to \$3.1BB. The total cost of power (capacity + energy) is about \$19 billion for both the DR and no DR cases. Thus, the standard deviation is rising from about 5.6% to 16.1%.

Similarly, on the production side, the shift of scarcity rents out of the capacity market and into the energy market creates a much larger distribution of expected profits. Figure 17 shows the average profits of a CCGT in the 500 Monte Carlo iterations.

Figure 17: Mean centered average CCGT profits.



The results of the monte carlo show that the introduction of demand response will increase the volatility in cost to load and profits to generators on the order of 4x in the tails of the distribution. Similarly, demand response raises the standard deviation of the distribution from \$7.1MM to \$31.3MM. Absent efficient hedging instruments, this increase in volatility will raise the cost of capital to power generation companies and the hedging costs of utilities, leading to increased costs that will likely be passed on to consumers.

Trying to quantify the magnitude of this cost is beyond the scope of the thesis. However, warning signs do exist about the ability of market participants to effectively manage risk. First, the liquidity of power derivatives tends to be limited (Oum, Oren, & Deng, 2006), indicating that procuring substantially larger quantities of these products could be difficult. Studies also found evidence of risk premiums in electricity forward contracts. Lucia and Enguix (2008) found significant evidence of positive risk premiums in short-term electricity future prices for the Nord Pool Market (Lucia & Enguix, 2008). Similarly, Feurio and Meneu (2010) also find strong evidence of forward risk premiums in the Spanish power market (Furió & Meneu, 2010). Additionally, Oum et al. (2006) argue that electricity markets are incomplete in that certain risk factors cannot be perfectly hedged. “In particular, the volumetric risks are not traded in the electricity markets. Thus, we cannot naively adopt the classical no-arbitrage approach” (Oum et al., 2006). Oum et al. (2006) propose a framework for optimally managing this non perfectly hedgable risk, but nevertheless some financial exposure will remain open. Over time, the suite of available risk management products will likely evolve to help manage a more volatile energy price. Unfortunately, fully understanding the state of electricity risk management markets remains difficult given that the majority of the contracts trade over the counter (OTC) without the transparency available to standardized exchange based products (Deng & Oren, 2006; Oum et al., 2006).

6.6: Reliability Option Analysis

One potentially efficient method of hedging would be for the regulator to use reliability options, as opposed to the reliability only product that is currently employed in PJM. In the PJM capacity market, the ISO procures only generation capacity. However, in a reliability option mechanism, the regulator procures both capacity and a hedge for load against prices rising above

a certain level. From the supply perspective, generators are selling off their right to collect energy market revenues above a certain price when they participate in the capacity market.

If prices are extremely stable and rarely rise above the marginal cost of a thermal unit, generator's would accept a low payment for forfeiting revenue from high energy prices. However, if prices are highly volatile, as they are under the demand response scenario, generators will require a much larger payment. The primary reason why reliability options represent an efficient hedging mechanism is that generators are naturally long power price and load is naturally short. For simplicity, assume a strike price of \$300/MWh. This strike price provides load a large degree of protection compared to prices rising to the \$1,500/MWh cap, but is also high enough that most thermal generation will be dispatched if prices reach this level. If a generator sells a reliability option equal to its production capacity, it is indifferent to power price movements above the strike price when it is online. Simply put, for every dollar the energy price rises above \$300/MWh, the generator makes an additional \$1/MWh in the energy market, and loses an additional \$1/MWh on the option. The reverse is also true for consumers, assuming that the regulator has purchased reliability options equal to or greater than the current load level. By eliminating the open power price exposure above the strike price, reliability options would decrease the revenue volatility of power plants, presumably lowering their cost of capital. In the context of a reliability option, the non-perfectly hedgable risk involves the generator entering a forced outage. If the generator trips during a high priced event, the natural hedge disappears and the generator has an open short on power price exposure. Thus, the generator will include some risk premium in its sale price of a reliability option based on its expected outage probability.

Altering the capacity market in the model to include a reliability option has a very small effect in Case 2 (no demand response entry). Note that the base model does not include

generation outages, so generators are bidding into the market in a risk neutral fashion. Incorporating the reliability option increases the average capacity clearing price over years $t=1$ to $t=15$ from \$7.9/MWh to \$8.4/MWh, roughly a 6% increase. The small premium for purchasing the option insurance in Case 2 is due to the coupling of a capacity mechanism with a thermal only system creating relatively stable prices. The results for the demand response entry scenario, however, are wildly different. Incorporating the reliability option led the average capacity market clearing price over the 15 years to rise from \$1.5/MWh to \$5.9/MWh, a 390% increase. Interestingly, implementing the reliability option, which increased the capacity price by a factor of 4x, made only a negligible difference on the generation mix. Figure 17 shows that with the reliability option, the distribution of CCGT profits is roughly inline with the previous no-demand response case. Thus, reliability options are an effective tool available to a regulator who would want to mitigate the increased revenue volatility generators face due to demand response entry. The downside to this approach is twofold. First, the ISO is forcing all generation and load to accept the same risk management instrument, regardless of their preferences. Certain consumers may want to purchase a more expensive hedge with a lower strike price, while other consumers may want to purchase no hedge. Second, the additional cost of the option component of the reliability product must be administratively recovered, and as was demonstrated with the generation outage example, this can send a distortionary price signal to consumers.

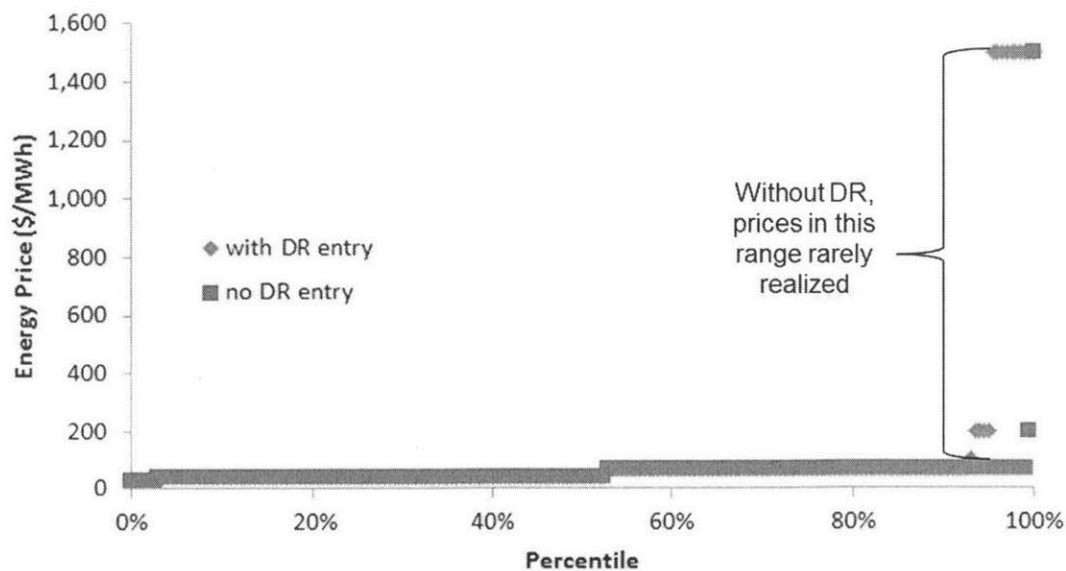
Chapter 7: Discussion and Policy Implications

This study demonstrated the initially counterintuitive finding that large scale demand response penetration in the form of a peaking product will increase energy price level and volatility. This occurs because, over time, demand response will displace natural gas turbines as

the extreme peaking technology. Recent evidence from PJM combined with previous econometric studies of VOLL indicate that demand response will likely bid at the price cap, far higher than the marginal cost of a gas turbine. Calling upon demand response to solve the system instead of a gas turbine will drastically increase the frequency of hours the energy price is set to the cap instead of the much lower marginal cost of a thermal plant. This will shift the market substantially towards an energy only framework. The model developed for this study finds that capacity market revenues will likely shrink by over 80%, with almost all of this value moving into the energy market.

The primary benefit provided by the shift towards an energy only market will be a more accurate energy price signal. In times of scarcity, the penetration of demand response will lead to higher price setting than in a thermal only system, sending a stronger and more transparent signal to load to reduce consumption.

Figure 18: Distribution of energy prices from the Monte Carlo in summer months (June-August) for hours 8 – 23.



The distribution in Figure 18 shows that in most hours, the no DR and DR cases produce identical prices because the same generating technology is marginal. However, in the case with DR entry, the top 5% of hours are at the price cap (primarily with DR setting price as opposed to unserved energy). In the case with no DR entry, prices are set above \$200/MWh under 0.5% of the time. This is a dramatic difference in the price signal sent to consumers during times of tight supply. These high prices may incentivize the development of more price sensitive demand as consumers see substantial opportunities for savings by cutting usage during peak periods.

Additionally, Faruqui and Palmer’s (2011) finding of a strong correlation between peak usage reductions and the ratio between peak to off-peak prices suggests that residential demand elasticity occurs at high electricity price levels. Table 7 summarizes the general trend that Faruqui and Palmer (2011) find.

Table 7: Peak reduction due to ratio of peak to off-peak prices.

Peak to Off-Peak Price Ratio	Peak Reduction
2	6%
4	11%
10	18%

The numerical values for on and off-peak prices are not given in Faruqui and Palmer (2011), but the table shows that a larger dynamic range leads to larger peak load reductions. By increasing the frequency of high priced hours, demand response could be a powerful force in developing more true demand elasticity. In a system without demand response, substantially more of the consumer power cost must be recovered through capacity charges, which are both less transparent than real time energy prices and largely divorced from current spot market conditions.

Due to the potential benefits of a much larger dynamic range in wholesale prices, regulators must carefully evaluate policies that may limit the range of realized prices. For example, after demand response set the LMP to the price cap during a few hours in summer 2013, PJM Vice President of Market Operations Stu Bresler commented that “the offer cap [\$1,800/MWh] for emergency DR is probably too high” (RTO Insider 2013). The \$1,000/MWh offer cap on generators, however, is premised on the notion that the marginal cost of a thermal unit will always be below this value.² The marginal cost or VOLL of demand response, however, is largely unknown to the ISO. Thus, the only limitation on its bid should be the price cap. Additionally, estimates of VOLL tend to be in the \$2,000 - \$15,000/MWh range, and thus it should be expected that demand response bids into the market at these high prices. Despite some initial apprehension about demand response setting the LMP to the price cap, PJM has not pursued tariff changes that would limit emergency demand response bids.

The potential negative consequences of the transition towards an energy only market is that the increased volatility will substantially raise the level of risk experienced by market participants. The cost of capital to finance new generation may increase and generators and utilities may be forced to purchase financial hedging products with substantial risk premiums. A purely free market view suggests that the electricity market will efficiently manage the risk. Indeed, many other industries such as automobiles and hotels operate without regulatory intervention to lower risk. However, Hobbs et al. (2007) finds capacity markets that lower volatility to risk averse investors create substantial savings to consumers. While several studies have investigated the various risk management products available to the power sector (Deng & Oren, 2006; Oum et al., 2006), a substantial literature review did not uncover any academic

² During January 2014 when regional gas price spikes raised the marginal cost of some thermal units above this level, PJM made an emergency filing with FERC to remove the \$1,000/MWh offer cap.

studies documenting industry practices. Given that many of these transactions occur over the counter through a web of brokers and financial firms, the data would likely be very difficult to compile.

Incorporating reliability options into the PJM capacity market would smooth generators' cash flows. However, simply purchasing reliability options to cover expected peak load will introduce the same price signal distortions to load that the capacity mechanism without DR entry produced. A large amount of the cost of power is moved out of the energy market and into the capacity market, where it must be administratively recovered. Additionally, all generators and consumers would be forced into purchasing the same risk management product regardless of their risk preferences.

A more efficient approach would likely recognize that reliability options are two separate products. They are a reliability product that maintains the desired reliability standard, and a financial hedging product. Beyond the transaction costs of running multiple auctions, there is little reason for why the two products should be packaged together, which forces the ISO to procure the same quantity of both products. The ISO could purchase the quantity of the reliability product required to maintain the NERC reliability standards, and then purchase adequate options for customers who choose to purchase the price protection. Multiple auctions could even be held to allow for options with different strike prices to provide finer grained risk management. More sophisticated customers would compare the price they were paying for the option verse the risk of leaving the electricity exposure open. Customers with greater demand elasticity would face lower risk from high prices because they can reduce their consumption during peak periods. Over time this would likely encourage the growth of demand elasticity while providing insurance to consumers less able to adjust their usage.

Several more operationally focused changes must also be made to ISO procedures to ensure the smooth operation of power markets under greater demand response penetration. Within the capacity market framework, demand response must provide a comparable capacity resource to generators. There are several reasons why demand response may be a less valuable form of capacity than a standard generator. Initial concerns focused on demand response simply not curtailing when the ISO required it. In 2012 PJM reported the compliance of the emergency demand response program as 104% (PJM 2012 SOM p.179). That is to say, demand response delivered 104% of the demand reduction it was asked to deliver. The 2012 State of the Market Report points out, however, that this calculation is potentially flawed. PJM set the value of the 13% of demand response that increased consumption when they were ordered to curtail to 0, as opposed to a positive value (Monitoring Analytics, 2013). Thus, the decrease in total power consumed was less than 104% of what PJM operators had curtailed, but the data to calculate the true value is not publicly available.

Another key difference between demand response and generation is that generators are required to bid into the Day Ahead market and are dispatched at a nodal level with extremely fine grained control from the central operator. Emergency demand response does not bid into the Day Ahead market, and is dispatched in rough blocks at the zonal level at the discretion of the system operator. As demand response becomes a greater fraction of the installed capacity, the imprecise nature of its dispatch could potentially lead to grid instabilities due to the ISO not having complete control over the volume or location of the curtailment.

Demand response should receive a lower capacity payment per MW than generation if it is not providing a capacity product that is truly fungible with generation capacity. PJM is already in the process of updating its capacity market clearing algorithm to allow for differential clearing

prices between demand response that is only available during the summer and year round resources. The auction could be updated to allow all demand resources to clear at a lower price if generation capacity is determined to be more valuable than demand response capacity. A second approach would be to incorporate demand response more fully into the Real Time and Day Ahead algorithms, but it will place larger requirements on demand response providers to more finely tune their reductions to the ISO instructions.

Perhaps most importantly, regulators must realize that this change in volatility will probably occur in the next few years. In the 2015/2016 capability year, demand response will comprise over 9% of the installed capacity. The risk management protocols of utilities and retailers are likely based on historical levels of market volatility and may be unprepared to manage the risk brought on by demand response price setting. Various capital adequacy and collateral requirements should be reviewed to ensure that a wave of bankruptcies do not occur due to an unexpected increase in volatility and energy price level. Oren (2004) notes that a retailer in Texas went bankrupt due to an ice storm driving unexpectedly high prices, and more generally the protection provided by Chapter 11 bankruptcy provides some incentive to under-hedge. A financial debacle brought on by increased demand response could create a regulatory environment hostile to future demand side participation. This could have long term consequences in that greater demand side participation, particularly the ability to shift load towards periods of high renewable output, has been considered an important component of integrating variable output renewables.

This study has highlighted the need for further research in a few areas. First and foremost, more research is needed assessing the apparent disconnect between the highly flexible and cheap demand response often assumed to exist verse the VOLL estimates and actual

payments that emergency demand response currently receives. The future balance of energy and capacity prices and energy price volatility likely depends on the prices required to decrease consumption. If historical estimates of the VOLL prove accurate, a paradigm of increased volatility will likely persist. However, a true smart grid that shifts appliance use to off-peak periods at very low cost would result in much lower volatility. The drop in demand response participation for the 2016/2017 planning year raises real concerns that demand response providers have VOLLs well above the price cap. On the other hand, the 109 pilot studies evaluated by Faruqui and Palmer (2011) suggest that at least some subgroup of residential customers decrease peak consumption in response to realized prices. More research is required to better understand the sources of these inconsistencies and the true potential for demand response and price sensitive demand.

Additional research is also required into the risk management processes of market participants in liberalized electricity markets. There appears very little literature on the risk management practices currently utilized in the electricity industry, and the degree to which spot market risk is hedged through long term contracts. The power generation industry's vocal support for capacity markets based on their revenue smoothing properties indicates that industry does prefer the guaranteed payments of capacity markets over volatile energy payments. From a societal perspective, some increase in capital costs due to demand response driven volatility may be an acceptable tradeoff for sending more accurate price signals to consumers. The current literature on both demand response and risk management, however, is not currently adequate to facilitate this evaluation.

The entry of demand response will shift PJM towards an energy only market. This transition is due to the high marginal cost but low fixed cost structure of demand response. The much

larger dynamic range in power prices driven by demand response may create the incentive required for development of true price sensitive demand that responds only to the energy price and not ISO curtailment orders. However, this shift towards an energy only market will increase the volatility and hence risk experienced by many market participants. ISO's could smooth this transition by creating a separate option product that provides both a hedge to consumers that desire insurance and a fixed revenue stream to the generators selling the product. Importantly, if any regulatory intervention is required it should be with the aim providing risk management without distorting the price signal.

Bibliography

- Albadi, M., & El-Saadany, E. (2007). Demand response in electricity markets: An overview. *Power Engineering Society ...*, 1–5. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4275494
- American Public Power Association. (2012). Power Plants Are Not Built On Spec: An Analysis of New Electric Generation Projects Constructed in 2011, (March).
- Aydin, B. O., Graves, F., & Celebi, M. (2013). *Coal Plant Retirements: Feedback Effects on Wholesale Electricity Prices*.
- Battle, C., Vázquez, C., Rivier, M., & Pérez-Arriaga, I. J. (2007). Enhancing power supply adequacy in Spain: Migrating from capacity payments to reliability options. *Energy Policy*, 35(9), 4545–4554. doi:10.1016/j.enpol.2007.04.002
- Besser, J. G., Farr, J. G., & Tierney, S. F. (2002). Long-Term Generation Adequacy : Why an ICAP Mechanism is Needed as Part of Standard Market Design, 6190(September).
- Borenstein, S., & Bushnell, J. (2000). Electricity restructuring: deregulation or reregulation. *Regulation*, 23(2), 46–52. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/rcatorbg23§ion=30
- Borenstein, S., Jaske, M., & Rosenfeld, A. (2002). Dynamic pricing, advanced metering, and demand response in electricity markets. *Recent Work, Center for the Study of Energy Markets, University of California Energy Institute, UC Berkeley*. Retrieved from <http://escholarship.org/uc/item/11w8d6m4.pdf>
- Brattle Group. (2011). Second Performance Assessment of PJM's Reliability Pricing Model. ... *Group for PJM ...*. Retrieved from <http://www.energycollection.us/Energy-Capacity/Second-Performance-Assessment.pdf>
- Bushnell, J., Hobbs, B. F., & Wolak, F. a. (2009). When It Comes to Demand Response, Is FERC Its Own Worst Enemy? *The Electricity Journal*, 22(8), 9–18.
- Cappers, P., Goldman, C., & Kathan, D. (2010). Demand response in US electricity markets: Empirical evidence. *Energy*, 35(4), 1526–1535. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0360544209002461>
- Caves, D., Eakin, K., & Faruqui, A. (2000). Mitigating price spikes in wholesale markets through market-based pricing in retail markets. *The Electricity Journal*, 6190(April), 13–23. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1040619000000920>
- Centolella, P. (2010). The integration of Price Responsive Demand into Regional Transmission Organization (RTO) wholesale power markets and system operations. *Energy*, 35(4), 1568–1574. doi:10.1016/j.energy.2009.06.046

- Chao, H., & DePillis, M. (2013). Incentive effects of paying demand response in wholesale electricity markets. *Journal of Regulatory Economics*, 43(3), 265–283. doi:10.1007/s11149-012-9208-1
- Cramton, P., Ockenfels, A., & Stoft, S. (2013). Capacity Market Fundamentals. *Economics of Energy & Environmental Policy*, 2(2), 1–21. doi:10.5547/2160-5890.2.2.2
- Cramton, P., & Stoft, S. (2006). The Convergence of Market Designs for Adequate Generating Capacity with Special Attention to the CAISO 's Resource Adequacy Problem.
- Davis, L. (2013). *Comments of Lee Davis, President of NRG's East Region*. Federal Energy Regulatory Commission.
- De Nooij, M., Lieshout, R., & Koopmans, C. (2009). Optimal blackouts: Empirical results on reducing the social cost of electricity outages through efficient regional rationing. *Energy Economics*, 31(3), 342–347. doi:10.1016/j.eneco.2008.11.004
- Deng, S., & Oren, S. (2006). Electricity derivatives and risk management. *Energy*, 31(6-7), 940–953. doi:10.1016/j.energy.2005.02.015
- Eager, D., Hobbs, B., & Bialek, J. (2012). Dynamic modeling of thermal generation capacity investment: application to markets with high wind penetration. *Power Systems, IEEE ...*, 27(4), 2127–2137. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6179363
- Faruqui, A., & Palmer, J. (2011). Dynamic Pricing of Electricity and its Discontents. *The Brattle Group, Inc.*
- Federal Energy Regulatory Commission. (2012). Order on Compliance Filing.
- Federal Energy Regulatory Commission. (2014). Order on Proposed Tariff Changes.
- Furió, D., & Meneu, V. (2010). Expectations and forward risk premium in the Spanish deregulated power market. *Energy Policy*, 38(2), 784–793. doi:10.1016/j.enpol.2009.10.023
- Holodak, J. G. (2013). *Statement of James G. Holodak, Jr. Vice President, Regulatory Strategy & Integrated Analytics National Grid USA*. Federal Energy Regulatory Commission.
- Hummon, M., Palchak, D., Denholm, P., Jorgenson, J., Olsen, D. J., Kiliccote, S., ... Ma, O. (2013). *Grid Integration of Aggregated Demand Response , Part 2 : Modeling Demand Response in a Production Cost Model Grid Integration of Aggregated Demand Response , Part 2 : Modeling Demand Response in a Production Cost Model*.
- Joskow, P. (2006). Competitive electricity markets and investment in new generating capacity. *AEI-Brookings Joint Center Working Paper*, (April). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=902005

- Joskow, P. (2006). Markets for Power in the United States: An Interim Assessment. *Energy Journal*, 27(1). Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=01956574&AN=19449973&h=GrcrdlNKHj7GIdIDKUH9x7H%2BAMjRNjkNGd2ugSb2z3a5bKVpnz4uTdY5tam1BqAney9RIfX6D5t8rTAcOamhSg%3D%3D&crl=c>
- Joskow, P., Bohi, D., & Gollop, F. (1989). Regulatory failure, regulatory reform, and structural change in the electrical power industry. *Brookings Papers on Economic Activity*. Retrieved from <http://www.jstor.org/stable/2534721>
- Joskow, P. L. (2012). Creating a Smarter U.S. Electricity Grid. *Journal of Economic Perspectives*, 26(1), 29–48. doi:10.1257/jep.26.1.29
- Joskow, P., & Tirole, J. (2004). Reliability and competitive electricity markets. *Center for the Study of Energy Economics, UC Berkeley*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1756-2171.2007.tb00044.x/abstract>
- Joskow, P., & Tirole, J. (2007). Reliability and competitive electricity markets. *The Rand Journal of Economics*, 38(1), 60–84. doi:10.1111/j.1756-2171.2007.tb00044.x
- Kirschen, D., & Strbac, G. (2000). Factoring the elasticity of demand in electricity prices. *Power Systems, ...*, 15(2), 612–617. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=867149
- Leahy, E., & Tol, R. S. J. (2011). An estimate of the value of lost load for Ireland. *Energy Policy*, 39(3), 1514–1520. doi:10.1016/j.enpol.2010.12.025
- Lucia, J., & Enguix, H. i. (2008). Short-term electricity futures prices: evidence on the time-varying risk premium. *Working Paper, Available at http://ssrn.com/abstract=1014035*. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=2602554&orden=158619&info=link>
- Monitoring Analytics, L. (2013). *2012 State of the Market Report for PJM - Volume 2 (Vol. 2)*.
- Neuhoff, K., & Vries, L. De. (2004). Insufficient incentives for investment in electricity generations. *Utilities Policy*, 12(4), 253–267. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0957178704000682>
- North American Reliability Council. (2012). *2012 Long-Term Reliability Assessment*.
- Oum, Y., Oren, S., & Deng, S. (2006). Hedging quantity risks with standard power options in a competitive wholesale electricity market. *Naval Research Logistics*, 53(7), 697–712. doi:10.1002/nav.20184

- Palmintier, B., & Webster, M. (2014). Heterogeneous Unit Clustering for Efficient Operational Flexibility Modeling. *IEEE Transactions on Power Systems*, 29(3). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6038963
- Pérez-Arriaga, I., & Meseguer, C. (1997). Wholesale marginal prices in competitive generation markets. *Power Systems, IEEE ...*, 12(2). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=589661
- PJM. (2010). Reliability Pricing Model LSE Obligations.
- PJM. (2013a). Emergency Demand Response (Load Management) Performance Report.
- PJM. (2013b, May 23). Shortage Pricing Refresher Training.
- Rassenti, S. J., Smith, V. L., & Wilson, B. J. (2003). Controlling market power and price spikes in electricity networks: Demand-side bidding. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), 2998–3003. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=151455&tool=pmcentrez&rendertype=abstract>
- Rodilla, P., & Batlle, C. (2013). Security of Generation Supply in Electricity Markets. In I. J. Pérez-Arriaga (Ed.), *Regulation of the Power Sector* (pp. 581–622). London: Springer London. doi:10.1007/978-1-4471-5034-3
- RTO Insider (9/3/2013). PJM: Demand Response Price Cap Too High. <http://www.rtoinsider.com/pjm-demand-response-price-cap-high/> (last accessed on 5/7/2014).
- Sener, A., & Kimball, S. (2007). Reviewing progress in PJM’s capacity market structure via the new reliability pricing model. *The Electricity Journal*, (August 2005), 40–53. Retrieved from <http://www.sciencedirect.com/science/article/pii/S104061900700139X>
- Simshauser, P. (2009). Capital adequacy , ETS and investment uncertainty in the Australian power market. *The Electricity Journal*, 23(1), 1–18.
- Singh, K., Padhy, N., & Sharma, J. (2011). Influence of price responsive demand shifting bidding on congestion and LMP in pool-based day-ahead electricity markets. *Power Systems, IEEE ...*, 26(2), 886–896. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5580103
- Testa, D. (2014, January 27). UPDATE: As PJM leans on coal plants in cold, AEP questions market rules. SNL. Retrieved from <http://www.snl.com/InteractiveX/Article.aspx?cid=A-26653967-11813>
- US Energy Information Administration. (2010). *Updated Capital Cost Estimates for Electricity Generation Plants*.

Ventosa, M., Linares, P., & Pérez-arriaga, I. J. (2013). Power System Economics. In I. J. Pérez-Arriaga (Ed.), *Regulation of the Power Sector*. London: Springer London. doi:10.1007/978-1-4471-5034-3

Wu, L. (2013). Impact of price-based demand response on market clearing and locational marginal prices. *IET Generation, Transmission & Distribution*, 7(10), 1087–1095. doi:10.1049/iet-gtd.2012.0504

Appendix 1: Clustered Unit Commitment

The classical unit commitment problem is formulated as a mixed integer program as follows:

Sets

i generator index
 h hour index

Parameters

C_i^{marg} marginal cost of running generator
 C_i^{start} generator start up cost
 V_i^{max} gen max output
 V_i^{min} gen min output
 $ramp_i^{up}$ gen ramp up limit
 $ramp_i^{down}$ gen ramp down limit
 min_up_i minimum number of hours gen must stay on for
 min_down_i minimum number of hours gen must stay off for
 D_h demand in hour h
 W_h wind output in hour h

Variables

Z total cost
 $X_{i,h}$ output of generator i in hour h
 eu_h unserved energy
 $curt_h^{wind}$ wind curtailment factor
 $U_{i,h}^{state}$ binary, 1 if generator is on
 $U_{i,h}^{up}$ binary, 1 if generator start up
 $U_{i,h}^{down}$ binary, 1 if generator shut down

Objective Function

$$\min_X Z = \sum^{i,h} X_{i,h} * C_{i,h} + \sum^{i,h} U_{i,h}^{up} * C_{i,h}^{start} + \sum^h eu_h * 1.500$$

s.t.

system balance constraint

$$D_h = \sum_1^i X_{i,h} + W_j * curt_h^{wind} + eu_h$$

gen max output constraint

$$X_{i,h} + X_{i,h}^{regup} + X_{i,h}^{secup} \leq V_i^{max} * U_{i,h}^{state}$$

gen min output constraint

$$X_{i,h} - X_{i,h}^{regdown} - X_{i,h}^{secdown} \geq V_i^{min} * U_{i,h}^{state}$$

generator ramp up limit

$$X_{i,h} - X_{i,h-1} \geq ramp_i^{up} + (1 - U_{i,h}^{state}) * 500$$

generator ramp down limit

$$X_{i,h} - X_{i,h-1} \geq -ramp_i^{down} - (1 - U_{i,h}^{state}) * 500$$

relationship between generator state, start, and shut down dummy variables

$$U_{i,h}^{state} = U_{i,h-1}^{state} + U_{i,h}^{up} - U_{i,h}^{down} \forall h > 1$$

generator min up time constraint

$$\sum_{h-min_{up_i}+1}^h U_{i,h}^{up} \leq U_{i,h}^{state}$$

generator min down time constraint

$$\sum_{h-min_{up_i}+1}^h U_{i,h}^{down} \leq 1 - U_{i,h}^{state}$$

Clustering Formulation

The individual generator index i is replaced with a cluster identifier \hat{i} . Additionally, the binary variables $U_{i,h}^{state}$, $U_{i,h}^{up}$, and $U_{i,h}^{down}$ are replaced with integer variables $\hat{U}_{\hat{i},h}^{state}$, $\hat{U}_{\hat{i},h}^{up}$, and $\hat{U}_{\hat{i},h}^{down}$. Most of the equations only require this change in index, however, the following equations also must be adjusted.

generator ramp up limit

$$X_{\hat{i},h} - X_{\hat{i},h-1} \geq \hat{U}_{\hat{i},h-1}^{state} * ramp_{\hat{i}}^{up} + \max(V_{\hat{i}}^{min}, ramp_{\hat{i}}^{up}) * \hat{U}_{\hat{i},h}^{up}$$

generator ramp down limit

$$X_{\hat{i},h-1} - X_{\hat{i},h} \geq \hat{U}_{\hat{i},h-1}^{state} * ramp_{\hat{i}}^{down} + \max(V_{\hat{i}}^{min}, ramp_{\hat{i}}^{down}) * \hat{U}_{\hat{i},h}^{down}$$

For the minimum down time equation, the difference between the number of units in each cluster $n_{\hat{i}}$ (as opposed to 1 in the standard formulation) and currently committed must be found.

$$n_{\hat{i}} - \hat{U}_{\hat{i},h-1}^{state} \geq \sum_{i-\min_down_i}^i \hat{U}_{i,h}^{down}$$