Title
Economic Analyses of Three Energy Policy Problems

Author
Accordino, Megan H.

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Economic Analyses of Three Energy Policy Problems

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Megan Henderson Accordino

2015

Abstract of the Dissertation

Economic Analyses of Three Energy Policy Problems

by

Megan Henderson Accordino

Doctor of Philosophy in Economics

University of California, Los Angeles, 2015

Professor Hugo Andres Hopenhayn, Chair

The essays included in this dissertation analyze three policy issues that have been frequently discussed in recent years. Chapter One analyzes the effects and likelihood of a particular type of manipulation with which the Federal Energy Regulatory Commission has become increasingly concerned and illustrates a potential screen for such manipulation. Chapter Two analyzes the effect a federal climate policy might have given the many state climate policies that are already in place. Finally, Chapter Three examines whether it is economically justified to encourage investment in energy storage (instead of flexible natural gas generation) to compensate for the increasing variability in wind generation.

Chapter 1: I examine the incentives for and detection of manipulation in a spot market with a related futures market. I find that the manipulability of spot prices depends on the amount of hedging in the futures market relative to the amount of non-strategic speculation. When spot prices are manipulated, the price impact of a trade declines because manipulation increases the relative amount of unpredictable uninformative trading. I employ this implication of the model as a screen for manipulation and test its efficacy with data from a recent case against BP who allegedly manipulated natural gas prices in Texas in 2008. Using several measures of the price impact of a trade, I find that in the period in which the allegedly manipulative strategy was profitable, price impact in the allegedly manipulated market was lower than in other periods and lower than in other

nearby markets.

Chapter 2: We analyze the effect of various combinations of state and national emissions policies on national emissions of a global pollutant, specifically, greenhouse gas emissions. We highlight the effect of unintended increases in out-of-state emissions on the efficacy of overlapping state policies. We show that emission taxes do not necessarily prevent a completely offsetting increase in out-of-state emissions when states add a state-level emissions tax to the national emissions tax. In particular, states small relative to their market will be unable to reduce national emissions with a state-level CO_2 tax or a system of tradable permits. However, under a national cap-and-trade regime that allows states to be carved out, a state of any size can reduce national emissions by setting a tighter state cap. This combination yields a lower total cost than the equivalent combination of national and state CO_2 taxes (when possible) but increases the cost to consumers outside the market.

Chapter 3: Since 2000, a majority of U.S. states have implemented Renewable Portfolio Standards (RPS) mandating a share of electricity consumption that must be served by generation from qualifying renewable resources. Of the 116,000 GWh increase in renewable generation (excluding hydroelectric) in the U.S. since 2002, 94 percent was generated by wind turbines. However, generation from wind turbines is variable and difficult to predict. As supply must always equal demand in electricity markets, other generators must adjust their output to compensate for the variation in the generation from wind. To increase the ability of the grid to absorb the unpredictable intermittency of wind generation, many have suggested an increase in energy storage capacity. In this paper, I compare the cost-effectiveness of increasing energy storage versus increasing flexible natural gas-powered generation to compensate for wind generation's increasingly large fluctuations in supply using a structural partial-equilibrium model of the electricity energy market simulated using data from the Pennsylvania Jersey Maryland (PJM) Interconnection. I find that due to the much higher capital cost of energy storage compared with natural gas generation, energy storage is cost-effective only if natural gas prices are predicted to be high (e.g. at 2008 levels of $9 per mmBtu).

The dissertation of Megan Henderson Accordino is approved.

Deepak Rajagopal

Pierre Olivier-Weill

Connan Andrew Snider

Hugo Andres Hopenhayn, Committee Chair

University of California, Los Angeles

2015

To Nick: thank you for always believing I could finish this dissertation despite my many doubts and protestations. You are now allowed to say "I told you so!" And to my parents who predicted I would love economics before I even knew what it was.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to especially thank my advisors Hugo Hopenhayn and Connan Snider for going through my endless list of ideas for how to analyze manipulation in commodity markets and for their suggestions about how to market Chapter One to a wider audience. Additionally, I would like to thank Pierre Olivier-Weill for his encouraging advice on my theoretical model and his suggested revisions to Chapter One. I am extremely grateful to Deepak Rajagopal, my co-author on Chapter Two, not only for his assistance in writing Chapter Two but also for introducing me to the publication process and encouraging me to attend and present at conferences. For guiding me through my first attempt at a research paper, Chapter Three, I would like to thank Mauricio Mazzocco. And last of all, special thanks go out to all my classmates who encouraged me to keep going and listened to my woes, research ideas, and presentations.

Chapter Two was co-authored with Deepak Rajagopal, who directed the project. The final version of the paper is forthcoming in the *Energy Journal* as Accordino, M. and D. Rajagopal (2015) When a National Cap-and-Trade Policy with Carve-out Provision May Be Preferable to a National CO_2 Tax. *Energy Journal* Vol. 36 Number 3. DOI: `http://dx.doi.org/10.5547/01956574.36.3.macc`. We thank our anonymous referees for insightful comments that helped improve the paper significantly. Megan Accordino was supported by the National Science Foundations Integrative Graduate Education Research Traineeship on Clean Energy for Green Industry (DGE-0903720) and the California NanoSystems Institute. We also thank the University of California Center for Energy and Environmental Economics and University of California Los Angeles Faculty Research Grants for partial financial support.

Vita

Education

2012-2015 Ph.D. Candidate, University of California, Los Angeles

2010-2012 M.A., Economics, University of California, Los Angeles

2003-2007 B.A., Economics and International Studies, Case Western Reserve University. *Summa Cum Laude*

2005-2006 The General Course, London School of Economics

Professional Experience

2013-2014 Statistical Consultant, 20/20 Business Vision

2009-2010 Energy Analyst, Economists Incorporated

2007-2009 Research Associate, Economists Incorporated

2008 Visiting Analyst, RBB Economics

2006 Data Analysis Intern, Ohio Board of Regents

2005 Lecture Programs Intern, Cleveland Council on World Affairs

Honors, Scholarships and Fellowships

2011-2013 Fellow, Clean Energy for Green Industry Integrative Graduate Education and Research Traineeship (National Science Foundation)

2010 Honorable Mention, National Science Foundation Graduate Research Fellowship Program

2007 Winner, Case Public Policy Competition

2007 H.W. Kneisner Prize for Excellence in Economics

2006-2007 Phi Beta Kappa (elected as a Junior)

2003-2007 Provost Scholarship and Ohio Leadership Award

CHAPTER 1

Incentives for and Detection of Manipulation with Spot and Futures Markets

1.1 Introduction

In the last several years, the Federal Energy Regulatory Commission (FERC) and the Commodity Futures Trading Commission (CFTC) have investigated a number of allegations of attempted market manipulation including complaints against Amaranth,[1] Energy Transfer Partners,[2] Constellation Energy Commodities Group,[3] Barclays Bank,[4] and BP America.[5] These cases have focused on one particular type of manipulation in which the defendant is accused of manipulating the spot price of an asset in order to increase the company's profit on a related cash-settled futures contract.[6] To ensure manipulation attempts are correctly identified, it is desirable to understand the circumstances under which this manipulation can arise and to develop economically well-founded tools that can be used to detect it and confirm it occurred.

Two recent papers, Ledgerwood and Carpenter (2012) and Ledgerwood and Pfeifenberger (2013), suggest that manipulation will always be a temptation when there are

[1] 120 FERC ¶61,085

[2] 120 FERC ¶61,086

[3] 138 FERC ¶61,168

[4] 141 FERC ¶61,084

[5] 144 FERC ¶61,100

[6] To clarify, a physical futures contract is a standardized contract in which the seller agrees to deliver an asset on a specific delivery date in the future. Under a cash-settled futures contract, rather than deliver the physical asset, the seller agrees to pay the buyer the value of the asset on the future delivery date, which will be determined by the spot price of the asset on the delivery date. The spot price of an asset is the price charged for delivering the asset immediately. The spot prices utilized to settle cash-settled futures contracts are standardized and generally determined by the weighted average trade price within a certain window immediately prior to delivery of the asset.

spot markets with a related cash-settled futures contract, but the authors do not fully consider the equilibrium responses of other market participants. The only full equilibrium model of this type of manipulation, Kumar and Seppi (1992), also indicates that this manipulation would generally be profitable in expectation and that it would reduce the informativeness of spot prices.

In this paper, I extend the existing equilibrium model of this type of manipulation and demonstrate that as one condition of the model relaxes in a sensible way, the expected profit from manipulation goes to zero, as does its effect on the spot price, suggesting that manipulation may not be as alluring as previously thought. Using the model, I also analyze the effect of a successful manipulation on the market and propose a screen for manipulation that utilizes publicly obtainable data. In the second half of the paper, I apply the screen to a recent case in which BP America was accused of manipulating the spot price of natural gas in Texas in 2008. To implement the screen, I exploit a dataset rarely available to academic researchers (though it is commonly available to regulators) and employ two types of estimators. In the period in which the alleged manipulation was supposedly profitable, both types of estimates confirm the hypothesis that manipulation may have occurred. In a later period in which manipulation was unprofitable, the results of the screen are inconclusive.

Ledgerwood and Carpenter (2012) and Ledgerwood and Pfeifenberger (2013) understand the manipulation in the following way. Suppose that a trader buys one cash-settled futures contract. By buying a cash-settled futures contract, he pays a fixed futures price, $\$F$, now, and on the delivery date, he will receive the spot price, $\$S$. The higher the price in the spot market relative to the futures price paid, the more profit the trader makes. Realizing this, the trader would like to find a way to make the spot price higher. But of course, this can easily be done by purchasing the asset in the spot market, the trader thinks. After all, sellers will certainly be happy to sell to him at an inflated price. Though the trader is likely to take a loss on his purchases, if his futures position is sufficiently large relative to the amount purchased in the spot market, the gain on the futures contract will outweigh the loss from buying the physical asset at an inflated price.

It is not clear however, that this would be profitable in an equilibrium with rational expectations, as noted also in Lo Prete and Hogan (2014). For instance, if other traders have a good estimate of the quantity the manipulator will buy or sell and how he will make his trades, then expectations of price may not be affected by the manipulator's trades, causing any change in price to be temporary if one can even be achieved. Alternatively, if a manipulator's increased buying activity draws more sellers to the market and therefore generates more competition among sellers, it may be difficult to create a sufficiently large increase in price. Furthermore, the manipulator's purchases in the futures market may raise the futures price, reducing any profit the manipulator would receive, and could cause the traders on the other side of the manipulator's futures position to trade in the spot market, thereby counteracting the manipulator's effect on the spot price.

Current understanding of this manipulation in an equilibrium setting with rational traders rests on a paper by Kumar and Seppi from 1992 which shows that, under the conditions of their model, manipulation is always profitable. The key condition identified by the authors that is required for successful manipulation is that trading in the futures market must be significantly less informative about the true value of the asset than trading in the spot market. If this condition is met, the price impact of a trade in the futures market will be lower than the price impact of a trade in the spot market, allowing the manipulator to take a large futures position without significantly moving the futures price and make spot market trades that are smaller in size but which have a greater impact on price. Together, these elements ensure that the expected profit on the futures position will be greater than the expected loss on the spot position. As trading in the spot markets that determine the settlement prices of cash-settled futures contracts often appears to be much thinner than trading in the futures markets, this condition often seems to be met and it would seem that little can be done to prevent this manipulation other than harshly punishing the manipulators that are caught to try to deter future manipulators.

However, I demonstrate that spot prices can only be manipulated if there are non-strategic speculators in the futures market who take a cash-settled futures position and do not trade in the spot market even though this strategy generates a loss on average. If all futures traders were hedging or speculating strategically, spot prices would be not be manipulable and the expected profit from manipulation would be zero, which corroborates a recent discussion in a 2009 working paper by Kyle, in which he argues that if all traders are strategic, their joint spot and futures market trades are not manipulative. I also show that as the proportion of non-strategic speculators to hedgers falls, so does the expected profit (and presumably the probability) of manipulation. As demand for futures contracts in commodity markets is generally driven by the inelastic demand and supply needs of risk-averse hedgers, this result suggests that manipulation may be much less appealing and less prevalent than previously thought.

Ideally, one could screen for the manipulability of a market by examining the spot market trades of those who held a position in the futures market to expiration. The share of non-strategic speculators could be measured as the share of the open positions in the futures contract at expiration that cannot be matched with corresponding trades in the spot market. While this information could potentially be gathered by surveying traders, it would be very costly and difficult to obtain a complete picture of the proportion of trades driven by hedging versus non-strategic speculation. Recognizing these difficulties, I instead suggest and test a screen for attempted manipulation that utilizes obtainable data from an allegedly manipulated market. This screen can be employed when manipulation is suspected to corroborate other evidence of manipulation and as a first pass at identifying irregularities in market activity.

Both the original and extended versions of the Kumar and Seppi (1992) model suggest that when spot prices can be manipulated by a strategic trader, manipulation will reduce the equilibrium price impact of a trade. This occurs because manipulation increases the amount of unpredictable uninformative trading and therefore reduces the amount of information about the true value of the asset revealed by trading on average. Thus, if market conditions cause participants to believe that the likelihood or scale of manipulation has increased, the price impact of a trade may fall. I apply this screen to a recent case in which BP America was accused of manipulating natural gas prices in Texas after

Hurricanes Gustav and Ike in 2008.

To test the proposed screen for manipulation, I obtained and analyzed data from the Intercontinental Exchange that contains the bids, offers, and trades from the market and time period in which traders from BP America allegedly manipulated the spot price of natural gas. In particular, BP America has been accused of manipulating the spread (or price difference) between the spot price of natural gas at Houston Ship Channel and the spot price at Henry Hub. As their alleged manipulation occurred after a pair of hurricanes that damaged natural gas production and transportation infrastructure and therefore reduced the correlation between the value of natural gas at Houston Ship Channel and Henry Hub, the expected profit from manipulating the spread likely increased relative to periods unaffected by hurricanes. If other market participants recognized the probability and profitability of manipulation had increased, the model suggests that the average price impact of a trade would be reduced relative to other time periods and relative to other hubs at which manipulation was not suspected.

To test the hypothesis that price impact was lower during the allegedly manipulated period, I develop two types of estimators for the price impact of a trade. The first is a measure of the resilience of the market, that is the speed and frequency with which market prices return to their pre-trade levels between trades. The less informative trades are about the true value of the asset, the more likely prices will bounce back and the quicker they will bounce back after a trade, since each trade reveals little information. The second estimator is a direct measure of the price impact of trading from a vector error correction model in which I separately identify the effect of a sale versus the effect of a purchase on the available bids and offers.

The results from both estimators suggest that in the middle of the allegedly manipulated period, when the allegedly manipulative strategy was profitable, the price impact of a trade was lower than expected. The estimates suggest that in this period, the price impact of a trade in the allegedly manipulated market was lower than in any other period in the dataset and lower than in other nearby markets, providing solid support for the hypothesis that price impact is lower when manipulation is suspected and for the hypothesis that manipulation may have occurred. I exclude the first few days of the alleged manipulation as the increase in uncertainty caused by the hurricanes makes it difficult to separately identify any effect of manipulation. At the end of the investigative period, when the allegedly manipulative strategy generated a loss, the results are inconclusive with some measures of price impact and market resilience indicating that the price impact of a trade may have returned to more usual levels. If BP's actions became more predictable in the later period, this may explain the increase in price impact as market participants could adjust to their trades in other ways.

The paper is organized as follows. Section 1.2 provides a simple intuitive model of when a manipulator might believe manipulation would be profitable. Section 1.3 discusses why manipulation is profitable in equilibrium in the model of Kumar and Seppi (1992) and extends the model to include the possibility that some or all futures noise trading is due to hedging demand from risk-averse traders. Section 1.3 also demonstrates that the same manipulation could be profitable with a physical futures contract rather than a cash-settled futures contract. In Section 1.4, I propose and test a screen for manipulation using BP America's alleged manipulation of natural gas prices in Texas as my test case.

1.2 The Profit Maximizing Strategy of a Manipulator

To understand the incentives manipulation, it is helpful to begin with a simple model. Suppose that there are two markets related to a physical asset, a spot market, where the physical asset can be traded, and a cash-settled futures market. Trading in the futures market occurs prior to trading in the spot market and traders in the futures market have no private information about the true value of the asset causing prices to be unaffected by trades. The manipulator is risk-neutral and able to trade in both the spot market and the futures market. When solving his profit maximization problem, he makes an assumption about the strategy of other traders in the futures market. In particular, the manipulator must consider the strategy of the trader who took the other side of his futures position.

I assume the spot price is determined as follows. First, some traders in the spot market do have private information about the asset's true value and therefore trades are expected to carry information on average and will affect spot prices. Second, traders assume that the price impact of a trade is constant throughout the trading period. Third, traders assume that the price impact of a trade is linear in the size of their trade and that their trading strategies do not affect the price impact of a trade.

Let μ represent trader i's expected spot price if he does not trade in either the futures or the spot market and other traders do not expect him to trade. I assume that trader i has no reason to trade other than his expectation that trading in both the spot and futures market may yield a profit and that he has no private information about the value of the asset. Let λ be trader i's expectation of the price impact of a trade in the spot market.

Let $y_{f,i}$ be the futures position of trader i. If there are N traders in the futures market, then $\sum_{i=1}^{N} y_{f,i} = 0$ since every trade must have a buyer and a seller. Let $y_{s,i}$ be trader i's spot market trade. To determine trader i's expected spot price if he trades in the futures market and then in the spot market, trader i must make an assumption about how the trader who took the other side of his futures trade will act. Trader i's futures trade causes another trader in the futures market to have a futures position $-y_{f,i}$. Suppose that trader i assumes the other trader will trade $-ky_{f,i}$ in the spot market if trader i traded $y_{f,i}$ in the futures market, where k is a constant. For instance, if trader i assumes that the other trader will not trade in the spot market in reaction to his futures trade, then $k = 0$. If other traders are hedging, then to obtain a physical futures position when only a cash-settled futures contract exists, hedgers will need to buy cash-settled futures contracts, for which they will receive S and pay F, and then buy the same amount in the spot market, for which they will pay S in exchange for the physical asset. On net, they pay F and receive the asset. Therefore, if trader i assumes all other futures traders are hedging, he will expect $k = 1$. Given k, trader i's expected spot price is:

$$\begin{aligned} E_i[S|y_{s,i}, y_{f,i}] &= \mu + \lambda\Big(y_{s,i} + k\Big(\sum_{j=1}^{N} y_{f,j} - y_{f,i}\Big)\Big) \\ &= \mu + \lambda(y_{s,i} - ky_{f,i}) \end{aligned} \tag{1.1}$$

where $\sum_{j=1}^{N} y_{f,j} - y_{f,i}$ is the net futures position of all other futures traders. Since $\sum_{j=1}^{N} y_{f,j} = 0$ by definition, the first term vanishes in the simplified equation.

Suppose that trader i's expectation of the spot price without him, μ, is the futures price, F, because he has no private information on the value of the asset. This also implies that trader i's expectation of the true value of the asset will be μ.

Trader i's problem at the opening of the spot market is:

$$\begin{aligned} \max_{y_{s,i}} E[\Pi|y_{fi}] &= \max_{y_{s,i}} E[y_{f,i}(S(y_{f,i}, y_{s,i}) - F(y_{f,i})) + y_{s,i}(v - S(y_{f,i}, y_{s,i}))|y_{f,i}] \\ &= \max_{y_{s,i}} y_{f,i}(\mu + \lambda(y_{s,i} - ky_{f,i}) - \mu) + y_{s,i}(\mu - \mu - \lambda(y_{s,i} - ky_{f,i})) \\ &= \max_{y_{s,i}} \lambda(y_{f,i} - y_{s,i})(y_{s,i} - ky_{f,i}) \end{aligned} \tag{1.2}$$

From equation (1.2), one can see that if $k \in [0, 1)$, then trader i can earn a positive profit if $y_{f,i} > y_{s,i}$ and $y_{s,i} > ky_{f,i}$. Figure 1.1 illustrates trader i's profit graphically if trader i has a long futures position, that is $y_{f,i} > 0$. When trader i trades $y_{s,i}$ in the spot market, he moves the spot price by $dS = \lambda(y_{s,i} - ky_{f,i})$. If $y_{s,i} > ky_{f,i}$, then trader i increases the spot price above the true value of the asset. He therefore expects to take a loss of $-dS * y_{s,i}$ on his spot market position. However, if $y_{f,i} > y_{s,i}$, then since he earns $dS * y_{f,i}$ but only loses $-dS * y_{s,i}$, his net profit is positive.

The first order condition of trader i's problem confirms that trader i's optimal spot trade is halfway between $ky_{f,i}$ and $y_{f,i}$ as it appears in Figure 1.1:

$$y_{s,i} = \tfrac{1}{2}(1 + k)y_{f,i} \tag{1.3}$$

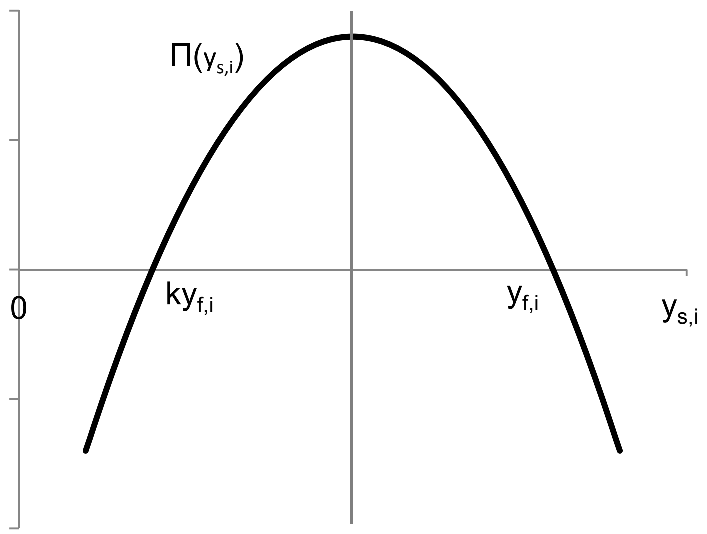


Figure 1.1: Trader i's Profit as a Function of his Spot Market Trade

Suppose trader i assumes other traders are not trading in the spot market, e.g. $k = 0$, then trader i will make a spot trade in the same direction as his futures trade but it will be half the size of his futures trade, e.g. if he bought 5 units in the futures market, he will buy 2.5 units in the spot market. If all traders with futures positions act strategically upon arriving in the spot market, that is they choose a spot trade to maximize their expected profit given their futures position, they would make the same decision as trader i, i.e. make a spot trade half the size of their futures trade, if they expected $k = 0$. Thus, if trader i believes the other traders are strategic and rational, it cannot be an equilibrium for trader i to assume that none of the others will trade in the spot market when he clearly has incentive to do so.

At the other extreme, if other traders exactly replicate their futures trade in the spot market, that is if they trade $y_{sj} = y_{fj}$ so that $k = 1$, then so will trader i by equation (1.3). If trader i trades $y_{si} = y_{fi}$ in the spot market, he earns zero profit and can do no better by making any other trade. This is his optimal choice because when all other traders choose $y_{sj} = y_{fj}$, trader i would have to make a bigger trade in the spot market than he made in the futures market, $y_{s,i} > y_{f,i}$, in order to increase the spot price. However, if his spot trade is larger than his futures trade, then his loss in the spot market, $-dS * y_{s,i}$, will be bigger than his gain in the futures market, $dS * y_{f,i}$, resulting in a net loss. Thus, $k = 1$ is an equilibrium because each trader, believing that other traders will exactly replicate their futures trades in the spot market, will do the same.

Note that there is no sense in which the manipulator could surprise the market with a manipulation attempt and expect to earn a profit if all other traders make the same trade in the spot market as they made in the futures market. For instance suppose he buys in the futures market: $y_{f,i} > 0$ and chooses a smaller spot trade: $y_{s,i} < y_{f,i}$. This would cause the spot price to go down since $y_{s,i} - y_{f,i} < 0$. But, he only makes money when the spot price goes down if his spot trade was bigger than his futures trade: $y_{s,i} > y_{f,i}$. Thus, any spot trade that is different from his futures trade will generate an expected loss for a rational manipulator.[7]

Returning to the expected spot price, we see that if $k = 1$, then the expected spot price will be unmoved by trader i's trade since $y_{s,i} = y_{f,i}$. As a result, trader i's expected profit will be zero and he will be indifferent between trading in the futures market or not. Furthermore, if the price impact coefficient, λ, is set based upon a linear prediction of the asset's true value given observed trading patterns and if each trader with a futures positions trades $y_{s,i} = y_{f,i}$ in the spot market, then the spot trades of those with futures positions will create equal pressure on the buy and sell sides of the spot market and will not affect perceptions of net order flow (a.k.a. net demand) causing the price impact to be unaffected. Thus, the expectation that everyone is acting in their best interest in fact prevents manipulation of the spot price; Adam Smith's proverbial invisible hand in action.

Small players in the market, of course, may fail to account for their (proportionally small) effect on the market and may wish to speculate. If some traders hold futures positions to expiration and do not trade in the spot market, causing $k < 1$, then the trades of strategic traders are likely to affect the realized spot price and may reduce its

[7] If the manipulator bought in the futures market, but failed to take into account that the trader who sold him his futures position will make the same sale in the spot market, then he might expect the spot price will go up (and yield him a profit) when actually it will go down (and cause him to lose money on average).

efficiency as a signal if the amount of uninformed trading is perceived to increase (as will be discussed in Section 1.3). This could be especially problematic if the aggregate trading of small buyers in the futures market is different from the aggregate trading of small sellers. For instance, if all buyers in the futures market are small and all sellers are large, then the large sellers' optimal strategies would move the price downward. In this case, the small buyers would consistently lose money on their futures trades since the spot price would be less than the futures price due to the spot market sales of the large futures sellers. That is: $S = \mu + \lambda y_{sL} < \mu = F$, where $y_{sL} < 0$ is the aggregate spot market sale of the large futures sellers. To prevent manipulation from occurring, one could advise small players to sell their speculative futures positions prior to the closure of the futures market, which would allow them to profit from any change in expectation since their position was purchased while preventing losses due to manipulation.

Spot market traders have an additional weapon not heretofore considered: their expectation of the spot trades of futures traders. If all sales are made by a few larger sellers in the futures market and all purchasers are small buyers, then if spot market traders can observe the amount of trading in the futures market, they can guess the aggregate spot market trade of the large futures sellers and filter out that portion of spot market trading when determining the price at which they are willing to trade in the spot market; a key reason why manipulation may not be effective in equilibrium. The better spot market traders can guess how a manipulator will trade, the less effect he will have on the realized spot price. To gain a clearer understanding of how equilibrium forces under rational expectations may limit or prevent manipulation, I re-examine and extend the full equilibrium model of Kumar and Seppi (1992).

1.3 Manipulation in Equilibrium

1.3.1 The Kumar and Seppi (1992) Model

Under the assumptions of Kumar and Seppi's 1992 model, there is an equilibrium with manipulation and manipulation will degrade the informativeness of prices in the spot market. Further, even if many players attempt to manipulate prices, Kumar and Seppi demonstrate that though the profit from manipulation goes to zero, the spot price is always less informative under manipulation. In this section, I review their model and demonstrate that if we consider more carefully the identity of the noise traders in the futures market, we find that more accurate assumptions about the behavior of the futures noise traders could render manipulation unprofitable and the spot price signal untainted.

Kumar and Seppi's 1992 model consists of three periods: in period 1, a cash-settled futures contract is traded, and in period 2, a spot contract is traded. In period 3, the true value of the asset is revealed and anyone with an open position in the spot contract receives or pays the true value of the asset. The cash-settled futures contract is structured such that the buyer of the contract pays the cash-settled futures price, F and receives the settlement price of the spot contract, S. Only the distribution of the true value of the asset, v, is known in period 1. It is assumed that v is distributed normally with mean μ and variance σ_v^2. In period 2, there is a risk-neutral informed trader who learns v at the beginning of the period, while all other players remain uninformed. The model is essentially a two-period version of the Kyle (1985) model except that there is an informed trader only in the second period of trading in Kumar and Seppi's model, while in Kyle's model there is a trader who knows the value v in every period. This information differential between periods is one of the features of the model that allows for an equilibrium with manipulation.

In the period 1 futures market, there are noise traders who place an order e drawn from a normal distribution with mean 0 and variance σ_e^2. e is independent of the true value. A risk-neutral potential manipulator may trade in the futures market if it is profitable, but it is assumed for simplicity that the informed trader does not trade in the futures market.[8] The potential manipulator submits an order Δ which is combined with the noise traders' order e to generate the net futures order flow $y_f = \Delta + e$. There are at least two risk-neutral market makers who observe y_f and compete *à* la Bertrand to set the price of the futures contract F given the net order flow y_f. As traders holding a futures contract will receive the spot price, the futures price, F, will equal the expected spot price given y_f, $F = E[S|y_f]$.

In the period 2 spot market, the informed trader submits an order X knowing v and y_f. The manipulator submits an order z knowing his own futures order, Δ and y_f (and consequently $e = y_f - \Delta$ as well). There are also noise traders in the spot market who submit an order u drawn from a normal distribution with mean 0 and variance σ_u^2. u is independent of v and e. The three orders in the spot market are combined into the net spot order flow $y_s = X + u + z$. A set of risk-neutral market makers observe y_f and y_s and again compete *à* la Bertrand to set the price of the spot contract. As anyone who holds a position in the spot contract at the end of period 2 will receive the true value v in period 3, the spot price S will be equal to the expected value of the asset, v, given y_s and y_f: $S = E[v|y_s, y_f]$.

The profit of the strategic players and the properties of the equilibrium are given by Proposition 1.1. For the full proof of the proposition, I refer the reader to the Appendix of Kumar and Seppi (1992). Here I will provide the intuition.

Proposition 1.1 (Linear Equilibrium of Kumar and Seppi (1992)) *There is an equilibrium in which (i) the strategy that maximizes the profit of manipulator,*

$$\max_{\Delta} E_e\Big\{ \max_{z} E_{v,u,X}\Big[\Delta\big(S(X+z+u, e+\Delta) - F(e+\Delta)\big) + z\big(v - S(X+z+u, e+\Delta)\big)\Big|e\Big]\Big\}$$
$$s.t.\ |\Delta| \leq |W| \tag{1.4}$$

is to randomize his futures order Δ between $|W|$ and $-|W|$ with equal probability and

[8] If he is allowed to trade in the futures market, he competes with the manipulator to manipulate the market.

trade

$$z = \tfrac{1}{2}[(1+k)\Delta + ke] \tag{1.5}$$

$$\textit{where } k = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_e^2}. \tag{1.6}$$

The manipulator's expected profit is:

$$E[\Pi_M|\Delta] = \tfrac{1}{4}\lambda(1-k)^2\Delta^2 + k^2\sigma_e^2 > 0 \tag{1.7}$$

(ii) the strategy that maximizes the profit of the informed trader,

$$\max_X E_{u,z}\Big[X\big(v - S(X+z+u, e+\Delta)\big)\Big|v, y_f\Big] \tag{1.8}$$

is to trade

$$X = \frac{1}{2\lambda}(v-\mu), \tag{1.9}$$

(iii) the futures price is $F = \mu$, and (iv) the spot price is set using the rule

$$S = \mu + \lambda(y_s - E[y_s|y_f]) \tag{1.10}$$

$$\textit{where } \lambda = \frac{\sigma_v}{2\sqrt{\sigma_u^2 + \sigma^2(z|y_f)}} \tag{1.11}$$

$$\sigma^2(z|y_f) = \tfrac{1}{4}k\sigma_e^2 \tag{1.12}$$

$$E[y_s|y_f] = E[z|y_f] = ky_f \tag{1.13}$$

The goal of the market makers in the spot market is to extract the signal of the true value, X, the trade of the informed trader which is based upon v, from the noise generated by the uninformed noise traders and the uninformed manipulator. To do so, they utilize their knowledge of the strategies of each trader and the observed order flows to form an estimate of the amount of each order flow that contains information and the correlation between that part of the order flow and the true value. In equilibrium all trades are linear combinations of the normally distributed (and independent) random variables, e, u, v, and W. Therefore, the order flows in the spot and futures markets, y_s and y_f, are also normally distributed. As a result, the spot price, which is equal to the expected value of the asset given the order flows, takes a linear form

$$\begin{aligned} S = E[v|y_f, y_s] &= E[v|y_f] + \frac{Cov(v, y_s|y_f)}{Var(y_s|y_f)}\big(y_s - E[y_s|y_f]\big) \\ &= \mu + \lambda(y_s - ky_f) \end{aligned}$$

Neither the manipulator's order nor the other futures traders' orders are correlated with the true value v, and therefore y_f does help to predict the true value of the asset. Consequently, the market makers' expectation of the value of the asset, v, given y_f is simply μ, the expectation of v.

The second term contains $y_s - E[y_s|y_f]$, which is the market makers' expectation of the part of the spot market order flow that contains information about the true value. The market makers know that traders in the futures market do not have information about the true value of the asset and that y_f is therefore uninformative about the true value. However, y_f does predict the amount that the uninformed manipulator will trade in the spot market. As y_f does not predict how much the spot noise traders will trade or how much the informed trader will trade, $E[y_s|y_f] = E[z|y_f]$. Thus, when observing any order flow y_s, the market makers first filter out their best estimate of the spot market order of the uninformed manipulator and the spot price will only move if y_s is different from their expectation of the uninformed manipulator's order.

After filtering out their best estimate of the manipulator's trade, the market makers in the spot market choose λ, which determines how much the spot price moves given the filtered order flow. The price impact, λ, will be based upon the covariance between the filtered order flow and the true value of the asset relative to the total variance of the filtered order flow. As can be seen in the equilibrium value of λ, shown in equation (1.11), the less precise the public information about the true value (the larger σ_v is) and the smaller the amount of noise from the manipulator and spot noise traders ($\sigma_u^2 + \sigma^2(z|y_f)$), the more informative the filtered order flow becomes and the larger the price impact per unit of filtered order flow.

Note that if the market makers in the spot market know the manipulator's trades with certainty either because the manipulator always chooses the same trades or because there are no noise traders in the futures market to camouflage the manipulator's futures position, then the market makers can perfectly filter out the manipulator's spot market trade and the spot price will be unaffected by the manipulator's trades. If the manipulator's order does not move the spot price, the manipulator earns no profit.

However, if the manipulator plays a mixed strategy in the futures market, randomizing between several possible trades, and if there are noise traders who trade only in the futures market, camouflaging the manipulator's futures position, market makers in the spot market can estimate the manipulator's futures position and spot market order (which will depend on his futures position), but will not know their exact values. Since the manipulator's spot order is both uncertain and known to be uninformative about the true value of the asset, the manipulator's spot market order increases the amount of uninformative trading in the spot market. The larger the variation in uninformative trading relative to the variation in the informed player's order (which is derived from the possible variation in the true value), the harder it is for the market maker to distinguish the exact volume traded by the informed trader which signals the true value of the asset and the lower the price impact of an order, as seen in equation (1.11). Thus, by increasing the possible variation in uninformed trading in the spot market, the manipulator decreases the informativeness of the price signal in the spot market. With the market maker unable to perfectly filter out the manipulator's spot trade when setting the spot price, the manipulator's spot trade will affect the spot price and manipulation can be profitable.

k indicates the manipulator's share of the variance of the futures order flow and therefore measures his size relative to the futures noise traders. Holding the variance of the futures noise trades constant, the closer k is to one, the larger the manipulator is and the lower the price impact of a trade in the spot market becomes because, when the manipulator is large relative to the futures noise, the variance in his uninformative spot market order given the futures order flow is also large, which increases the noisiness of the spot order flow and decreases its informativeness.

By iterated expectations, the futures price, F, will always be equal to the baseline expected value of the asset, μ. This result arises because the market makers in the spot market have more information than the market makers in the futures market. As discussed above, the market makers in the spot market carefully filter out the trade of the manipulator given both y_f and y_s. Furthermore, the futures order flow, y_f, contains no information about the true value, and only enters the spot price because it helps predict and filter out uninformative trading. Therefore, since y_f does not help predict the true value v and will be filtered out of the spot order flow y_s when forming the spot price, the expectation of the spot price knowing only y_f will not be influenced by y_f and will be equal to the expected value of the asset, μ. In other words, from the perspective of the market makers in the futures market who know only y_f, the expected spot price is not affected by y_f, and therefore the futures price, F, which equals the expected spot price given y_f is not affected by y_f.

If the futures order flow did reveal information about the true value of the asset, then the futures price will respond to trades in the futures market because they are informative. Although it is beyond the scope of this paper to delve deeply into the implications of information in the futures market, preliminary findings correlate the assertion of Kumar and Seppi that adding an informed trader to the futures market in their model makes manipulation unprofitable. This result appears to arise both because the manipulator's futures trades now move the futures price adversely and because the manipulator has a more precise signal of the uninformative trading in the futures market than the market makers in the spot market have causing the manipulator's spot market trade to be informative about the true value of the asset.

Returning to the original model, the informed trader, picks his trade X knowing v and y_f. Since the spot noise traders and manipulator do not have private information about the true value of the asset, their trades do not depend on v. Therefore, the manipulator has the same information as the market maker about how the spot noise traders and the manipulator will trade. This means that expected trades of the manipulator and spot noise traders will not affect the informed trader's expectation of the spot price:

$$
\begin{aligned}
E[S|v, y_f] &= \mu + \lambda(X + E[u + z|v, y_f] - E[X + u + z|y_f]) \\
&= \mu + \lambda(X + E[u + z|y_f] - E[X + u + z|y_f]) \\
&= \mu + \lambda(X - E[X|y_f])
\end{aligned}
$$

Maximizing his profit given his expected profit and his monopoly on information about the true value of the asset, shows that he will trade $X = \frac{1}{2\lambda}(v - \mu)$, which would move the price halfway to the true value if the net order flow of others were zero. This also implies that the expected order of the informed trader, given knowledge of y_f alone is zero since $E[v|y_f] = \mu$.

Turning to the manipulator's spot trade, we see that his optimal trade is

$$
z = \tfrac{1}{2}[(1 + k)\Delta + ke] = \tfrac{1}{2}[\Delta + ky_f]
$$

which has a nearly identical form to his optimal trade in the simple model of Section 1.2. This can be understood by considering his expected profit:

$$
\begin{aligned}
E[\Pi|e, \Delta, z] &= E\Big[\Delta\big(S(X + z + u, e + \Delta) - F(e + \Delta)\big) \\
&\quad + z\big(v - S(X + z + u, e + \Delta)\big)\Big|e, \Delta, z\Big] \\
&= \lambda(\Delta - z)(z - E[z|y_f])
\end{aligned}
\tag{1.14}
$$

The market makers' expectation of the manipulator's spot trade, z, given the futures order flow y_f is equal to ky_f where k indicates the manipulator's share of the variance of the futures order flow. As in the earlier model, the manipulator can increase price if $z > E[z|y_f] = ky_f$ and will make a profit if his loss $-dS * z = -\lambda(z - ky_f)z$ is smaller than his expected gain, $dS * \Delta$, which requires $z < \Delta$. To have $z < \Delta$ and $z > ky_f$

requires also that $ky_f < \Delta$. If $ky_f > \Delta$, the manipulator would like to decrease price by setting $z < ky_f$ and will earn a profit if $z > \Delta$.

Suppose $\Delta > 0$ to simplify the intuition. If the market believes that the manipulator's trade will be smaller than his actual futures position, $ky_f < \Delta$, because the futures noise order e was small or negative, the manipulator can place a spot order z smaller than Δ but larger than ky_f and move the spot price upward while still making a profit since $\Delta > z$. This is illustrated in Case A in Figure 1.2. If the market believes that the manipulator's trade will be bigger than his actual futures position, $ky_f > \Delta$, because the futures noise order e was very large, the manipulator would like to buy a small amount z between Δ and ky_f, which will keep price below its expected value μ. Since $z < ky_f$, the manipulator will earn a profit in the spot market, in this case, and since $z > \Delta$, his spot profit will be larger than his loss in the futures market. This case is illustrated in Case B in Figure 1.2.

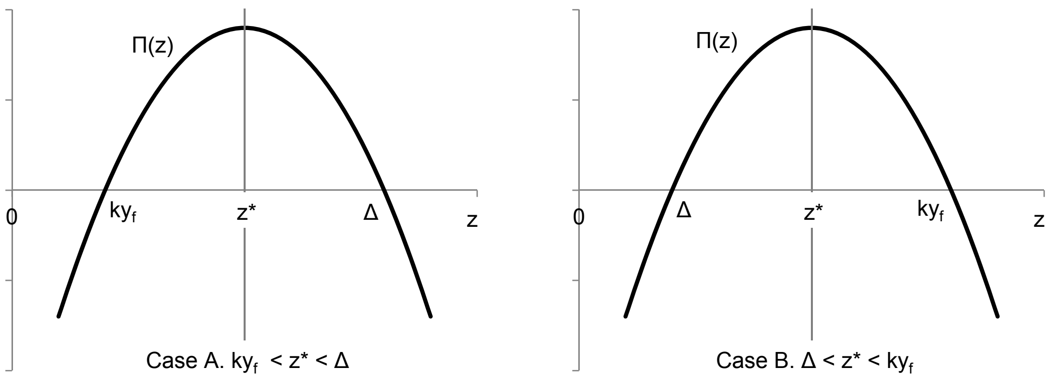


Figure 1.2: The Manipulator's Profit as a Function of his Spot Market Trade

As the futures noise traders camouflage the manipulator's futures order, market makers will be unable to estimate his exact futures position given only the futures order flow if the manipulator randomizes his futures trade. Since the market makers know that the manipulator's spot trade depends on both the net order of the futures noise traders and the manipulator's random futures order, they will be unable to exactly predict the manipulator's spot order. To achieve a linear equilibrium, Kumar and Seppi make some additional assumptions. As shown in Proposition 1.1, the manipulator's expected profit given his futures position takes the form $\frac{1}{4}\lambda(1-k)^2\Delta^2 + k^2\sigma_e^2$. Thus as long as $1-k \neq 0$, which holds as long as there is some noise trading in the futures market, the manipulator would wish to take an infinite futures position, but is indifferent between buying and selling. To prevent the manipulator from taking an infinitely large position, Kumar and Seppi assume that the manipulator is subject to an initial wealth constraint $|W|$, and only the distribution of W ($N(0, \sigma_w^2)$) is known to other players. W is independent of v, u, and e. To submit an order Δ, the manipulator must deposit $|\Delta|$ in a margin account thereby preventing him from taking an infinite position.

Under these assumptions, the manipulator receives some initial wealth $|W|$ and, since he is indifferent between trading $|W|$ and $-|W|$ he randomizes between trading $|W|$ and $-|W|$ with equal probability, which gives Δ the same normal distribution as W. In this way, the market makers are uncertain of whether he bought or sold as well as the amount he traded and therefore cannot perfectly filter out his spot market trade when setting the spot price, generating the potential for profit from manipulation.

The key feature of the equilibrium to note in the Kumar and Seppi (1992) model is that the manipulator is able to make a profit because he has an information advantage over the market makers in predicting the spot market order flow. This can be seen by rearranging equation (1.5) into: $z = \Delta + \frac{1}{2}(e - E[e|y_f])$. From this version, one sees that the strategy of the manipulator can be characterized as trading his full futures position, Δ, plus half of his information advantage, $e - E[e|y_f]$. If the market makers knew e, then the manipulator's optimal strategy would be to trade $z = \Delta$, in which case he earns zero profit since the futures price F is always equal to the expected value of the asset.

The natural question is thus, under what circumstances would the manipulator lose his information advantage? Obviously, if there were no noise traders in the futures market, then the futures market order flow would perfectly reflect the manipulator's order. Alternatively, if the manipulator always made the same trade in the futures market, then the futures order flow would perfectly reflect the aggregate trades of the noise traders, e. It turns out that there is also a third situation in which the manipulator loses his advantage: the case in which all noise trading in the futures market is executed as a hedge for planned trades in the spot market.

1.3.2 Hedging versus Non-Strategic Speculation

In the current model, the aggregate profit of the noise traders in the futures market is $e(S-F)$ whose expected value at the beginning of the futures market is always negative, $-\frac{1}{2}\lambda k\sigma_e^2$. As the futures product they have traded yields only a transfer of cash upon settlement, the noise traders in the futures market in the model of Kumar and Seppi (1992) are essentially non-strategic speculators who either do not notice or do not care that, on average, they will lose money by trading futures contracts. They are neither acquiring the physical asset nor hedging against any real risks in this model, so there is no reason for them to trade unless their expected profit from the futures trade alone is non-negative. A more compelling type of noise trader for the futures market, especially in the context of a commodity such as natural gas, would be a risk-averse hedger, whose trades are generated by actual (and potentially inelastic) demand for or supply of the asset in question. Since the futures contract assumed in this model is a cash-settled futures contract, to actually acquire or dispose of the physical asset, the same quantity that is purchased in the futures market as a hedge, h, must also be purchased in the spot market, in which case the profit of a hedger is $\Pi_h = h(S-F)+h(v-S) = h(v-F)$. Note that if he buys h as a hedge in the futures market, he also *buys* h in the spot market. If a physical futures contract were available, he could achieve the same outcome with only one transaction in the physical futures market.

Suppose all noise trading in the futures market is due to hedging and represent it by h. In this case, the spot market order flow is comprised of the orders of the informed trader, the spot noise traders, the manipulator and the hedgers. The strategies of the informed trader and spot noise traders are not affected by the manipulator and hedgers' strategies as they have the same information as the market maker in the spot market. Therefore, taking their strategies from Proposition 1.1 and the prior assumptions, the expected spot trades of the informed and spot noise traders, given the futures order flow, are zero. Consider now the manipulator's first order condition for z, where $y_f = \Delta + h$ and $y_s = X + u + z + h$:

$$z = \tfrac{1}{2}(\Delta - E[X + u + h|h] + E[X + z + u + h|y_f])$$

Since $E[X|y_f] = E[X|h] = E[u|y_f] = E[u|h] = 0$

$$z = \tfrac{1}{2}(\Delta - h + E[z + h|y_f])$$

Taking the expectation of both sides given y_f reveals that $E[z|y_f] = E[\Delta|y_f]$. Therefore

$$\begin{aligned} z &= \tfrac{1}{2}(\Delta - h + E[\Delta + h|y_f]) \\ &= \tfrac{1}{2}(\Delta - h + y_f) \\ &= \Delta \end{aligned}$$

As also occurs in the case with only non-strategic speculators, the market makers' best guess of the manipulator's trade, z, given the futures order flow is the same as their best guess of Δ given the futures order flow. However, since both the futures market noise traders and the manipulator trade in the spot market and since the futures market noise traders trade exactly the same quantity in the spot market as they traded in the futures market, the market makers' best guess of the spot market order flow that will be caused by uninformed futures traders is now just the futures order flow, that is $E[y_s|y_f] = E[h+\Delta|y_f] = y_f$. Knowing that the market makers expect that the spot order flow due to uninformed futures trading will be y_f, it turns out that the manipulator's optimal spot market trade is same as his futures market trade, $z = \Delta$.

With hedgers, the noise from the other futures traders works against the manipulator, unlike in the case of the non-strategic speculators, because it appears both in the spot and futures market order flow. Given an arbitrary spot trade z, the manipulator's expected spot price is $S = \mu + \lambda(z + h - y_f) = \mu + \lambda(z - \Delta)$ since $y_f = h + \Delta$. His expected profit is therefore:

$$
\begin{aligned}
E[\Pi|h,\Delta,z] &= E\Big[\Delta\big(S(X+z+u+h,h+\Delta)-F(h+\Delta)\big)\\
&\quad + z\big(v-S(X+z+u+h,h+\Delta)\big)\Big|h,\Delta,z\Big]\\
&= \lambda(\Delta-z)(z-\Delta)\\
&= -\lambda(z-\Delta)^2<0 \qquad (1.15)
\end{aligned}
$$

With hedgers, the only way to increase the spot price is to make a bigger trade in the spot market than he made in the futures market $z>\Delta$. However, if $z>\Delta$, his expected profit decreases because the increase in the spot price (dS) hurts his spot position by $-dS*z$ but only benefits his futures position by $dS*\Delta$ and $\Delta<z$. Therefore, his expected profit is always negative in equilibrium, since λ must be positive for the manipulator's and informed trader's second order conditions to be satisfied. As a result, the best he can do is to set $z=\Delta$ and earn zero profit in expectation.

Furthermore, since the uninformative spot trades of the futures traders can be determined exactly by the market makers in the spot market, they can filter them out exactly when setting the spot price, and so the spot price is not affected by the manipulator's actions even if he does decide to trade (despite an expectation of zero profit). Or in other words, when all of the other traders in the spot market trade their entire futures positions in the spot market, the manipulator loses his information advantage over the market makers in predicting the uninformed portion of the spot market order flow.

Of course, there's no reason that there couldn't be speculators in the futures market along with the hedgers, but in local commodity markets, such as I examine in this paper, I expect that the speculators are likely to be strategic, like the manipulator, as they should recognize that an isolated futures trade would be expected to lose money in the presence of strategic speculators. Speculators who do not wish to have a physical position at the conclusion of the spot market can trade in their futures position prior to the close of the futures market, and earn a profit on any change in the futures price that occurred since they obtained the position without introducing inefficiency into the spot price.

If all speculators trade in their cash-settled futures position for a physical futures position, it would not mean that there are no speculators in the market, only that there are no non-strategic speculators. If the traders in the futures market are either strategic speculators or hedgers, I show in Proposition 1.2 that given a futures position Δ, the optimal spot market order of a strategic speculator is $z = \Delta$, in which case all strategic speculators expect to earn zero profit from any manipulation attempt. Thus as long as all those who demand liquidity in the futures market are either (1) strategic speculators, or (2) hedgers, then the expected profit from strategic speculation will be zero and the spot price signal will be untainted. As in the simple model examined earlier, if all other futures traders are either strategic or hedging, the manipulator could not even surprise the market and expect to make a profit since any difference between his spot and futures market will move price in the wrong direction given the gap between his futures and spot positions.

Proposition 1.2 (Equilibrium when All Speculators are Strategic) *If the futures noise trading is due to hedging and/or there are multiple strategic speculators in the market, in equilibrium, all strategic speculators will choose $z = \Delta$, the expected profit from speculation will be zero and the spot price signal will be unaffected.*

Proof. Suppose that a particular strategic speculator, call him the manipulator, takes others' strategies as given, assuming that whatever their net futures trade, ε, is, the other strategic speculators will trade $\zeta = R_1\varepsilon + R_2 y_f$ in the spot market. If there are hedgers, they trade h in both the spot and futures markets. Thus the futures order flow, y_f equals $\Delta + h + \varepsilon$ and the spot order flow, y_s is $X + u + h + z + \zeta$. The manipulator's profit maximization problem is the same as in (1.4) with the new definitions of y_f and y_s. His first order condition for z is therefore:

$$z = \tfrac{1}{2}(\Delta - E[X + u + h + \zeta|e] + E[z + X + u + h + \zeta|y_f]) \qquad (1.16)$$

As before, taking the expectation of both sides of (1.16) with respect to y_f yields, $E[z|y_f] = E[\Delta|y_f]$ by iterated expectations. Taking into account $E[X + u|e] = E[X +$

$u|y_f] = 0$ and $\zeta = R_1\varepsilon + R_2 y_f$, his first order condition becomes

$$z = \tfrac{1}{2}(\Delta - E[h + R_1\varepsilon|e] + E[\Delta + h + R_1\varepsilon|y_f]) \tag{1.17}$$

Suppose that $R_1 = 1$, then

$$\begin{aligned} z &= \tfrac{1}{2}(\Delta - E[h + \varepsilon|e] + E[\Delta + h + \varepsilon|y_f]) && (1.18)\\ &= \tfrac{1}{2}(\Delta - e + y_f) && (1.19)\\ &= \Delta && (1.20) \end{aligned}$$

Thus, if all other speculators trade their entire futures position in the spot market, the optimal trade of the manipulator is to also trade his entire futures position in the spot market. Plugging $z = \Delta$ and $E[v|e] = \mu = F$ into the manipulator's expected profit shows that the manipulator's expected profit will be zero. As before, if the full futures order flow is traded in the spot market, the market makers in the spot market know exactly the uninformative spot market order of the futures traders and can subtract it from the spot order flow causing the spot market price to be unaffected by the strategic speculators. ■

In Appendix 1.A, I demonstrate that any amount of non-strategic speculation in the futures market will result in a positive expected profit for a manipulator and a compromised price signal, though the expected profit declines as non-strategic noise trading declines. The price impact of a trade is always smaller with non-strategic noise traders, but is generally hook-shaped in the share of futures noise trading due to non-strategic speculators, with a minimum somewhere between zero and 100% non-strategic speculators and a maximum when there are no non-strategic speculators, in which case the price impact parameter is the same as the price impact without manipulation. Thus, the more non-strategic speculators there are, the greater the profit from manipulation and the more likely a manipulator would attempt it.

To summarize the implications of the model thus far, manipulation becomes less likely in an equilibrium setting because other traders will be aware of the presence and strategy of potential manipulators. Knowing the strategy of potential manipulators given the parameters of the markets in which they trade, traders can (1) attempt to infer the spot market trade(s) of the manipulator(s) from the futures market order flow and adjust their expectations of how much volume in the spot market is likely to be traded by uninformed futures traders, (2) adjust downward the price impact coefficient if they believe the share of trading that is uninformative about the true value of the asset has increased in order to prevent prices from straying too far from their expectations, and (3) adjust their own trading strategies in response to the manipulators. In richer models, traders would have even more options available to counteract the effects of a manipulator, although manipulators would undoubtedly gain new tricks as well.

Furthermore, the results demonstrate that if manipulation is profitable, the spot price will be manipulable. In the presence of manipulation the realized spot price will, on average, stray further from the asset's true value (making it a less efficient and informative signal) because the manipulator has increased the variance of uninformative trading. That is, if the market makers cannot perfectly filter out the manipulator's trade from the spot market order flow, then the manipulator's trading in the spot market will increase the amount of unpredictable uninformed volume relative to the amount of informed volume, reducing the ability of the market makers to infer the true value of the asset. Conversely, if manipulation ies expected to yield zero profit because all futures traders are expected to repeat their futures trades in the spot market, the spot price will not be manipulated even if the manipulator trades because the market makers can perfectly filter out the uninformative orders of all of the futures traders when computing the spot price.

1.3.3 Kumar and Seppi (1992) with a Physical Futures Contract

In the introduction to their 1992 paper, Kumar and Seppi state that "If futures accounts are closed through 'cash settlement' rather than 'physical delivery,' then manipulative spot trading can improve the settlement price." By "improve the settlement price," they mean improve the settlement price for the manipulator, in fact, the quality of the settlement price as a signal of the asset's value declines. However, once the same assumptions utilized in the cash-settled futures model are applied to the physical futures setting, one finds that if manipulation is profitable with a cash-settled futures contract, then it should also be profitable with a physical futures contract.

The key to demonstrating that a physical futures contract generates a model identical to that generated by a cash-settled future is to ensure that all traders follow the same strategy regardless of the type of futures contract offered. Note first that if only one type of contract is offered, hedgers will attempt to hedge using whichever is available, and speculators would speculate using whichever futures contract is available. Secondly, under one key assumption identified in Kyle (2009): that it is possible to convert a cash-settled futures contract to a physical futures contract (or vice versa) with zero transactions cost, a condition met in this model, a physical futures contract is exactly equivalent to a cash-settled futures contract. To convert a cash-settled futures contract to a physical futures contract, one needs to be able to purchase or sell in the spot market at exactly the settlement price of the spot contract. For instance, if I own a cash-settled futures contract, then I owe the futures price and will receive the settlement price of the spot contract, $E[\Pi] = E[S - F]$. To convert this position to a physical futures position, if I can buy a spot contract at exactly the spot market settlement price with no transaction cost, then I will own the asset and pay the settlement spot price, making my total expected profit $E[\Pi] = E[S - F + v - S] = E[v - F]$, which is exactly the expected profit from owning a physical futures contract. Thus, in the model of Kumar and Seppi (1992), to purchase a physical futures position, one must both purchase a cash-settled futures contract and a spot contract.

Continuing on, in the model of Kumar and Seppi (1992), the futures noise traders are expected to speculate non-strategically by buying a cash-settled futures contract and holding their position to expiration, which generates an expected profit of $E[\Pi|q_c] = E[q_c(S - F)]$, where q_c is the speculator's cash-settled futures trade. Buying a physical futures contract and selling in the spot market would generate a profit of $E[\Pi|q_p, q_s] = E[q_p(v-F)+q_s(v-S)|q_p, q_s]$, where q_p represents the quantity of physical futures traded and q_s indicates the quantity traded in the spot market. If the trader buys the same amount in the physical futures market as he sells in the spot market, $q_s = -q_p$, then $E[q_p(v - F) + q_s(v - S)|q_p, q_s] = E[q_p(S - F)|q_p]$. Therefore, to exactly convert the model of Kumar and Seppi (1992) to one with a physical futures contract instead of a cash-settled futures contract, it must be assumed that the non-strategic speculators of Kumar and Seppi's model would trade e in the physical futures market and $-e$ in the spot market.

Under the original model, the strategy of the manipulator is to take a cash-settled futures position Δ and trade z in the spot market, which generates an expected profit of $E[\Pi|\Delta, z] = E[\Delta(S - F) + z(v - S)|\Delta, z]$. Under a physical futures contract, this strategy can be replicated by trading Δ in the physical futures market, and $z - \Delta$ in the spot market, which would generate an expected profit of $E[\Pi|\Delta, z] = E[\Delta(v - F) + (z - \Delta)(v - S)|\Delta, z] = E[\Delta(S - F) + z(v - S)|\Delta, z]$. In equilibrium, the market makers will know that this is the strategy of the manipulator and will take it into account when computing their expected spot and futures prices.

Assuming, as before, Bertrand competition among market makers and no information in the futures market, the spot price will be the expected value of the asset given the spot and futures orders:

$$S = E[v|y_s, y_f] = E[v|y_f] + \lambda(y_s - E[y_s|y_f]) = \mu + \lambda(y_s - E[y_s|y_f]) \tag{1.21}$$

To see that the spot price will end up the same whether there is a physical futures contract or a cash-settled futures contract, equation (1.21) shows that one must determine if the difference between actual and expected volume, $y_s - E[y_s|y_f]$, is the same under each type of contract. If $y_s - E[y_s|y_f]$ is the same in both situations, then by the normality of the random variables, $\lambda = Cov(v, y_s|y_f)/Var(y_s|y_f)$ will be the same under both contracts. $y_s - E[y_s|y_f]$ represents the market maker's best estimate of the net order flow that is driven by information. Therefore, if the market maker's estimate of informed trading does not change when switching from a cash-settled futures contract to a physical futures contract, keeping all traders' strategies the same, then the spot price will not change either.

Assuming that the futures noise traders are non-strategic speculators, the spot order flow due to futures traders under a cash-settled futures contract would be z. Under a physical futures contract, it would be $z - \Delta - e$, since both the manipulator and the noise traders trade in their physical futures position to get cash-settled positions. Since the total spot order flow from futures traders is $z - \Delta - e = z - y_f$ and the futures order flow y_f is known, the inference problem of the market makers in the spot problem is the same as with a cash-settled futures contract, they need to guess z from the futures order flow. This can be seen in the following comparison, which demonstrates that $y_s - E[y_s|y_f]$ is the same under either contract.

Variable	Cash-Settled Future	Physical Future
y_f	$\Delta + e$	$\Delta + e$
y_s	$X + u + z$	$X + u + z - \Delta - e$
$y_s - E[y_s\|y_f]$	$X + u + z - E[z\|y_f]$	$X + u + z - \Delta - e - E[z - \Delta - e\|y_f]$
		$X + u + z - y_f - E[z - y_f\|y_f]$
		$X + u + z - E[z\|y_f]$

If we instead assume that futures noise traders are hedging and that the only futures contract available is a cash-settled futures contract, the hedgers would convert their cash-settled futures positions to physical futures positions in the spot market by trading the same quantity in the spot market as was traded in the cash-settled futures market. This causes the spot market order flow driven by futures traders to be $z + e$. If the available futures contract, were a physical futures contract, the hedgers would not trade in the spot market, while the manipulator would make his spot trade z and convert his physical futures position to a cash-settled one by trading $-\Delta$. Therefore, under a physical futures contract, the spot market order flow driven by futures traders would be $z - \Delta$. In the cash-settled case, the market makers must infer $z + e$ from the futures order flow, while in the second case, they must infer $z - \Delta$. However, since they know the futures order flow y_f and that $y_f = e + \Delta$ guessing e given y_f is the same as guessing $-\Delta = y_f + e$ given y_f. This can be seen in the following comparison demonstrating that $y_s - E[y_s|y_f]$ will be the same under either contract:

Variable	Cash-Settled Future	Physical Future
y_f	$\Delta + e$	$\Delta + e$
y_s	$X + u + z + e$	$X + u + z - \Delta$
$y_s - E[y_s\|y_f]$	$X + u + z + e - E[z + e\|y_f]$	$X + u + z - \Delta - E[z - \Delta\|y_f]$
		$X + u + z - y_f + e - E[z - y_f + e\|y_f]$
		$X + u + z + e - E[z + e\|y_f]$

If only some futures noise traders hedge, while the other speculate non-strategically, the same argument can be used to show that the spot price will be the same under either type of contract.

Given that the formula for the spot price will be the same under either contract type, the strategy of the manipulator will be the same under either contract as will his profit and effect on the market. Therefore, if manipulation is profitable and moves the spot price because there are non-strategic speculators with a cash-settled future, the same strategy would also work with a physical futures contract, all else equal. Thus, Kumar and Seppi (1992) find that manipulation is profitable with a cash-settled futures contract but unprofitable with a physical futures contract because they fail to notice that when they switch to a physical futures contract, if they continue to assume that futures noise traders do not trade in the spot market, then they are actually assuming that the strategy of the futures noise traders has changed from non-strategic speculation to hedging.

1.4 Screening for Manipulation

The conditions for manipulation in the theoretical model presented above suggest two potential screens for manipulability. Recall that the first condition identified for profitable manipulation was that the price impact of a trade in the futures market must be significantly lower than the price impact of a trade in the spot market. Otherwise, the manipulator's large position in the futures market would move futures prices adversely, diminishing and perhaps eliminating the profit from manipulation. Thus a screen for manipulability in a pair of markets would compare the price impact in the futures market against the price impact in the related spot market. The more similar the price impact in the futures market is to the price impact in the spot market, the less profit manipulation will generate and the less likely manipulation will occur. Unfortunately, this is not easily done in the case I examine here due to limitations in the available data and the nature of the many futures contracts utilized by the accused manipulator.

The second condition for profitable manipulation which I identify is that some traders in the futures market must consistently speculate non-strategically by taking a cash-settled futures position and not trading in the spot market. Thus, one could screen for manipulability by examining the spot market trades of those who held a position in the futures market to expiration. The share of the open positions in the futures market at expiration that cannot be matched with corresponding trades in the spot market would indicate the share of non-strategic speculation. While this information could potentially be gathered by surveying traders, it would be very costly and difficult to obtain a complete picture of the proportion of trades driven by hedging versus non-strategic speculation.

A more useful screen would utilize data that is easily obtainable and simple to analyze. I therefore suggest and test a screen for attempted manipulation that utilizes obtainable data from an allegedly manipulated market. This screen can be employed when manipulation is suspected to corroborate other evidence of manipulation and as a first pass at identifying irregularities in market activity.

The theoretical model of manipulation presented above suggests that if manipulation is successful in moving the spot price, then manipulative trading increases the amount of trading that is both uninformative about the spot price and indistinguishable from the informative trading. This reduces the average informativeness of trading and causes other traders to infer less about the true value of the asset from each trade, meaning that the price impact of a trade will be lower. This suggests that if one suspects manipulation has affected prices during a particular period of time, one can compare estimated price impacts from potentially manipulated periods against price impacts in other periods or in other similar markets. A finding that price impacts were lower than would be expected given the comparison group would corroborate the evidence that the market may have been manipulated in that period.

Note that the model implies that the equilibrium given any non-zero amount of non-strategic speculation will involve manipulation. For price impact to decrease, it must be that some other force has increased the wealth of the manipulator, the number of manipulators per Kumar and Seppi (1992), or changed the amount of non-strategic speculation. Thus, this screen is most likely to be effective if an exogenous force has changed market conditions. As the company accused of manipulation was allegedly trying to manipulate the price difference (or spread) between two spot markets in the case I am about to examine, a third reason to expect increased manipulation arises: that the correlation between the value of the assets in the two markets falls. When the correlation between asset values in two markets falls, it increases the likelihood that a trade will move the price in one market more than the price in the other market, thereby increasing the ability of a manipulator to move the spread between the prices via strategic trading. If the value of the assets in the two markets were perfectly correlated, then any trade in either market would have the same effect on both prices, moving the spread would be nearly impossible, and manipulation would be unprofitable.

I propose three estimates of price impact that can used for such a screen. The first measures the resilience of the market, that is, the frequency and speed with which, after a trade, market prices return to their pre-trade levels before the next trade. When trading is perceived to be less informative or is less correlated with new information, price will be more likely to return to its prior level after a trade and will return more quickly. The second measure is the immediate price impact of a trade estimated from a vector error correction model (VEC). Work by Sandås (2001) and Frey and Grammig (2006) suggests that prices at each level of the order book will be based on the bidder or offerer's expectation of price if a trade were to take their full bid or more than their full bid. For instance, suppose that there are 10 units available at the best bid, 5 at the second best bid, and 20 at lower-priced bids. The price for the best bid will be based on that bidder's expectation of price given a trade that is 10 units *or larger*, while the price of the second best bid will be based on that bidder's expectation of price given a trade that is 15 units *or larger*. Therefore the price at each level is an overestimate of the revision in expectations given a trade of a particular size, e.g. $E[p|q > 10] > E[p|q = 10]$. Thus, an estimate of the immediate price impact of a trade, where the immediate impact is defined as the part of the change in price between the second just before the trade and the end of second in which the trade occurred that was caused by the trade, is likely to be an overestimate of a trader's new expectation of price after the trade occurs.

The final estimate, the permanent price impact of a trade, is derived from the impulse response function of the VEC model. This estimator makes use of the full evolution of the market after a trade. Although the estimators for the immediate and permanent price impacts are likely to overstate the causal effect of a trade on price if trades are correlated with new public information, the fact that manipulation reduces both the correlation between trading and information and the causal impact of a trade suggests that if estimates of the permanent price impact of a trade decline during an allegedly manipulative period, then the change in the price impacts corroborates the evidence for manipulation.

1.4.1 FERC's Allegations Against BP

On September 1, 2008 at 10:00 am, Hurricane Gustav made landfall between Morgan City and Port Grand Isle, Louisiana leaving 1 million customers in Louisiana without power, shutting off 92% natural gas production and causing damage to at least 21 natural gas pipelines in the Gulf of Mexico. As Louisiana began to recover from Gustav, reports of another hurricane arrived. Although Hurricane Ike formed on September 1st far out in the Atlantic Ocean, by September 7th it reached Cuba and then continued into the Gulf of Mexico, headed for Texas. By September 13th, when Ike reached land, just south of Galveston, Texas, its strength was reduced to a category 2 hurricane, but the damage it inflicted was severe. Nearly 4 million customers in Texas, Louisiana, Oklahoma, Arkansas, and Missouri were left without power, 22 natural gas pipelines reported significant damage and without functioning pipelines, 40% of natural gas production in the Gulf of Mexico remained off-line for over a month. Electricity was not fully restored until 3 weeks later.

Given the severe disruptions to both supply and demand for natural gas, it is not surprising that the usual price relationships between various locations in Texas and Louisiana broke down during this period. Henry Hub, the delivery point for the New York Mercantile Exchange (NYMEX) physical natural gas futures contract and the central location for natural gas trading in the U.S., was not fully operational for over a month. In this period, natural gas flowed from West Texas to East Texas to Louisiana and then up to the Northeast and Midwest U.S. Disruptions on the pipelines between Texas and Louisiana due to the hurricanes therefore backed up natural gas supplies in Texas while severe reductions in supply from the Gulf of Mexico reduced supply flowing into Henry Hub from the south. In combination, these phenomena lowered prices in East Texas relative to those in Louisiana, the key Louisiana price being that at Henry Hub and the key East Texas price being the Houston Ship Channel (HSC) price.

Figure 1.3 shows the relationship between the Henry Hub and HSC spot prices for 2008 through 2009. The spot prices in the figure are the weighted average prices of trades executed in the next-day fixed-price markets at each location (gas traded in these markets is delivered the next day). The dotted black line shows the spread, or price difference, between the two locations. As can be seen, the spread increased dramatically after Hurricane Ike and did not fully recover to pre-hurricane levels until April 2009. From

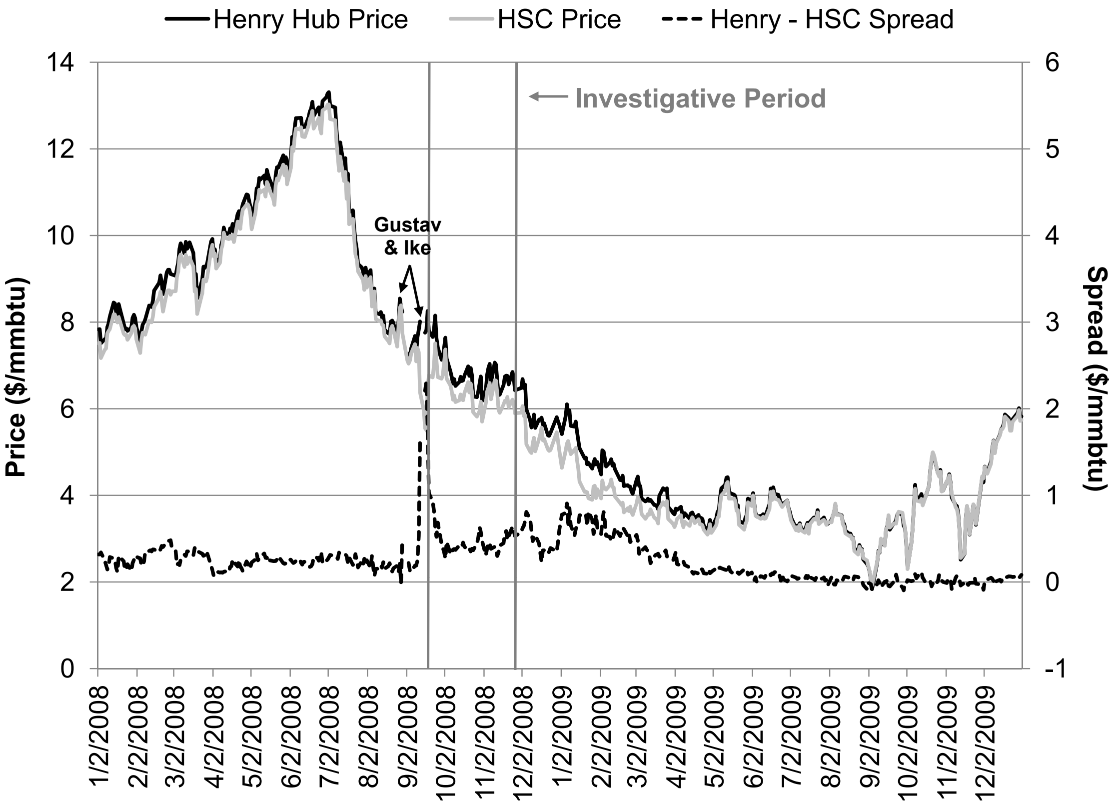


Figure 1.3: Next-Day Natural Gas Prices at Henry Hub and HSC
Source: ICE.

September 18, 2008 to November 30, 2008, FERC alleges that BP diverted supplies from Katy (just a few miles west of Houston) to HSC in order to lower the spot price of natural gas at HSC relative to Henry Hub (widening the spread), thereby increasing BP's profit on a cash-settled futures spread position that benefited from lower prices at HSC relative to Henry Hub.[9]

Entering the month of September 2008, among other positions, BP held a long position in natural gas at Katy (i.e., they owned natural gas at Katy), 200,000 mmBtu of pipeline capacity capable of moving natural gas from Katy to HSC, and a cash-settled futures spread position between HSC and Henry Hub. The futures position had been established

[9] BP's cash-settled position was technically in forward contracts, non-standard futures contracts, which are subject to fewer regulations. Under the Dodd-Frank law, the contracts utilized by BP are now designated as futures contracts. To keep terminology simple and consistent, I refer to these contracts as futures contracts. The differences between forward and futures contracts are not relevant to the analysis.

in January 2008 and benefited when the spread widened, i.e. the price at HSC went down relative to the price at Henry Hub. Enjoying the massive profit earned when the spread widened immediately following Hurricane Ike and recognizing that sales at HSC generally move prices down, FERC alleges that BP attempted to increase their profit on their futures spread position by utilizing their pipeline capacity and their long position in natural gas at Katy to move gas from Katy to HSC in the days following Hurricane Ike, to suppress the price at HSC relative to Henry Hub.

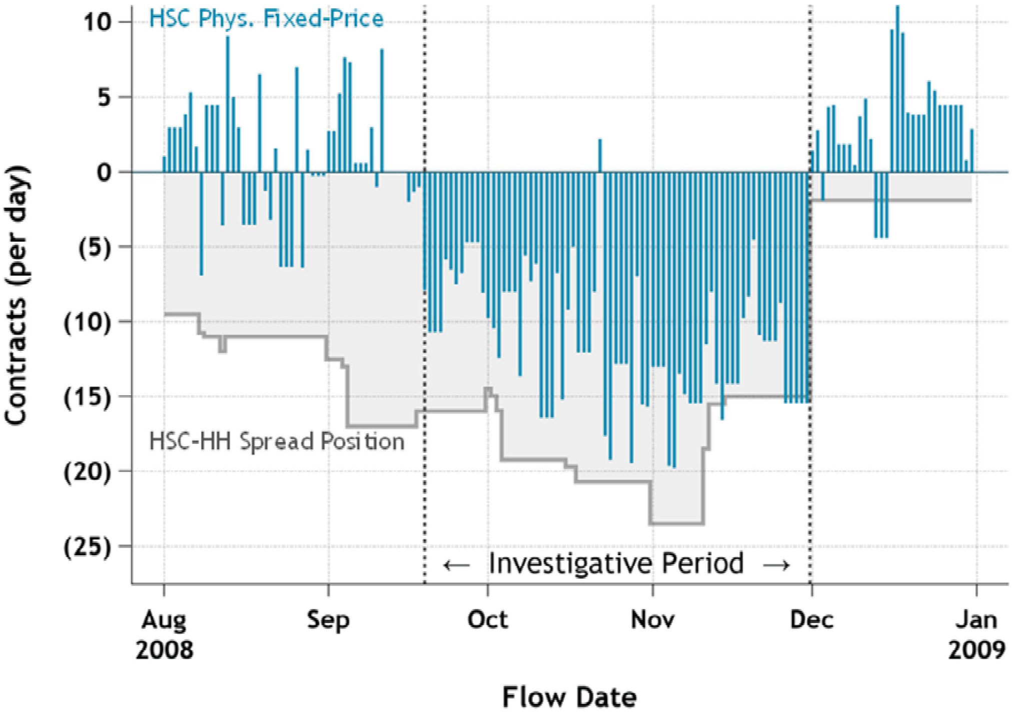


Figure 1.4: BP Futures Spread Position and Physical Fixed Price Trading
Source: BP Show Cause Order.

Observing that their strategy seemed to have been effective in the last weeks of September 2008, FERC alleges that BP then created an even larger futures spread position between HSC and Henry Hub for the months of October and November 2008 and increased their long position at Katy in anticipation of further manipulation attempts. Figure 1.4, which comes from the BP Show Cause Order issued by FERC, shows BP's trading in the fixed price market and their HSC-Henry Hub futures spread position. The dark blue bars illustrate BP's fixed-price trades, with negative values indicating sales, and the light grey shaded area illustrates BP's spread position, where the negative values indicate that the company held a short spread position and would therefore pay the HSC price and receive the Henry Hub price. From this figure, one can see that during the investigative period, BP sold natural gas in the next-day fixed-price market at HSC nearly every day during the investigative period and that the quantity of their sales was correlated with the size of their short position in the HSC-Henry Hub spread.

1.4.2 Data

To investigate the effect of BP's trading on the next-day fixed-price natural gas markets in Texas and Louisiana, I obtained data from the Intercontinental Exchange (ICE) which hosts the trading platform on which the allegedly manipulative trades took place. ICE provides a centralized electronic exchange for trading standardized over-the-counter contracts for physically-delivered natural gas (among many other commodities and products). Although a significant portion of natural gas trades are executed bilaterally or via brokers, virtually all next-day fixed-price natural gas is traded on ICE. Thus, the data I will utilize to test my screen encompasses nearly all of the trades in the relevant markets.

The data obtained from ICE contain all bids to buy, offers to sell, and consummated trades that occurred from 2008 to 2009 in the next-day fixed-price natural gas markets at HSC, Henry Hub, and Katy. In particular, the data contain one observation each time a bid or offer was initiated, cancelled, temporarily withdrawn, changed, or consummated (e.g. another trader accepted the bid or offer and a trade occurred). The bids and offers are linked by order identification numbers, such that a particular bid or offered can be tracked from the moment it is initiated until it is cancelled or consummated. The data do not contain any identifying information about the traders who initiate the bids, offers, and trades, nor is it possible to group bids, offers, and trades by trader. If a trader accepts a bid, the trade is classified as a sale, as the initiator of the trade was a seller. Similarly, if a trader accepts an offer, the trade is classified as a purchase.

The data contains 483 trading days. The main trading hours for a particular delivery date are 7:30 am to 12:30 pm on the business day prior to delivery, although trading for each delivery date begins at 12:31 pm two business days prior to delivery. At the three hubs in this dataset, approximately 95% of trades for a particular delivery date occur between 8:00 am and 10:30 am. The smallest price increment or tick-size, is $0.0025. The results will present price changes and price impacts in ticks rather than dollars to facilitate interpretation.

Using the raw list of bids and offers, I transformed the data for each location into an estimate of the order book that existed at the end of each second.[10] An order book contains the list of active bids and offers that could be accepted by other traders at a particular moment in time. In the order book, active bids and offers are arranged in ascending order and traders must generally accept the best bid (a.k.a. the highest bid) if they would like to sell and the best offer (a.k.a. the lowest offer) if they would like to buy. Differing from futures markets where anyone can trade with anyone else, in the ICE next-day fixed-price natural gas markets, traders are constrained to only trade with others with whom they have pre-arranged credit. To alert traders whether they can accept a particular bid or offer, each individual trader sees the order book on his screen with a green mark indicating that he can take a particular bid or offer because he has credit with the other trader, or a red mark indicating that he does not have credit with that bidder or offer and therefore cannot accept that bid or offer.[11] As a result, traders are only required to accept the best bid or offer within the group of bids and offers that were initiated by those with whom they have credit. This frequently causes bids and offers at higher levels of the order book to be consummated before the best bid or offer. It also leads to periods where the best bid is above the best offer, but no trade occurs, causing the bid-offer spread to be negative.

Table 1.1 displays a few summary statistics about each market. The average bid-offer spread between 7:30 am and 12:30 pm, when one exists is larger at HSC than at the other

[10] As the precision of the time variable in my dataset was only to the nearest second, I often had to guess the sequence of bids and offers within a second using information from earlier and later periods. When it was not possible to guess the sequence with any accuracy, I chose the bid or offer that would make the order book the steepest. Differences between the steepest possible order book and the flattest possible order book were generally small.

[11] The order book is otherwise anonymous.

	HSC	Henry	Katy
Avg. Bid-Offer Spread	\$0.094	\$0.067	\$0.074
% of Seconds with Spread	67.0%	77.9%	69.1%
Avg. No. of Quotes	4.6	7.9	9.3
Avg. No. of Trades/Day	66.9	119.5	98.1
Avg. No. of Traders	21.8	36.4	35.7

Table 1.1: Summary Statistics by Market

two hubs. HSC is also somewhat less likely to have both a bid and an offer available at any moment in time. On average, there are 4-5 quotes (bids and offers) available at HSC, or roughly 2.5 bids and 2.5 offers, while Henry and Katy typically have closer to 8 or 9 quotes. HSC also has the fewest trades and the fewest traders per day. Traders represents the number of unique people who bought or sold in a day.

1.4.3 Resilience

The first measure of price impact that I utilize is the resilience of the order book, where resilience is defined as the frequency and speed with which, after a trade, bids or offers return to their pre-trade levels before the next trade on the same side of the book. As traders who accept all available bids or offers at a particular price level have an instantaneous effect on the price at which the next trader could sell or buy, these trades instantaneously change prices even if they have no permanent effect on beliefs or prices. When trading is perceived as less informative or is less correlated with information, prices will be more likely to recover after a trade and will recover more quickly. As BP allegedly manipulated the market by selling, I focus the analysis on the bid side of the order book.

Utilizing the raw list of active bids and sales, I measure the share of sales after which a new bid with a price greater than or equal to the sale's price arrives prior to the next sale. Given that a better or equally-priced bid arrives before the next sale, I also measure the median time between the trade and the arrival of the better or equally-priced bid. As multiple bids are often accepted by sellers in the same second, I look for a new bid with a price greater than or equal to the highest-priced bid that was accepted in each second.

Many trades only accept part of a bid, and therefore have no instantaneous effect on price. These trades are dropped from the analysis.[12]

By design, the order book only records prices at the end of each second, and therefore, any fleeting bids lasting less than one second are not displayed in the order book. Using the raw list of active bids allows me to observe the fleeting bids and avoids any sequencing inaccuracies that occurred during my recreation of the order book. Many sales also occur at higher levels of the order book, that is, the seller accepts the second or third best bid because he does not have credit with the buyer associated with the first or second best bid. Looking for new bids at equal or better prices in the raw list of bids allows me to include these trades in the analysis with the same weight as a trade at the best bid or offer.

1.4.4 Vector Error Correction Model of Price Impact

Trading at HSC, Katy, and Henry Hub is strongly entwined together. Consequently, to accurately model the evolution of prices and trading at HSC, the activity in the other markets should be included. To isolate the effect of a sale on the available bids, I model bids and offers and purchases and sales separately at each hub. As BP sought to sell natural gas at HSC, they either submitted offers or accepted bids. Thus, the bids were not controlled by BP and reflect reactions to BP's sales activity. To improve identification, I include variables for the total quantity available at the best bid and best offer, and the price of the second best bid and second best offer.

Hasbrouck (1991a,b) was the first to suggest modeling the price impact of a trade with a dynamic system and showed that ignoring the indirect impacts of trades could seriously bias the estimated price effect of a trade. For instance, if a purchase increases the price and causes a sale to occur, the immediate effect of the purchase overestimates

[12] If there are multiple active bids with the same price and a seller does not accept all of the bids at that price, this would also not change the price. However, the current methodology does not allow me to flag these trades and they are therefore included in the analysis. These trades do reduce the quantity available at the sale's price, however, meaning that new bids arriving at that price or better would still indicate a return of the order book to its pre-sale state.

the trade's permanent effect on price. His original model was a vector autoregression (VAR) containing two endogenous variables, the change in the midpoint between the bid and offer (m_t), and the signed volume of a trade (x_t).

$$\Delta m_t = \lambda_0 x_t + \sum_{s=1}^{L} \lambda_s x_{t-s} + + \sum_{s=1}^{L} \gamma_s \Delta m_{t-s} + \varepsilon_t$$
$$x_t = \sum_{s=1}^{L} \beta_s x_{t-s} + \sum_{s=1}^{L} \alpha_s \Delta m_{t-s} + \eta_t$$

This model also accounts for his earlier observation, in Hasbrouck (1988), that expected trades cannot hold any private information and therefore it is the unexpected part of trades the influences prices. If other traders commonly believe there is a high probability that a purchase of one unit will occur shortly, that expectation will already be incorporated into the price. If the one unit purchase actually arrives, the price should not change significantly. Therefore, to estimate the price impact of a trade, one must also include an estimate of the expected trade. The VAR model implicitly incorporates the expectation of a trade, though it is limited to the expectation that can be formed given the variables included in the model. That is, if the expectation of future trade quantities is based on past trades and past midpoints, then since those variables are already included in the model for the midpoint of the bid and offer prices, the expectation of the future trade quantity is incorporated as well.

In the years since the Hasbrouck papers were published, a number of authors have built on his methodology. In Engle and Patton (2004), the authors estimated the price impact of trades using a VEC model with the bid and offer modeled separately, rather than taking the midpoint. They motivate this approach through the empirical results of Jang and Venkatesh (1991) which illustrate that after no change at all, the most common effect of a buy is to move only the offer, and the most common effect of a sale is move only the bid. Modeling only the midpoint obscures this valuable information and the dynamic effects of a trade. Furthermore, they note that although the bid and offer are generally non-stationary variables, the difference between them, the bid-offer spread, is stationary. That is, the bid-offer spread is relatively constant over time. Therefore, the bid and offer are cointegrated variables, where the cointegration vector is the spread, and the appropriate model is not a VAR in differences but a VEC model. However, to simplify their analysis, they do not model trading as endogenous and therefore are unlikely to correctly identify the effect of a trade.

Building on Hasbrouck (1991a,b) and Engle and Patton (2004), Escribano and Pascual (2006) makes trading endogenous and separately estimates the impact of a buy and a sale on the bid and offer in a VEC model. Mizrach (2008) then demonstrated empirically that ignoring the other levels of the order book, i.e. the higher offers and lower bids, may bias the estimated price impacts. Several papers since then, such as Hautsch and Huang (2012) and Fleming and Mizrach (2009), have made use of these innovations to examine how trading effects prices in various markets.

Modifying the derivation of the VEC model of Pascual et al. (2006), which examines a security traded in many markets, and of Escribano and Pascual (2006), which illustrates the method for a model including bids, offers, and trading information, I derive the VEC model used to analyze whether the price impact of a sale at HSC was lower during the investigative period than would otherwise be expected. Let v_t be the expected true value of next-day natural gas at HSC given all information available at time t on day d. I suppress the subscript d for simplicity. Suppose that the expectations of the true values at Katy (v_t^K) and Henry Hub (v_t^{HH}) differ from that at HSC only by a constant plus long-lived noise, representing transportation costs and supply and demand conditions that vary over long periods of time as infrastructure, technology, and tastes adjust. The noise can also account for medium-term changes from one day to the next driven by weather, etc. the effects of which are likely to be fairly accurately anticipated by traders before trading for a day begins. That is:

$$v_t^{HSC} = v_t$$
$$v_t^{K} = v_t + c^K + \varepsilon_d^K$$
$$v_t^{HH} = v_t + c^{HH} + \varepsilon_d^{HH}$$

c^K and c^{HH} represent the constant price differences and ε_d^K and ε_d^{HH} the noise, which varies across days but not within a day.

Suppose that the expectations of the true value adjust within a day due to shocks to common knowledge, such as new weather reports, and inference from observed trading, which reveals private values and private information. Then the innovation process can be represented as:

$$v_t = v_{t-1} + I_t + \lambda\xi_t \tag{1.22}$$

Here, I_t represents new public information, available to traders at all hubs, and ξ_t is a 6x1 vector representing unexpected purchases and sales at each hub, a.k.a. the trading innovation. As I assume that the difference between the expectations of the true values in each market is constant across a day, new public information within a day affects only expected price levels, not the expected spread between prices, causing the expectation of the true value at any particular hub to evolve according to the same process. This assumption seems to hold in general in the data as the spreads between hubs in each of the periods I define later are stationary variables according to a Dickey-Fuller test with 15 lags. In the periods during and shortly after the hurricane, deviations from the average spread are corrected more slowly than in other periods but all of the price series still generally move together.

I follow Hasbrouck (1991a,b) and Escribano and Pascual (2006) in assuming that the bid and offer generating process at each hub can be represented by:

$$o_t^i = v_t^i + \alpha^i(o_{t-1}^i - v_{t-1}^i) + \sum_{s=0}^{l} A_s^i x_{t-s} + \eta_t^{o,i}$$
$$b_t^i = v_t^i + \beta^i(v_{t-1}^i - b_{t-1}^i) + \sum_{s=0}^{l} B_s^i x_{t-s} + \eta_t^{b,i}$$

o_t^i and b_t^i represent the offer and bid available at the *end* of time t in market i. x_t is a 6x1 vector of the purchases and sales made at time t. $\eta_t^{o,i}$ and $\eta_t^{b,i}$ are idiosyncratic errors caused by market inefficiencies and model misspecifications. This model assumes the bids and offers are driven by the expected true value, inventory considerations (represented by the $A_s^i x_{t-s}$ and $B_s^i x_{t-s}$ terms which indicate that trades affect bids and offers for reasons other than the new information they reveal), and competition between the traders setting the bids and offers which causes bids and offers to be drawn toward the expected value of the asset (represented by adjustment terms $o_{t-1}^i - v_{t-1}^i$ and $v_{t-1}^i - b_{t-1}^i$).

Rewriting each equation as

$$(1 - \alpha^i L)(o_t^i - v_t^i) = \sum_{s=0}^{l} A_s^i x_{t-s} + \eta_t^{o,i}$$
$$(1 - \beta^i L)(b_t^i - v_t^i) = \sum_{s=0}^{l} B_s^i x_{t-s} + \eta_t^{b,i}$$

where L indicates the lag operator, and then dividing by the lag coefficient on the left-hand side and applying the difference operator yields:

$$\Delta o_t^i = \Delta v_t^i + \tilde{A}(L)^i x_t + \tilde{\theta}^{o,i}(L)\eta_t^{o,i} \quad (1.23)$$

$$\Delta b_t^i = \Delta v_t^i + \tilde{B}(L)^i x_t + \tilde{\theta}^{b,i}(L)\eta_t^{b,i} \quad (1.24)$$

The $\tilde{A}(L)^i$, $\tilde{B}(L)^i$, $\tilde{\theta}^{o,i}(L)$, and $\tilde{\theta}^{b,i}(L)$ operators are collections of all the lag operators and coefficients entering the equations. The change in the expected true value can be removed using equation (1.22):

$$o_t^i = o_{t-1}^i + \tilde{A}(L)^i x_t + \tilde{\theta}^{o,i}\eta_t^{o,i} + I_t + \lambda\xi_t \tag{1.25}$$

$$b_t^i = b_{t-1}^i + \tilde{B}(L)^i x_t + \tilde{\theta}^{b,i}\eta_t^{b,i} + I_t + \lambda\xi_t \tag{1.26}$$

Similarly the trading process at each hub is described by:

$$x_t^{B,i} = \gamma_B^i(o_{t-1}^i - v_{t-1}^i) + \xi_t^{B,i}$$

$$x_t^{S,i} = \gamma_S^i(v_{t-1}^i - b_{t-1}^i) + \xi_t^{S,i}$$

$x_t^{B,i}$ and $x_t^{S,i}$ represent the quantity bought and sold at hub i. Note that the error terms, $\xi_t^{B,i}$ and $\xi_t^{S,i}$ are elements of the ξ_t vector determining the innovation in the expected value. Here, the decision to buy is based upon the distance between the best offer existing when the trader makes his decision and the expected true value, as to buy immediately one must accept an offer. Correspondingly, the decision to sell is based upon the difference between the best bid and the true value. Finally, an identifying assumption has been made: that the decision to buy or sell is not related to new public information, I_t. I will discuss the validity of this assumption in the next section. Using equations (1.23) and (1.24) before the difference operator was applied, one can substitute out the expected value of the natural gas:

$$x_t^{B,i} = \gamma_B^i \hat{A}(L)^i x_{t-1} + \gamma_B^i \hat{\theta}^{o,i}(L)\eta_t^{o,i} + \xi_t^{B,i} \tag{1.27}$$

$$x_t^{S,i} = \gamma_S^i \hat{B}(L)^i x_{t-1} + \gamma_S^i \hat{\theta}^{b,i}(L)\eta_t^{b,i} + \xi_t^{S,i} \tag{1.28}$$

To eliminate the trade related shocks, ξ_t, in equations (1.25) and (1.26), one can substitute from equations (1.27) and (1.28). Stacking all of the bid, offer, and trade equations, a VARMA model is obtained as the error terms contain a moving average (MA) of the errors from prior periods. By inverting the MA coefficients on the error term and truncating at lag p, a VAR(p) model in levels is obtained.

$$y_t = \begin{bmatrix} o_t \\ b_t \\ x_t \end{bmatrix} = \begin{bmatrix} C_0^o \\ C_0^b \\ 0 \end{bmatrix} x_t + \sum_{s=1}^{p} C_s y_{t-s} + u_t \tag{1.29}$$

The bids and offers are non-stationary and also contain a single common trend v_{t-1} that causes the non-stationarity. Therefore, since there are 6 prices, and one trend, there are 5 cointegrating vectors, which are known: the bid-offer spread at each hub, the spread between HSC and Katy and the spread between HSC and Henry (the spread between Henry and Katy is redundant). I define the spread between HSC and Katy and HSC and Henry using the bids at each hub. The model can then be rewritten using Granger's Representation Theorem from Engle and Granger (1987) in vector error correction form:

$$\tilde{y}_t = \begin{bmatrix} \Delta o_t \\ \Delta b_t \\ x_t \end{bmatrix} = \begin{bmatrix} C_0^o \\ C_0^b \\ 0 \end{bmatrix} x_t + \sum_{s=1}^{p-1} \tilde{C}_s \tilde{y}_{t-s} + D_p x_{t-p} + E s_{t-1} + u_t \tag{1.30}$$

s_{t-1} represents the 5 spreads in $t-1$. Note that, as x_t is already stationary I keep it in levels.

The immediate price impact of a trade can be measured by the coefficient on x_t in each equation. To obtain the long-run permanent price impact of a trade requires computation of the impulse response function, which illustrates how a shock reverberates through the market over time. For instance, a trade shock has an immediate impact on prices measured by C_0. In the next period, the change in prices last period affects both how prices change this period and how much is traded this period. To calculate the impulse response functions, Lütkepohl and Reimers (1992), advise converting the VEC model back into a VAR model in levels. Therefore, in Appendix 1.B, I explain how to map the coefficients from the VEC model into the VAR(p) model. Given the matrix C_s which contains all of the coefficients for lag s, the impulse response function r periods after the shock is computed as:

$$\Phi_r = \sum_{q=1}^{r} \Phi_{r-q} C_q \text{ for } r = 1, 2, \ldots \tag{1.31}$$

$$\text{where } \Phi_0 = \begin{bmatrix} I & 0 & C_0^o \\ 0 & I & C_0^b \\ 0 & 0 & I \end{bmatrix} \text{ and } C_q = 0 \text{ for } q > p \tag{1.32}$$

The permanent price impact of a sale at HSC, for instance, can be measured by assuming a one unit shock to the HSC quantity sold equation occurs and that there are no shocks to any other variables. The shock vector is then multiplied by the impulse response coefficients Φ_r for each period r after the shock.

1.4.4.1 Identification

To identify the impact of a trade from the preceding VEC model requires that (1) trades are not caused by new public information that would cause price to move even in the absence of the trade, and (2) that trades are not autocorrelated due to exogenous pre-planned trading strategies such as order-splitting. As Hasbrouck (1991a) points out, (1) does not require that trades be entirely independent of information, it does require, however, that information affect trades only through past prices. That is, price increases caused by observation of a demand shock can cause trading, but the demand shock itself cannot affect trading directly. Many theoretical models of financial markets justify this assumption by arguing that prices should always perfectly reflect all available public information and therefore trading on public information would yield no financial benefit. In commodity markets, where demand shocks can cause trading because more of the commodity must be acquired, this seems a difficult assumption to maintain. Further, if market frictions exist or if traders differ in the speed with which they process information, stale orders, with inaccurate prices, may remain available and may be picked-off by other traders. In this case, it is not the trade that is causing the price to move, but the trade's correlation with public information. Thus the price impacts coefficients from the current model are likely to overestimate the effect of trades on price.

To perfectly identify the potential effect of manipulative trading, uncorrelated with information about the asset's true value, would require either knowledge of which trades were manipulative (and not information-motivated), which may be known to investigators of a particular case, or the ability to separate trades into those motivated by new information versus those motivated by private liquidity needs. Note that this model cannot prove whether a trade was manipulative or not, it only produces estimates of price impact *assuming* the trades are not motivated by information. However, my purpose is to identify whether manipulative trading may have occurred. If manipulative trading is occurring it has two effects, as discussed in the theoretical model. First, it increases the amount of uninformative trading, which reduces the correlation between trading and information. Second, it lowers the causal effect of trade on price. As both the correlation and causal impact move in the same direction if manipulation occurs, a lower price impact coefficient during a period under investigation corroborates the evidence that manipulation may have occurred.

A number of empirical papers, such as Biais et al. (1995), have documented strong positive autocorrelation in trading activity, meaning that buys tend to follow buys and sales to follow sales. One explanation for such a pattern is that traders who need to trade a large quantity divide the total quantity they need to trade into many small trades to reduce their price impact on the market, a strategy known as order-splitting. Theoretical models by Obizhaeva and Wang (2013) and Alfonsi et al. (2010) confirm that if part or all of the price impact of a trade is transient, that is, if price and the quantity of outstanding bids and offers are expected to recover after a trade, then order-splitting is the optimal strategy. Thus a past trade would not be the cause of the current trade, it would just be correlated because of a trader's pre-planned strategy. This may also be a problem for estimating the effect of trades at one hub on the trades at another if two trades are executed to profit from a price divergence across hubs, but are not executed simultaneously. If order-splitting and arbitrage are the main reasons for the correlation between current and past trades, this will bias upward the estimates of price impact from the impulse response functions. However, if the upward bias across periods and locations is similar, because the manipulator's trades also generated autocorrelation, then it may not bias the comparisons over time and across hubs that I use to infer whether price impact declined during the investigative period.

If publicly observable features of the order book influence both trade and price changes, then excluding them from the model introduces unnecessary bias in the estimates of price impact. That is, suppose the price of the second best bid just decreased, getting farther from the first best bid. A trader may take this as a sign that prices are decreasing and he should trade now to get the best price. If other bidders also see the decrease in the second best bid as a sign the price should decrease, then the price of the best bid would go down even if the trade does not occur. Thus, the best bid decreases mainly because the second best bid decreased and failing to control for the second best bid in the model would yield estimates that overstate the effect of the trade. Therefore, to improve identification, I augment the model with the second level of prices in the order book, that is the second best bids and offers and the quantity available at the best bid and best offer. Adding the second level of prices adds 6 more spreads to the model, the spread between the second best bid and first best bid, and the spread between the second best offer and first best offer at each hub. As the quantity at the best bid and best offer is stationary according to a Dickey-Fuller test with 15 lags, I add it to the model in levels.

The basic identifying assumptions of the model are that past trades and features of the order book are not correlated with the current error term. Additionally, I assume that the bid (offer) side of the order book at the hub at which a sale (buy) occurs *is* causally affected by a same second sale (buy). The other side of the order book at that hub and the order book at other hubs can only be causally affected by that trade with a lag. For example, of the contemporaneous trades at the three hubs, only a sale at HSC is included in the equation and is therefore permitted to affect the contemporaneous bids at HSC (as a sale at HSC generally will have an instantaneous and automatic effect on the bids at HSC even if the bidders take no action post-trade). Sales at HSC are therefore assumed to be uncorrelated with the error terms in the equations for the bids at HSC. All other contemporaneous trades are not included in the regressions for the bids at HSC. Similarly, only contemporaneous *purchases* at HSC are included in the *offer* regressions for HSC. This assumption is justified if it takes time to observe and digest how a trade changes the expectations of those who are not involved in the trade. If allowing trades at other hubs or purchases at HSC to affect the contemporaneous bid at HSC would generate non-zero estimates of their price impact, I assume it is due to the correlation between trading and new information and does not indicate a causal relationship.

1.4.4.2 Estimation and Testing

The data for the next-day natural gas markets at HSC, Katy, and Henry Hub have a problem that does not occur in the data for equity and other financial markets that are generally utilized in studies of trading activity. Namely, in ICE next-day natural gas markets there are many periods when there is simply no bid or offer available at one or all of the hubs. It is not that I face the usual sort of missing data problem, where there are bids and offers and I just don't see them. The problem is that sometimes no one is offering to buy or sell. A trader who wants to buy natural gas at HSC is completely unable to do so in a second when there is no offer available. The fact that no one is offering to buy or sell is therefore meaningful. Ignoring observations where a price is missing at one or more hubs eliminates valuable information and can cause inaccurate estimates of the dynamics the system. For instance, estimating the model with only the observations for which all variables and all necessary lags exist yields estimates that indicate there is no cointegrating relationship between hubs, despite the strong co-movement in prices that can be seen from casual observation of the price series.

To account for the missing data requires several adjustments to the usual estimation strategy. Note first that the trade quantity and the quantity at the best bid and offer do not have a missing data problem, as if there are no bids and offers, the quantity variables are just zero. Second, except for the spread variables, which I treat differently, the prices enter the model as price changes. If a price was missing last period and is still missing now, I record it as a zero price change since the missing-ness of the price did not change. Thus, the price changes are only missing if the price was not missing last period but is now missing (it went to missing) or if it was missing last period but now is not missing (it came from missing).

Third, I make two versions of each of the price change variables. In one version, which is used when that variable is the independent variable, the to-missing and from-missing observations remain missing. In the second version, which is used when that variable is a dependent variable, I fill in all missing observations with zeros. I then generate two dummy variables for each variable indicating in one if the variable came from missing and in the other if the variable went to missing.

In each equation, one lag of each spread appears. The spread is based upon price levels and therefore if one of the two prices that make up the spread is missing, the spread is missing. Note that if a price came from missing, its price change will be missing but the spread may not be. To account for the effect of all the reasons a spread may be missing, I generate several different variables. First, I generate a dummy variable indicating whether each price was missing, e.g. I generate one dummy variable indicating if there was no bid at HSC, and another indicating if there was no offer available at HSC, etc. I also create dummy variables indicating, for each hub, if the bid-offer spread was missing, if the HSC-Henry spread was missing, and if the HSC-Katy spread was missing. This allows the effect of a non-existent spread to vary depending on the reason that it does not exist. Note that I define the HSC-Henry and HSC-Katy spreads using the difference between the best bids at each hub as these are least likely to be missing.

Therefore in each price change regression, the independent variable has missing observations whenever it came from missing or went to missing. The dependent variables are the adjusted versions of each variable, with zeros where they were missing and dummy variables indicating that they were missing and why. In this way, I only lose the observations in each equation where the independent variable was missing and can measure and control for the effect of other variables being missing. Since every disappearance of a bid or offer is matched by a reappearance, it is equally likely that a price change will be missing because the bid disappeared as because a bid reappeared. Therefore, I assume that the expectation of the error term is still zero in all of the price change equations. That is, if a price becomes missing because of a bad shock and it comes back from missing because of a good shock, both the top and bottom of the distribution of the error term are truncated and the expectation of the error will not change. This strategy may still underestimate some of the dynamics that pull a price back from missing or send it to missing, but is a vast improvement over totally ignoring the missing data. For simplicity, I treat all of the dummy variables indicating whether a variable was missing as exogenous. In the impulse response analysis, I estimate the effect of a shock assuming nothing was missing and that the shock does not cause anything to become missing.

As each of the 24 equations uses different observations and dynamics are severely biased if any are lost, I currently estimate the model equation-by-equation to obtain the coefficients for the impulse response function. Following much of the literature, to reduce the size of the data and account for the differences in duration between events that arise due to differences in trading activity across a day, I drop any second in which none of the variables change at any of the hubs.

I divide the data into six periods based upon the hurricane and manipulation periods and their effect on the spread between HSC and Henry Hub prices. The first period, the pre-hurricane period, runs from the beginning of the data, January 1, 2008, to August 31, 2008 and has roughly 221,000 observations. As Hurricane Gustav made landfall on September 1, 2008 and Hurricane Ike on September 13, 2008, I mark the month of September 2008 as the Hurricane period. It has roughly 22,600 observations. BP's attempts to manipulate allegedly continued until the end of November 2008. As market conditions seem to have been quite different in October and November 2008, I define October 2008 as one period and November 2008 as another. They have roughly 29,000 and 25,000 observations respectively. As the spread between Henry Hub and HSC did not recover until mid-April 2009, and as trading patterns seem to have shifted considerably toward the end of 2009, I define December 1, 2008 to April 15, 2009 as the next period. It has roughly 95,000 observations. The final period is April 16, 2009 to December 31, 2009, which has roughly 212,000 observations.

To account for variation in trading across the trading day, I include dummy variables in each equation indicating the hour.[13] To account for any time trends, I also include a trend variable. As the various spreads do not have a zero mean and because they vary by hour and across days, the cointegrating vector has a non-zero intercept. To account for this, I regress each spread on dummy variables for the hour of the day and the time trend. When including multiple time periods in a specification, I also include dummy variables for the period. If only one period is being used in a specification, I use only data for that period to estimate the intercept of the cointegrating vector. I then utilize the residuals in the VEC model, which are the spreads less the time trend and the hourly and period-specific means.

To determine the number of lags to include in the model (and keep them consistent across time periods), I estimate the model using the entire dataset with dummy variables for each time period, compute the residuals from each equation and utilize the Akaike Information Criteria (AIC), which indicates that 7 lags of each variable are sufficient.[14] With 7 lags of each variable, the constant, the time trend and the dummy variables for each hour, there are 335 variables in each regression (334 in the regressions for quantity sold and bought which do not contain any contemporaneous variables) for a total of 8,034 coefficients to be estimated.

I estimate the model separately for each time period to compute the impulse response functions, which allows for dynamics to be completely different across periods. To test whether the immediate price impact of a trade at HSC differs across hubs in a particular period, I estimate pairs of equations via seemingly unrelated regression (SUR). That is, to find out if a sale at HSC has a greater immediate impact on the best bid at HSC than a sale at Henry has on the best bid at Henry, I estimate the two best bid equations for HSC and Henry jointly via SUR. This retains as many observations as possible, while allowing for computation of the test statistic. To test whether the immediate price

[13] I combine the few observations for 12:00-12:30 pm with those for 11:00 am-12:00 pm.

[14] Note that to compute the likelihood used in the AIC, I lose the 11% of the 605,000 observations where one or more of the price variables is missing.

impact of a trade varies between periods, I generate two versions of every variable for the two time periods to be compared and re-estimate each of the best bid and offer regressions combining the two time periods and including both versions of each variable. This allows for all of the dynamics to be different between time periods, while permitting me to test for differences in the immediate price impact coefficients. Estimates of the immediate price impact coefficients are virtually unchanged in these auxiliary regressions. Estimates of the standard errors when regressing one equation at a time are computed using White's heteroskedasticity robust estimator and are generally higher than those estimated via SUR.

1.4.5 Results

The following figures and tables illustrate the estimated resilience and price impact of sales at HSC, Henry Hub, and Katy for each of six time periods: pre-hurricane (January - August 2008), hurricane (September 2008), October 2008, November 2008, post-hurricane 1 (December 1, 2008 - April 15, 2009), and post-hurricane 2 (April 16, 2009 - December 31, 2009). As the hurricanes created significant uncertainty, any effect BP's manipulative activity had in the first few days post-hurricane is extremely difficult to separately identify. I therefore focus my analysis on whether the price impact in October and November 2008 was lower than would otherwise have been expected.

To determine if a price impact is lower than expected, I utilize two types of comparisons, (1) over time, and (2) across locations. If the hurricanes created a lingering increase in uncertainty that dissipated gradually, one would expect that resilience would increase gradually and price impact would decrease gradually as time passed. I therefore compare the price impacts in October and November 2008 against all periods not affected by the hurricane or alleged manipulation (termed non-hurricane for brevity) and against the post-hurricane 1 period, during which the spread between Henry Hub and HSC remained wider than it was pre-hurricane. The wider spread in the post-hurricane 1 period indicates that the recovery from the hurricanes was not yet complete and therefore provides the most similar comparison group. In comparing across locations, I examine the similarity between resilience and price impact estimates for sales at HSC, sales at Henry Hub, and sales at Katy within the same period. Note that the price impacts presented are the absolute values of the estimated price impacts (which are negative since I examine sales) to facilitate interpretation.

Figure 1.5 displays graphically the estimated resilience at the three locations. While resilience in October 2008 at Henry Hub and Katy was roughly similar to resilience during the hurricane period, resilience at HSC increased dramatically from September 2008 to October 2008. In fact, resilience at HSC in October 2008, was higher on average than during any other period of time, a fact confirmed by the t-test assuming unequal variance displayed in Table 1.2. In the months following October 2008, resilience at Henry Hub rose somewhat as might be expected during the recovery from a shock, while resilience at HSC fell steeply in November 2008, then rose again in the post-hurricane 1 period. Estimates of resilience at Katy are highly correlated with those at HSC, excluding the jump in October 2008, which is expected given the proximity of the two locations. Thus, the October 2008 estimates of resilience support the hypothesis that price impact was lower in October 2008 than might have otherwise been expected, while the November 2008 estimates do not support the hypothesis.

Alternative Hypothesis	**Oct/Nov Avg.**	**Non-Hurricane Avg.**	**T-Stat.**	**P-value**
Oct-08 > Non-Hurr.	55.1%	49.6%	-2.679	0.004
Nov-08 > Non-Hurr.	45.3%	49.6%	1.789	0.963

Table 1.2: T-Tests of Differences in Resilience over Time at HSC

Figure 1.6 displays the time to price recovery given that recovery occurs before the next sale. This figure shows that the median time to price recovery was faster in October 2008 and November 2008 than in the post-hurricane 1 period. A t-test of the average time to recovery, assuming unequal variance, supports the hypothesis at a 90% confidence level that both the October and November 2008 periods had faster price recovery on average than the post-hurricane 1 period (see Table 1.3). The χ^2 test implemented by

Stata to test whether the median of two distributions is the same indicates the median time to recovery was significantly different in October 2008 than in the post-hurricane 1 period, but not in November 2008 (see Table 1.4). Comparing across locations, the median October 2008 recovery time was faster than at Henry Hub, but the November 2008 recovery time was slower than at Henry Hub. These observations are confirmed by hypothesis tests in Tables 1.5 and 1.6. Thus, the October 2008 results are more supportive than the November 2008 results of the hypothesis that price impact was lower than expected during the alleged manipulation.

Alternative Hypothesis	**Oct/Nov Avg.**	**Post-Hurr. 1 Avg.**	**T-Stat.**	**P-value**
Oct-08 $<$ Post-Hurr. 1	114	152	1.502	0.067
Nov-08 $<$ Post-Hurr. 1	112	152	1.319	0.094

Table 1.3: T-Tests of Differences in Mean Time to Recovery at HSC

Alternative Hypothesis	**Oct/Nov Med.**	**Post-Hurr. 1 Med.**	**Chi^2 Stat.**	**P-value**
Oct-08 $\neq$ Post-Hurr. 1	26	40.5	13.64	0.000
Nov-08 $\neq$ Post-Hurr. 1	32	40.5	0.459	0.498

Table 1.4: Tests of Differences in Median Time to Recovery at HSC

Alternative Hypothesis	**HSC Avg.**	**Henry Hub Avg.**	**T-Stat.**	**P-value**
Oct-08, HSC $<$ Henry	114	131	-0.686	0.246
Nov-08, HSC $<$ Henry	112	78	2.070	0.981

Table 1.5: T-Tests of Differences between Mean Time to Recovery at HSC vs. Henry Hub

Alternative Hypothesis	**HSC Med.**	**Henry Hub Med.**	**Chi^2 Stat.**	**P-value**
Oct-08, HSC $\neq$ Henry	26	39	8.847	0.003
Nov-08, HSC $\neq$ Henry	32	15	16.95	0.000

Table 1.6: Tests of Differences between Median Time to Recovery at HSC vs. Henry Hub

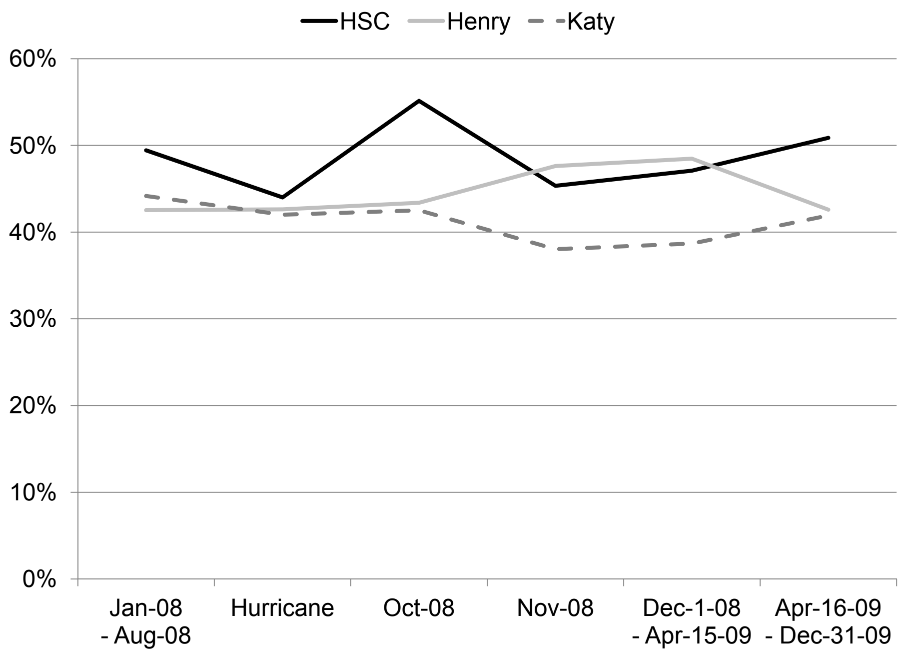


Figure 1.5: Share of Sales where Price Recovers Before the Next Sale

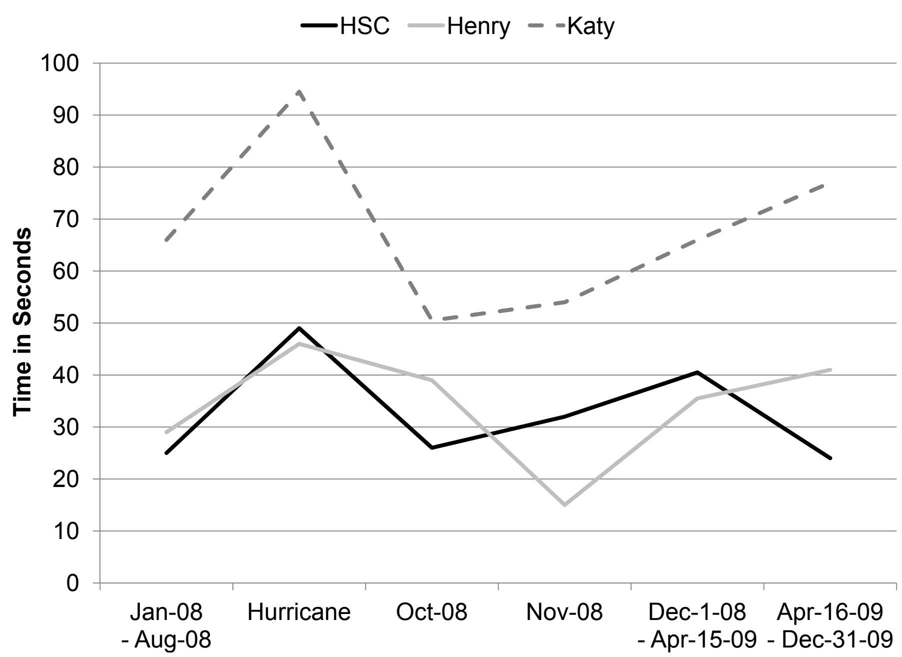


Figure 1.6: Median Time in Minutes to Price Recovery, Given Recovery Occurs

From Figure 1.7, one can see that the immediate price *change* after a sale, that is the change in the best bid from one second before a sale to the end of the second in which a sale occurred, is lower in October 2008 and November 2008 than in the post-hurricane 1 period. Additionally, the average immediate price change in October 2008 appears lower than in any other period in the dataset. Hypothesis tests, displayed in Table 1.7, illustrate that although the point estimates are indeed lower, one cannot reject any of the hypotheses at the 95% confidence level. The null hypotheses that the October 2008 immediate price change is lower on average than in other periods can be rejected at lower confidence levels than the November 2008 hypotheses, but at best it can only be rejected at the 90% confidence level indicating weaker support for the hypothesis of smaller price impacts in the manipulative period. Comparing across locations and excluding the hurricane period, the immediate price change in the bid at HSC due to a sale at HSC appears to always be greater than the immediate price change in the bid at Henry Hub (Katy) due to a sale at Henry Hub (Katy).

Alternative Hypothesis	**Oct/Nov Avg.**	**Non-Hurricane Avg.**	**T-Stat.**	**P-value**
Oct-08 < Non-Hurr.	3.63	4.15	0.882	0.189
Nov-08 < Non-Hurr.	3.79	4.15	0.478	0.317
Oct-08 < Post-Hurr.	3.63	4.39	1.264	0.103
Nov-08 < Post-Hurr.	3.79	4.39	0.791	0.215

Table 1.7: Tests of Differences in Immediate Price Change over Time at HSC

Turning to the results of the VEC model, Figure 1.8 depicts the immediate price impacts of sales at each location on their respective bids. The immediate price impact estimate for HSC sales in October 2008 again appears lower than in any other period, and the November 2008 immediate price impact estimate is again lower than in the post-hurricane 1 period, but higher than in October 2008 and other non-hurricane periods. F-tests, shown in Table 1.8, indicate that the October 2008 results are significant at the 99% confidence level, while the November 2008 immediate price impact estimate is only significantly lower than the post-hurricane 1 period at the 75% confidence level.

Table 1.9 compares the estimated immediate price impact at HSC against the estimated immediate price impacts at Henry Hub and Katy. In the pre and post-hurricane periods, the immediate price impact of a sale at HSC is significantly higher than the immediate price impact of sales at Henry Hub and Katy. However, in October 2008 (but not in November 2008), the immediate price impact of a sale at HSC was significantly lower than the immediate price impacts of sales at the other locations. Thus, comparing across time periods and locations, the October 2008 results are again more supportive than the November 2008 results of the hypothesis that price impacts were lower than their expected level, although the November 2008 results are not entirely inconsistent with the hypothesis.

Alternative Hypothesis	**Oct./Nov. Coef.**	**SE**	**Other Coef.**	**SE**	**F-Stat.**	**P-value**
Oct-08 < Pre-Hurr.	1.66	0.22	2.56	0.17	10.45	0.00
Oct-08 < Post-Hurr. 1			3.37	0.25	26.38	0.00
Oct-08 < Post-Hurr. 2			2.61	0.20	10.26	0.00
Nov-08 < Pre-Hurr.	2.88	0.68	2.56	0.17	0.21	0.68
Nov-08 < Post-Hurr. 1			3.37	0.25	0.45	0.25
Nov-08 < Post-Hurr. 2			2.61	0.20	0.15	0.65

Table 1.8: Tests of Differences in Immediate Price Impact over Time at HSC

Finally, Figures 1.9 and 1.10 illustrate the longer term responses of bids to sales at HSC. Figure 1.9 compares the response of the bid at HSC to a sale at HSC as the time since the sale increases for each of the six time periods. Although in a few of the initial periods the price impact of a sale at HSC is larger for October or November 2008 than in the post-hurricane 1 period, the medium-term price impacts appear lower (and almost identical) in October 2008 and November 2008 than in the post-hurricane 1 period. The permanent price impact of a trade, that is, the price impact after all bids and offers at all three hubs have converged to a stationary level, is displayed in Figure 1.10. This figure indicates that the permanent price impact of a trade at HSC may have been lower in October 2008 than in the post-hurricane 1 period, while the November 2008 price impact is similar to that in the post-hurricane 1 period. Furthermore, in October 2008,

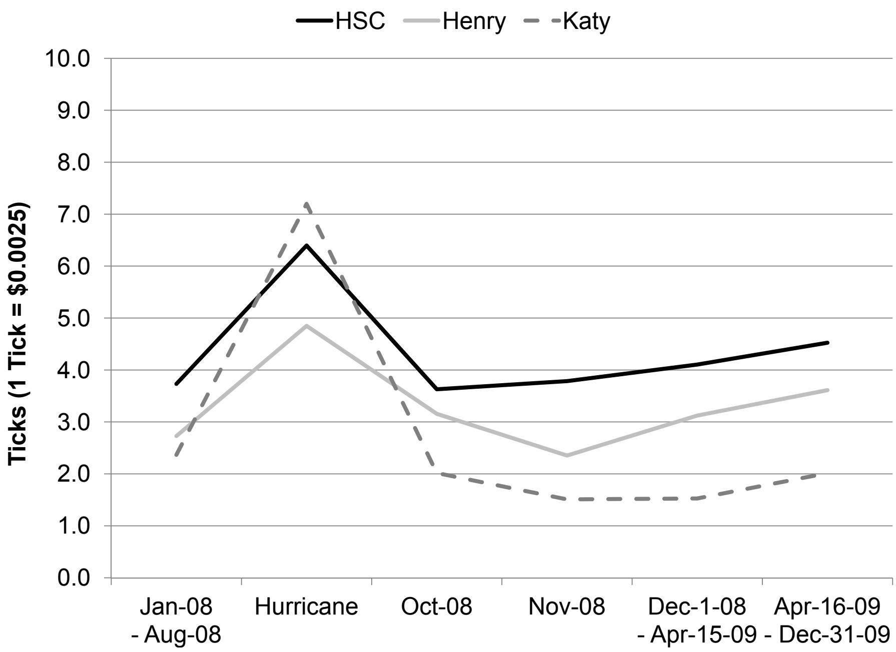


Figure 1.7: Immediate Price Change Post-Sale

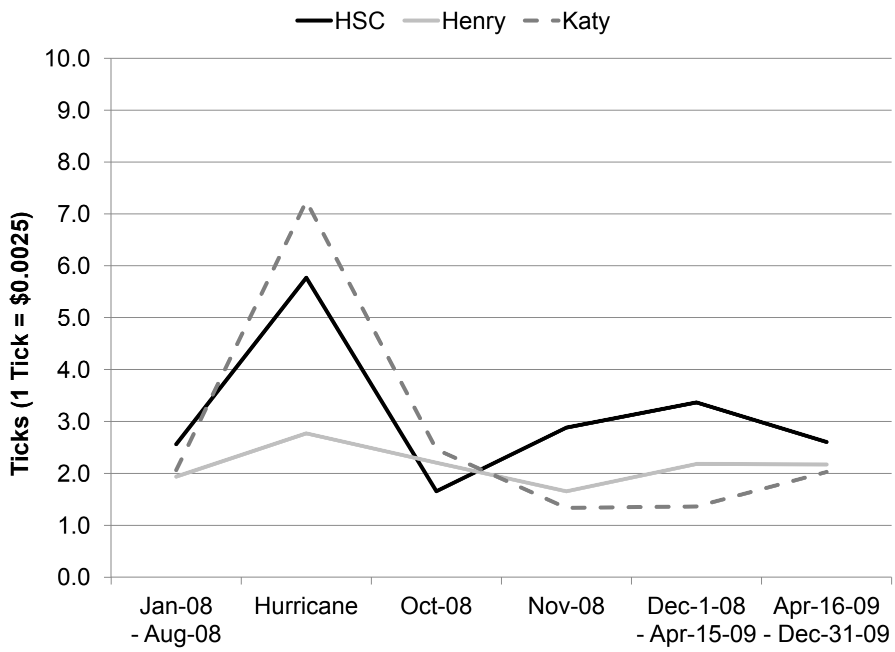


Figure 1.8: Immediate Price Impact

Period	Alternative Hypothesis	HSC Coef.	SE	Other Hub Coef.	SE	Chi^2 Stat.	P-value
Pre-Hurr.	HSC > Henry	2.56	0.05	1.93	0.07	50.05	0.00
	HSC > Katy			2.06	0.11	15.85	0.00
Hurricane	HSC > Henry	5.76	0.35	2.77	0.13	64.05	0.00
	HSC > Katy			7.28	0.66	4.14	0.98
Oct-08	HSC < Henry	1.66	0.13	2.21	0.12	9.71	0.00
	HSC < Katy			2.45	0.18	12.39	0.00
Nov-08	HSC < Henry	2.88	0.16	1.65	0.15	31.49	1.00
	HSC < Katy			1.34	0.15	50.39	1.00
Post-Hurr. 1	HSC > Henry	3.37	0.10	2.18	0.06	109.92	0.00
	HSC > Katy			1.36	0.06	301.50	0.00
Post-Hurr. 2	HSC > Henry	2.60	0.08	2.16	0.06	21.32	0.00
	HSC > Katy			2.02	0.05	39.45	0.00

Table 1.9: Tests of Differences in Immediate Price Impact across Locations

the permanent price impact of a sale at HSC appears to be lower than that of a sale at Henry Hub, although in November 2008, the estimated permanent price impact of a sale at Henry Hub is lower.[15]

The estimates of price impact considered here display a distinct pattern. The price impact of a sale at HSC in October 2008 appears to be lower than in other time periods, especially the post-hurricane 1 period. The price impact of a sale at HSC also appears to be lower than the price impact of a sale at Henry Hub in October 2008, while in other periods, most estimates of the price impact of a sale at HSC are higher than the estimated price impact of a sale at Henry Hub. Thus, the screen for October 2008 seems to indicate that price impact at HSC was lower than would otherwise be expected, which suggests that manipulation may indeed have occurred. In November 2008, the results of the screen are inconclusive. Some estimates indicate that the November 2008 price impacts were weakly lower than in the post-hurricane 1 period, while others indicate they may have been higher. This could be explained if BP's actions had become more predictable and traders adjusted to them in other ways. Thus, to fully understand the effectiveness of this screen at identifying manipulation, it would be useful to apply the

[15]In future work, I hope to confirm these observations with hypothesis tests.

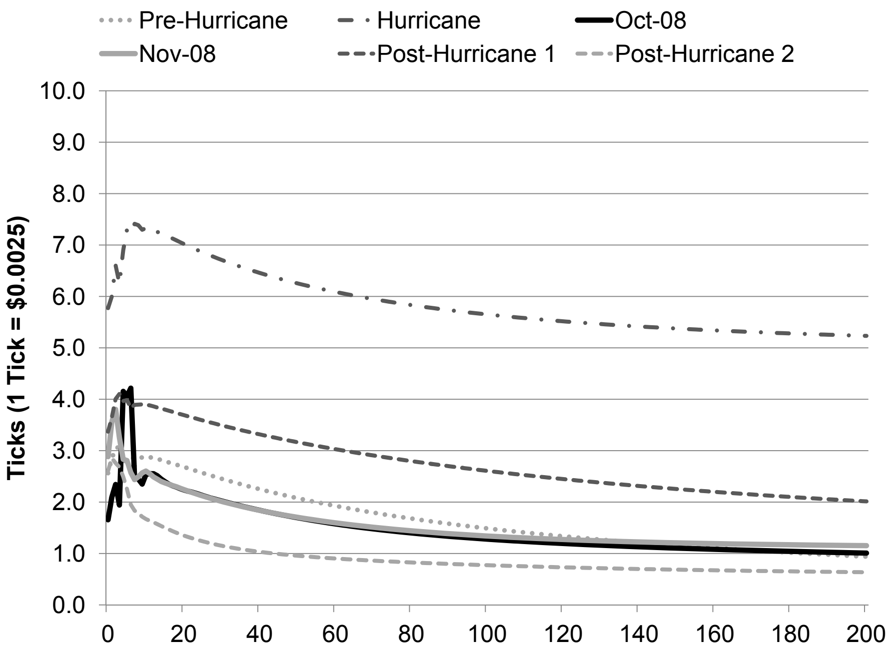


Figure 1.9: Response of Bid at HSC to Sale at HSC

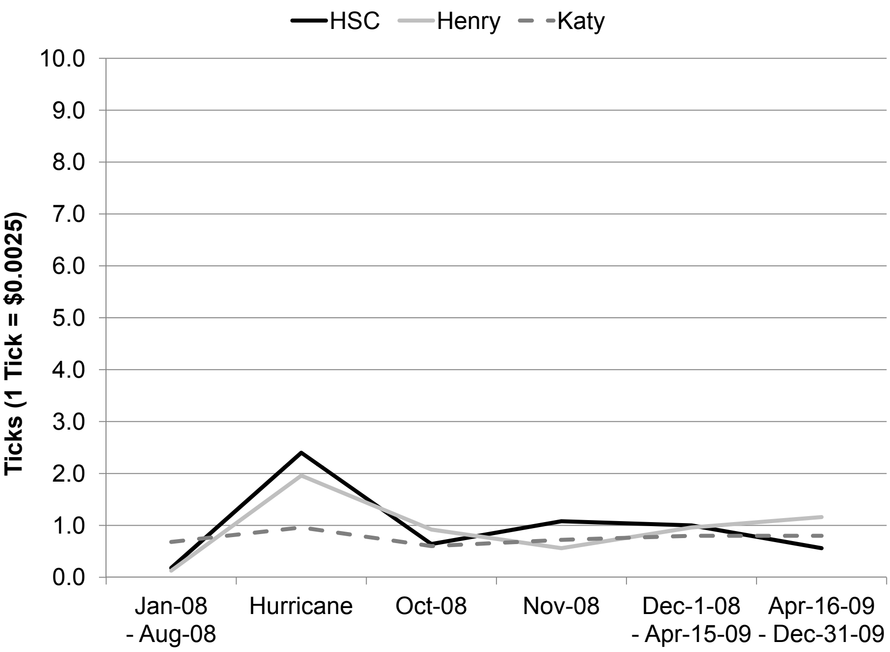


Figure 1.10: Permanent Price Impact

screen to other cases in which manipulation is believed to have occurred.

1.5 Conclusion

This work amends the conditions required for a manipulator to be able to move prices in a spot market to increase the profit on a related futures position. It also suggests two potential screens for manipulability and a screen for attempted manipulation. Previous results indicated that for this type of manipulation to be profitable, the price impact of a trade in the futures market must be significantly lower than the price impact of a trade in the spot market. The results of this paper suggest that there is a second condition: that some traders in the futures market must speculate non-strategically by obtaining cash-settled futures positions and not trading in the spot market. The more non-strategic speculators there are in the futures market, the more profit manipulation can generate and the more likely the spot price will be manipulated. Conversely, if all traders with futures positions are acting strategically or seeking to hedge planned transactions in the spot market, the net effect of their trading will not reduce the informativeness of the spot price, since their trades will cancel out in the spot market, and therefore the spot price will be not be manipulated. This does not have to rule out non-strategic speculation entirely, as traders in real markets who do not wish to take on a physical position can close their futures position before the end of the futures market, which would limit the manipulability of the spot price. These two conditions can be used to screen pairs of spot and futures markets for manipulability.

The model also suggests that, when manipulation is expected to be profitable, the price impact of a trade falls as the expected amount of manipulation rises, which provides a potential screen for manipulation. To test this screen for manipulation, I obtained and analyzed data from the Intercontinental Exchange that contains the bids, offers, and trades from the market and time period in which traders from BP America allegedly manipulated the spot price of natural gas in Texas. In particular, BP America has been accused of manipulating the spread (or price difference) between the spot price of natural gas at Houston Ship Channel and the spot price at Henry Hub. As their alleged manipulation occurred after Hurricanes Gustav and Ike which damaged natural gas production and transportation infrastructure and therefore reduced the correlation between the value of natural gas at Houston Ship Channel and Henry Hub, the expected profit from manipulating the spread likely increased relative to periods unaffected by hurricanes. If other market participants recognized the probability and profitability of manipulation had increased at Houston Ship Channel, the model suggests that the average price impact of a trade would be reduced relative to other time periods and relative to other hubs that were not manipulated.

Estimates of the price impact of a trade from a vector error correction model and estimates of market resilience, which indicate how often and how quickly prices recover after a trade, suggest that in the middle of the allegedly manipulated period, the price impact of a trade was lower than expected and manipulation may have occurred. I exclude the first few days of the alleged manipulation as the increase in uncertainty caused by the hurricanes makes it difficult to separately identify any effect of manipulation. At the end of the investigative period, the measures of price impact and market resilience indicate that the price impact of a trade may have returned to more usual levels. If BP's actions became more predictable in the later period, this may explain the increase in price impact as market participants could have adjusted to their trades in other ways.

In future work, I would like to survey companies that participate in natural gas and other markets about their joint spot and futures market trading to determine the share of hedgers versus non-strategic speculators in the futures market. With a sufficiently complete dataset, firm evidence could be obtained about whether manipulation could actually be effective at moving spot prices and generating positive profit on average. I also hope to test the screen for attempted manipulation on other markets in which manipulation is suspected. Finally, I have begun analyzing several additions to the manipulation strategy considered here to evaluate the robustness of my theoretical results. Preliminary results indicate that if, for instance, a manipulator buys an asset in advance and then sells it later in order to push a particular spot price down, the effect of the attempted manipulation depends on the relative price effect of the initial purchases versus the later sales. If the initial purchases caused a price increase equal in size to the price decrease caused by the later sales, manipulation would be unprofitable.

Appendices

1.A Equilibrium with Non-Strategic Speculators, Hedgers, and a Manipulator

The set-up of the model is the same as in Kumar & Seppi (1992) except that some futures noise trading is generated by hedgers and some by non-strategic speculators. Let e represent the total futures noise trading, h represent the net volume traded by hedgers, and ε represent the net volume traded by non-strategic speculators. h and ε are independent with zero means and variances, σ_h^2 and σ_ε^2. The total variance of the futures noise trading is $\sigma_e^2 = \sigma_h^2 + \sigma_\varepsilon^2$.

Proposition 1.3 (Linear Equilibrium with Non-Strategic Speculators, Hedgers, and a Manipulator) *There is an equilibrium in which (i) the strategy that maximizes the profit of manipulator,*

$$\begin{aligned}
&\max_{\Delta} E_e\Big\{ \max_{z} E_{v,u,h,X}\Big[\Delta\big(S(X+z+u+h, e+\Delta) - F(e+\Delta)\big) \\
&\quad + z\big(v - S(X+z+u+h, e+\Delta)\big)\Big| e\Big]\Big\} \\
&\quad s.t.\ |\Delta| \le |W| \qquad (1.A.1)
\end{aligned}$$

is to randomize his futures order Δ between $|W|$ and $-|W|$ with equal probability and trade

$$z = \tfrac{1}{2}[(2-k_\varepsilon)\Delta + (C_\varepsilon - k_\varepsilon)e] \qquad (1.A.2)$$

$$\text{where } k_\varepsilon = \frac{\sigma_\varepsilon^2}{\sigma_w^2 + \sigma_e^2} \text{ and } C_\varepsilon = \frac{\sigma_\varepsilon^2}{\sigma_e^2}. \qquad (1.A.3)$$

The manipulator's expected profit is:

$$E[\Pi_M|\Delta] = \tfrac{1}{4}\lambda\big(k_\varepsilon^2\Delta^2 - (C_\varepsilon^2 - k_\varepsilon^2)\sigma_e^2\big) + \tfrac{1}{2}\lambda\big(C_\varepsilon - k_\varepsilon\big)\sigma_\varepsilon^2 \qquad (1.A.4)$$

(ii) the strategy that maximizes the profit of the informed trader,

$$\max_X E_{u,h,z}\Big[X\big(v - S(X+z+u+h, e+\Delta)\big)\Big|v, y_f\Big] \tag{1.A.5}$$

is to trade

$$X = \frac{1}{2\lambda}(v-\mu), \tag{1.A.6}$$

(iii) the futures price is $F = \mu$, and (iv) the spot price is set using the rule

$$S = \mu + \lambda(y_s - E[y_s|y_f]) \tag{1.A.7}$$

$$\textit{where } \lambda = \frac{\sigma_v}{2\sqrt{\sigma_u^2 + \sigma^2(z|y_f)}} \tag{1.A.8}$$

$$\sigma^2(z|y_f) = \Big[\tfrac{1}{4}k(2-C_\varepsilon)^2 - k(1-C_\varepsilon)^2 + (1-k)C_\varepsilon(1-C_\varepsilon)\Big]\sigma_e^2 \tag{1.A.9}$$

$$E[y_s|y_f] = E[z+h|y_f] = (1-k_\varepsilon)y_f \tag{1.A.10}$$

To derive the manipulator's strategy, first assume that the formulas determining the prices will be:

$$F = \mu \tag{1.A.11}$$

$$S = \mu + \lambda(y_s - E[y_s|y_f]) \tag{1.A.12}$$

Since the manipulator is uninformed, his expectation of the true value of the asset is μ.

Next derive the strategy of the informed trader in the spot market:

$$\max_X X\Big(v - \mu - \lambda\big(X + E[z+u+h|y_f] - E[X+z+u+h|y_f]\big)\Big)$$
$$= \max_X X\Big(v - \mu - \lambda\big(X - E[X|y_f]\big)\Big)$$

His first order condition is:

$$X = \frac{1}{2\lambda}(v - \mu) + \frac{1}{2}E[X|y_f]$$

Taking the expectation of both sides given y_f yields:

$$E[X|y_f] = \frac{1}{\lambda}E[(v - \mu)|y_f] = 0$$

since y_f reveals nothing about the true value of the asset. Therefore, the informed trader's optimal trade is:

$$X = \frac{1}{2\lambda}(v - \mu)$$

Then simplify the manipulator's profit maximization problem in the spot market, given his futures position, to:

$$\max_z (\Delta - z)\lambda\big(z + E[X + u + h|e] - E[X + z + u + h|y_f]\big)$$

Note that I have used the fact that since the manipulator knows Δ and y_f, he knows e. Since the expected value of the spot noise trading and informed trader's trade is zero, this simplifies to:

$$\max_z (\Delta - z)\lambda\big(z + E[h|e] - E[z + h|y_f]\big)$$

The manipulator's first order condition is:

$$z = \frac{1}{2}\Big[\Delta + E[z|y_f] - (E[h|e] - E[h|y_f])\Big]$$

Taking the expectation of both sides given y_f yields:

$$E[z|y_f] = E[\Delta|y_f]$$

This can be seen by computing $E[h|e]$ and $E[h|y_f]$ assuming as before that the manipulator will randomize between trading $|W|$ and $-|W|$ in the futures market where W is distributed $N(0, \sigma_w^2)$ and therefore so is Δ. Since all the trades in the futures market are based on normally distributed random variables with zero mean:

$$\begin{aligned}
E[h|e] &= \frac{\sigma_h^2}{\sigma_e^2} e = C_h e \\
E[h|y_f] &= \frac{\sigma_h^2}{\sigma_w^2 + \sigma_e^2} y_f = k_h y_f
\end{aligned}$$

Taking the expectation of $E[h|e]$ given y_f yields:

$$\begin{aligned}
E\big[E[h|e]\big|y_f\big] &= \frac{\sigma_h^2}{\sigma_e^2} E[e|y_f] \\
&= \frac{\sigma_h^2}{\sigma_e^2} \frac{\sigma_e^2}{\sigma_w^2 + \sigma_e^2} y_f \\
&= \frac{\sigma_h^2}{\sigma_w^2 + \sigma_e^2} y_f \\
&= E[h|y_f]
\end{aligned}$$

Therefore $E\big[E[h|e] - E[h|y_f]\big|y_f\big] = 0$.

Inserting $E[z|y_f]$ into the manipulator's optimal spot trade yields:

$$z = \frac{1}{2}\Big[\Delta + E[\Delta|y_f] - (E[h|e] - E[h|y_f])\Big]$$

At this point is useful to note a few more coefficients:

$$\begin{aligned}
E[\varepsilon|e] &= \frac{\sigma_\varepsilon^2}{\sigma_e^2}e = C_\varepsilon e \\
e &= E[\varepsilon|e] + E[h|e] = (C_\varepsilon + C_h)e \\
E[\varepsilon|y_f] &= \frac{\sigma_\varepsilon^2}{\sigma_w^2 + \sigma_e^2}y_f = k_\varepsilon y_f \\
E[\Delta|y_f] &= \frac{\sigma_w^2}{\sigma_w^2 + \sigma_e^2}y_f = k_\Delta y_f \\
y_f &= E[\Delta|y_f] + E[h|y_f] + E[\varepsilon|y_f] = (k_\Delta + k_h + k_\varepsilon)y_f
\end{aligned}$$

Armed with the additional coefficients, z can be rewritten

$$\begin{aligned}
z &= \frac{1}{2}\Big[\Delta + k_\Delta y_f - (C_h e - k_h y_f)\Big] \\
&= \frac{1}{2}\Big[\Delta + (k_\Delta + k_h)y_f - C_h e\Big] \\
&= \frac{1}{2}\Big[\Delta + (1 - k_\varepsilon)(\Delta + e) - C_h e\Big] \\
&= \frac{1}{2}\Big[(2 - k_\varepsilon)\Delta + (C_\varepsilon - C_h)e\Big]
\end{aligned}$$

Plugging the manipulator's optimal spot trade into his profit function, simplifying, and taking the expectation with respect to Δ yields

$$E[\Pi|\Delta] = E\Big[-\tfrac{1}{2}\lambda\big((C_\varepsilon - k_\varepsilon)e - k_\varepsilon\Delta\big)\Big(\tfrac{1}{2}\big((C_\varepsilon + k_\varepsilon)e + k_\varepsilon\Delta\big) + X + u - \varepsilon\Big)\Big|\Delta\Big]$$

Remembering that h, ε, and Δ are independent with mean zero, his expected profit given Δ simplifies to:

$$E[\Pi|\Delta] = \tfrac{1}{4}\lambda\big(k_\varepsilon^2\Delta^2 - (C_\varepsilon^2 - k_\varepsilon^2)\sigma_e^2\big) + \tfrac{1}{2}\lambda\big(C_\varepsilon - k_\varepsilon\big)\sigma_\varepsilon^2 \qquad (1.A.13)$$

Since $\frac{1}{4}\lambda k_\varepsilon^2$ is always positive, the manipulator's expected profit is always increasing in the absolute size of his futures trade, and therefore the manipulator would want to make an infinitely-sized trade if he was not wealth-constrained. Since his expected profit is a function of Δ^2, he is indifferent between buying and selling in the futures market and so a strategy of buying and selling with equal probability is feasible in equilibrium.

To solve for the price impact coefficient, $\lambda = \frac{Cov(v,y_s|y_f)}{Var(y_s|y_f)}$, one must solve for $Cov(v, y_s|y_f)$ and $Var(y_s|y_f)$.

$Cov(v, y_s|y_f)$ is:

$$\begin{aligned} Cov(v, y_s|y_f) &= E[(y_s - E[y_s|y_f])(v - \mu)|y_f] \\ &= E[(X + u + h + z - E[h + z|y_f])(v - \mu)|y_f] \\ &= E[\frac{1}{2\lambda}(v - \mu)^2|y_f] \\ &= \frac{1}{2\lambda}\sigma_v^2 \end{aligned}$$

The second line follows from the first because the expected values of the spot noise traders' order and the informed trader's order give y_f is zero. The third line follows because the informed trader is the only trader whose trade is correlated with v.

And $Var(y_s|y_f)$ is:

$$\begin{aligned} Var(y_s|y_f) &= E[(y_s - E[y_s|y_f])^2|y_f] \\ &= E[(X + u + h + z - E[h + z|y_f])^2|y_f] \\ &= Var(X|y_f) + Var(u|y_f) + Var(z + h|y_f) \\ &= \frac{1}{4\lambda^2}\sigma_v^2 + \sigma_u^2 + Var(z + h|y_f) \end{aligned}$$

where

$$Var(z + h|y_f) = Var(z|y_f) + Var(h|y_f) + Cov(z, h|y_f)$$

$$\begin{aligned}
Var(z|y_f) &= E[(\Delta - E[\Delta|y_f] + \tfrac{1}{2}(C_\varepsilon e - k_\varepsilon y_f))^2|y_f] \\
&= E[(\Delta - E[\Delta|y_f] + \tfrac{1}{2}C_\varepsilon(e - (1-k)y_f))^2|y_f] \\
&= E[(\Delta - E[\Delta|y_f] + \tfrac{1}{2}C_\varepsilon(e - E[e|y_f]))^2|y_f] \\
&= Var(\Delta|y_f) + \tfrac{1}{4}C_\varepsilon^2 Var(e|y_f) + C_\varepsilon Cov(\Delta, e|y_f) \\
&= Var(\Delta|y_f) + \tfrac{1}{4}C_\varepsilon^2 Var(e|y_f) - C_\varepsilon Var(\Delta|y_f) \\
&= \tfrac{1}{4}k\sigma_e^2(2 - C_\varepsilon)^2
\end{aligned}$$

To go from the second to last line to the last line, I use the formula for determining the variance of a normal random variable (x) given another normal random variable (y): $(1-\rho_{x,y}^2)\sigma_x^2$. Here $\rho_{x,y}^2 = k_x$. I also use the fact that $k\sigma_e^2 = (1-k)\sigma_w^2 = \frac{\sigma_e^2\sigma_w^2}{\sigma_e^2+\sigma_w^2}$.

Next:

$$Var(h|y_f) = (1-k_h)\sigma_h^2 = (1-k_h)C_h\sigma_e^2$$

Then, $Cov(z, h|y_f)$ can be determined by writing the four random variables, $(h, \Delta, \varepsilon, y_f)$ as a vector which is distributed:

$$N\left(0, \begin{bmatrix} \sigma_h^2 & 0 & 0 & \sigma_h^2 \\ 0 & \sigma_w^2 & 0 & \sigma_w^2 \\ 0 & 0 & \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \\ \sigma_h^2 & \sigma_w^2 & \sigma_\varepsilon^2 & \sigma_h^2 + \sigma_w^2 + \sigma_\varepsilon^2 \end{bmatrix}\right)$$

Partitioning $(h, \Delta, \varepsilon, y_f)$ into $f = (h, \Delta, \varepsilon)$ and y_f, $Var(f|y_f) = \Sigma_{f,f} - \Sigma_{f,y_f}\Sigma_{y_f,y_f}^{-1}\Sigma_{y_f,f}$:

$$Var(f|y_f) = \begin{bmatrix} \sigma_h^2(1-k_h) & -\sigma_h^2 k & -\sigma_h^2 k_\varepsilon \\ -\sigma_h^2 k & \sigma_w^2(1-k) & -\sigma_w^2 k_\varepsilon \\ -\sigma_h^2 k_\varepsilon & -\sigma_w^2 k_\varepsilon & \sigma_\varepsilon^2(1-k_\varepsilon) \end{bmatrix}$$

Therefore,

$$
\begin{aligned}
Cov(z,h|y_f) &= E[(\Delta - E[\Delta|y_f] + \tfrac{1}{2}C_\varepsilon(e-(1-k)y_f))(h - E[h|y_f])|y_f] \\
&= Cov(\Delta, h|y_f]) + \tfrac{1}{2}C_\varepsilon Cov(e,h|y_f) \\
&= Cov(\Delta, h|y_f]) + \tfrac{1}{2}C_\varepsilon(Var(h|y_f) + Cov(\varepsilon, h|y_f)) \\
&= -k\sigma_h^2 + \tfrac{1}{2}C_\varepsilon k\sigma_h^2 \\
&= (\tfrac{1}{2}C_\varepsilon - 1)kC_h\sigma_e^2
\end{aligned}
$$

To go from the third line to the fourth, I used the fact that $1 - k_h - k_\varepsilon = k$.

And finally,

$$
\begin{aligned}
Var(z+h|y_f) &= Var(z|y_f) + Var(h|y_f) + Cov(z,h|y_f) \\
&= \left[\tfrac{1}{4}k(2-C_\varepsilon)^2 - k(1-C_\varepsilon)^2 + (1-k)C_\varepsilon(1-C_\varepsilon)\right]\sigma_e^2
\end{aligned}
$$

Note that the variance of the manipulator and hedgers' trades does not depend on the price impact coefficient λ.

Solving the first line below for λ, yields the equilibrium price impact coefficient shown in the second line:

$$
\begin{aligned}
\lambda &= \frac{\frac{1}{2\lambda}\sigma_v^2}{\frac{1}{4\lambda^2}\sigma_v^2 + \sigma_u^2 + Var(z+h|y_f)} \\
&= \frac{\sigma_v}{2\sqrt{\sigma_u^2 + Var(z+h|y_f)}}
\end{aligned}
$$

1.B Moving between the VAR and VEC Representations

In this appendix, I demonstrate how the VEC coefficients in equation (1.30) can be converted back to the VAR coefficients in equation (1.29). Decompose the vector $\tilde{y}_t$ into $(\Delta p_t; x_t)$, where Δp_t represents the vector of differenced bids and offers. The vector of spreads, s_{t-1} in equation (1.30), can be decomposed into Mp_{t-1} where p_{t-1} represents the vector of bids and offers (in levels) in $t-1$. M is the matrix that when multiplied by

p_{t-1} computes the spreads. Decompose each matrix $\tilde{C}_s$ into $[\tilde{F}_s\ D_s]$ where $\tilde{F}_s$ is the 12x6 matrix of coefficients that multiplies Δp_{t-s} and D_s is the 12x6 matrix of coefficients that multiplies x_{t-s}. In equation (1.29), decompose each matrix C_s into $[F_s\ D_s]$ where F_s is the matrix of coefficients that multiplies p_{t-s}. The matrix D_s is the same in equations (1.29) and (1.30). The following equations illustrate how to convert $\tilde{F}_s$ and Es_{t-1} back to F_s:

$$C_1 = \tilde{C}_1 + EM + [I_6; 0_6] \tag{1.B.1}$$

$$C_r = \tilde{C}_r - \tilde{C}_{r-1} \text{ for } r \in 2, \ldots, p-1 \tag{1.B.2}$$

$$C_p = -\tilde{C}_{p-1} \tag{1.B.3}$$

CHAPTER 2

When a National Cap-and-Trade Policy with Carve-out Provision May Be Preferable to a National CO2 Tax

with Deepak Rajagopal

2.1 Introduction

Greenhouse gas (GHG) emissions are widely regarded as a textbook case of a global externality warranting coordinated global action (Oates, 2001). However, what appears to be emerging from international negotiations is a weaker agreement whereby countries set their own targets for emission reduction (Diringer, 2013). One impediment to a stronger global commitment is the lack of national consensus within some large industrialized countries including the United States and Canada (Rabe et al., 2005; Bulkeley, 2010). In such countries (and elsewhere too), lower levels of government are undertaking various measures to reduce GHG emissions (Rabe, 2008). The range of policy measures includes carbon dioxide (CO_2) taxes (e.g., the province of British Columbia in Canada and the city of Boulder, Colorado in the U.S.), tradable emission permits, henceforth referred to as cap-and-trade, (e.g., the state of California and the Regional Greenhouse Gas Initiative by states in the north-eastern U.S.), emission intensity standards (e.g., the province of Alberta in Canada and the state of California in the U.S.) and renewable energy policies (e.g., state-level Renewable Portfolio Standards, feed-in-tariffs, and various forms of subsidies).

While economic theory suggests that emission pricing, either directly through a CO_2 tax or indirectly through a cap-and-trade program, is the cost-effective approach, renewable energy policies appear the more popular approach for state-level action. Justifications for renewable energy include the local economic benefits of "home-grown" energy resources for long-term economic development and the benefits of reducing (or even simply aiming) to reduce GHG emissions (Rabe, 2008). Bushnell et al. (2008) argue that in a market comprised of many states which are not subject to a unified climate policy and which do not have state-level CO_2 reduction programs, if one state decides to reduce its own emissions, then this goal may be achieved by simply reshuffling pollution within the market such that the state with the policy consumes "cleaner" products while the rest of the market consumes the "dirtier" products. For instance, electricity is susceptible to reshuffling because wholesale purchases of electricity are financial arrangements which are not tied to the physical exchange of electrons. Thus, if "clean" products already have a significant market share, the policy can be satisfied with no change in production or emissions. Indeed, in many electricity markets in the U.S., sizeable zero-carbon electricity generating capacity in the form of nuclear and hydroelectric power exists which may prevent policies targeting CO_2 emissions from being effective in many states.

The goal of this paper is to formally model the interaction of policies at multiple levels of jurisdiction, specifically at the federal and state level, in order to identify the effect on pollution and the relative costs and benefits of CO_2 taxes vis-a-vis cap-and-trade at the federal level when combined with overlapping state-level climate policies (specifically, CO_2 taxes[1] or renewable portfolio standards (RPS)).[2] This research is motivated by the premise that in countries where national opinion on climate change is divided, in the near to medium-term, any national agreement, should it be achieved, would likely be viewed by some states as insufficiently stringent and such states would likely pursue overlapping state-level policies. While an emission tax and a cap-and-trade program are *ex ante*

[1] At the level of lower jurisdiction, CO_2 taxes and cap-and-trade are equivalent but we will show that these two policies exhibit some differences at the level of the higher jurisdiction

[2] For a detailed discussion of the motivation for state-level policies for addressing climate change we refer to Rabe (2008).

equivalent (Jaffe et al., 2003), we show that when states enact additional emission control policies, the two national policies could yield different results. In any case, the cost of a given reduction in national emissions is always lower under a unified national emissions policy than under differentiated and/or overlapping policies by multiple jurisdictions.

Several authors have analyzed the effect of combining state and federal emissions reduction policies (Bushnell et al., 2008; McGuinness and Ellerman, 2008; Burtraw and Shobe, 2009; Goulder and Stavins, 2011a,b; Williams, 2012). One common conclusion in these studies is that under a national cap-and-trade regime, additional state policies have little to no effect on national emissions as any additional emission reduction at the state or local-level, beyond that which would have resulted under the national policy alone, only allows emissions from the rest of the nation to rise back to the level of the national cap. However, by developing innovative policies and infrastructure, state and local regulators could help lower the cost of achieving national emission goals (Burtraw and Shobe, 2009). Another set of papers analyzes the effect of renewable energy policies operating under the European Union (EU) Emissions Trading System (ETS). See Fischer and Preonas (2010) for a summary of this literature. These articles conclude that overlapping national renewable energy policies raise the cost of national cap-and-trade policies without affecting national emissions and may benefit the dirtiest fuels.

The offsetting increase in consumption outside the state under a national cap could, however, be avoided by either "carving-out", i.e. exempting states from the national policy provided they set a stricter state policy, or through price-based regulations, e.g. a CO_2 tax (Goulder and Stavins, 2011a).[3] Contrary to Goulder and Stavins (2011a), we show that a price-based regulation, specifically a CO_2 tax, does not necessarily prevent a completely offsetting increase in emissions elsewhere when states adopt an additional CO_2 tax on top of the national CO_2 tax. Consequently, we also show that, for small states (relative to their market, see Section 2.3 for definition) that are subject to a national CO_2 tax, a state-level renewable energy policy is able to further reduce national

[3]As noted by Goulder and Stavins (2011a), there is ample precedent for carve-out provisions in the context of fuel economy and emissions standards. However, we are not aware of any carve-out provisions associated with cap-and-trade policies.

emissions while a state-level emissions policy cannot. However, if a carve-out provision is added to a national cap-and-trade program, allowing states to exempt themselves from the national policy provided they set a tighter cap, and a state decides to set a tighter cap, emissions must decline regardless of the size of the state as the sum of permitted national and state emissions is now lower. Furthermore, because any reshuffling[4] or leakage[5] of emissions within the market caused by a tighter state cap would increase the national emissions permit price (in order to keep emissions outside the state constant), a cap-and-trade policy with a carve-out provision limits reshuffling and leakage within the market and reduces the cost of achieving a given reduction in emissions with a state policy relative to the cost under a national CO_2 tax coupled with an additional state CO_2 tax. However, a tighter state cap under a national cap-and-trade policy with a carve-out raises electricity costs for consumers outside the market relative to the costs before the tighter state cap was implemented and relative to those under equivalent national and state CO_2 taxes, which may impede support for state carve-outs from the national regime.

Our findings result from the following key features of our model: (i) the commodity (or commodities) under consideration can be produced with inputs (say, energy) from different sources or using different technologies resulting in different emissions per unit of output and at least one such input or process results in a zero emission product. In our example, the commodity is electricity derived from coal, natural gas, nuclear, hydro and renewable resources, the latter three being considered zero emission resources; (ii) the commodity is traded at negligible transportation cost within a specified geographic region that spans multiple policy jurisdictions. In our example, it refers to the free flow of electricity within a regional interconnected grid; (iii) under any state-level climate policy, retailers are accountable for emissions *attributable* to final in-state sales regardless of where emissions actually arise in the supply chain, which may be outside the policy

[4] Defined here as the reallocation of existing emissions across jurisdictions.

[5] Defined here as an increase in emissions outside the state caused by an increase in consumption of carbon-intensive resources outside the state due to the reduction in demand for carbon-intensive resources within the state.

jurisdiction. In our example, this implies that even though electricity consumption is emission-free, regulated state retailers are accountable for CO_2 emitted during generation of the electricity imported into the state.

Relaxing the above assumptions affects our findings as follows. Without a zero-carbon resource, pure reshuffling of output would not be sufficient to avoid the state CO_2 tax and thus a state CO_2 tax would be effective even for small states. As the cost of reshuffling increases, the ability of state-level policies to affect national emissions increases for any given state size. Thus the higher the transportation costs (or any other costs associated with shuffling the distribution of the final good), the more effective state-level policies will be at reducing national emissions. The implications of state policies directed at targets other than the emissions attributable to final in-state sales, say, extraction of primary fossils fuels, are discussed in Section 3.5.

While our mathematical and numerical illustrations are for a single commodity, specifically electricity, the simplified model allows for more general conclusions about emission policies spanning multiple economic activities. As the scope of the policy at either the state-level, the national-level or both widens to include emissions from multiple sectors, so does the scope for reshuffling and leakage, causing the efficacy of state-level emissions policies to depend on how the size of the state changes relative to the broader market(s) across which resources can be reshuffled. The comparisons of the various state and national policy combinations are, however, unaffected. Given the global effects of CO_2 emissions, our results also speak to the interactions that occur when global policies overlap national policies or state policies overlap local policies and product markets are larger than the smaller jurisdiction. Finally, although we focus on only three policies - CO_2 taxes, cap-and-trade programs, and RPS, our framework can be extended to consider many other policies such as emission intensity standards, subsidies for renewable energy and border adjustment policies.

2.2 Model

To demonstrate how national and state climate policies might interact, we construct a model comprised of three regions: the nation, a regional market embedded within the nation, and a state comprising a portion of the regional market. A market should be interpreted as an integrated wholesale market in which electricity can flow freely within the market and in which a centralized body clears wholesale transactions and manages power flows. A national policy applies to all regions unless it has a carve-out provision and the state has enacted a sufficiently stringent state-level policy, in which case the state follows its own policy while the rest of the nation is subject to the national policy. We later discuss how policies in the rest of the market (outside the state) might affect our results.

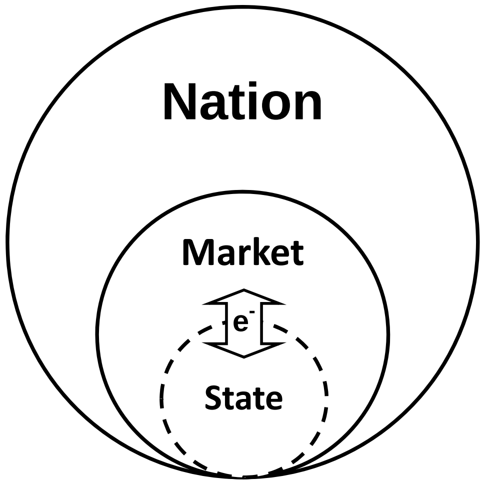


Figure 2.1: The Model

We analyze the interaction of different state and national policy regimes in a static partial equilibrium framework assuming perfect competition. As illustrated in Figure 2.1, power can flow freely between the state and market, but does not flow into or out of the market. We assume there are four types of fuels available to generate electricity: coal, natural gas, qualifying renewable, and non-qualifying zero-carbon. Qualifying renewable fuels represent those that would qualify as renewable under existing RPS policies. Non-qualifying zero-carbon fuels represent nuclear and large hydroelectric facilities, which do not generally qualify as renewable under current state RPS policies. Given the significant environmental and regulatory hurdles to building new nuclear or large hydro generation capacity in addition to their high capital cost (CBO, 2011), we assume that the capacity of non-qualifying zero-carbon resources is fixed. There is one firm operating each generation technology which converts the input (fuel) into output (electricity) and emissions in fixed proportions.

Within the market, the firm may sell power either to the state or to the rest of the market at the electricity price in that region but may be required to pay an explicit or implicit tax or may receive an implicit subsidy based upon the policy (policies) in place in the region(s) to which it sells (not based upon the location of the producer). Electricity sold outside the market is produced separately from the in-market electricity and is therefore subject to a separate supply curve.

2.2.1 Mathematical Formulation of the Model

Let p denote price, q denote quantity of electricity, and the subscripts c, g, r, and z denote coal, gas, qualifying renewables, and non-qualifying zero-carbon resources respectively. A representative consumer in each region, R, maximizes a quasi–linear utility function, $u^R(q^R) - p^R q^R$.[6] $u^R(\cdot)$ is the consumer's utility of consuming electricity in region R. A state comprises a fraction ρ of the market and therefore consumes ρ of the electricity pre-policy, which will be reflected in the preferences of the representative consumers.

Each producer seeks to maximize profit, which is defined as the sum of the revenue

[6] With this utility function, we implicitly assume that the cost of emissions to a consumer is additively separable from the consumer's utility of electricity and money.

from electricity sold in each region less the cost of generating the electricity:

$$\max_{q_f^s, q_f^m, q_f^n} \quad (p^s + x_f^s + x_f^N)q_f^s + (p^m + x_f^N)q_f^m - c_f^M(q_f^s + q_f^m) + (p^n + x_f^N)q_f^n - c_f^n(q_f^n) \tag{2.1}$$

$$\text{s.t. } q_f^s \geq 0,\ q_f^m \geq 0,\ q_f^n \geq 0 \tag{2.2}$$

f indicates the fuel utilized by the producer. s indicates the state, m indicates the rest of the market, M indicates the market as a whole, n indicates the rest of the nation (outside the market), and N indicates the nation as a whole. Market clearing conditions ensure that the sum of production from each fuel for each region equals total consumption in each region ($q_c^R + q_g^r + q_r^R + q_z^R = q^R$).

$p^s + x_f^s + x_f^N$ is the price received by the producer in the state for the electricity sold, q_f^s. p^s represents the price of electricity in the state, x_f^s and x_f^N represent the explicit or implicit tax or subsidy per MWh of electricity produced by each fuel f from the state policy and national policy respectively. The values of x_f^s and x_f^N under each policy are discussed in Section 2.2.2. Similarly, $p^m + x_f^N$ is the price received by the producer in the rest of the market and $p^n + x_f^N$ is the price received by the producer in the rest of the nation.

$c_f^M(\cdot)$ and $c_f^n(\cdot)$ represent the cost of generating electricity from each fuel, f, in the market and the rest of the nation respectively. The assumption of separate cost curves for the market and the rest of the nation ensures that supply to one market does not affect supply in another market.[7] Non-qualifying zero-carbon generation is assumed to have zero marginal cost but to face a capacity constraint such that $q_z^s + q_z^m \leq Q_z^M$ where Q_z^M is the total existing capacity of non-qualifying zero-carbon generation in the market.

[7] Even when technical or economic factors limit the exchange of a commodity (say electricity or biofuel) to a specified geographic region, the intermediate inputs used to produce the commodity (such as coal or crops) need not face such constraints, in which case, the cost function, and therefore the supply, in one market will be related to that in the other. For instance, a reduction in demand for the input in one market, increases the supply of inputs to the other market. This is yet another mechanism of leakage, one that manifests via input markets as opposed to the output market which is the focus of our illustration and does not weaken our findings.

In the rest of the nation, non-qualifying zero-carbon generation is fixed at Q_z^n.

2.2.2 Mathematical Formulation of Each Policy

2.2.2.1 Renewable Portfolio Standard

An RPS dictates that qualifying renewable generation must be a specified share of total generation, α. The RPS requirement is represented by:

$$\frac{q_r}{q_c + q_g + q_z + q_r} \geq \alpha \quad \text{or} \quad q_r \geq A(q_c + q_g + q_z) \text{ where } A = \frac{\alpha}{1-\alpha} \tag{2.3}$$

Under an RPS, suppliers of electricity must demonstrate that at least α percent of the electricity sold to end-users was generated from qualifying renewable resources. To do this they can either generate electricity using renewable resources or purchase renewable energy credits (RECs) at price γ from other producers generating electricity using renewables. Thus, a supplier of electricity from renewable resources receives an implicit subsidy γ, while a supplier of electricity from conventional resources pays an implicit tax $A\gamma$ for electricity sold in region(s) with an RPS.

2.2.2.2 Tradable pollution permits or Cap-and-Trade

A cap-and-trade program specifies a maximum level of CO_2 emissions by giving away or auctioning a number of permits equal to the cap. The cap-and-trade requirement is represented by:

$$e_c q_c + e_g q_g \leq \bar{E} \tag{2.4}$$

We assume that each producer must purchase one carbon credit per of emissions at price τ for electricity sold in the region(s) with cap(s). The price is set by competition for the limited supply of credits.[8] e_c and e_g are the tonnes of CO_2 emissions per MWh

[8]Whether credits are distributed for free or auctioned, the price of a carbon credit will end up the

of electricity generated by coal and natural gas.

2.2.2.3 CO_2 Tax

With a CO_2 tax, the regulator selects a tax of $\$T$/tonne of CO_2 emissions to achieve a given emissions reduction. Producers using coal to generate electricity, which emits e_c tonnes of CO_2 per MWh, pay e_cT \$/MWh of generation, and producers using natural gas pay e_gT \$/MWh for electricity sold in region(s) with a CO_2 tax.

2.2.2.4 Carve-Out Provision

To implement the carve-out provision, we assume that a state's pre-state-policy share of emissions is equal to its pre-state-policy share of consumption. For example, prior to the state policy, the national cap on emissions is 1,000 tonnes. If the market emits 100 tonnes and the state is 25 percent of the market, then the state is assumed to emit 25 tonnes of CO_2 pre-state-policy and the national cap with the state carved out of the policy would be set to 975 tonnes and the state's cap must be less than or equal to 25 tonnes.

2.2.3 A classification of states based on their market share

A state's ability to affect national emissions using a state policy is determined largely by its size, measured in terms of its share of consumption (or emissions). We classify a state as small if its total consumption is less than the quantity of zero-carbon resources in the market. To understand the typical size of states relative to their markets in the U.S., we examine data on electricity consumption and fuel mix for each state relative to its relevant wholesale market.

In the U.S. there are seven wholesale electricity markets: Independent System Operator New England (ISO-NE), the New York ISO, the Pennsylvania Jersey Maryland (PJM) Interconnection, covering much of the Mid-Atlantic and Midwest, the Midwest

same though firms' profits will differ based upon the number of credits granted for free.

ISO (MISO), the Electric Reliability Council of Texas (ERCOT), the Southwest Power Pool (SPP) (as of January 2014) and the California ISO. As our results pertain mainly to states participating in a wholesale market with other states, we exclude those states that do not participate in a wholesale market as well as Texas, for which the bulk of the electricity grid is isolated from the rest of the U.S.[9] We also exclude California and New York as these states operate their own wholesale markets but trade extensively with neighboring regions, making it difficult to acquire the statistics needed for our analysis. After these exclusions, there are 34 states in the U.S. in which at least some utilities participate in a larger wholesale market (ISO-NE, PJM, MISO, or SPP).

We obtained data on state-level electricity consumption from the EIA's Electric Sales, Revenue, and Average Price Report for 2011[10] while data for market-level fuel mix was obtained from the website for each wholesale electricity market. The data indicate that of the 34 states that participate in a larger wholesale market, only eight could be considered large. Of these eight, four are located in MISO, which had only 12.8 percent of generation from zero-carbon resources in 2011. Three are located in SPP which receives only 14 percent of its power from zero-carbon resources in 2012.[11] The remaining large state in 2011 was Massachusetts, which made up 50.5 percent of ISO-NE in terms of consumption, while ISO-NE had 41.3 percent of generation from zero-carbon resources. In PJM, the largest state by consumption is Virginia with 25.3 percent of consumption, but generation from zero-carbon resources in 2011 was 35 percent. Therefore no state in PJM could be considered large in 2011. Consequently, the majority of states that participate in larger markets should be considered small.[12]

[9] For states that do not trade with other states, a national cap-and-trade with a carve-out provision and a national CO_2 tax will lead to the same outcome if states set more stringent emissions policies as will be discussed below.

[10] See `http://www.eia.gov/electricity/data.cfm#sales` for state-level data

[11] 2012 is the only year for which data was available.

[12] This analysis excludes each market's import capability, but if it were included in the analysis, the share of zero-carbon resources deliverable into the market would increase while the relative size of the state would diminish, rendering states even more likely to be small by our definition.

2.3 Results and Discussion

2.3.1 National CO_2 Tax or No National Policy

2.3.1.1 Small States

When there is either a national CO_2 tax or no national climate policy and other states in the market do not have climate policies, a state-level CO_2 tax will be unable to reduce national emissions if the state is small (i.e., the state's pre-policy consumption is less than the quantity of zero-carbon generation already present in the market).[13] This occurs because when the state is small, there are sufficient zero-carbon resources already existing in the market for state consumers to trade all generation from fossil fuels for generation from zero-carbon fuels with no net change in production, emissions, or tax burden. This reallocation of existing production across consumers is what we henceforth refer to as reshuffling.

To illustrate this result more clearly, suppose there are two fuels, zero-carbon and coal. Pre-state-policy all producers using coal pay a national CO_2 tax of $\$T^n$/tonne of CO_2 while producers using zero-carbon resources pay no CO_2 taxes. After the state adds a CO_2 tax of $\$T^s$/tonne of CO_2 on emissions from generation sold in-state, suppose total production does not change but producers sell only generation from zero-carbon resources to in-state consumers to avoid paying the state tax. Since we have assumed that the state is small, there are sufficient zero-carbon resources to serve all in-state demand and so consumption in-state need not change. Any excess zero-carbon generation and all coal generation is sold to out-of-state customers. Since total production did not change, total production less in-state zero-carbon generation is equal to the rest of the market's pre-state policy consumption, meaning that the rest of the market's consumption does not change post-state-policy. Finally, since all coal generation is sold in the rest of the market, generators using coal still pay $\$T^n$/tonne of CO_2. With production, consumption, and

[13] We assume that in formulating their emissions policies, states are primarily concerned with their effect on national emissions as off-setting increases out-of-state from a global pollutant reduce the benefit of the state policy.

total tax paid unchanged, the price of electricity will not change either and so there is no reason to change production or consumption. Note that this result requires only that trade between states is possible, that reshuffling costs are negligible, and that zero-carbon resources already exist and are available to shuffle. As the cost of reshuffling increases, the ability of state-level policies to affect national emissions increases for any given relative state size.

When some of the pre-existing zero-carbon fuels would qualify as renewables under a state RPS while others would not, a small state may be able to achieve emissions reduction under a state RPS even though a state CO_2 tax would be ineffective.[14] This is because an RPS policy requires a more specific type of zero-carbon resource meaning that there are, by definition, fewer qualifying resources already present in the market to reshuffle. As a result, even for a small state, a real change in production will be necessary to meet a sufficiently stringent RPS policy and emissions within the market will decline. Therefore, since emissions outside the market are not affected by the state's RPS policy and emissions in the market fall, the state RPS policy will reduce emissions in the nation as a whole. See Appendix 2.A for the mathematical proof.

Considering that about two-thirds of the states in the U.S. participate in larger markets, and that among those states, the majority can be considered small relative to their market, our results suggest that should such a state decide to adopt unilateral measures to reduce CO_2 emissions, an RPS approach is more likely to allow them to have an impact on national emissions either in the absence of national policy or in the presence of a national CO_2 tax. This is one possible rationale for the current U.S. climate policy landscape, in which there is no national climate policy and 29 states have adopted RPS policies.[15]

[14] If a state is so small that the pre-state-policy consumption in-state is less than pre-existing generation from qualifying renewable fuels, a state RPS policy will also be ineffective.

[15] Database of State Incentives for Renewables & Efficiency, `http://www.dsireusa.org`

2.3.1.2 Large States

Now consider the case of a large state: because existing zero carbon resources cannot satisfy its full demand, some producers will have to pay the state CO_2 tax in addition to any national CO_2 tax, which increases their tax burden. The increase in tax burden caused by the state's CO_2 tax will therefore generally cause a reduction in national emissions. However, if there are fuels that are zero-carbon but do not qualify as renewables under an RPS, large states will be able to achieve larger national emissions reductions with a state RPS than with a state CO_2 tax, provided there is at least one other state in the market. See Appendix 2.B for the mathematical proof.

Intuitively, this result follows because under the most stringent state RPS, the state would consume only qualifying renewable fuels, ceding all existing zero-carbon generation to the rest of the market and reducing demand for coal and natural gas in the rest of the market relative to the demand under an infinitely high state CO_2 tax, under which only coal and natural gas would be available to the rest of the market as the state would consume *all* zero-carbon resources. Therefore, total emissions under the most stringent RPS will be lower than under an infinitely high CO_2 tax. If a state does not participate in a larger market and therefore does not trade with other states, then a state CO_2 tax and a state RPS will be able to achieve the same maximum reduction in emissions. When both a state-level CO_2 tax and a state-level RPS can achieve a given reduction in emissions, intuition suggests that a state CO_2 tax will be more cost-effective than an RPS at reducing emissions as it targets emissions directly and provides more flexibility in the options for compliance. This hypothesis is confirmed by our simulation results in Section 3.3.

2.3.2 National Cap-and-Trade Policy with a Carve-Out Provision

We next consider the effect of a national cap-and-trade program with a carve-out provision. Recall that under a national cap-and-trade program without a carve-out provision, state policies cannot induce additional emission reductions because any reduction in emissions caused by a state policy will lower the national emissions permit price and therefore cause a corresponding increase in emissions outside the state up to the level of the national cap. The carve-out provision allows any state to become exempt from the national cap provided that it implements a tighter state cap.[16] Therefore, if a state decides to set a tighter cap, the level of emissions outside the state does not change while emissions within the state decrease. Together this implies that, regardless of the size of the state with the tighter cap, the state cap will cause national emissions to decline.

Under a national cap-and-trade with carve-out, when other states in the market do not have climate policies, a tighter state cap pushes carbon-intensive resources out of the state, increasing emissions in the rest of the market. To compensate, emissions outside the market must decline. Unlike under a national CO_2 tax where reshuffling may have no effect on prices and the total tax burden, in this case, reshuffling of resources within the market would raise emissions in the rest of the market and force consumers outside the market to make costly reductions in emissions. Thus, while a state CO_2 tax or RPS in addition to a national CO_2 tax (or no national policy) leaves consumers outside the market unaffected, a tighter state cap under a national cap-and-trade with carve-out imposes additional costs on consumers outside the market when the state with the tighter cap participates in a larger market. If the state does not participate in a larger market, then a state emissions policy would not cause reshuffling regardless of the national policy (since there is no region to reshuffle with), and therefore a state CO_2 tax that causes a reduction in emissions beyond that caused by a national CO_2 tax would lead to the same outcome as a tighter state cap and a national cap-and-trade program with a carve-out that mandated the same total reduction in emissions.

[16] In our model, a state CO_2 tax set above the level of the current national carbon credit price would generate the same outcome as a tighter state cap. Also, note that carving a state out of a national CO_2 tax policy provided it sets a higher tax yields the same outcome as adding a state tax on top of a national tax since the total tax would be the same.

2.3.3 Comparison of Costs under Various Policy Combinations

Under a national CO_2 tax, if a state policy is enacted that is sufficiently stringent to cause national CO_2 emissions to decline and that state also participates in a larger market with other states that do not have climate policies, then resources in the market will be reshuffled and the price paid to carbon-intensive resources must fall to reduce generation from carbon-intensive resources and thereby reduce CO_2 emissions. If the price of carbon-intensive resources falls, then consumption of carbon-intensive resources in the rest of the market will rise, the phenomenon known as leakage. Consequently, for a state to reduce national emissions by X tonnes it will have to reduce its own emissions by more than X tonnes to account for the increase in emissions outside the market. However, a national cap-and-trade policy with a carve-out provision limits both reshuffling and leakage within the market because any increase in emissions in the rest of the market forces consumers outside the market to make costly reductions in emissions to ensure the national cap is met, which increases the national emissions permit price and prevents large consumption increases in the rest of the market. Thus, a reduction in emissions caused by altering consumption within the state by a given amount is met by a smaller increase in emissions in the rest of the market than if a national CO_2 tax were in place and therefore, the net emissions reduction is larger for a given change in state consumption patterns. As a result, the cost of a given reduction in emissions achieved by a state policy is lower under a national cap-and-trade with carve-out than under a national CO_2 tax.

To illustrate the degree of difference between the costs of the three national-state policy combinations (National CO_2 Tax + State CO_2 Tax, National CO_2 Tax + State RPS, and National Cap-and-Trade with Carve-out + State Cap) we perform numerical simulations for two relative state sizes: 25 percent of the market and 75 percent of the market. A state that is 25 percent of its market is small in our simulations as generation from zero-carbon resources was 32 percent of total consumption. A state that is 75 percent of its market is therefore large. For each simulation, we calculate the surplus accruing to the consumers in each region and the surplus accruing to each type of producer. The sum of the consumer surplus, producer surplus and the total tax or emissions permit revenue paid by each region yields the national surplus under each scenario. The reduction in national surplus due to a particular policy is the cost of the policy. Details on the data and calibration can be found in Appendix 2.C.

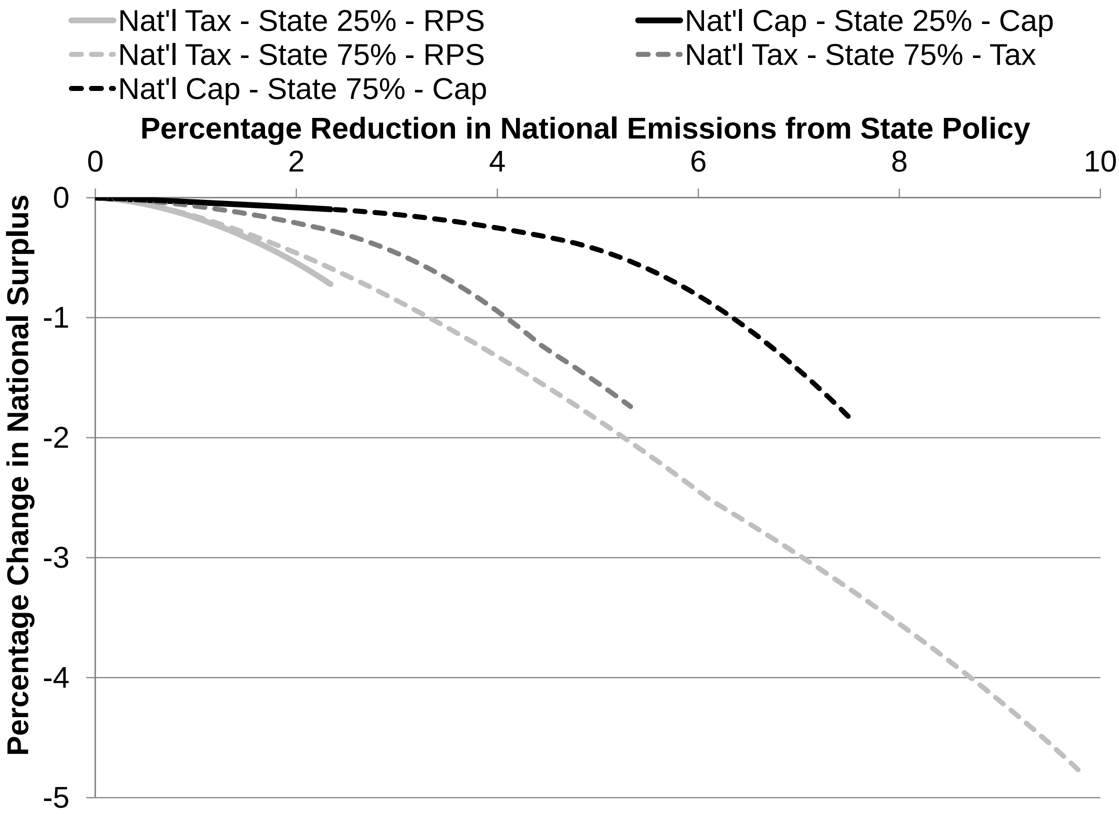


Figure 2.2: The figure depicts the relationship between national surplus changes and national emissions reductions for various combinations of national and state policies. Each line represents the relationship for a fixed national policy and a state policy gradually increasing in stringency.

Figure 2.2 plots the percentage change in national surplus against the percentage reduction in national emissions caused by each national-state policy combination under our baseline parameters for two relative state sizes, 25 percent and 75 percent. Note that cap-and-trade polices at both the national and state levels *without* state carve-out and national and state CO_2 taxes with a small state result in the state having no incremental impact on national emissions and are therefore not shown. In the figure, the national policy is held fixed at \$20 (or at the national cap that is equivalent when there is no state policy) while the state policy increases in stringency. The lines terminate at the maximum achievable emissions reduction given the policy combination and relative state size.

The figure verifies our intuition that the cost to achieve a given reduction in CO_2 emissions using a national cap with a carve-out and tighter state cap is less than the cost using a national CO_2 tax with a state RPS or CO_2 tax. The figure also confirms that the cost of a national CO_2 tax with a state CO_2 tax is less than the cost of a national CO_2 tax with state RPS when both are feasible and that a state RPS can achieve larger reductions in emissions than a state CO_2 tax under a national CO_2 tax. The exact differences between the costs of the different policy combinations depend on the parametrization of the model, but the ordering of the policies in terms of cost-effectiveness is invariant over a broad range of elasticities of regional demand and fuel supply, region sizes and fuel mix.

In summary, when there is a national CO_2 tax or no national emissions policy and a state is small or a state is large but desires a large reduction in emissions, a state RPS may be able to achieve the state's desired emission reduction while a state CO_2 tax would fail. For smaller emissions reductions in large states, either a state RPS policy or a state CO_2 tax will be able to reduce national emissions, though the state CO_2 tax will be more cost-effective. When there is a national cap-and-trade with a carve-out, a state of any size can cause a reduction in national emissions by setting a tighter state cap. Furthermore, the cost of a given reduction in emissions under a national cap-and-trade with a carve-out and a state cap will be lower than the cost under a national CO_2 tax with a state RPS or CO_2 tax. Thus, a national cap-and-trade policy with a carve-out provision may be preferable to a national CO_2 tax.

However, the distribution of the costs from each national policy differs. Under a national CO_2 tax, a state policy only affects the market in which the state participates and therefore only affects costs for consumers and producers inside the market. Under a national cap-and-trade with a carve-out, a state policy also increases costs for consumers and producers outside of the market. Thus, although a national cap-and-trade policy with a carve-out is less costly than a national CO_2 tax when states initiate stricter policies, a national CO_2 tax ensures that the state pays for the majority of the costs of their policy-making. In other words, a national CO_2 tax allows large states to reduce emissions if they so desire, while a national cap-and-trade program allows a state of any size to reduce national emissions, but also imposes additional costs on other states. This may cause many states to oppose carve-out provisions.

2.3.4 Effect of Policies by Other States Within the Market

We now consider the implications of overlapping policies in other states within the market. We build on the intuition developed in the prior sections to outline how different combinations of policies across states might interact under each of the national policies considered here.

Under a national CO_2 tax or no national policy, if at least one other state in the market has a binding RPS (targeting in-state consumption of renewables rather than in-state production), a state of any size will be able to implement an RPS policy that binds and reduces national emissions. This is because the states that already have RPS policies will be consuming all qualifying renewable resources in the market and will be unwilling to relinquish them. Therefore, to satisfy a new RPS policy in a state with no prior climate policy, production from qualifying renewable resources must increase, which will cause prices and production from all resources to adjust. Since qualifying renewable generation is increasing, generation from coal and natural gas will decrease.

To be able to affect national emissions via a state-level CO_2 tax when there is a national CO_2 tax or no national policy and other states in the market have RPS policies, a state's pre-state-policy consumption must be larger than the existing non-qualifying zero carbon resources. The size barrier a state must exceed to be able to implement an effective CO_2 tax is now lower because the other states with RPS policies will not be willing to trade qualifying renewable resources for fossil fuels with no change in prices, as this trade would increase the cost of complying with their RPS policies. However, the other states with RPS polices are willing to trade non-qualifying zero-carbon resources because these fuels pay the same implicit tax as fossil fuels under an RPS policy.

If other states have CO_2 taxes that reduce national emissions in the presence of a national CO_2 tax or no national policy, the effect of a state CO_2 tax in a new state will depend on the level of the new CO_2 tax relative to existing state CO_2 taxes in the market as well as the relative size of the state adding the new CO_2 tax. For instance, suppose there is one other state, state A, with a CO_2 tax, T^A. To be effective at reducing national emissions, state A must be large. If state B sets a new CO_2 tax, $T^B < T^A$, it will not be able to draw in any zero-carbon resources from state A because the value of zero-carbon resources is higher in state A. However, generation and emissions will change in this case because all generation sold in state B will be carbon-intensive and subject to the tax T^B.

If state B's electricity consumption is less than the quantity of generation from zero-carbon resources prior to the initiation of state B's policy (or in other words, if state B is small), then setting T^B above T^A will generate the same emission reduction as setting T^B equal to T^A. This occurs because when $T^B \geq T^A$, zero carbon resources are at least as valuable in state B as in state A, which will cause zero-carbon resources to be shuffled to state B. When state B is small, all demand in state B can be satisfied with existing zero-carbon resources. If state B is large, then increasing state B's CO_2 tax from T^A to a higher level will generate additional reductions in emissions.

If there is a national cap-and-trade policy with a carve-out and states can be carved out of the national policy only if they set a tighter state cap, then any other type of state policy would overlap with the national cap (rather than supersede it) and would therefore be unable to affect national emissions. States may nevertheless have other types of climate policies if they believe they are correcting other externalities. In this case, the other states' policies will generally further limit the reshuffling of resources that can occur when a state carves itself out of the national cap to set a tighter cap, but they will not affect the state's ability to reduce emissions. If other states in the market are already carved-out, mathematically, total emissions must decline when an additional state is carved out since the sum of the caps is now lower.

In summary, if other states have RPS policies that are binding, then a state of any size will able to reduce emissions with an RPS policy since all qualifying renewable resources are being utilized. Under a national CO_2 tax, increasing the state CO_2 tax within one state up to the level of the maximum state CO_2 tax in the market will always bring about additional emissions reductions. Depending on the size of the state, CO_2 taxes above level of the maximum state CO_2 tax may or may not have an additional effect. Under a national cap-and-trade program with a carve-out, when other states in the market have been carved out, carving out an additional state brings about additional emissions reductions.

2.3.5 Vertical Targeting of State Policies

We modeled the state-level policy as targeting emissions that are attributable to in-state consumption as opposed to emissions from in-state production. Under the assumption that one motivation for unilateral state-level (national) policies targeting global public goods is a concern for the common good rather than just the economic impacts to the state (nation), it is consistent for states (nations) that are net importers of pollution (i.e., the emissions embodied in the goods imported for domestic consumption exceed that embodied in what they export) to target emissions from in-state (national) consumption. In fact, many of the states in the U.S. and several of the nations in the European Union that support strong policies to limit CO_2 emissions are net importers of energy and/or energy-intensive goods and services.

Another alternative is to tax fossil fuels directly based on their carbon content rather than taxing emissions from use of fossil fuels. Again, states could either tax producers of the fossil fuels based on their production within the state or based upon their sales in the state. For states with little production of fossil fuels, taxing in-state production would have little effect. Taxing out-of-state producers based on their sales of fossil fuels in-state would also, in effect, raise the cost of in-state production of energy intensive commodities and eventually lead to relocation of such activities, which is a type of leakage.

2.4 Conclusion

We analyze the effect of two different state-level policies – a CO_2 tax (or an equivalent cap and trade system) and an RPS, on national emissions of a global pollutant under different national policy regimes – no national policy, a national CO_2 tax and a national cap-and-trade program with state carve-out.[17,18] We highlight the effect of pollution shuffling and leakage on the ability of state-level policies to reduce national emissions and the cost they impose on the rest of the nation.

We find that the effectiveness of a state RPS or CO_2 tax at reducing national emissions will be influenced by a set of common factors whether there is a national CO_2 tax or no national climate policy. In both cases, a state whose consumption is less than the quantity of qualifying renewable generation within the larger market that can be reallocated to that state, will not be able to affect overall emissions with either an RPS or CO_2 tax at the state-level. A state whose consumption is greater than the existing qualifying renewable generation but less than the existing zero-carbon generation in the market (small states) can affect national emissions by adopting a state-level RPS policy, while a state-level CO_2 tax will not be able to reduce emissions due to reshuffling of zero-carbon resources. We thus show that an emission tax at the national-level does not guarantee that overlapping state-level policies are immune to complete leakage. For large states subject to a national CO_2 tax or no national climate policy, modest emission reduction goals can be achieved with either a state-level CO_2 tax or a state-level RPS, though the cost should be lower under a state-level CO_2 tax. The maximum feasible reduction in

[17] That a state-level policy is unable to affect national emissions under a national emissions cap *without* carve-out is well known (See Burtraw and Shobe, 2009; Goulder and Stavins, 2011a, e.g.)).

[18] The one possible combination among these policies that we do not analyze is a state RPS with national cap and trade with state carve-out. This option is excluded because it would be very difficult in practice to determine what the minimum stringency of a state's RPS ought to be to qualify the state to be carved out of a national cap and trade regime. Conversely, when the state's policy is an emissions tax or lower emissions cap, the requirements for a state to be carved out are clear.

national emissions, however, is higher for a state-level RPS compared to a state-level CO_2 tax.

Under a national cap-and-trade program with a carve-out provision, a state of any size can achieve a reduction in national emissions by setting a tighter state cap because the sum of the national and state emissions caps has been reduced. Examining cost-effectiveness, we find that a national cap-and-trade policy with a carve-out provision will cost less than a national CO_2 tax when states pursue more stringent overlapping policies. This occurs because reshuffling or leakage of emissions within the market raises the national emissions permit price (in order to keep emissions in the rest of the nation constant). Consequently, for any given reduction in national emissions, the increase in emissions in the rest of the market is smaller under a national cap-and-trade program with a carve-out than under a national CO_2 tax and therefore the cost of achieving that reduction in emissions is lower under the national cap-and-trade program with a carve-out. If a tighter state cap does cause an increase in emissions in the rest of the market, then emissions outside the market must fall, increasing prices for consumers outside the market. Under a CO_2 tax, emissions and cost to consumers outside the market are both unaffected by the state policy. Thus, while a national cap-and-trade policy with a carve-out is less costly for the nation as a whole when states implement tighter caps than the equivalent national and state CO_2 taxes, a national cap-and-trade with a carve-out will lead to higher costs for consumers outside the market than would a national CO_2 tax, which could create political opposition to allowing individual states to be a carved out of a national cap-and-trade program.

Extending the model to consider overlapping policies in multiple states within a market together with a national-level policy, we find that, holding the total quantity of renewable generation within the market fixed, the size threshold for a state to affect national emissions through a CO_2 tax diminishes as more states within the market adopt targets for renewable energy consumption. If other states in the market have CO_2 taxes and there is a national CO_2 tax, a state of any size could cause additional emissions reduction by adding a state CO_2 tax. If there is a national cap-and-trade policy with a carve-out provision, the climate policies of other states do not affect the ability of the state to reduce national emissions by setting a tighter cap. Extending to a multi-sectoral or economy-wide context, we conclude that the efficacy of a state-level policy in reducing national emissions will change depending on how the relative size of the state changes with the widening scope of the policy. Given the global effects of CO_2 emissions, our results also speak to the interactions that could take place when global policies overlap national policies or state policies overlap local policies and product markets are larger than the smaller jurisdiction. Our framework can additionally be extended to consider other policies such as emission intensity standards, subsidies for renewable energy and border adjustment policies.

Appendices

2.A Proof of Proposition 2.1

Proposition 2.1 *For any pre-policy generation portfolio with fossil fuels, RPS-qualifying renewables, and non-qualifying near-zero carbon resources, there exists a range of relative state sizes such that a sufficiently stringent state RPS policy may reduce CO_2 emissions, but a state CO_2 tax cannot when there is no national policy or when there is a national CO_2 tax.*

Proof. Assume for simplicity that the utility functions are continuous, increasing, and strictly concave and the cost functions are continuous, increasing, and strictly convex.

Let $\rho \in [0, 1]$ indicate the size of the state relative to the market. As state size is measured by the state's share of the market's electricity consumption pre-state policy, in-state consumption is $q^{s0} = \rho q^{M0}$, the 0 superscript indicating pre-state-policy. Let $\underline{\mathrm{R}}$ represent the share of market consumption that could be met with existing qualifying renewable generation alone, $\underline{\mathrm{R}} q^{M0} = q_r^{M0}$. Then $\underline{\mathrm{R}}$ also represents the largest state size such that the state can consume only qualifying renewables *and* satisfy all pre-policy demand, i.e. if $q^{s0} \leq \underline{\mathrm{R}} q^{M0}$ then all demand in state can be served by existing qualifying renewables, q_r^{M0}. Let $\bar{R}$ be such that $\bar{R} q^{M0} = q_r^{M0} + q_z^{M0}$ and in-state consumption be such that $q^{s0} \leq \bar{R} q^{M0}$. Then $\bar{R}$ is the share of zero-carbon resources in the market as well as the largest state size such that the state can consume only zero-carbon resources *and* satisfy all pre-policy demand.

As explained in section 2.3.1.1, when $\rho < \bar{R}$, a state CO_2 tax cannot reduce emissions when there is a national CO_2 tax $T^n \geq 0$ as there are sufficient zero-carbon resources in the market to satisfy demand in the state with no change in production or emissions and thus no change in tax burden. Also, if the national policy is a CO_2 tax or if there is no policy, the rest of the nation's production and emissions will be unaffected by any state policy.

The remainder of the proof will proceed in three steps. In step 1, we demonstrate that for a sufficiently large state, $\rho > \underline{\mathrm{R}}$, there exists an RPS stringency, α, such that the qualifying renewable resources required by the policy if state consumption remained unchanged would be larger than existing quantity of qualifying renewables, $\alpha\rho q^{M0} > q_r^{M0}$. In step 2, we note that if α is such that $\alpha\rho q^{M0} > q_r^{M0}$, then the RPS constraint binds and the shadow price on the constraint γ^s (a.k.a. the REC price) must be larger than zero. In step 3, we demonstrate that given $\alpha\rho q^{M0} > q_r^{M0}$, the RPS policy will cause the price in the rest of the market, p^m to decline which will reduce the quantity of generation from coal and natural gas and therefore reduce emissions.

Step 1: Prove that there exists an $\alpha \in (0,1)$ such that $\alpha\rho q^{M0} > q_r^{M0}$ when $\rho > \underline{\mathrm{R}}$: **Proof.** At $\rho = \underline{\mathrm{R}}$, $\underline{\mathrm{R}}q^{M0} = q_r^{M0}$. For $\rho > \underline{\mathrm{R}}$, $\rho q^{M0} > q_r^{M0}$. For α sufficiently close to one, $\alpha\rho q^{M0} > q_r^{M0}$. ■

Thus, there exists an RPS requirement, α, that will force the state to either increase generation from renewable resources beyond what was produced in the market pre-policy or reduce consumption.

Step 2: If α is such that $\alpha\rho q^{M0} > q_r^{M0}$ and $\alpha < 1$, then existing renewables cannot satisfy the RPS constraint and the RPS constraint will bind, causing the shadow price (price of a REC in-state), γ^s, to be positive.

Step 3: Prove that $p^m < p^0$ for $\alpha > \hat{\alpha}$ where $\hat{\alpha}$ is such that the RPS just binds ($\hat{\alpha}\rho q^{M0} = q_r^{M0}$): **Proof.** If $\alpha < 1$, $\gamma^s > 0$ and $q^s > 0$, then the binding RPS constraint causes non-qualifying fuels (coal, natural gas and non-qualifying zero-carbon fuels, defined collectively as q_{nq}) to be used in-state, $q_{nq}^s = q_c^s + q_{ng}^s + q_z^s > 0$ (otherwise the RPS constraint which requires $q^s = Aq_{nq}^s$, $A = \frac{\alpha}{1-\alpha}$, would dictate that $q^s = 0$ also). From the first order conditions, if generation from renewable fuels is sold both in-state and to the rest of the market, $q_r^s > 0$ and $q_r^m > 0$, then it must receive the same price in both: $p^s + \gamma^s = p^m$, where $p^s + \gamma^s$ is the price received for renewable generation in-state and p^m is the price that all types of generation receive out-of-state. If $p^s + \gamma^s = p^m$ then $p^s - A\gamma^s < p^m$ since $A > 0$. Since $p^s - A\gamma^s$ is the price received in-state for all non-qualifying fuels, $p^s - A\gamma^s < p^m$, implies producers would only want to sell non-qualifying fuels to the rest of the market where the price is higher so $q^s_{nq} = 0$ and $q^m_{nq} > 0$, but this is a contradiction. Thus, $q^s_r > 0$, $q^m_r = 0$, $q^s_{nq} > 0$ and $q^m_{nq} > 0$ and $p^s - A\gamma^s = p^m$, or in other words, if non-qualifying resources are consumed in both regions, they must receive the same price in each region. Thus if p^m declines, then the strictly increasing and convex cost curves ensure that generation from coal and natural gas in the market declines.

With $\alpha > \hat{\alpha}$, to meet the RPS constraint either consumption declines, $q^s < \rho q^{M0}$, renewable generation increases, $q^s_r > q^{M0}_r$, or both. Suppose that consumption and renewable generation increase, $q^s \geq \rho q^{M0}$ and $q_r > q^{M0}_r$. If $q_r > q^{M0}_r$, $p^s + \gamma^s \geq p^0$ by strict convexity of costs. If $q^s \geq \rho q^{M0}$, then $p^s \leq p^0$ by strict concavity of utility. Together, it must be that $p^0 \geq p^s$, which implies $p^0 > p^s - A\gamma^s = p^m$.

In the other case, if consumption declines, $q^s < \rho q^{M0}$, then $p^s > p^0$ by strict concavity of utility. $\gamma^s > 0$ implies $p^s + \gamma^s > p^0$, so it must be that renewable generation increases $q_r > q^{M0}_r$. At $\hat{\alpha}$ where the RPS just binds, $\rho q^{M0} - q^{M0}_r = \hat{q}^s - q^{M0}_r = \hat{q}^s_{nq}$, $\hat{q}^m_{nq} = q^m_0$, and $\hat{q}^s_{nq} + \hat{q}^m_{nq} = q^{M0}_{nq}$. $\hat{q}$ indicates the equilibrium quantity at $\hat{\alpha}$.

We have assumed that $q^s < \rho q^{M0}$ and $q_r > q^{M0}_r$, which imply $q^s_{nq} = q^s - q_r < \rho q^{M0} - q^{M0}_r = \hat{q}^s_{nq}$. Suppose $q^m_{nq} \leq \hat{q}^m_{nq} = q^m_0$. Since $q^m = q^m_{nq}$, $p^m \geq p^0$ by strict concavity of utility. But if $p^m \geq p^0$, then $q^s_{nq} + q^m_{nq} < q^{M0}_{nq}$, which implies $p^m < p^0$ by strict convexity of costs, a contradiction. Therefore, $q^m = q^m_{nq} > \hat{q}^m_{nq} = q^m_0$ and, as long as $q^M_{nq} < q^{M0}_{nq}$, we have $p^m < p^0$ by strictly increasing and concave demand and strictly increasing and convex costs, as desired. ■

With $p^m < p^0$, generation from coal and natural gas decreases ($q^M_c < q^{M0}_c$ and $q^M_g < q^{M0}_g$) by strictly increasing and convex costs. Therefore CO_2 emissions in the market decline. ■

2.B Proof of Proposition 2.2

Proposition 2.2 *If there is no national climate policy in place or if there is a national CO_2 tax, and if there are zero-carbon resources that are not RPS-qualifying renewables,* *then the maximum reduction in emissions that can be achieved by a state RPS will exceed the maximum reduction that can be achieved by a state* CO_2 *tax for states participating in markets with other states.*

Proof. Assume for simplicity that the utility functions are continuous, increasing, and strictly concave and the cost functions are continuous, increasing, and strictly convex.

If there is a national CO_2 tax or if there is no national climate policy (i.e. $T^n = 0$), the rest of the nation's production and emissions will be unaffected by any state policy. If a state RPS demands 100 percent renewables, then coal, natural gas, and non-qualifying zero-carbon fuels are used only in the rest of the market and the following first order conditions determine their output:

$$p^m = \frac{\partial}{\partial q_c^m} u^m(q_c^m + q_g^m + q_z^m) = \frac{d}{dq_c^m} c_c(q_c^m) + e_c T^n \tag{2.B.1}$$

$$p^m = \frac{\partial}{\partial q_g^m} u^m(q_c^m + q_g^m + q_z^m) = \frac{d}{dq_g^m} c_g(q_g^m) + e_g T^n \tag{2.B.2}$$

$$p^m = \frac{\partial}{\partial q_z^m} u^m(q_c^m + q_g^m + q_z^m) = \psi \tag{2.B.3}$$

ψ is the Lagrange multiplier on the constraint $q_z^m + q_z^s \leq Q_z^M$. If the price in the rest of the market is positive, $p^m > 0$, then the full capacity of non-qualifying zero-carbon generation will be utilized, $\psi > 0$ and $q_z^m = Q_z^M$. If not, $p^m = 0$ and $q_z^m \leq Q_z^M$. If we assume (i) $\frac{d}{dq_c^m} c_c(0) > 0$ and (ii) $\frac{d}{dq_g^m} c_g(0) > 0$, then when $p^m = 0$, $q_c^m = 0$ and $q_g^m = 0$.

If a state CO_2 tax is sufficiently high, then only zero-carbon resources will be used in-state and only coal and natural gas will be used in the rest of the market. Generation of coal and natural gas are then determined by the first order conditions:

$$p^m = \frac{\partial}{\partial q_c^m} u^m(q_c^m + q_g^m) = \frac{d}{dq_c^m} c_c(q_c^m) + e_c T^n \tag{2.B.4}$$

$$p^m = \frac{\partial}{\partial q_g^m} u^m(q_c^m + q_g^m) = \frac{d}{dq_g^m} c_g(q_g^m) + e_g T^n \tag{2.B.5}$$

If we assume (iii) $\frac{\partial}{\partial q_c^m} u^m(0) > \frac{d}{dq_c^m} c_c(0) > 0$ and (iv) $\frac{\partial}{\partial q_g^m} u^m(0) > \frac{d}{dq_g^m} c_g(0) > 0$, then generation from coal and natural gas will always occur, $q_c^m > 0$ and $q_g^m > 0$.

If $q_c^m = 0$ and $q_g^m = 0$ under the most stringent state RPS, CO_2 emissions in the market will be zero by assumptions (i) and (ii). Conversely, even with a very high state CO_2 tax, emissions will always be positive because, in that case, $q_c^m > 0$ and $q_g^m > 0$ by assumptions (iii) and (iv).

If $q_c^m > 0$ and $q_g^m > 0$ under the most stringent RPS, then $q_z^m = Q_z^M$. If q_c^m and q_g^m solve equations (2.B.1) and (2.B.2) and we were to remove the non-qualifying zero-carbon generation being used, then $\frac{d}{dq_c^m} c_c(q_c^m) + e_c T^n < \frac{\partial}{\partial q_c^m} u^m(q_c^m + q_g^m)$ and $\frac{d}{dq_g^m} c_g(q_g) + e_g T^n < \frac{\partial}{\partial q_g^m} u^m(q_c^m + q_g^m)$ by strict concavity of utility. Note that except for the inequality, these equations are equations (2.B.4) and (2.B.5). By continuity, concave strictly increasing utility, and convex strictly increasing costs, there exists a $\hat{q}_c^m > q_c^m$ and a $\hat{q}_g^m > q_g^m$ such that equations (2.B.4) and (2.B.5) are satisfied. Thus, the most stringent RPS induces less generation from fossil fuels and therefore fewer emissions than would an infinite CO_2 tax because, unlike under an infinitely high state CO_2 tax, zero-carbon generation is available to out-of-state consumers under the most stringent RPS, which replaces much of their demand for generation from coal and natural gas. ■

2.C Data and calibration

We assume the supply and demand curves are linear and represent a long term response to long-term price trends in the market. Thus, the demand curve represents the average consumer response to price changes over the long term, and the supply curve is modeled as a long-term adjustment by producers who may be investing in new generation capacity. To ensure our demand and supply functions have the required interpretation, we utilize data from the Annual Energy Outlook 2011 (AEO2011) published by the U.S. Energy Information Administration (EIA) which focuses on the factors that shape the U.S. energy system over the long term. Our baseline pre-policy scenario utilizes 2009 data. Using the reference case, the high demand growth and the low demand growth side cases, we compute the elasticity of supply implied by the difference between the reference case and the side cases. We compute the elasticity of demand using the reference case and a side case developed to examine the effect of a clean energy standard for Senator Jeff Bingaman.

Parameter	**Interpretation**	**Value**
p^0	Initial Price (\$/KWh)	0.098
q^0	Initial Total Generation (KWh)	3.98E+12
q_z^0	Initial Non-Qualifying Zero-Carbon Gen. (KWh)	1.13E+12
q_r^0	Initial Renewable Generation (KWh)	1.45E+11
q_g^0	Initial Natural Gas Generation (KWh)	9.31E+11
q_c^0	Initial Coal Generation (KWh)	1.77E+12
ε	Demand Elasticity	-0.2
ε_r	Renewables Elasticity	1.49
ε_g	Natural Gas Elasticity	2.57
ε_c	Coal Elasticity	1.10

Table 2.C.1: Parameters of the Model

The baseline price, quantities, and elasticities used to calculate the parameters of the supply and demand curves are shown in Table 2.C.1. As our model contains three regions, state, rest of the market, and rest of the nation, we compute three demand curves. In all scenarios, we assume that a market consumes 10 percent of the national electricity consumed. We consider two possible sizes of the state relative to the market, 25 and 75 percent. If the state is 25 percent of the market, then 25 percent of the pre-policy market generation is consumed in the state. The price and elasticity of demand are assumed to be the same in each region. There are also two supply curves, one for the market and one for the rest of the nation. We assume resources are uniformly distributed across the nation and that the elasticity of supply is the same across the nation.[19]

Lastly, the Environmental Protection Agency indicates that CO_2 emissions are approximately 1.125 tonnes/MWh of coal generation and 0.5625 tonnes/MWh of natural gas generation.[20] Thus, the CO_2 emissions per MWh of coal generation are roughly double the CO_2 emissions per MWh of natural gas generation.

[19]The uniform distribution of resources only affects our results in that it sets the share of market-wide electricity a state must consume to be considered large (See Section 2.2.3).

[20]`http://www.epa.gov/cleanenergy/energy-and-you/affect/air-emissions.html`

CHAPTER 3

Storage or Natural Gas: Which Best Complements Increasing Wind Generation?

3.1 Introduction

Since 2000, 29 U.S. states, the District of Columbia, and 2 U.S. territories have implemented Renewable Portfolio Standards (RPS).[1] These policies mandate a share of electricity consumption that must be served by generation from qualifying renewable resources. The types of resources that qualify as renewables under each state's RPS vary, but generally included are wind, solar, geothermal, and small hydroelectric facilities as well as biomass incineration facilities. Of the 116,000 GWh increase in renewable generation (excluding hydroelectric) in the U.S. since 2002, 94 percent was generated by wind turbines, 1 percent was generated by solar technology, and the remaining 5 percent by geothermal and biomass.[2] Thus, wind has been the primary respondent to the RPS policies.

However, generation from wind turbines is dependent upon the prevailing winds and is therefore difficult to predict with 100 percent accuracy. As supply must always equal demand in electricity markets, other generators must adjust their output to compensate for the variation in the generation from wind. While these types of adjustments occur even in the absence of wind, due to the variation in demand, there is concern in the industry that the large amounts of wind generation being added in response to RPS policies may strain the flexibility of the electric power grid (Denholm et al., 2010). To

[1]Database of State Incentives for Renewables & Efficiency, `http://www.dsireusa.org`

[2]Energy Information Administration, Table 1.1.A. Net Generation by Other Renewables, 2002-June 2012, `http://www.eia.gov/electricity/data.cfm#generation`

increase the ability of the grid to absorb the unpredictable intermittency of wind generation, many have suggested an increase in energy storage capacity. As storage can both absorb excess wind generation and substitute for wind generation when the wind dies down, storage seems to be an ideal technology to smooth the output of a wind generator.

Nonetheless, flexible generators can also smooth wind generation by decreasing production when wind generation is high and increasing production when wind generation is low. A flexible generator is one which can quickly adjust its output at minimal additional expense. The most prominent examples of flexible generation are natural gas-fired combustion turbines and natural gas-fired combined cycle units. Combustion turbines are the traditional technology used on peak demand days as they can start quickly, have low start-up costs, and can adjust their output quickly. However, the efficiency of these units is low, making them costly to operate. Combined cycle units are a newer technology in which one or more combustion turbines is connected to a steam engine which captures the heat generated by the combustion turbines to generate additional electricity, thus increasing the efficiency and lowering the cost of generation. These units are utilized to follow demand, operating at nearly full capacity during the day and reducing output or shutting off at night. Either type of unit can adjust its output to the output of a wind turbine and both operate more flexibly than a coal or nuclear power plant.

Therefore, the question asked in this paper is: which is the more economical alternative to increase the flexibility of the electric power grid in the face of increasing quantities of intermittent wind generation, storage or natural gas? To answer this question, I utilize a simple partial-equilibrium model of the electricity energy market[3] with three periods, a night period, characterized by low demand, and a day period, characterized by high demand, and a shoulder period, characterized by middling demand. In each period, electricity may be generated, and in the night period, when demand is low, electricity can be stored for use in the day period when demand is high. For simplicity, costs and demand are assumed to be separable between periods and there are no transmission constraints

[3]The market for electric power, as opposed to other markets operated simultaneously that are used to procure reserves in case of outages and power quality maintenance services.

or transportation costs. Additionally, as I seek to understand which technology is more beneficial and cost-effective for society as a whole, I assume perfect competition and solve the model as a social planner.

Using data on existing generation and storage capacity, generation costs, and demand from the Pennsylvania Jersey Maryland (PJM) Interconnection, a market covering much of the Mid-Atlantic and Midwest U.S., I set up a base case scenario and solve for the optimal generation and storage in each period. The initial scenario assumes no uncertainty, while subsequent scenarios add uncertainty about demand and uncertainty about wind generation. For each scenario I calculate the market surplus given the existing generation and storage capacity, the market surplus from adding a storage facility, and the market surplus from adding a natural gas combined cycle unit. Comparing the change in market surplus from adding a storage facility with the change from adding a combined cycle unit, I am able to ascertain which technology is more valuable to the market.

The results indicate that the most important determinant of the value of additional storage capacity relative to the value of additional combined cycle generation capacity is the cost of natural gas. At high natural gas prices, as seen in 2008, the net lifetime benefit from additional storage capacity exceeds that of additional combined cycle generation capacity. However at the lower natural gas prices seen in 2011 and 2012, the net lifetime benefit of storage is significantly less than the new lifetime benefit of new combined cycle capacity and may even be negative. As the low natural gas prices seen today are likely to continue at least in the near future due to the recent boom in production (Yergin and Inesin, 2009), it is likely that storage will continue to be a poor investment relative to new natural gas combined cycle capacity. Conversely the effect of the variance of wind generation has a much smaller effect on the value of additional storage or generation capacity in this model. Increasing the efficiency of either technology would increase the value of new capacity from that technology but leave the other unaffected.

However, as the quantity of wind generation in the market grows, the value of combined cycle capacity falls rapidly, regardless of fuel prices, as less generation from fossil fuels is required to meet demand. However, the value of storage capacity may rise or fall depending on the steepness of the marginal cost curve. For instance, when fuel prices are high, the marginal cost curve becomes steeper and the price difference between day and night increases, increasing the value of storage. Therefore, the value of new storage capacity rises relative to the value of new combined cycle capacity when the quantity of wind generation in the market increases. Were wind (and/or solar) generation to become the predominant source of power in the U.S., storage would be a very important tool to balance the uncertainty inherent in these power sources, while combined cycle generation capacity would be virtually obsolete.

3.2 Literature Review

Much of the prior literature has calculated the value of an additional energy storage unit by simply utilizing historical prices and the operating characteristics of the unit. These analyses use different time periods and different markets and find differing results. For instance, Drury et al. (2011) analyze four U.S. electricity markets using prices from 2002 to 2009 and find that compressed air energy storage may be a profitable investment as do Sioshansi et al. (2011) who analyze the PJM market from 2002 to 2008. Figueiredo et al. (2006) analyze 14 markets in the U.S. and abroad using price data from 1996 to 2001, and find that in 6 of the markets storage could never generate an adequate return on investment, while in the remaining markets it may be a feasible investment. Graves et al. (1999) analyze a variety of U.S. and foreign markets using price data from 1997 and 1998 and find that in three markets there is a strong evidence that storage would be a profitable investment while in the others it is less likely. Sioshansi et al. (2009) analyze the effects of fuel prices, transmission constraints, efficiency of storage, and the capacity of storage and the fuel mix on the potential revenue that storage could earn in the PJM market from 2002 to 2007. Finally, Walawalkar et al. (2007) analyze the New York market using price data from 2001 to 2004 and find strong evidence that storage would be a profitable investment in the New York City area.

While many of these analyses do make an effort to tease out the effect of the fuel costs and other variables on their results, few utilize a structural model. An exception to this rule is Crampes and Moreaux (2010) who investigate analytically the optimal usage of storage as a function of the costs of generation and demand. They additionally emphasize that the net change in market surplus should be the driver of the usage of storage. A second exception is the analysis of Mokrian and Stephen (2006) who create a structural model in which a firm maximizes profit by operating a storage facility. The authors explore the differing results of a linear programming model, a dynamic programming model, and a stochastic programming model in which prices are either known, are generated according to a Markov process or are generated according to an ARMA process, respectively. Their models indicate that a compressed air energy storage facility may be a profitable investment while a smaller sodium-sulfur battery would not be. However, while their model incorporates uncertainty into the estimation of the value of storage, it does not allow the prices to respond to the introduction of storage but instead assumes they arrive exogenously. Although Crampes and Moreaux (2010) have some discussion of the difference between storage and generation, none of the analyses discussed here compare the economics of investing in additional storage capacity against the economics of investing in additional generation capacity.

3.3 The Model

The model consists of three periods: a night period with low demand, a shoulder period with middling demand, and a day period with high demand. I assume that the cost of electricity generation and demand are separable across the periods. The cost curves are assumed to be increasing and convex and the demand curves to be decreasing and concave. The economy is endowed with existing generation capacity, consisting of generators using nuclear, renewable, coal, natural gas, and oil resources, and existing storage capacity. A social planner maximizes welfare in the market.

Prior to the start of the night period, the social planner may choose to add additional natural gas combined cycle capacity and/or energy storage capacity subject to the cost of the additional capacity, also assumed to be increasing and convex. Combined cycle generators are designed to operate most efficiently at full capacity, causing the marginal cost curve for an individual generator to be decreasing. To avoid the problems this would cause in solving the model, the marginal cost of generating electricity from a new combined cycle unit is assumed to be constant. At prices below the marginal cost of the new combined cycle capacity, the cost function for generation is unaffected. At higher prices, the cost function for generation shifts laterally to the right according to the amount of capacity added. As a result, adding combined cycle capacity either leaves unchanged or reduces the cost of generating a given quantity of electricity, in addition to increasing the total generation capacity in the market.

It is assumed that all storage facilities face the same cost to store energy, and, therefore, adding additional storage capacity increases the amount of energy that can be stored but does not affect the cost of storage. As storage units are not perfectly efficient, not all of the electricity stored can be recovered. The ratio of the amount of electricity that can be recovered to the amount that was stored is known as the round-trip efficiency, e. As a result, given the rate at which a storage facility can fill or empty, it requires $\frac{1}{e}$ hours to store enough energy in the night period to have one hour of electricity available in the day period from a storage facility. Therefore, the cost per unit of energy stored is $\frac{1}{e} - 1 = \frac{1-e}{e}$ multiplied by the price of electricity in the night period.

Dividing a 24 hour day into three periods, I assume that the day period is 8 hours long, the night period is $\frac{1}{e}$*8 hours long and the shoulder period, in which storage would neither fill nor empty, is $(2 - \frac{1}{e})$*8 hours long. In the night period, the social planner chooses the quantity of electricity to generate and the quantity of electricity to store for use in the day period subject to the cost of generation and storage and the available generation and storage capacity. In the shoulder period, the planner chooses how much electricity to generate. In the day period, the planner chooses the quantity of electricity to generate for use in the day period subject to the cost of generation and the generation capacity available.

In the initial version of the model, there is no uncertainty. The second version of the model adds uncertainty about the demand functions. When choosing additional generation or storage capacity, the planner does not know the demand functions. In the night period, the planner knows the night period demand function, but does not know the shoulder or day period demand function when choosing generation for the night period and the amount of energy to store. The decisions in the night period are linked to the decisions in the day period by storage, but are unaffected by the outcome of the shoulder period. The uncertainty about demand in the shoulder period is resolved before the shoulder period generation decision is made. In the day period, all uncertainty is resolved before the planner chooses generation for the day period. The third version of the model adds uncertainty about wind generation, where the uncertainty about wind generation is resolved at the same time as uncertainty about demand. It is assumed that wind generation is uncorrelated with demand conditions.

3.3.1 Version 1 - No Uncertainty

The purpose of this version of the model is to develop an intuition for what drives the decision to invest in additional generation or storage capacity. As will be seen in versions 2 and 3 of the model, adding uncertainty may increase the value of additional storage or generation capacity, but the increase in value occurs through the same mechanism as in version 1.

To find the optimal generation, storage, and amount of capacity to build, the planner solves the following problem:

$$\max_{Q^N,,Q^{Sh},Q^D,q_s,X_g,X_s} \tfrac{1}{e}u^N(Q^N - q_s) + (2 - \tfrac{1}{e})u^{Sh}(Q^{Sh}) + u^D(Q^D + q_s) - \tfrac{1}{e}c(Q^N, X_g)$$
$$- (2 - \tfrac{1}{e})c(Q^{Sh}, X_g) - c(Q^D, X_g) - k_g(X_g) - k_s(X_s)$$

$$
\begin{aligned}
\text{s.t.} \quad & 0 \le Q^N \le C_g + X_g && \gamma_g^N, \lambda_g^N \\
& 0 \le Q^{Sh} \le C_g + X_g && \gamma_g^{Sh}, \lambda_g^{Sh} \\
& 0 \le Q^D \le C_g + X_g && \gamma_g^D, \lambda_g^D \\
& 0 \le q_s \le C_s + X_s && \gamma_s, \lambda_s \\
& 0 \le Q^N - q_s && \delta
\end{aligned}
$$

Q^N, Q^{Sh} and Q^D are the quantities of electricity generated in the night, shoulder, and day periods. q_s is the quantity of electricity stored in the night period, to be used in the day period. $Q^N - q_s$ is the quantity of electricity consumed by consumers at night and $Q^D + q_s$ is the quantity of electricity consumed during the day. All energy the planner chooses to store in the night period will be used in the day period because there is no uncertainty and storage is costly. $u^N(\cdot)$, $u^{Sh}(\cdot)$, and $u^D(\cdot)$ are the demand functions and $c(Q, X_g)$ is the cost function for generating electricity, which is dependent on the amount of electricity generated, Q, and the generation capacity added by the planner, X_g. X_s is the amount of additional storage capacity the planner chooses to build. $k_g(\cdot)$ and $k_s(\cdot)$ are the curves that determine the cost of building generation or storage capacity.

C_g is the existing generating capacity and C_s is the existing storage capacity in the market. The first four constraints ensure that generation and storage are non-negative and less than or equal to the capacity of generation and storage respectively. The γ variables are the Lagrange multipliers on the non-negativity constraints, and the λ variables are the Lagrange multipliers on the capacity constraints. Negative storage would imply that the planner would like to move electricity from the day period to the night period. However, because the night period is assumed to take place before the day period, storage is constrained to be positive. The final constraint ensures that the quantity generated in the night period is greater than or equal to the amount stored plus the electricity used in the storage process. This constraint is only necessary if the quantity stored is positive. It is not necessary to constrain the planner to only build positive quantities of storage or generation capacity as there is no cost to hold capacity, and therefore no reason to destroy capacity.

The cost functions and the demand functions are governed by the following assumptions:

Assumption 3.1 $c_1(C_g, X_g) > c_1(0, X_g) = 0,\ c_{11}(Q, X_g) \geq 0$

Here C_g is total available generation capacity. This first part of the assumption states that the marginal cost of generating the last possible unit of electricity is greater than the marginal cost of generating the first unit of electricity, which is equal to zero, given the amount of additional generation built by the planner, X_g. The second part of the assumption ensures that the cost curve is convex in Q so that the marginal cost curve is weakly increasing in the quantity generated.

Assumption 3.2 $k_s(0), k_g(0) > 0,\ k_s'(\cdot), k_g'(\cdot) \geq 0,\ k_s''(\cdot), k_g''(\cdot) \geq 0$

The first part of this assumption indicates that there are fixed costs to installing generation or storage capacity. The second part specifies that the cost of additional generation or storage capacity be increasing and convex.

Assumption 3.3 $u^{t\,\prime\prime}(Q) \leq 0,\ t \in \{N, Sh, D\},\ u^{D\,\prime}(Q) > u^{Sh\,\prime}(Q) > u^{N\,\prime}(Q)$
$u^{D\,\prime}(0), u^{Sh\,\prime}(0), u^{N\,\prime}(0) > 0$

The first part of the assumption guarantees that the utility function in each period t is concave so that marginal utility is weakly decreasing in the quantity of electricity consumed. The second part of the assumption ensures that the marginal utility of a given quantity of electricity, Q, in the day period is always greater than the marginal utility of the same quantity of electricity in the should period which is greater than the marginal utility of that quantity in the night period. As a result, for the same price, consumers in the day period would demand more electricity than consumers in the night period. This assumption is relaxed in the uncertainty version of the model. The third portion of the assumption ensures that the marginal utility at zero is always positive.

As described in the introduction to the model, I assume that adding generation capacity affects the cost curve for generation by shifting it right for quantities such that the marginal cost of generating that quantity is greater than the assumed marginal cost of the additional combined cycle generating capacity. For quantities of generation such that the marginal cost of generating that quantity is less than the marginal cost of the new generating capacity, the cost curve is unaffected. Thus, defining B as the quantity at which the marginal cost of existing generation equals the marginal cost of the new generating capacity, and m as the constant marginal cost of the new capacity, the cost curve $c(Q, X_g)$ can be written:

Assumption 3.4

$$c(Q, X_g) = \begin{cases} c(Q) & \textit{for } 0 \leq Q \leq B \\ c(B) + m(Q - B) & \textit{for } B < Q < B + X_g \\ mX_g + c(Q - X_g) & \textit{for } B + X_g \leq Q \leq C_g + X_g \end{cases}$$

3.3.1.1 Analytical Results

Combining the first order conditions of the planner's problem, shown below, and Assumptions 3.1 - 3.4, the conditions under which the planner would choose to invest in additional generation or storage can be determined.

$$\mathbf{Q^N}\text{: } \tfrac{1}{e}u^{N\,\prime}(Q^N - q_s) = \tfrac{1}{e}c_1(Q^N, X_g) + \lambda_g^N - \gamma_g^N - \delta \tag{3.1}$$

$$\mathbf{Q^N}\text{: } (2 - \tfrac{1}{e})u^{Sh\,\prime}(Q^{Sh}) = (2 - \tfrac{1}{e})c_1(Q^{Sh}, X_g) + \lambda_g^{Sh} - \gamma_g^{Sh} \tag{3.2}$$

$$\mathbf{Q^D}\text{: } u^{D\,\prime}(Q^D + q_s) = c_1(Q^D, X_g) + \lambda_g^D - \gamma_g^D \tag{3.3}$$

$$\mathbf{q_s}\text{: } \lambda_s - \gamma_s + \delta = u^{D\,\prime}(Q^D + q_s) - \tfrac{1}{e}u^{N\,\prime}(Q^N - q_s) \tag{3.4}$$

$$\begin{aligned}\mathbf{X_g}\text{: } k_g'(X_g) = (>)\lambda_g^N + \lambda_g^{Sh} + \lambda_g^D - \tfrac{1}{e}c_2(Q^N, X_g)& \\ - (2 - \tfrac{1}{e})c_2(Q^{Sh}, X_g) - c_2(Q^D, X_g) & \qquad \text{if } X_g > (=)0\end{aligned} \tag{3.5}$$

$$\mathbf{X_s}\text{: } k_s'(X_s) = (>)\lambda_s \qquad \text{if } X_s > (=)0 \tag{3.6}$$

Beginning with the problem of determining the amount of additional generation capacity to add, equation (3.5) shows that the first order conditions set the marginal cost of additional generation capacity equal to the sum of the Lagrange multipliers on the generation capacity constraints less the derivatives of the generation cost functions for each period with respect to additional capacity. The derivative of the cost function with respect to additional capacity in either period is represented by:

$$c_2(Q, X_g) = \begin{cases} 0 & \text{for } 0 \le Q < B + X_g \\ m - c_1(Q - X_g) & \text{for } B + X_g \le Q \le C_g + X_g \end{cases}$$

As B is the quantity at which the marginal cost of generation is m, for any quantity greater than $B + X_g$, the marginal cost, $c_1(Q - X_g)$, is greater than m. Further, in the case that the quantity of generation Q is greater than $B + X_g$, $c_1(Q - X_g) = c_1(Q, X_g)$. Therefore, equation (3.5) can be rewritten as:

$$\begin{aligned}k_g'(X_g) = &\lambda_g^N + \lambda_g^{Sh} + \lambda_g^D + \tfrac{1}{e}max(c_1(Q^N, X_g) - m, 0) \\ &+ (2 - \tfrac{1}{e})max(c_1(Q^{Sh}, X_g) - m, 0) + max(c_1(Q^D, X_g) - m, 0)\end{aligned} \tag{3.7}$$

As the marginal cost of the first unit of electricity is 0, by assumption 3.1, and the marginal utility of the first unit of electricity is positive, by assumption 3.3, the non-negativity constraints, γ, not bind. Setting price equal to marginal utility and substituting equations (3.3), (3.1), and (3.2) into (3.7) yields:

$$k_g'(X_g) = \frac{1}{e}max(P^N + e\delta - m, 0) + (2 - \frac{1}{e})max(P^{Sh} - m, 0) + max(P^D - m, 0) \quad (3.8)$$

δ greater than zero indicates that there is no consumption in the night period because the marginal cost of generating the electricity the planner wishes to store for day is greater than the marginal utility of the first unit of electricity. This should occur rarely, if ever. Thus, as long as δ is zero, the marginal value of additional combined cycle capacity is equal to the weighted sum of the maximum of the equilibrium price less the marginal cost of the new generation capacity and zero in each period.

If the quantity of generation chosen in all three periods is strictly less than the threshold level, B, plus a positive amount of additional capacity, X_g, then: (1) generation in the three periods is less than the existing generation capacity plus the additional capacity, $Q < C_g + X_g$, and (2) the marginal benefit of the additional generation on the cost function is zero ($c_2(Q, X_g) = 0$). When the first item occurs, the capacity constraints on generation do not bind, leading the Lagrange multipliers, λ_g^N,λ_g^{Sh}, and λ_g^D, to be zero. The second item is true because if generation in all periods is less than $B + X_g$, some of the added capacity is going unused and therefore should not have been installed. In this situation, equation (3.6) reduces to $k_g'(X_g) = 0$. As the marginal cost of additional capacity, $k_g'(X_g) \geq 0$, by Assumption 3.2, this indicates that the amount of additional generation may need to be reduced.

If generation in any period is greater than the threshold level plus the amount of capacity added, $Q > B + X_g$, but generation in all periods is less than the total market capacity, $Q < C_g + X_g$, then the marginal benefit of additional generation capacity will be $-c_2(Q, X_g) = c_1(Q - X_g) - m > 0$, or the difference between the marginal cost of the electricity generated less the marginal cost of generation from the new combined cycle capacity. Therefore, the higher the marginal cost of the total quantity of electricity generated, the higher the value of additional capacity and the more additional capacity will be installed. If generation in any period is equal to the total market capacity, $Q = C_g + X_g$, then the capacity constraint in that period will bind and the marginal benefit of additional generation capacity will increase further by the value of the Lagrange multipliers on the capacity constraints: $\lambda_g^N + \lambda_g^{Sh} + \lambda_g^D$. Thus the planner will want to increase generation capacity if the market is capacity constrained *or* if the marginal cost of the electricity demanded by consumers is significantly greater than the marginal cost of new capacity.

The relationship between the marginal cost of electricity at the equilibrium quantity and the marginal cost of new generation capacity is determined by two factors. First, the elasticity and level of demand for a given generation cost function determines the price in each period. The greater the willingness to pay for each unit of electricity, the higher the price of electricity and the higher the marginal cost of electricity at the equilibrium quantity. The more inelastic demand is, the more sensitive the electricity price will be to the cost of generation.

More importantly, the cost of fuel and other variable generation costs will determine where on the marginal cost (or supply) curve the new natural gas-fired combined cycle capacity will be placed. New natural gas combined cycle generators are generally more efficient per unit of energy than existing coal generation, but when natural gas prices are high relative to coal prices, the new combined cycle unit is likely to be more expensive than existing coal and will be placed further up the supply curve. The more similar coal and natural gas prices are, the more likely it is for the cost of the new natural gas capacity to be below the cost of existing coal units. This places the new capacity lower on the supply curve and causes the new capacity to be used more often, increasing its value.

Additionally, a larger amount of zero or near-zero cost generation, such as nuclear and renewable generation, shifts the supply curve of the remaining generation to the right, increasing the quantity of electricity in equilibrium and decreasing the price. The decrease in price lowers the price the new generation capacity would receive and also the probability the new generation capacity will be used.

Turning to the decision to build additional storage capacity, equation (3.6) shows that the planner would choose to build additional capacity if the value of the Lagrange multiplier on the storage capacity constraint, λ_s is sufficiently large. Equation (3.4) indicates that the value of the multiplier λ_s is determined by the difference between the marginal utility of an additional unit of electricity consumption during the day, $u^{D\,\prime}(Q^D + q_s)$, less the marginal utility of electricity consumption at night, $u^{N\,\prime}(Q^N - q_s)$, less the cost of storing the energy, $\frac{1-e}{e}u^{N\,\prime}(Q^N - q_s)$. Setting price equal to marginal utility, the value of storage is equal to the day price less the night price less the cost of storage $P^D - P^N - \frac{1-e}{e}P^N = P^D - \frac{1}{e}P^N$. Therefore the value of additional storage capacity is determined by the difference in price between night and day and the efficiency of storage. If the difference is negative, no additional storage capacity is necessary, when the difference is positive additional storage would be valuable.

In turn, the day and night prices are determined by the amount of generation available, generation costs, the utility functions, and the efficiency of existing storage capacity. If insufficient generation capacity is available in the day period to serve demand, this will raise the value of storage above the difference in marginal costs between periods. This results is obtained because the price in the day period is dependent both on generation costs and whether the market is capacity constrained, as can be seen in equation (3.3). Conversely, insufficient generation in both periods may have a negative effect on the value of storage as consumers would be less willing to give up the limited amount of electricity available in the night period for use in the day period.

Figure 3.1 illustrates the effects of changing the cost of generating electricity. For instance, in Figures 3.1.A and 3.1.B the slope of the generation cost curve is reduced. The steepness of the generation cost curve is determined by fuel prices and the efficiency of existing generators, as higher fuel prices magnify the difference in efficiency between neighboring generators on the supply curve. When fuel prices fall, this reduces the slope of the generation cost curve and therefore the difference in price between two quantities. However, the change in the day-night electricity price difference is also affected by the

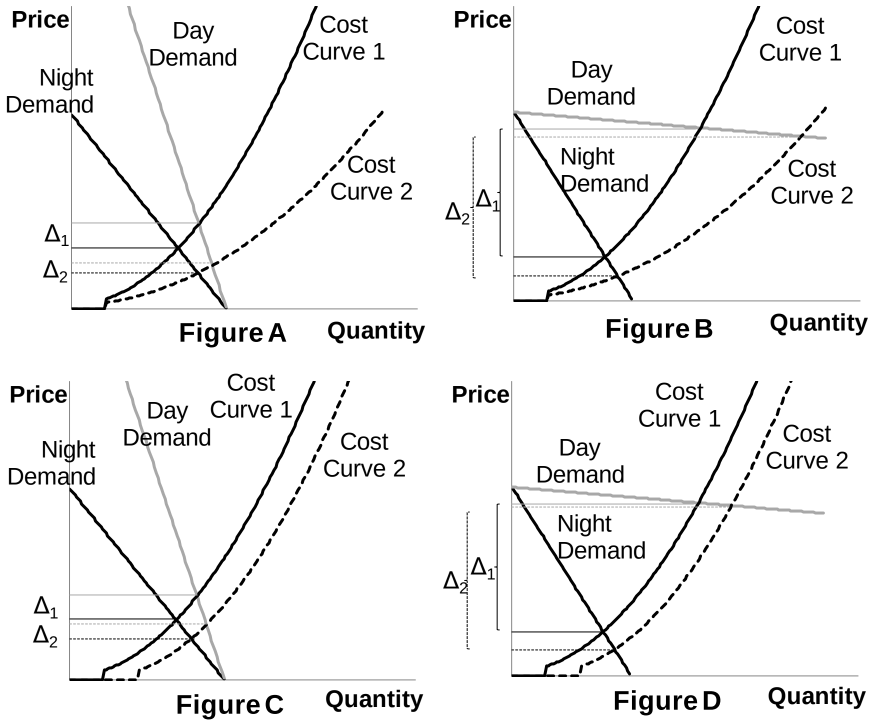


Figure 3.1: Effect of the Cost Function on the Value of Additional Storage

responsiveness of day and night demand to price and the level of the difference between day and night demand at each price.

In Figure 3.1.A, the slope of the demand curve for the day period is steeper than the slope of the demand curve for the night period, indicating that demand during the day is less responsive to price than is demand during the night. This causes the value of a large quantity of electricity to be similar in both periods, but the value of a small quantity of electricity to be significantly larger in the day period than in the night period. When the slope of the cost curve decreases, the quantity of electricity demanded during the day increases by less than the quantity demanded at night, but the price falls by much more in the day period than in the night period, due to the steeper slope of the day demand function. Therefore, falling generation costs decrease the price difference between day and night. Figure 3.1.B, however, shows that if demand during the day is significantly more responsive to price than demand at night, the decrease in the slope of the generation cost curve can increase the price difference between day and night.

Figures 3.1.C and 3.1.D show the effect of an increase in the quantity of generation that can be generated at zero cost, for instance an increase in wind generation. As the cost of generating electricity from other fuels does not change, this shifts the cost curve to the right. This reduces the cost of electricity for a given quantity and therefore affects the day-night electricity price difference in a similar way as a decrease in fuel costs. If demand is significantly more responsive to price in the day period than in the night period, the day-night price difference will grow, as in Figure 3.1.D, while a day demand function with a steeper slope that intersects the x-axis at the same quantity as the night demand function causes the day-night price difference to fall, as in Figure 3.1.C. However, this effect is also dependent on the position of the day demand curve relative to the night demand curve. For instance, shifting the day demand curve right in Figure 3.1.C, causes a larger day-night price difference. Thus, the effect of a reduction in the cost of generating electricity, either due to a decrease in fuel costs or an increase in the capacity of zero (or simply low) marginal cost generation, on the value of storage is difficult to predict.

Further, from equation (3.4), the value of storage is dependent not on the raw price difference but on the difference between the price storage will receive during the day less the cost of the energy stored at night, or $P^D - \frac{1}{e}P^N$. Therefore the day price must be $\frac{1-e}{e}$ percent greater than the night price. As the night price grows smaller, the absolute difference between the day and night prices required for storage to be valuable declines. When prices go to zero in the night period, any positive price in the day period will present an opportunity to utilize storage. As a result, greater wind generation in the night period than in the day period will increase in the value of storage and the larger the difference between night and day wind generation, the larger the value of storage.

The effect of changing the day demand curve relative to the night demand curve, while keeping generation costs constant, is much easier to predict, as the night price remains constant. Figures 3.2.A and 3.2.B illustrate the potential effects of increasing the slope of the demand function in the day period relative to the demand function in the night period. In each figure, the solid grey line indicates the original demand function for the day period, which has the same slope as the night demand function. If the slope of

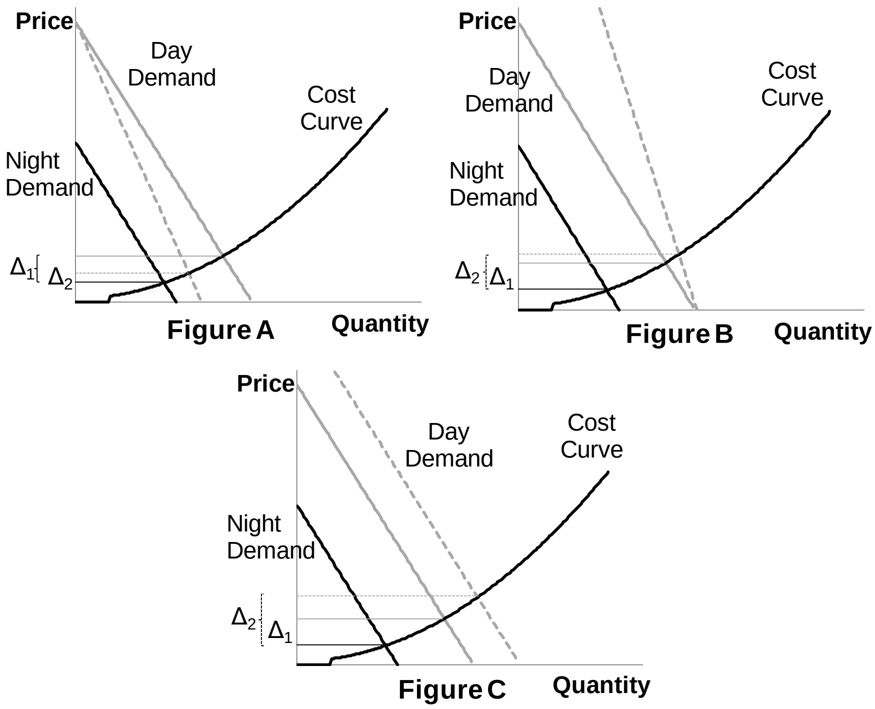


Figure 3.2: Effect of the Demand Functions on the Value of Additional Storage

the day demand curve increases, but the intercept stays constant, the willingness to pay for a given quantity of electricity in the day period falls, and therefore the equilibrium quantity falls, as can be seen in Figure 3.2.A. This reduces the price of electricity in the day period and therefore the day-night price difference. Conversely, if the slope and the intercept of the demand function for the day period increase, as in Figure 3.2.B, the willingness to pay for a given quantity of electricity increases and the equilibrium quantity of electricity increases, increasing the day-night price difference. If the demand curve shifts right, as in Figure 3.2.C, the same result occurs. Thus, if willingness to pay for a given quantity of electricity rises in the day period relative to the night period, the day-night price difference will rise as will the value of storage.

Finally, the cost of storing electricity is most dependent upon the efficiency of the storage device, e. Increasing the efficiency of the storage device decreases the cost of storing energy and therefore decreases the required difference between day and night electricity prices that must occur for storage to be useful. This in turn increases the number of hours in a year that using storage would be profitable and therefore increases the value of adding storage capacity.

Comparing the origins of the value of additional generation capacity versus the value of additional storage capacity, indicates that which investment provides more value to the market depends upon the demand and supply conditions in the market. Additional generation capacity is valuable whenever the market price of electricity is above the marginal cost of the new generator(s) in any period. Conversely, the value of storage depends on the difference between the night and day electricity price. Therefore, including the shoulder period is important to the accuracy of the comparison between the incentive to invest in additional storage versus generation capacity because the shoulder period provides several hours in each day in which the new natural gas combined cycle generator(s) may be operating, while new storage capacity would be idle. Comparing the value of new capacity from generation versus storage in the day and night period, a higher price of electricity at night relative to the price of electricity during the day would decrease the value of storage, but would increase the value of additional generation, while a higher price during the day relative to the night price increases the value of both types of capacity.

On the other hand, increasing fuel costs decrease the quantity of generation demanded and may shift the location of the new combined cycle capacity upward on the supply curve, and therefore may reduce the amount of time new combined cycle capacity would be used. However, as shown in Figure 3.1.A, the increase in the slope of the generation cost curve may cause an increase in the day-night electricity price difference and therefore increase the value of storage. Thus, an increase in fuel costs may decrease the value of new combined cycle generation capacity but increase the value of storage capacity, which would make it more likely for storage capacity to be a better investment than combined cycle capacity.

Increasing the quantity of wind generation shifts the cost curve to right and therefore increases the amount of electricity demanded, but also decreases the equilibrium price of electricity. This causes the return to new combined cycle capacity to fall, regardless of whether the new combined cycle capacity is utilized less in equilibrium due to the increase in generation from wind. The fall in the price of electricity in the night period, however, reduces the absolute difference in price between periods required to make storage useful. This may increase the value of storage, especially if the output of the new wind generation is greater at night than during the day. An increase in the value of storage due to increased wind generation, however, is by no means certain and depends upon demand conditions and the distribution of the wind generation across periods.

3.3.2 Version 2 - Add Demand Uncertainty

In this version of the model, I assume that prior to the beginning of each period, the planner knows only the distribution of the possible demand functions for the upcoming period(s). Therefore, when choosing additional generation or storage capacity prior to the beginning of the night period, the planner does not know the night or day demand functions. In the night period, the planner learns the night period demand function, prior to choosing generation for the night period and the amount of energy to store. In the shoulder period, the planner learns the shoulder period demand function, prior to choosing generation for the shoulder period. In the day period, all uncertainty is resolved before the planner chooses generation for the day period. Although it makes the exposition of the model somewhat more confusing, I replace Q^D, representing daytime generation, with $\hat{Q}^D = Q^D + q_s$, representing daytime consumption, because it makes solving the model more straightforward.

Prior to the beginning of the night period, the planner maximizes the sum of expected utility in each period by choosing the amount of additional generation capacity, X_g, and storage capacity, X_s to build subject to the cost of new capacity, $k_g(X_g)$ and $k_s(X_s)$, and the expected cost of generation in each period:

$$\max_{X_g,X_s} \mathbb{E}_N \Big[\tfrac{1}{e} u^N \big(Q^N(X_g, X_s) - q_s(X_g, X_s) \big) + \mathbb{E}_{Sh|N} \big[(2 - \tfrac{1}{e}) u^{Sh} \big(Q^{Sh}(X_g) \big) \big]$$
$$+ \mathbb{E}_{D|N} \big[u^D \big(\hat{Q}^D(q_s(X_g, X_s), X_g) \big) \big] - \tfrac{1}{e} c\big(Q^N(X_g, X_s), X_g \big)$$
$$- \mathbb{E}_{Sh|N} \big[(2 - \tfrac{1}{e}) c\big(Q^{Sh}(X_g), X_g \big) \big] - \mathbb{E}_{D|N} \big[c\big(\hat{Q}^D(q_s(X_g, X_s), X_g) - q_s(X_g, X_s) \big), X_g \big) \big] \Big]$$
$$- k_g(X_g) - k_s(X_s)$$

The choice of generation in the night period, Q^N, generation in the night period, Q^{Sh}, consumption in the day period, $\hat{Q}^D$, and storage, q_s, will depend upon the choices of generation and storage capacity as well as the realization of demand. $\mathbb{E}_N$ represents the expectation of demand conditions in the night period, while $\mathbb{E}_{Sh|N}$ and $\mathbb{E}_{D|N}$ represents expectation of demand conditions in shoulder and day periods given the demand conditions in the night period.

In the night period, the planner maximizes utility in the night period and expected utility in the day period by choosing the quantity of electricity to generate in the night period, Q^N, and the quantity of electricity to store for use in the day period, q_s, subject to the available generation and storage capacity, the cost of generation in the night period, and the expected cost of generation in day period. The shoulder period is left out of the night period's decisions because there is no storage decision in the shoulder period and therefore the shoulder period is not linked to the night and day periods and does not affect the decisions in these periods.

$$\max_{Q^N,q_s} \tfrac{1}{e} u^N(Q^N - q_s) + \mathbb{E}_D \big[u^D(\hat{Q}^D(q_s, X_g) \big]$$
$$- \tfrac{1}{e} c(Q^N, X_g) - \mathbb{E}_D \big[c(\hat{Q}^D(q_s, X_g) - q_s, X_g) \big]$$

$$\text{s.t.} \quad 0 \le Q^N \le C_g + X_g \qquad \gamma_g^N, \lambda_g^N$$

$$0 \le q_s \le C_s + X_s \qquad \gamma_s, \lambda_s$$

$$0 \le Q^N - q_s \qquad \delta$$

The level of consumption in the day period, $\hat{Q^D}$, will depend on the choice of storage, q_s, made in the night period, as stored electricity can be utilized in the day period at zero marginal cost. The first constraint in the above prevents generation from being negative or greater than the available capacity. The second constrains storage to be positive and prevents storage from exceeding the available capacity. The final constraint prevents the electricity stored from exceeding the amount generated in the night period.

The quantity of generation capacity added affects the solution by increasing the amount of capacity available and potentially decreasing the cost of generation, depending on the where the generation enters the supply curve and the equilibrium quantity of electricity generated in the night period. Indirectly, but via the same mechanism, it also affects the amount of electricity that can be stored. Additional storage capacity increases the amount that can be stored.

In the shoulder period, the planner maximizes utility by choosing the quantity of electricity to be consumed, subject to the cost of generation and the generation capacity constraint.

$$\max_{Q^{Sh}} (2 - \tfrac{1}{e})\big(u^{Sh}(Q^{Sh}) - c(Q^{Sh}, X_g)\big)$$

$$\text{s.t. } 0 \le Q^{Sh} \le C_g + X_g \qquad \gamma_g^{Sh}, \lambda_g^{Sh}$$

The quantity of additional generation capacity enters the problem by reducing the cost of generation and increasing the available capacity. The quantity of additional storage capacity does not enter this problem.

In the day period, the planner maximizes utility by choosing the quantity of electricity to be consumed, subject to the amount of stored energy, the cost of generation, and the generation capacity constraint.

$$\max_{\hat{Q}^D} u^D(\hat{Q}^D) - c(\hat{Q}^D - q_s, X_g)$$

$$\text{s.t. } 0 \leq \hat{Q}^D \leq C_g + X_g + q_s \qquad \gamma_g^D, \lambda_g^D$$

Energy storage enters the problem for the day period in two ways. First it increases the amount of generation available at zero marginal cost, therefore shifting the cost curve right and reducing costs. Second, it increases the maximum possible level of consumption. Here consumption is constrained to be positive, $\hat{Q}^D \geq 0$, but the planner is not required to use all of the stored energy. Therefore, $\hat{Q}^D - q_s$, may be negative. In the case that $\hat{Q}^D - q_s$ is negative, the cost of generation, $c(\hat{Q}^D - q_s, X_g)$, is assumed to be zero. The quantity of added storage capacity enters the problem only through the amount of energy stored. The quantity of added generation capacity enters the problem through the amount of energy stored, as it may decrease the cost of generating electricity to store and increase the possible amount that can be stored. It also enters the problem by increasing the amount of electricity that can be generated in the day period and possibly reducing the cost of generation.

3.3.2.1 Analytical Results

To find the optimal generation, storage, and additional capacity to build, the planner solves the problem by backward induction. In the day period, the first order condition for the quantity of electricity to consume is:

$$\mathbf{\hat{Q}^D}\text{: } u^{D\,\prime}(\hat{Q}^D(q_s, X_g)) = c_1(\hat{Q}^D(q_s, X_g) - q_s, X_g) + \lambda_g^D - \gamma_g^D \tag{3.9}$$

First, note that if consumption in the day period, $\hat{Q}^D$, were replaced with generation in the day period, Q^D, this expression would be identical to the expression in the model without uncertainty, equation (3.3). Second, as the marginal cost of utilizing the energy from storage is zero, the marginal cost for the first q_s units of electricity is zero. By assumption 3.3, demand in the day period is never so low that the marginal utility of zero consumption, $u^{D\,\prime}(0)$, is less than zero. Therefore the non-negativity constraint on consumption, with Lagrange multiplier γ_g^D, never binds, and γ_g^D is always zero.

The night period decisions and the choice of additional storage and generation capacity will depend upon the first derivative of day period consumption, $\hat{Q}^D(q_s, X_g)$ with respect to q_s and X_g. In forming the cost curve, I assume that it is possible that there is a jump in cost between the zero-marginal cost generation and the costly generation. If there is a discontinuity in the cost curve at this point, which I label C_{g_0}, the capacity constraint, λ_g^D, will bind and the quantity of electricity consumed, $\hat{Q}^D(q_s, X_g)$, will be the sum of the zero marginal cost generation and the amount of electricity stored, $C_{g_0} + q_s$. Therefore, if λ_g^D binds, the quantity of electricity consumed will either be the sum of zero marginal cost generation and electricity stored, $C_{g_0} + q_s$, or the sum of all generation capacity and the quantity of electricity stored, $C_g + X_g + q_s$. In either case, the derivative of $\hat{Q}^D(q_s, X_g)$ with respect to storage, q_s, will be 1, while the derivative of $\hat{Q}^D(q_s, X_g)$ with respect to additional generation capacity, X_g, will be 0 in the first case and 1 in the second case. If λ_g^D does not bind, the derivatives of $\hat{Q}^D(q_s, X_g)$ cannot be determined without further assumptions.

In the shoulder period, the first order condition for the quantity of electricity to consume is:

$$\mathbf{Q^{Sh}}\text{: } (2 - \tfrac{1}{e})u^{Sh\,\prime}(Q^{Sh}(X_g)) = (2 - \tfrac{1}{e})c_1(Q^{Sh}(X_g), X_g) + \lambda_g^{Sh} - \gamma_g^{Sh} \tag{3.10}$$

This expression is identical to the expression in the model without uncertainty, equation (3.2). Again, by assumption 3.3, it will never be optimal to consumer zero electricity, and therefore γ_g^{Sh} is always zero. The choice of additional generation capacity will depend on first derivative of $Q^{Sh}(X_g)$ with respect to the additional generation capacity, X_g. When λ_g^{Sh} is positive, generation, $Q^{Sh}(X_g)$, is either C_{g_0}, the quantity of zero marginal cost generation, or $C_g + X_g$, the total quantity of generation available. In the first case, the derivative of $Q^{Sh}(X_g)$ will be 0 and in the second case it will be 1. If λ_g^{Sh} is zero, the derivative is unknown.

In the night period, the first order conditions of the maximization problem are:

$$\mathbf{Q^N}\text{: } \tfrac{1}{e}u^{N\,\prime}(Q^N(X_g,X_s) - q_s(X_g,X_s)) = \tfrac{1}{e}c_1(Q^N(X_g,X_s),X_g) + \lambda_g^N - \gamma_g^N - \delta \quad (3.11)$$

$$\begin{aligned}\mathbf{q_s}\text{: } \lambda_s - \gamma_s + \delta = \mathbb{E}_D\big[&u^{D\,\prime}(\hat{Q}^D(q_s(X_g,X_s),X_g))\hat{Q}_1^D(q_s(X_g,X_s),X_g) \\ &- c_1(\hat{Q}^D(q_s(X_g,X_s),X_g) - q_s(X_g,X_s))(\hat{Q}_1^D(q_s(X_g,X_s),X_g) - 1)\big] \\ &- \tfrac{1}{e}u^{N\,\prime}(Q^N(X_g,X_s) - q_s(X_g,X_s)) \end{aligned} \quad (3.12)$$

Note that equation for the choice of generation in the night period, Q^N, is identical to the expression in the model without uncertainty, equation (3.1). With a little work, the equation determining the choice of q_s will also become very similar to that seen in the first version of the model, equation (3.4).

The first order condition for the choice of storage can be rewritten as:

$$\begin{aligned}&\lambda_s - \gamma_s + \delta = \\ &\quad \mathbb{E}_D\Big[\big(u^{D\,\prime}(\hat{Q}^D(q_s(X_g,X_s),X_g)) - c_1(\hat{Q}^D(q_s(X_g,X_s),X_g) - q_s(X_g,X_s))\big)\hat{Q}_1^D(q_s(X_g,X_s),X_g) \\ &\quad + c_1(\hat{Q}^D(q_s(X_g,X_s),X_g) - q_s(X_g,X_s))\Big] - \tfrac{1}{e}u^{N\,\prime}(Q^N(X_g,X_s) - q_s(X_g,X_s))\end{aligned} \quad (3.13)$$

Using the observations from the night period, the first term on the right-hand side of equation (3.13) can be replaced with a simpler expression. The expectation operator can be thought of as summing up all the possible cases that could occur in the day period. Referring to the term $u^{D\,\prime}(\hat{Q}^D(q_s(X_g,X_s),X_g)) - c_1(\hat{Q}^D(q_s(X_g,X_s),X_g) - q_s(X_g,X_s))$ as A, in the case that neither constraint binds, $A = 0$. In the case that λ_g^D binds, A can be replaced with λ_g^D based upon equation (3.9). Also, as discussed above, the derivative of consumption in the day period with respect to storage, $\hat{Q}_1^D(q_s, X_g)$, is 1. Therefore, equation (3.13) can be rewritten as:

$$\begin{aligned}\lambda_s - \gamma_s + \delta =& \mathbb{E}_D\Big[\lambda_g^D + c_1(\hat{Q}^D(q_s(X_g, X_s), X_g) - q_s(X_g, X_s))\Big] \\ &- \tfrac{1}{e} u^{N\,\prime}(Q^N(X_g, X_s) - q_s(X_g, X_s))\end{aligned} \tag{3.14}$$

Replacing $\lambda_g^D + c_1(\hat{Q}^D(q_s(X_g, X_s), X_g) - q_s(X_g, X_s))$ with $u^{D\,\prime}(\hat{Q}^D(q_s(X_g, X_s), X_g))$ using equation (3.9), would yield an expression that matches that in the model without uncertainty, equation (3.4), although there is now an expectation over the day period marginal utility function. Thus, the equations determining the choice of generation and storage are nearly identical to the model without uncertainty. The effect of the uncertainty is simply to change the expected marginal utility and costs in the day period when viewed from the night period.

To choose the optimal generation and storage capacity additions, it is necessary to determine the first derivatives of Q^N and q_s with respect to additional generation and storage capacity, X_g and X_s. I assume that it will never be the case that generation in the night period will be zero, which means that the non-negativity constraint for generation at night will never bind and γ_g^N will always be zero. If no electricity was generated in the night period, no energy could be stored in the night period. This assumption is slightly different than assuming that consumption in the night period will never be zero. Consumption in the night period can be zero even if generation is not zero if all generation is stored for the day period. If the marginal cost for the energy stored is greater than the marginal utility of zero consumption, consumption will be zero. In this case, the third constraint in the night period maximization problem binds and δ is positive but, γ_g^N is zero.

If the capacity constraint on generation in the night period, λ_g^N, is positive, either all zero marginal cost generation is utilized but no other resources are utilized, or all available generation is utilized. In the first case, the quantity of generation in the night period, $Q^N(X_g, X_s)$ would be C_{g0}, in the second case, $Q^N(X_g, X_s)$ would be $C_g + X_g$.

Therefore, $Q_1^N(X_g, X_s)$, the derivative with respect to additional generation is 0 in the first case and 1 in the second. The derivative with respect to storage, $Q_2^N(X_g, X_s)$, would be zero in both cases. If neither λ_g^N nor δ is positive the derivatives are unknown.

If δ is positive, then storage in the night period equals the amount of generation, $q_s(X_g, X_s) = Q^N(X_g, X_s)$. Therefore the derivative of q_s with respect to additional capacity, X_g or X_s, will be $Q_1^N(X_g, X_s)$ and $Q_2^N(X_g, X_s)$ respectively. Secondly, it is possible that δ and λ_g^N could bind at the same time. If both bind, then the derivatives of $Q^N(X_g, X_s)$ are as before. Otherwise the derivatives are unknown.

If λ_s is positive, storage is filled to capacity or $q_s(X_g, X_s) = C_s + X_s$. Therefore, the derivative of $q_s(X_g, X_s)$ with respect to additional generation capacity, X_g, is 0 and the derivative with respect to additional storage capacity, X_s, is 1. If γ_s is positive, storage is not utilized and $q_s(X_g, X_s) = 0$. This causes the derivatives of $q_s(X_g, X_s)$ with respect to X_g and X_s to be zero. If neither λ_s, γ_s, nor δ are positive, the derivatives are unknown.

Finally, the first order conditions for the planner's choice of generation and storage capacity are (omitting the variables that Q^N, Q^{Sh}, q_s and $\hat{Q}^D$ depend on for brevity and legibility, and indicating the derivative with respect to the first argument with a subscript 1 and the derivative with respect to the second argument with a subscript 2):

$$\begin{aligned}\mathbf{X_g}\text{: } k_g'(X_g) =& \mathbb{E}_N\Big[\tfrac{1}{e}\big(u^{N\prime}(Q^N - q_s) - c_1(Q^N, X_g)\big)Q_1^N - \tfrac{1}{e}c_2(Q^N, X_g) \\ &+ \mathbb{E}_{D|N}\big[c_1(\hat{Q}^D - q_s, X_g) - \tfrac{1}{e}u^{N\prime}(Q^N - q_s)\big]q_{s1} \\ &+ \mathbb{E}_{D|N}\big[\big(u^{D\prime}(\hat{Q}^D) - c_1(\hat{Q}^D - q_s, X_g)\big)(\hat{Q}_1^D q_{s1} + \hat{Q}_2^D) - c_2(\hat{Q}^D - q_s, X_g)\big]\Big] \\ &+ \mathbb{E}_{Sh|N}\big[(2 - \tfrac{1}{e})\big(u^{Sh\prime}(Q^{Sh}) - c_1(Q^{Sh}, X_g)\big)Q_1^{Sh} - (2 - \tfrac{1}{e})c_2(Q^{Sh}, X_g)\big]\Big]\end{aligned} \tag{3.15}$$

$$\begin{aligned}\mathbf{X_s}\text{: } k_s'(X_s) =& \mathbb{E}_N\Big[\tfrac{1}{e}\big(u^{N\prime}(Q^N - q_s) - c_1(Q^N, X_g)\big)Q_2^N \\ &+ \mathbb{E}_{D|N}\big[c_1(\hat{Q}^D - q_s, X_g) - \tfrac{1}{e}u^{N\prime}(Q^N - q_s)\big]q_{s2} \\ &+ \mathbb{E}_{D|N}\big[\big(u^{D\prime}(\hat{Q}^D) - c_1(\hat{Q}^D - q_s, X_g)\big)\hat{Q}_1^D q_{s2}\big]\Big]\end{aligned} \tag{3.16}$$

Thinking of the expectation operators as summing all the possible cases together, using the derivatives computed above, where known, and substituting from equations (3.11), (3.14), (3.9), and (3.10), these expressions can be simplified considerably. The first term in each equation, $\frac{1}{e}u^{N\,\prime}(Q^N - q_s) - \frac{1}{e}c_1(Q^N, X_g)$ equals $\lambda_g^N - \delta$, by equation (3.11). In the second line of each equation, $\lambda_s - \gamma_s + \delta - \mathbb{E}_{D|N}[\lambda_g^D]$ can replace $\mathbb{E}_{D|N}[c_1(\hat{Q}^D - q_s, X_g) - \frac{1}{e}u^{N\,\prime}(Q^N - q_s)]$, by equation (3.14). In the third line of each equation, $u^{D\,\prime}(\hat{Q}^D) - c_1(\hat{Q}^D - q_s, X_g)$ can be replaced with λ_g^D using equation (3.9). In the fourth line of the first equation, $(2 - \frac{1}{e})\big(u^{Sh\,\prime}(Q^{Sh}) - c_1(Q^{Sh}, X_g)\big)$ can be replaced with λ_g^{Sh}, by equation (3.10). Plugging in the derivatives, where known, yields the following simplified forms for the first order conditions of X_g and X_s:

$$\begin{aligned}
\mathbf{X_g}\colon\ k_g'(X_g) =& \mathbb{E}_N\Big[\lambda_g^N - \delta Q_1^N - \tfrac{1}{e}c_2(Q^N, X_g) \\
& - \mathbb{E}_{D|N}\big[\lambda_g^D q_{s1}\big] + \delta Q_1^N \\
& + \mathbb{E}_{D|N}\big[\lambda_g^D(q_{s1}+1) - c_2(\hat{Q}^D - q_s, X_g)\big] \\
& + \mathbb{E}_{Sh|N}\big[\lambda_g^{Sh} - (2 - \tfrac{1}{e})c_2(Q^{Sh}, X_g)\big]\Big] \qquad (3.17) \\
\mathbf{X_s}\colon\ k_s'(X_s) =& \mathbb{E}_N\Big[-\delta Q_2^N \\
& + \lambda_s - \mathbb{E}_{D|N}\big[\lambda_g^D q_{s2}\big] + \delta Q_2^N \\
& + \mathbb{E}_{D|N}\big[\lambda_g^D q_{s2}\big]\Big] \qquad (3.18)
\end{aligned}$$

Noting that several terms cancel, these expressions can be simplified further:

$$\begin{aligned}
\mathbf{X_g}\colon\ k_g'(X_g) =& \mathbb{E}_N\big[\lambda_g^N - \tfrac{1}{e}c_2(Q^N, X_g) + \mathbb{E}_{D|N}[\lambda_g^D - c_2(\hat{Q}^D - q_s, X_g)] \\
& + \mathbb{E}_{Sh|N}[\lambda_g^{Sh} - (2 - \tfrac{1}{e})c_2(Q^{Sh}, X_g)]\big] \qquad (3.19) \\
\mathbf{X_s}\colon\ k_s'(X_s) =& \mathbb{E}_N\big[\lambda_s\big] \qquad (3.20)
\end{aligned}$$

With the exception of the expectation operators and the change in variable from generation in the day period to consumption in the day period, these expressions are identical to the expressions found in the certainty version of the model, equations (3.5) and (3.6). Consequently the decision to invest in additional generation and/or storage capacity occurs for the same reasons, although uncertainty may raise or lower the values of each term.

3.3.3 Version 3 - Add Wind Generation Uncertainty

In the third version of the model, I add uncertainty about how much generation will be produced by wind in each period. The uncertainty about the wind generation is resolved at the same time as uncertainty about demand. As a result, the equations derived for version 2 of the model hold equally under version 3, although the expectation is now over possible realizations of wind generation as well as (or instead of) realizations of demand. Wind generation is assumed to have zero marginal cost and therefore shifts the marginal cost curve laterally as it varies.

3.4 Numerical Simulation

To understand how the interlacing forces of supply and demand affect the value of additional combined cycle capacity relative to the value of additional storage capacity, I numerically simulate a variety of possible situations in the Pennsylvania Jersey Maryland (PJM) Interconnection market. The PJM Interconnection market covers many of the Mid-Atlantic and Midwestern U.S. states, extending up the coast Virginia to Pennsylvania and across to Ohio and parts of Illinois and Indiana. The market is operated by an independent system operator (ISO) which operates a competitive wholesale electricity market and manages the high-voltage electricity grid to ensure the reliable movement of power from generators to customers. PJM was the first ISO in the country and has one of the most easily accessible and reliable datasets on market operations of any electricity market in the country.[4]

For the purposes of this study I focus on pumped hydroelectric storage technology, which is the most mature and widely-used storage technology as well as the simplest to model. Unfortunately, I was able to obtain only a range for the potential average cost

[4] `http:\www.pjm.com`

per mega-watt (MW) of pumped storage capacity for small units and a range for large units. Additionally, because the cost of pumped storage capacity is highly dependent on the location and only three pumped storage facilities have been built in the U.S. in the last 20 years, compared with 370 new combined cycle generation facilities, costs are highly uncertain.[5] As a result, I do not have enough information to trace out the function governing the capital cost of installing new storage capacity.

I therefore estimate the relative value of adding a single 400 MW combined cycle generator versus adding a storage facility with the capacity to store or generate 400 MWh of electricity per hour, and compare the value against the capital cost of building the storage or generation capacity. The 400 MW size was chosen because this is the size of the most efficient combined cycle generator for which I have capital cost information.

To compute the value to the market of adding new generation or storage capacity, I begin by parametrizing the model and estimating the value of each parameter. I then calculate the market surplus given existing generation and storage capacity and the market surplus with the additional 540 MW of generation or storage capacity. Computing the difference between the scenario with additional generation or storage capacity with the scenario with only the pre-existing generation and storage capacity yields the increase in market surplus due to the new capacity and therefore the value to the market of the new capacity.

3.4.1 Parametrization

For ease of exposition and solution, I assume that marginal utility (demand) is linear. Specifically:

$$\begin{aligned} u^{N\,\prime}(Q) &= a^N + b^N Q \\ u^{Sh\,\prime}(Q) &= a^{Sh} + b^{Sh} Q \\ u^{D\,\prime}(Q) &= a^D + b^D Q \end{aligned}$$

[5] Energy Information Administration, 2010 Form 860

Here $a^t > 0$ and $b^t < 0$, where t indicates the night (N), shoulder (Sh), or day (D) period. The price, P, of a given quantity of electricity, Q, is defined as the marginal utility at Q units of electricity. To compute the intercepts, a^t, and the slopes, b^t, of the demand functions, I begin with an estimated elasticity of demand, ε, and a baseline price and quantity demanded, P_0^t and Q_0^t, for the night, shoulder and day periods. I assume the same elasticity in all periods. The definition of the elasticity of supply or demand is: the percent change in quantity over the percent change in price or:

$$\varepsilon = \frac{\frac{dQ^t}{Q_0^t}}{\frac{dP^t}{P_0^t}} = \frac{dQ^t}{dP^t}\frac{P_0^t}{Q_0^t}$$

dP^t and dQ^t represent change in price and quantity from the baseline prices and quantities, P_0^t and Q_0^t. As $\frac{dQ^t}{dP^t}$ is also known as the derivative of quantity with respect to price, rearranging $P^t = a^t + b^t Q^t$ taking the derivative of Q^t with respect to P^t yields, $\frac{dQ^t}{dP^t} = \frac{1}{b^t}$. Plugging this fact into the above equation and solving for b^t gives the slope of the demand function in terms of the estimated elasticity and baseline prices and quantities:

$$b^t = \frac{P_0^t}{\varepsilon Q_0^t} \tag{3.21}$$

As the baseline prices and quantities are on the demand curve, plugging P_0^t, Q_0^t and the formula for b^t into $P^t = a^t + b^t Q^t$ and solving for a^t yields the intercept of the demand function:

$$a^t = P_0^t - b^t Q_0^t = P_0^t - \frac{P_0^t}{\varepsilon} \tag{3.22}$$

To model uncertainty about demand in the day period, I assume that the percentage difference between the day and night price varies according to a mean zero normal distribution with an estimated variance, or $P^D = P^N(1 + \rho + \mu)$. Here ρ is the average night-day price difference computed from the data and μ is the error term. As I assume that the marginal utility of zero consumption is always positive, I restrict the percentage difference, $\rho + \mu$, to be greater than -1, which eliminates at most 0.3 percent of the distribution. The prices given each value of μ and equation (3.22) are then used to compute the intercept of the demand function for the day period, causing the intercept of the demand function to vary stochastically while the slope stays constant. As the shoulder period exhibits significantly lower variance, and to keep the solution of the model manageable, I assume that shoulder period demand does not vary stochastically.

Because a shift upward (or downward) of the demand curve given a constant supply curve induces higher (or lower) equilibrium electricity consumption, $\hat{Q}^D$, the demand curve will intersect each portion of the supply curve for a particular range of μ. For instance, $\hat{Q}^D$ will be constrained by $C_g + X_g + q_s$ for high μ leading λ_g^D to be greater than zero. As μ decreases, the marginal cost of $\hat{Q}^D$ will decline and intersect the main body of the cost function. When the demand function has a very low intercept, it will cross the cost function in the zero marginal cost range. λ_g^D will also bind at any discontinuities in the cost curve until μ declines sufficiently to reach the next section of the cost curve.

The marginal cost function is assumed to be piecewise linear such that:

$$c_1(Q) = \begin{cases} 0 & \text{if } 0 \leq Q \leq C_{g_0} \\ \alpha_L + \beta_L(Q - C_{g0}) & \text{if } C_{g_0} < Q \leq C_{g_1} \\ \alpha_H + \beta_H(Q - C_{g_1}) & \text{if } C_{g_1} < Q \leq C_g \end{cases} \tag{3.23}$$

C_{g_0} represents the quantity of zero marginal cost generation. The next section of the marginal cost function is upward sloping, $\beta_L > 0$, and the intercept $\alpha_L - \beta_L C_{g_0}$ is greater than or equal to zero. If the intercept is positive, there will be a discontinuity in the marginal cost curve at C_{g_0}. The third section of the cost curve, in all but one of the cases considered, has a steeper slope, $beta_H > 0$, and, in all cases, intersects the second section of the marginal cost curve with no discontinuity at C_{g_1}. C_g is the total available generation in the market.

If additional generation capacity is added, the marginal cost function becomes:

$$c_1(Q) = \begin{cases} 0 & \text{if } 0 \leq Q \leq C_{g_0} \\ \alpha_L + \beta_L(Q - C_{g_0}) & \text{if } C_{g_0} < Q \leq B \\ m & \text{if } B < Q < B + X_g \\ \alpha_L + \beta_L(Q - C_{g_0} - X_g) & \text{if } B + X_g \leq Q \leq C_{g_1} + X_g \\ \alpha_H + \beta_H(Q - C_{g_1} - X_g) & \text{if } C_{g_1} + X_g < Q \leq C_g + X_g \end{cases} \tag{3.24}$$

B is the quantity at which the second section of the marginal cost curve, $\alpha_L+\beta_L(Q-C_{g_0})$, reaches the marginal cost of the new generation capacity, m. X_g is the quantity of new generation. In all cases considered, the marginal cost of the new generation intersects the cost curve in this lower portion of the cost curve.

As wind generation varies, the quantity of zero marginal cost generation varies, C_{g_0}. To incorporate uncertainty about wind generation in the model, I assume that the capacity factor of wind generation can assume either a high or low value. The capacity factor is the share of total capacity being utilized to generate electricity, e.g. if a wind turbine has a capacity of 10 MW and is currently generating 5 mega-watt hours (MWh) of electricity, the capacity factor is 50 percent. The probability of low or high wind in the day period depends on the whether the wind in the night period is low or high to account for the observed persistency in the level of wind generation across periods.

3.4.2 Data

For the baseline scenarios, I utilize the generation portfolio of PJM in 2011. The capacity of existing storage facilities in the PJM market was computed from PJM's 2010 Form 411 submitted to the Energy Information Administration (EIA). The demand and cost functions are generated using electricity price and demand data for 2008, 2011, and 2012 from PJM. The probability and level of high and low wind generation in each period are calculated from data on hourly wind generation in PJM during 2011. Capital costs for new storage facilities come from Rastler (2010) and capital costs for new combined cycle generators come from EIA (2010).

3.4.2.1 Storage Capacity and Costs

The existing storage facilities in the PJM market can generate approximately 5,000 MWh of electricity per hour (or store 5,000 MWh of electricity in an hour). The ability to generate 5,000 MWh of electricity per hour is also known as having 5,000 MW of generation power. This information comes from the 2010 Form 411 submitted by PJM to the EIA which contains a list of generators and storage facilities that participate in the PJM market. Rastler (2010), a white paper from the Electric Power Research Institute on storage technology, indicates that the typical pumped storage facility has an efficiency of 80 percent, 8 hours of storage capacity, and a lifetime on the order of 50 years. The cost of a storage facility, according to the report, ranges from 2,500-4,300 $/kW for a 280-530 MW facility to 1,500-2,700 $/kW for a 900-1,400 MW facility. Therefore, I assume that the added storage facility has the capacity to store or generate 400 MWh of electricity per hour, and therefore has 400 MW of power, and has the capacity to store 8 hours worth of electricity or 3200 MWh. Assuming an 80 percent efficiency it requires 10 hours to fill the 8 hours of storage capacity. The minimum cost of the 400 MW storage facility is 2,500 $/kW, or $1 billion.

3.4.2.2 Generation Capacity and Costs

The baseline generation capacity utilized in the model comes from PJM's capacity by fuel type information sheet for 2011. The data indicate that there was 178,172 MW of installed generation capacity in PJM in 2011, making adjustments for wind capacity as explained in Section 3.4.2.3. Of this capacity, 42,707 MW consisted of zero marginal cost generation capacity such as nuclear, hydroelectric, solar, and wind. The remaining capacity consisted of coal, natural gas, and various petroleum and waste products. In the spring/fall season, a significant portion of generators are removed from service for maintenance, as demand is generally quite low relative to winter and summer in the PJM region. The hourly demand data for 2011 indicate that the maximum load in spring/fall is approximately 4/5 of summer demand. I therefore assume that the available capacity in PJM is 4/5 of the total capacity in spring/fall (therefore reducing the zero-marginal cost generation by 4/5 as well).

To estimate the marginal cost function for the baseline scenario, I utilize hourly data from PJM on prices and demand for 2011. As demand for electricity is extremely inelastic and the PJM market operator is tasked with serving demand at minimum cost given generators' supply bids, the price roughly reflects the marginal cost of generation. However, the data contains two sets of prices for the PJM market for each hour. The first set of prices is the day-ahead price, which is computed at noon on the day prior given the bids posted by generators and expected demand. The second set is the real-time price, which reflects actual load and system operating conditions. As a result of transmission constraints, unexpected generator outages and other system operations issues, the real-time price does not clearly reflect the usual marginal cost of generation. I therefore utilize the day-ahead prices to calculate the marginal cost function.

As costs vary across the seasons due to the variation in demand, I compute a separate marginal cost function for each season. I begin by rounding demand to the nearest thousand and computing the average price for each category. The data for 2011 exhibit a convex cost function that is well-approximated by a two-part linear function. I therefore set an initial threshold quantity and utilize ordinary least squares (OLS) to compute the best linear approximation of the data below the threshold quantity and of the data above the threshold. I then minimize the mean-squared error of the approximation by adjusting the threshold quantity and recomputing the OLS estimate for each portion of the cost function. As originally estimated, the winter marginal cost curve crosses below zero at a quantity of generation greater than the assumed amount of zero marginal cost generation (42,707 MW) in the market, yielding a small negative portion of the marginal cost curve. To prevent this, I increase the intercept such that the marginal cost of 42,707 MWh of generation is zero, which increases the intercept by approximately $0.107 or 0.23

percent. Figure 3.3 illustrates the observed cost function for each season in 2011 versus the estimated cost function.

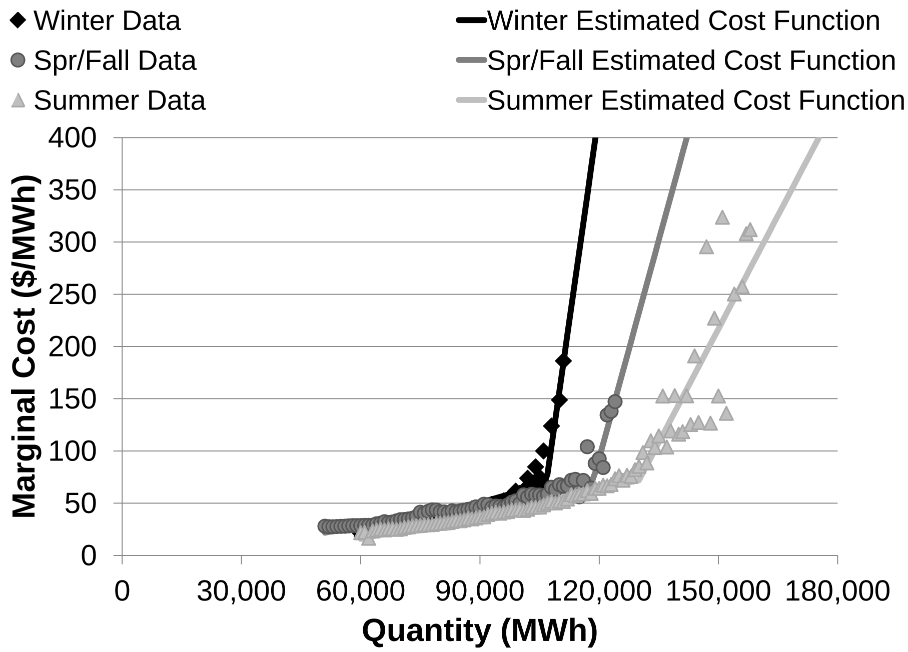
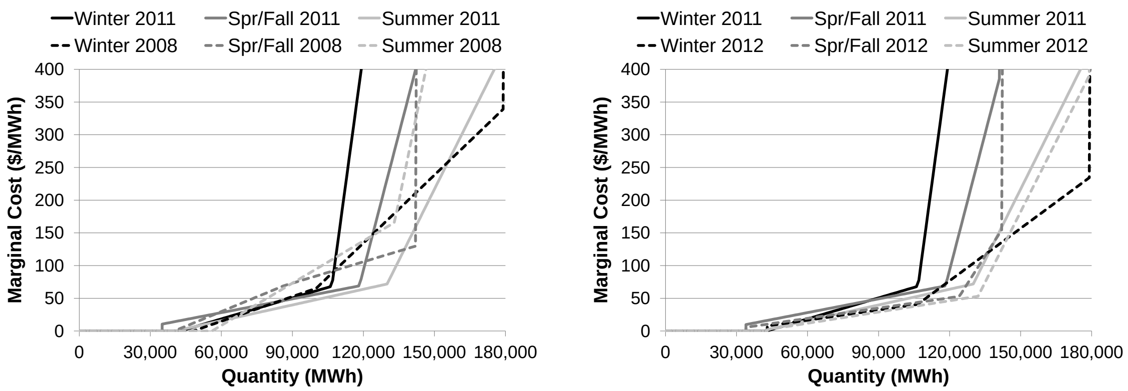


Figure 3.3: Observed and Estimated Cost Functions for 2011

To examine the effect of fuel costs on the relative value of storage versus new combined cycle generation capacity, I recompute the marginal cost functions using 2012 data, in which natural gas prices fell considerably against coal prices, and using 2008 data, the year in which natural gas prices peaked and in which coal and petroleum prices were relatively low. However, to account for the change in generation capacity from 2008 to 2011 to 2012, I stretch the curves over a proportionally larger capacity for 2008 and shrink the curve to cover a proportionally smaller capacity for 2012. To do this, I begin by utilizing the 2008 and 2012 capacity by fuel type information sheets, adjusting for wind capacity, to determine the total quantity of generation capacity in the market in each year and the quantity of zero marginal cost generation. I then calculate the ratio of 2008 or 2012 non-zero marginal cost capacity to 2011 non-zero marginal cost capacity. I assume that each portion of the curve stretches out from the same minimum price to the same maximum price but across the initial quantity multiplied by the ratio of 2008 or 2012 capacity to 2011 capacity. For the lower portion of the curve, the minimum price is the cost (according to the equation for the lower part of the cost curve) at the quantity of zero-marginal cost generation in that year and season, and the maximum price is the price at which the lower portion of the curve intersects the upper portion. For the upper portion of the curve, the minimum price is the price at the intersection of the two parts of the cost curve, and the maximum price is the cost of generating at full capacity.

While this methodology yields no problems for 2012, the 2008 data exhibit more curvature which causes the linear approximation to cross below zero at quantities greater than the assumed quantity of zero marginal cost generation, which would cause the cost curve to have a negative section after the zero marginal cost generation. To rectify this issue with as little damage as possible to the fit of the marginal cost function to the cost data, I increase the quantity of zero marginal cost generation. For winter, the quantity of zero marginal cost generation must increase by 5,435 MW (13 percent), for spring/fall it must increase by 5,938 (17 percent), and for summer it must increase by 13,322 MW (31 percent). Additionally, to generate a better match to the rest of the data, three outlying observations were excluded in estimating the winter 2008 marginal cost curve. Interestingly, the data indicate that the cost curve was somewhat concave for spring/fall in 2008 and therefore the slope of the upper portion of the cost curve is smaller than the slope for the lower portion.



(a) 2011/2008 Comparison (b) 2011/2012 Comparison

Figure 3.4: Comparison of Estimated Marginal Cost Functions

Figure 3.4 illustrates the marginal cost functions estimated for 2008 and 2012 as they compare to the 2011 marginal cost functions. The vertical sections indicate that the capacity constraint has been met. The 2008 curves fall below the 2011 curves at low levels of generation because coal prices were lower, which (up until 2012 at least) provide the cheapest fossil-fueled generation. The middle sections of the marginal cost functions for 2008 are situated above the 2011 curves because of the much higher price of natural gas in 2008. At high levels of generation, it is likely that the 2008 marginal cost functions indicate lower marginal costs than the 2011 marginal cost functions because prices for petroleum liquids, which are utilized by the highest cost generators, were lower. In 2012, the low natural gas price causes the marginal cost curves for 2012 to fall below the marginal cost curves for 2011. For reference, Table 3.1 shows the average price of coal, natural gas, and petroleum liquids paid by the electric power sector in the PJM region for a representative month in each year. These data are found in the EIA's Electric Power Monthly publication.

Fuel	**Oct-08**	**Apr-11**	**Apr-12**
Coal	$2.43	$2.76	$2.76
Natural Gas	$8.19	$5.32	$3.29
Petroleum Liquids	$14.11	$20.33	$23.67

Table 3.1: Average Price of Coal, Natural Gas, and Petroleum Liquids Paid by the Electric Power Sector in the PJM Region

To determine the marginal cost of new combined cycle generation capacity for 2008, 2011, and 2012, I compiled the prices of natural gas and total receipts of natural gas for electric power generation by region for each month of 2008, 2011, and 2012 from the EIA's Electric Power Monthly publication. To compute the weighted average cost of natural gas to the electric power sector in PJM, I compiled the prices and total receipts of natural gas for the Mid-Atlantic, South Atlantic, and East North Central regions which contain the states within the PJM region. I then computed the weighted average cost of natural gas for winter, spring/fall, and summer in each year. The calculated prices can be found in Table 3.2.

The cost of electricity generated by a natural gas combined cycle unit is determined

Season	2008	2011	2012
Winter	$9.35	$5.45	$3.95
Spring/Fall	$9.61	$5.04	$3.34
Summer	$11.84	$5.35	$3.48

Table 3.2: Weighted Average Price of Natural Gas Paid by the Electric Power Sector in the PJM Region

by the efficiency of the unit. Consulting the EIA's *Updated Capital Cost Estimates for Electricity Generation Plants* published in November 2010, the most efficient available combined cycle unit has an efficiency of 6.43 mmBtu/MWh (million British thermal units per MWh), which is the rate at which 1 mmBtu of fuel can be converted into 1 MWh of electricity. The price of natural gas is given in $/mmBtu. Multiplying the natural gas price by the efficiency of the combined cycle generator and adding variable operations and maintenance costs of $3.11/MWh, also from EIA (2010), yields the marginal cost of generation assumed for new combined cycle generation capacity. According to EIA (2010), the capital cost of the most efficient combined cycle generator available is approximately $1,003/kW of capacity, yielding a total cost of $401.2 million for a new 400 MW combined cycle generator. The average lifespan of a combined cycle generator is 30 years.

3.4.2.3 Wind Generation

Information on wind generation comes from PJM's hourly wind generation database for 2011. I assume that wind generation is uncorrelated with demand conditions. I begin by separating the hours of each day into night, day, and shoulder periods. As I assume that storage has 8 hours of capacity and requires 10 hours to charge, the shoulder period constitutes 6 hours of the day. Consulting the average price by hour for 2011 from PJM's day-ahead price data, I assign 10 pm-midnight on the day prior and midnight to 8 am to the night period as these are the 10 lowest priced hours on average. The hours 1 pm to 9 pm are assigned to the day period, as these hours have the highest average prices, and the remaining hours, 8 am to 1 pm and 9 pm to 10 pm, are assigned to the shoulder period.

Finding that the maximum generation from wind in PJM in a single hour in 2011 was 3,960 MWh, I assume the capacity of wind generation was 4,000 MW. Dividing the hourly wind generation by the 4,000 MW of capacity yields the hourly capacity factor. I then compute the average capacity factor for the night, shoulder, and day period of each day. Finding that the average capacity factor across the year is roughly the same for each period, I set the capacity factor for the low and high capacity factor categories to the same value for each season. The value for the low capacity factor category is the average capacity factor over all hours in 2011 when the capacity factor was less than 50 percent, and the value for the high category is the average capacity factor when the capacity factor was greater than 50 percent. Thus, the low category has a capacity factor of 22 percent, and the capacity factor of the high category is 65 percent.

Next, I compute the probability of having a high or low capacity factor in the night period, the probability of having a high or low capacity factor in the shoulder period, and the probability of having a high or low capacity factor in the day period given the night capacity factor category. The values are shown in Table 3.3. Because the decision in the shoulder period is not affected by the night period, the outcome in the shoulder period will be the same for a given level of shoulder period wind generation regardless of the night period. Combined with the fact that the market surplus for a day is simply the sum of market surplus in each period, the weighted average market surplus for the year can be computed without knowing the conditional probability of shoulder period wind generation given night period wind generation. For the day period, because there is persistence in the level of wind generation across a 24-hour period and the day period decision is contingent on the outcome in the night period, it is necessary to allow the day period wind generation to depend on the night period wind generation.

Finally, I compute the weighted average capacity factor given the probabilities and the values of the capacity factor in each category. This yields a weighted average capacity factor of 31.12 percent which roughly matches the average capacity factor when simply averaging over the hourly capacity factors in 2011 (31.15 percent). Given the assumption

Category	Shoulder	Night	Day After Low Night	Day After High Night
Low	80.0%	77.5%	90.5%	40.2%
High	20.0%	22.5%	9.5%	59.8%

Table 3.3: Probability of Low and High Wind Generation in each Period

of 4,000 MW of wind capacity in PJM, this yields an average hourly wind output of 1,247 MWh. The generation capacity by fuel type information sheet for 2011 indicates that there is 652.5 MW of firm wind capacity in PJM. The discrepancy is caused by the definition PJM utilizes for capacity when compiling the data. For the purposes of this modeling exercise, I prefer to use the average wind capacity of 1,247. This is the adjustment which yields 178,172 MW of capacity in PJM. To adjust the 2008 and 2012 capacity for wind generation, I compute the ratio of average wind generation, 1,247 MWh, to the reported wind capacity, 652.5 MW, and increase the value of the 2008 and 2012 reported wind capacity by this ratio.

3.4.2.4 Demand

To estimate the demand functions for each season, I utilize the PJM day-ahead price and demand data and a price elasticity of demand equal to -0.2. This elasticity is based upon estimates in Bernstein and Griffin (2005). They find that the short-run price-elasticity of demand for both residential and commercial consumers is approximately -0.2. I divide each day into night, shoulder, and day periods based upon price. The 8 hours with the highest prices each day are assigned to day, the 10 lowest priced hours of each day are assigned to night, and the remaining 6 hours to the shoulder period. I assume each day begins at 10 pm the night before as this is when demand generally collapses in the evening.

Demand for electricity follows clear seasonal trends and also varies significantly from day to day. To account for the variation in price across seasons, I estimate prices in each period for 3 seasons, winter, spring/fall, and summer. To account for variations within a season, I estimate the price at night for the 25th, 50th, 75th, and 90th percentiles of the price distribution in each season. For the shoulder period, I estimate the 50th percentile of the shoulder period price distribution for each season and each night-time price percentile. For the day period, I begin with an estimate of the daytime price for the 25th, 50th, 75th, and 90th percentiles of the daytime price distribution for each season and for each of the night-time price percentiles. These 12 different night prices, 12 different shoulder period prices, and 48 different day prices are the prices used in the certainty version of the model to calculate the demand functions.

For each of the 12 prices in the night and shoulder periods, I estimate the demand that occurs at that price from the PJM demand data. Based on the Rand From the 12 price-quantity pairs in each period, I estimate 12 intercepts and 12 slopes using equations (3.22) and (3.21). For day, I estimate the demand that occurs at the 25th daytime price percentile for each season and night percentile. From these 12 price-quantity combinations, I estimate 12 intercepts and 12 slopes. As the quantity demanded at the various day prices within one of the 12 season-night percentile bins varies little, I assume that the demand functions for the 50th, 75th, and 90th daytime price percentiles are simply a vertical shift upwards of the demand functions from the 25th daytime price percentile. Consequently, within a season-night percentile bin, the slope is constant, but the intercept varies based upon the price for each of the daytime price percentiles. Thus, there are 12 slopes and 48 intercepts for the day period in the certainty version of the model. Table 3.A.1 in Appendix 3-A contains the full set of demand parameters utilized in the baseline model without uncertainty.

The winter season accounts 25 percent of the year, the spring/fall season accounts for 50 percent of the year, and summer accounts for 25 percent of the year. Using the midpoint between percentile bins, I assume the 25th percentile of night prices accounts for 37.5 percent of each season, the 50th percentile of night prices accounts for 25 percent of each season, the 75th percentile of night prices accounts for 20 percent of each season, and the 90th percentile of night prices accounts for 17.5 percent of each season. The 25th, 50th, 75th, and 90th percentiles of day prices then account for 37.5 percent, 25 percent, 20 percent, and 17.5 percent respectively of each season-night percentile bin. In estimating the value of additional storage or generation, I compute the weighted average value across a year.

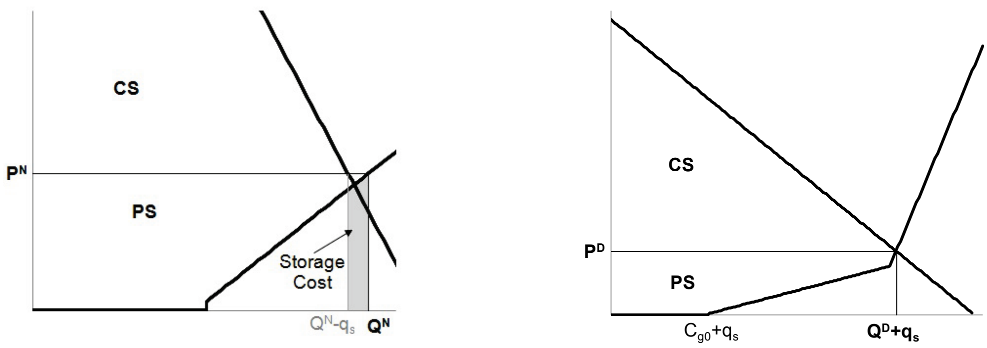
When including uncertainty about demand in the day period, I assume that the day price is a function of the night price or $P^D = P^N(1+\rho+\mu)$. ρ is the percentage difference between the price of electricity between each of the 48 day prices in the certainty version of the model, and the price of electricity in the corresponding night periods. μ is the error term, which I assume is normally distributed with mean zero and a positive variance. To estimate the variance, I begin by calculating the average night and day price for each day in 2011 from the day-ahead and real-time price data. The day-ahead prices are calculated based upon expected demand and supply conditions on the day prior to the delivery of the electricity. The real-time prices are calculated at the time of delivery and reflect the realizations of supply and demand. Therefore, I compute the expected percentage difference between day and night prices for each day from the day-ahead prices and the actual percentage difference between day and night prices from the real-time data. I then take compute the difference of the differences as the error term. As the observations of the error have some extreme values that make it difficult for the normal distribution to match the majority of the observed distribution, I calculate the variance of the observed errors for each season, excluding the top 5 percent and the bottom 5 percent of observations. This yields an estimated standard deviation of 0.25 for winter, 0.29 for spring/fall, and 0.54 for summer. Each value of μ generates a different estimate of price for day and a different intercept of the demand function for day.

3.4.3 Market Surplus Calculation

The metric I utilize to determine the value of installing an additional 400 MW storage facility versus an additional 400 MW combined cycle unit, is the change in market surplus caused by adding the capacity to the market. Market surplus is defined as the sum of the utility received by consumers from using electricity (consumer surplus) less the cost to producers of generating the electricity (producer surplus).

Consumer surplus (CS) is equal to the utility the consumers derive from the electricity consumed less the price of the electricity, which is the area under the demand curve (marginal utility curve) from zero to the total quantity of electricity consumed in that period minus price times quantity demanded, as can be see in Figure 3.5. In the night period, the quantity consumed is $Q^N - q_s$, in the shoulder period, the quantity consumed is Q^{Sh}, and in the day period the quantity consumed is $\hat{Q}^D = Q^D + q_s$. Therefore consumer surplus is simply:

$$CS = \frac{1}{e}\left(\int_0^{Q^N - q_s} u^{N\prime}(q)\mathrm{d}q - P^N(Q^N - q_s)\right) + \left(2 - \frac{1}{e}\right)\left(\int_0^{Q^{Sh}} u^{Sh\prime}(q)\mathrm{d}q - P^{Sh}Q^{Sh}\right) + \int_0^{Q^D + q_s} u^{D\prime}(q)\mathrm{d}q - P^D(Q^D + q_s)$$



(a) Night (b) Day

Figure 3.5: Market Surplus Calculations

As is hinted at in Figure 3.5, computing the producer surplus is more tricky because the night and day periods are linked by storage. In the night period, the producer surplus is equal to the price received for the electricity sold multiplied by the quantity sold, $Q^N - q_s$, less the cost of generating the electricity sold, $Q^N - q_s$, and the cost of the electricity stored, q_s, as shown in Figure 3.5a, multiplied by the night period's relative length, $\frac{1}{e}$:

$$PS^N = \frac{1}{e}\left(P^N(Q^N - q_s) - \int_0^{Q^N} c_1(q, X_g)\mathrm{d}q\right)$$

In the shoulder period, the producer surplus is simply the profit of the producer in the shoulder period multiplied by the shoulder period's relative length, $2 - \frac{1}{e}$ or:

$$PS^{Sh} = \left(2 - \frac{1}{e}\right)\left(P^{Sh}Q^{Sh} - \int_0^{Q^{Sh}} c_1(q, X_g)\mathrm{d}q\right)$$

In the day period, any storage shifts the quantity of zero-marginal cost generation, C_{g_0}, out by q_s, as shown in Figure 3.5b. The producer surplus is therefore equal to the price in the day period multiplied by the total quantity of generation sold, $Q^D + q_s$ less the cost of generation, Q^D.

$$PS^D = P^D(Q^D + q_s) - \int_0^{Q^D} c_1(q, X_g)\mathrm{d}q$$

Adding storage capacity simply increases the amount generators can store, while adding generation capacity shifts the cost curve out as described above.

3.5 Results

I begin, in Section 3.5.1, by discussing the goodness of fit between the model and the data which generate the parameters of the model for the certainty version of the model. Section 3.5.2 explores the effect of wind generation and demand uncertainty as well the effect of varying, but not uncertain, wind generation on the value of additional storage capacity and on the value of additional generation capacity. Using the lessons from these sections, I estimate the value of adding a 400 MW storage facility versus a 400 MW combined cycle unit as the level of wind generation in the market increases.

I then compare how the relative value of storage versus generation capacity changes as fuel costs, the efficiency of the new storage and combined cycle units, and the variance of wind output changes.

3.5.1 Goodness of Fit

Table 3.4 reports the equilibrium prices and quantities of electricity generated and stored for each of the 48 different scenarios in the certainty version of the model with average wind generation in each period. First note that as expected, generation is less in the night period, than in the shoulder period, which has lower generation than the day period. Correspondingly, prices in the day period are the highest, followed by the shoulder period, with the night period having the lowest prices. As the night prices increase, from the 25th to the 50th percentile, etc., prices increase in all periods. For the higher day price percentiles, the shoulder price stays constant, while the night price may increase if more storage is utilized than at a lower day price percentile. If all storage was utilized at the lower day price percentile, the price of electricity in the night period does not change as all storage will be utilized at higher day prices as well. In the summer period, which has the highest daytime demand and prices relative to the night period in the data, the full capacity of existing storage is utilized for each set of demand functions. At the 25th and 50th day price percentiles, storage is frequently not utilized to capacity because the day-night price difference is not large enough.

Table 3.5 presents the estimated marginal value of additional generation and storage calculated by the model versus the value of additional generation and storage calculated from the raw data. Recalling the formulas in Sections 3.3.1 and 3.3.2, the marginal value of additional storage is equal to the day price less the night price multiplied by $\frac{1}{e}$ or $P^D - \frac{1}{e}P^N$, when this value is positive, and zero otherwise. The marginal value of additional generation in each period is equal to the maximum of the price in each period less the marginal cost of the new generation and zero, multiplied by the relative length of each period, or $\frac{1}{e}max(P^N - m, 0) + (2 - \frac{1}{e})max(P^{Sh} - m, 0) + max(P^D - m, 0)$.

Night Price Percentile	Day Price Percentile	Season	Night Gen.	Shoulder Gen.	Day Gen.	Quantity Stored	Night Price	Shoulder Price	Day Price
25th	25th	Winter	76,563	81,533	85,027	954	$36.24	$41.56	$45.29
		Spring/Fall	66,101	76,994	77,639	1,910	$32.36	$40.00	$40.45
		Summer	78,591	93,034	97,220	5,000	$30.68	$42.24	$45.59
	50th	Winter	79,137	81,533	89,447	5,000	$38.99	$41.56	$50.03
		Spring/Fall	67,693	76,994	83,627	4,000	$33.48	$40.00	$44.65
		Summer	78,591	93,034	112,434	5,000	$30.68	$42.24	$57.78
	75th	Winter	79,137	81,533	95,402	5,000	$38.99	$41.56	$56.40
		Spring/Fall	67,693	76,994	90,085	4,000	$33.48	$40.00	$49.17
		Summer	78,591	93,034	116,623	5,000	$30.68	$42.24	$61.13
	90th	Winter	79,137	81,533	99,119	5,000	$38.99	$41.56	$60.38
		Spring/Fall	67,693	76,994	104,154	4,000	$33.48	$40.00	$59.04
		Summer	78,591	93,034	126,313	5,000	$30.68	$42.24	$68.89
50th	25th	Winter	79,360	83,595	88,523	2,018	$39.23	$43.76	$49.04
		Spring/Fall	68,823	80,706	81,042	2,428	$34.27	$42.60	$42.83
		Summer	83,151	98,502	108,273	5,000	$34.33	$46.62	$54.45
	50th	Winter	81,335	83,595	91,406	5,000	$41.34	$43.76	$52.12
		Spring/Fall	69,995	80,706	82,508	3,960	$35.09	$42.60	$43.86
		Summer	83,151	98,502	114,674	5,000	$34.33	$46.62	$59.57
	75th	Winter	81,335	83,595	94,363	5,000	$41.34	$43.76	$55.29
		Spring/Fall	70,026	80,706	90,702	4,000	$35.11	$42.60	$49.61
		Summer	83,151	98,502	120,929	5,000	$34.33	$46.62	$64.58
	90th	Winter	81,335	83,595	100,616	5,000	$41.34	$43.76	$61.98
		Spring/Fall	70,026	80,706	111,534	4,000	$35.11	$42.60	$64.21
		Summer	83,151	98,502	133,074	5,000	$34.33	$46.62	$94.20
75th	25th	Winter	83,361	89,979	93,525	1,711	$43.51	$50.59	$54.39
		Spring/Fall	71,003	81,617	83,767	1,547	$35.80	$43.24	$44.75
		Summer	88,200	109,287	113,396	5,000	$38.37	$55.26	$58.55
	50th	Winter	84,586	89,979	95,056	3,486	$44.82	$50.59	$56.03
		Spring/Fall	72,735	81,617	85,933	3,803	$37.01	$43.24	$46.26
		Summer	88,200	109,287	116,515	5,000	$38.37	$55.26	$61.05
	75th	Winter	85,632	89,979	101,819	5,000	$45.94	$50.59	$63.27
		Spring/Fall	72,886	81,617	97,915	4,000	$37.12	$43.24	$54.66
		Summer	88,200	109,287	135,122	5,000	$38.37	$55.26	$109.04
	90th	Winter	85,632	89,979	106,847	5,000	$45.94	$50.59	$73.58
		Spring/Fall	72,886	81,617	108,476	4,000	$37.12	$43.24	$62.07
		Summer	88,200	109,287	145,230	5,000	$38.37	$55.26	$182.28
90th	25th	Winter	87,596	95,771	98,818	2,490	$48.04	$56.79	$60.06
		Spring/Fall	71,878	81,931	84,861	62	$36.41	$43.46	$45.51
		Summer	93,448	119,629	128,065	5,000	$42.57	$63.54	$70.29
	50th	Winter	89,346	95,771	104,909	5,000	$49.92	$56.79	$66.57
		Spring/Fall	74,917	81,931	89,321	4,000	$38.54	$43.46	$48.64
		Summer	93,448	119,629	143,328	5,000	$42.57	$63.54	$168.51
	75th	Winter	89,346	95,771	109,839	5,000	$49.92	$56.79	$153.45
		Spring/Fall	74,917	81,931	94,945	4,000	$38.54	$43.46	$52.58
		Summer	93,448	119,629	164,966	5,000	$42.57	$63.54	$325.29
	90th	Winter	89,346	95,771	115,387	5,000	$49.92	$56.79	$301.55
		Spring/Fall	74,917	81,931	104,642	4,000	$38.54	$43.46	$59.38
		Summer	93,448	119,629	178,172	5,000	$42.57	$63.54	$880.12
		Weighted Average	76,481	86,989	98,489	3,718	$36.79	$45.01	$67.05

Table 3.4: Equilibrium Prices and Quantities in the Baseline Model with No Uncertainty

Night Price Percentile	Day Price Percentile	Season	Marginal Value of Added Generation	Marginal Value of Added Generation From Prices in the Data	Percentage Error	Marginal Value of Added Storage	Marginal Value of Added Storage From Day-Night Price Difference in Data	Percentage Error
25th	25th	Winter	$9.69	$0.00	Inf	$0.00	$0.00	0.0%
		Spring/Fall	$8.29	$2.91	184.9%	$0.00	$0.97	-100.0%
		Summer	$11.63	$8.76	32.8%	$7.25	$9.92	-26.9%
	50th	Winter	$15.47	$1.22	1167.1%	$1.29	$2.34	-45.1%
		Spring/Fall	$12.49	$6.76	84.8%	$2.80	$4.82	-41.8%
		Summer	$23.82	$16.41	45.2%	$19.43	$17.57	10.6%
	75th	Winter	$21.84	$4.24	415.2%	$7.66	$5.36	42.8%
		Spring/Fall	$17.02	$10.03	69.7%	$7.33	$8.09	-9.4%
		Summer	$27.17	$18.51	46.8%	$22.79	$19.67	15.8%
	90th	Winter	$25.82	$6.12	321.6%	$11.64	$7.25	60.6%
		Spring/Fall	$26.88	$17.16	56.7%	$17.19	$15.22	13.0%
		Summer	$34.93	$23.38	49.4%	$30.55	$24.54	24.5%
50th	25th	Winter	$16.43	$8.86	85.4%	$0.00	$4.12	-100.0%
		Spring/Fall	$12.63	$10.55	19.7%	$0.00	$4.64	-100.0%
		Summer	$23.77	$21.91	8.5%	$11.53	$17.93	-35.7%
	50th	Winter	$22.16	$11.84	87.2%	$0.44	$7.09	-93.8%
		Spring/Fall	$13.66	$12.02	13.6%	$0.00	$6.11	-100.0%
		Summer	$28.89	$25.41	13.7%	$16.66	$21.44	-22.3%
	75th	Winter	$25.33	$13.61	86.1%	$3.61	$8.86	-59.3%
		Spring/Fall	$19.40	$16.56	17.2%	$5.72	$10.65	-46.3%
		Summer	$33.90	$28.84	17.6%	$21.67	$24.86	-12.8%
	90th	Winter	$32.02	$17.34	84.6%	$10.30	$12.60	-18.2%
		Spring/Fall	$34.01	$28.05	21.2%	$20.32	$22.14	-8.2%
		Summer	$63.53	$38.81	63.7%	$51.29	$34.83	47.3%
75th	25th	Winter	$32.27	$25.05	28.8%	$0.00	$5.76	-100.0%
		Spring/Fall	$15.37	$17.11	-10.2%	$0.00	$6.06	-100.0%
		Summer	$35.42	$36.67	-3.4%	$10.58	$23.45	-54.9%
	50th	Winter	$35.54	$26.93	32.0%	$0.00	$7.64	-100.0%
		Spring/Fall	$18.40	$19.42	-5.2%	$0.00	$8.36	-100.0%
		Summer	$37.92	$38.55	-1.6%	$13.08	$25.33	-48.4%
	75th	Winter	$44.18	$32.17	37.3%	$5.84	$12.87	-54.6%
		Spring/Fall	$26.94	$26.47	1.8%	$8.27	$15.41	-46.4%
		Summer	$85.92	$55.26	55.5%	$61.08	$42.05	45.3%
	90th	Winter	$54.49	$36.33	50.0%	$16.15	$17.04	-5.2%
		Spring/Fall	$34.34	$32.60	5.3%	$15.67	$21.55	-27.3%
		Summer	$159.16	$72.21	120.4%	$134.32	$58.99	127.7%
90th	25th	Winter	$48.25	$49.45	-2.4%	$0.00	$14.17	-100.0%
		Spring/Fall	$17.06	$18.62	-8.4%	$0.00	$5.58	-100.0%
		Summer	$58.63	$65.13	-10.0%	$17.08	$37.85	-54.9%
	50th	Winter	$57.11	$55.45	3.0%	$4.18	$20.17	-79.3%
		Spring/Fall	$22.86	$23.20	-1.5%	$0.46	$10.16	-95.4%
		Summer	$156.85	$89.70	74.9%	$115.29	$62.42	84.7%
	75th	Winter	$143.98	$72.75	97.9%	$91.05	$37.46	143.1%
		Spring/Fall	$26.80	$26.58	0.8%	$4.41	$13.54	-67.4%
		Summer	$313.63	$127.46	146.1%	$272.08	$100.18	171.6%
	90th	Winter	$292.08	$100.60	190.3%	$239.15	$65.31	266.2%
		Spring/Fall	$33.60	$32.40	3.7%	$11.21	$19.36	-42.1%
		Summer	$868.46	$227.02	282.5%	$826.90	$199.75	314.0%
		Weighted Average	$38.61	$23.27	65.9%	$21.06	$15.19	38.6%

Table 3.5: Comparison of the Calculated Marginal Value of Additional Storage and Generation Capacity from the Baseline Model Versus the Value from the Data

To compute the marginal value of storage and generation from the data, I apply these formulas to the prices I put into the model to generate the demand functions (see Table 3.A.1 in Appendix 3-A).

Comparing the marginal values of additional generation capacity from the data versus the model in Figure 3.5, shows that the values I estimate for generation are almost always too high, sometimes more than 1000 percent above the value in the data. This leads the weighted average marginal value of generation to be nearly 66 percent greater in the model than in the data, which is clearly not a desirable attribute. Interestingly, however, the marginal value of storage is not over-estimated in a corresponding fashion. The marginal values of storage in the 25th percentile of night prices are not terribly far off, though in some cases the model yields a zero value for additional storage while the data exhibit a small positive marginal value. Marginal values for additional storage in the 50th and 75th percentiles of night prices are generally too low, although the 90th percentile of day prices in the summer season is too high. In the 90th percentile of night prices, at the 25th and 50th percentiles of day prices, the marginal value of storage is generally too low, but for the higher percentiles of night prices, the marginal value of storage is significantly above the marginal value indicated by the data. It is these very high values in the 90th percentiles that are driving the marginal value of storage in the model to be 39 percent above the marginal value of storage in the data.

The differing levels of overestimation for additional generation and storage capacity lead the marginal value of additional storage to be 45.4 percent lower than the marginal value of additional generation in the model, but only 34.7 percent below the marginal value of additional generation in the data. Finding a way to correct these shortcomings of the model, is therefore likely to improve the marginal value of additional storage relative to the marginal value of additional generation. Unfortunately there was no obvious way to fix these problems at this time, and therefore the results presented below should be taken as preliminary.

In the first row of Table 3.6, the results on the relative value of additional storage versus additional generation capacity in the baseline scenario are reported. The increase in surplus from adding a 400 MW combined cycle generator is \$14,603, while the increase in surplus from adding a 400 MW storage facility is \$8,172. Thus, in the baseline case using 2011 data, new storage capacity is worth 44 percent less than new generation capacity.

The increase in surplus represents the increase over one day period, lasting 1 unit of time, one shoulder period lasting $2 - \frac{1}{e}$ units of time, and one night period lasting $\frac{1}{e}$ units of time. As the data utilized to calculate the parameters of the model represent the average hourly price and demand, I assume a unit of time is an hour. Henceforth, I will term the increase in surplus reported in the first and third columns of data in the table the hourly value of generation or storage capacity, though it actually represents 3 hours. As I assume there are 8 hours of storage capacity and the efficiency of the storage facility, e, is 80 percent, multiplying the increase in surplus from additional storage or generation by 8 hours yields the value of storage or generation over 1 day. Multiplying the daily value by 365 days in the year yields the annual value of storage or generation. As noted in the section 3.4.2, storage facilities have an average lifetime of 50 years, while combined cycle generators have an average lifetime of 30 years. Multiplying the annual value of storage by 50 years and the annual value of generation by 30 years gives the lifetime benefit from new storage or generation capacity. The capital cost of a new storage facility is \$1 billion and the capital cost of a new combined cycle unit is \$401.2 million. Subtracting the capital cost of each type of capacity from its lifetime benefit yields the net lifetime benefit of new storage or generation capacity reported in the second to last and third to last columns.

After accounting for the differing lifespans and capital costs, the net lifetime benefit of a new 400 MW combined cycle generator is \$878 million, while the net lifetime benefit of a new 400 MW storage facility is \$193 million, which is 78 percent less than the net lifetime benefit of a new combined cycle unit. Thus, despite the longer lifespan of a storage facility, the greater capital cost of storage makes additional storage capacity a significantly less attractive investment than additional generation capacity given the assumptions of the baseline scenario.

As noted in Section 3.5.1, the marginal value of combined cycle generation is overstated by 66 percent relative to the data, and the marginal value of storage capacity is overstated by 39 percent relative to the data. Assuming this basic relationship holds when examining the value of 400 MW of additional capacity, and recalculating the overstatements of value to be in reference to the value from the model rather than the value in the data, the data would indicate that the value of generation capacity should be 40 percent lower and the value of generation should be 28 percent lower. After making these adjustments, the net lifetime benefit of new generation capacity would be \$370 million and the net lifetime benefit of new storage capacity would be -\$140 million. This revision makes storage seem even less attractive even though the absolute difference between the net lifetime benefits of the two options is reduced. If this relationship between the model and the data were to hold when adding uncertainty in demand, the net lifetime benefit of storage would improve to \$5.7 million and the value of generation to \$476 million. However, it is not clear how to compare the model with the data in scenarios other than the baseline. The important point here is simply that the shortcomings of the model cause storage to seem like a somewhat profitable investment, when it may not be, and probably increase the absolute difference between the net lifetime benefit of storage and the net lifetime benefit of generation.

Scenario Number	Scenario	Increase in Surplus from Adding a 400 MW CC Unit	Share of Baseline	Increase in Surplus from Adding 400 MW Storage	Share of Baseline	% Difference Between Storage and CC Value	Net Lifetime Benefit of a 400 MW CC (Millions of $)	Net Lifetime Benefit of 400 MW of Storage (Millions of $)	% Difference Between Storage and CC Net Benefit
1	No Uncertainty; Average Wind	$14,603		$8,172		-44.0%	$878.0	$193.0	-78.0%
2	No Uncertainty; Wind Varies	$14,608	0.034%	$8,178	0.083%	-44.0%	$878.5	$194.0	-77.9%
3	Wind Uncertainty	$14,608	0.036%	$8,177	0.072%	-44.0%	$878.5	$193.9	-77.9%
4	Demand Uncertainty; Avg. Wind	$16,625	13.849%	$9,543	16.779%	-42.6%	$1,055.2	$393.2	-62.7%
5	Demand Uncertainty; Wind Varies	$16,632	13.892%	$9,551	16.881%	-42.6%	$1,055.7	$394.4	-62.6%
6	Add 4000 Wind	$15,948	9.208%	$9,460	15.770%	-40.7%	$995.8	$381.2	-61.7%
7	Add 8000 Wind	$15,308	4.831%	$9,380	14.783%	-38.7%	$939.8	$369.4	-60.7%
8	Add 16000 Wind	$14,173	-2.942%	$9,247	13.160%	-34.8%	$840.4	$350.1	-58.3%
9	Add 32000 Wind	$12,292	-15.825%	$9,059	10.858%	-26.3%	$675.6	$322.6	-52.3%
10	Add 64000 Wind	$9,589	-34.338%	$8,908	9.013%	-7.1%	$438.8	$300.6	-31.5%
11	Add 96000 Wind	$7,776	-46.750%	$9,047	10.708%	16.3%	$280.0	$320.8	14.6%
12	Demand & Wind Uncertainty	$16,632	13.894%	$9,550	16.873%	-42.6%	$1,055.8	$394.3	-62.6%
13	Add 4000 Wind	$15,948	9.212%	$9,456	15.724%	-40.7%	$995.9	$380.6	-61.8%
14	Add 8000 Wind	$15,309	4.833%	$9,370	14.661%	-38.8%	$939.8	$368.0	-60.8%
15	Add 16000 Wind	$14,175	-2.933%	$9,219	12.818%	-35.0%	$840.5	$346.0	-58.8%
16	Add 32000 Wind	$12,298	-15.786%	$9,011	10.270%	-26.7%	$676.1	$315.6	-53.3%
17	Add 64000 Wind	$9,608	-34.202%	$8,808	7.786%	-8.3%	$440.5	$285.9	-35.1%
18	Add 96000 Wind	$7,806	-46.549%	$8,867	8.514%	13.6%	$282.6	$294.6	4.3%

Table 3.6: Comparison of Results under Different Types of Uncertainty and Different Levels of Wind Generation

3.5.2 Effect of Uncertainty

The remaining rows of Table 3.6 compare the value of additional storage or generation capacity against the baseline for a variety of different scenarios to understand the effect of the assumptions about uncertainty and the level of wind generation have on the results. Before adding uncertainty in wind generation, I begin by assuming that wind varies in the same way it would under uncertainty, but assume the planner knows the level of wind generation that will occur in future periods prior to making decisions in the night period. This is scenario 2 in Table 3.6. From the table, we see that adding variation in wind generation, keeping the weighted average generation constant, increases the hourly value of additional generation or storage capacity very slightly, but increases the hour value of storage by a somewhat greater amount. Storage receives a greater benefit because the weighted average price of electricity in the day period goes up while the weighted average price of electricity in the night period goes down slightly, thereby increasing the day-night price difference, but decreasing the value of additional generation in the night period. The change in the relationship between prices is the result of the nature of the variation in wind. Given the assumed capacity factors for the high and low wind generation categories and the probabilities assigned to each category in each period, the weighted average capacity factor for wind generators increases slightly in the night period and decreases slightly in the day period, which causes the observed changes in price.

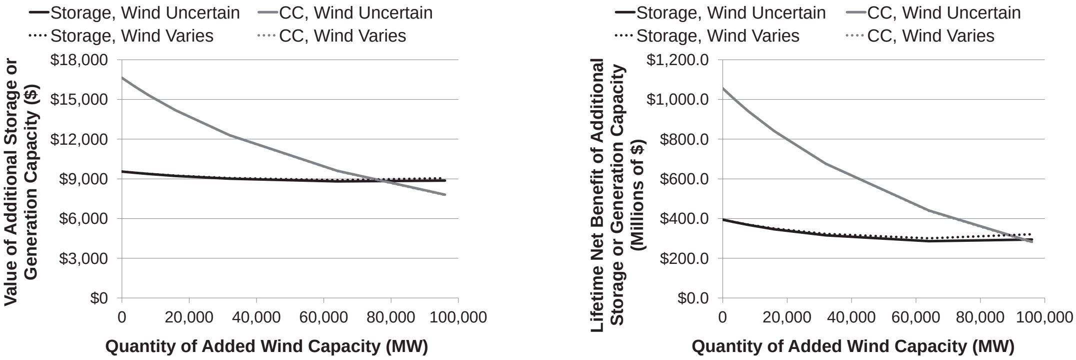
Adding uncertainty in wind generation, scenario 3, increases the value of additional generation further, but decreases the value of additional storage somewhat. This occurs because the weighted average price of electricity in the night period goes up slightly. The increase in the night price is caused by increased, but less efficient, use of storage due to the uncertainty.

Adding uncertainty in demand but keeping wind generation constant in each period at the weighted average level, scenario 5, causes the hourly value and lifetime net benefit of additional storage and additional generation to increase dramatically. This is a result of a significant increase in the expected day price due to the fact that prices are bounded below by zero but are unbounded above. The increase in the expected day price also increases the utilization of storage and therefore the weighted average price at night, which causes the value of additional generation to not increase as much on a percentage basis as the value of additional storage. Therefore, hourly value of additional storage relative to the hourly value of additional generation increases from 44 percent less to 42 percent less, while the net lifetime benefit of additional storage relative to the net lifetime benefit of additional generation increases from 78 percent less to 63 percent less. Adding variation or uncertainty in wind generation, scenarios 6 and 12, increases the relative value of additional storage versus additional generation in the same fashion as does variation and uncertainty in wind generation without demand uncertainty.

Figure 3.6 illustrates the change in the hourly value and lifetime net benefit of additional storage capacity and additional generation capacity as the capacity of wind in the market increases. Note that increasing the available capacity of wind increases the variance of wind generation because the difference between the low capacity factor, 22 percent, and the high capacity factor, 65 percent, is magnified by the increase in the quantity of capacity that multiplies the capacity factor to determine generation. As can be seen in the figure, even with very large increases in the capacity of wind generation, the difference between the certainty and uncertainty versions of the model is quite small. Although the uncertainty version of the model does decrease the net lifetime benefit of storage a bit more relative to the certainty version than it decreases the hourly value. The value of additional generation capacity is virtually unaffected.

3.5.3 Effect of Increasing Wind Capacity in the Market

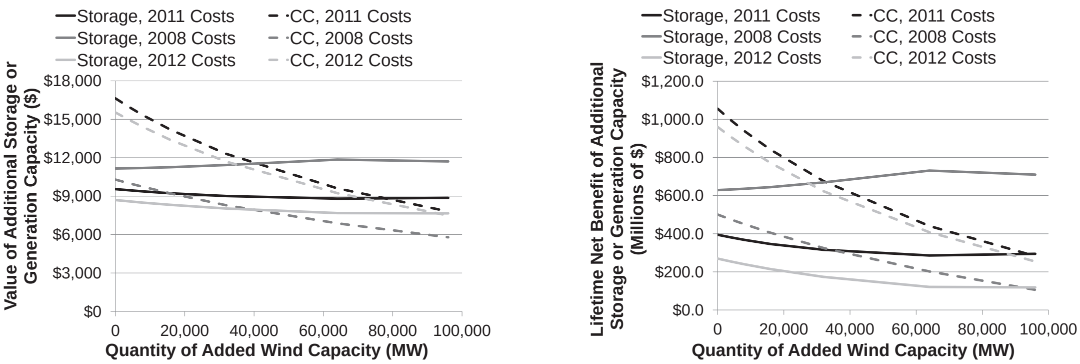
Figure 3.6 also reveals that as the level of wind capacity in the market increases, the hourly value of additional storage falls very little, and even begins to increase at the highest levels of wind capacity, while the hourly value of additional generation drops precipitously. Due to the difference in the capital cost, however, the net lifetime benefit of storage does not exceed the net lifetime benefit of generation until there is 100,000 MW



(a) Hourly Value (b) Net Lifetime Benefit

Figure 3.6: Hourly Value and Net Lifetime Benefit of Additional Storage or Generation Capacity as Wind Capacity Increases, 2011 Costs

of wind capacity in the market (96,000 MW of added capacity), a 2400 percent increase in wind capacity. Given the cost function utilized in the model, the electricity price at night does not drop to zero until 96,000 MW of wind capacity have been added. In reality, however, real-time electricity prices in the night period (those calculated given actual operating conditions) have already dropped below zero in some hours because much of the generation capacity utilized at night cannot be turned off at night if it needs to be run the next day due to long start-up times. This is known as a must-run requirement. Thus, a model that takes into account must-run requirements may be able to generate zero prices at lower quantities of additional wind generation.



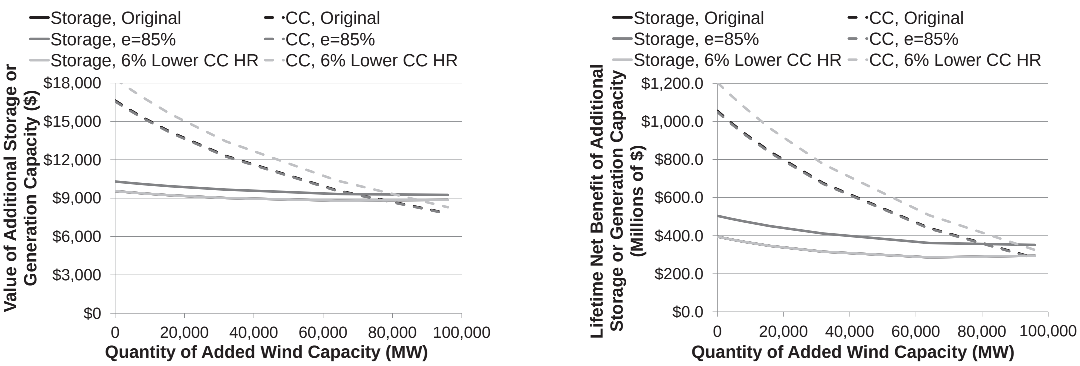
(a) Hourly Value (b) Net Lifetime Benefit

Figure 3.7: Hourly Value and Net Lifetime Benefit of Additional Storage or Generation Capacity as Wind Capacity Increases, 2008 Versus 2012 Costs

Figure 3.7 compares the hourly value and net lifetime benefit of additional generation or storage capacity as fuel costs change. The model incorporating uncertainty in both demand and wind generation was utilized to generate the data for this figure and the following figures. At 2008 fuel costs, the marginal cost curves are much steeper in the initial upward sloping portion, as shown in Figure 3.4a, but flatter in the second portion for spring/fall and winter than the marginal cost curves for 2011. The increase in the steepness of the cost curve due to the increased fuel costs, combined with a daytime demand curve that is steeper than the night-time demand curve increases the difference between night and day prices and therefore increases the value of storage, as predicted by Figure 3.1.A, for equilibria that were on the first portion of the marginal cost curve under 2011 prices. As wind generation increases, less generation from costly fossil fuels is needed and therefore it is more likely that the equilibrium quantity will lie on the first portion of the cost curve, where costs were higher in 2008 than in 2011. Thus, as wind generation increases, the hourly value and net lifetime benefit of additional storage increases, though for the largest quantity of added wind capacity the value declines somewhat. Conversely, the increased cost of fuel decreases the amount of electricity demanded and increases the marginal cost of generation from the new combined cycle capacity. Together, these forces reduce the value of new combined cycle capacity sufficiently to make storage a more cost-effective investment by either of the measures reported in Figure 3.7 and for any level of wind generation.

In 2012, natural gas became much cheaper while other fuels prices changed little. This caused the marginal cost curves to be much flatter than the marginal cost curves under 2011 fuel costs and the cost of generation at every quantity is less under 2012 fuel costs than under 2011 fuel costs, see Figure 3.4b. Therefore the price of electricity is lower at every level and equilibrium quantities are higher. Although the reduction in fuel costs reduces the marginal cost of generation from new combined cycle capacity, the reduction is not sufficient to make up for the lost revenue due to the decrease in price. Therefore, the hourly value and net lifetime benefit of new combined cycle capacity decreases somewhat relative to the values with 2011 fuel costs. The flattening of the supply curve, combined with a day demand curve that is steeper than the night demand curve decreases the difference between night and day prices and significantly decreases the hourly value and net lifetime benefit of additional storage relative to that under 2011. As wind generation increases, this effect grows more pronounced.

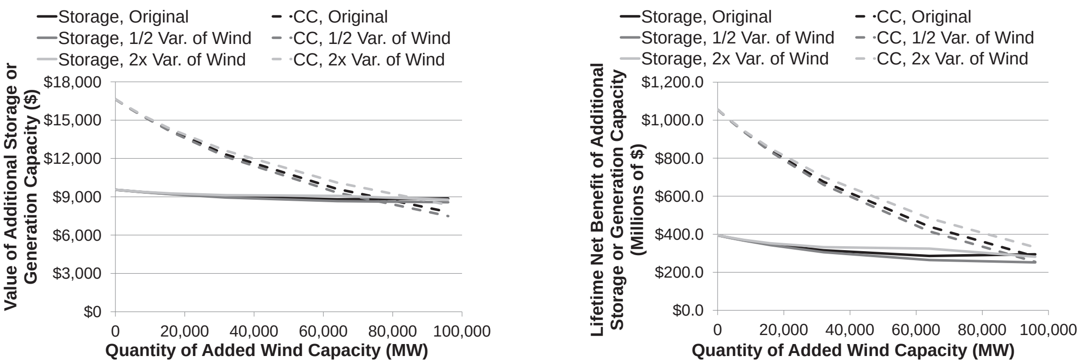
Thus, as gas prices fall, as they did in 2012, the value of additional storage capacity falls more significantly than the value of additional generation capacity. When natural gas prices rise, as in 2008, the value of new combined cycle capacity falls significantly while the value of new storage capacity rises. Thus, the model is able to support the common assertion that falling natural gas prices hurt the competitiveness of storage relative to natural gas capacity. However, the figures also show that combined cycle capacity loses value much more rapidly than does storage as wind generation increases and drives electricity prices lower. At extremely high volumes of wind capacity, e.g. 100,000 MW, the value of additional storage capcity is likely to exceed the value of additional combined cycle capacity regardless of fuel costs because prices at night are driven to zero in many periods. If prices are driven to zero more often in reality than in the model, this indicates that the value of storage should be relatively higher as combined cycle units would not be profitable to operate in these hours.



(a) Hourly Value (b) Net Lifetime Benefit

Figure 3.8: Hourly Value and Net Lifetime Benefit of Additional Storage or Generation Capacity as Wind Capacity Increases, Increase Efficiency of Storage and Combined Cycle Unit

Figure 3.8 illustrates the effect of increasing the efficiency of storage, e, from 80 percent to 85 percent efficiency (a roughly 6 percent increase) versus lowering the heat rate (HR) of the new combined cycle capcity by 6 percent, which increases the efficiency by 6 percent. Unfortunately the model is not currently able to accommodate two groups of storage capacity at different efficiencies and therefore in increasing the efficiency of the new storage capacity, I was forced to also increase the efficiency of the existing capacity. As can be seen in the figure, increasing the efficiency of storage has only a small effect on the value of additional storage capacity and the difference between the value of additional storage capacity at 80 percent efficiency and the value at 85 percent efficiency falls as wind generation grows. The change in the efficiency of storage has very little impact on the value of new combined cycle capacity because the quantity of storage capacity in the market is small. Increasing the efficiency of the new generation capacity has no effect on the value of new storage capacity, but does moderately improve the hourly value and net lifetime benefit of installing an additional combined cycle unit. Thus increasing the efficiency of either type of capacity would increase that technology's competitiveness relative to the other.



(a) Hourly Value (b) Net Lifetime Benefit

Figure 3.9: Hourly Value and Net Lifetime Benefit of Additional Storage or Generation Capacity as Wind Capacity Increases, Adjust Variance of Wind Generation

As the capacity of wind generation in a market grows, the variance of the total wind generation in the market will fall if the wind generation becomes more geographically dispersed and thus subject to less correlated weather patterns. Conversely, if the potential locations for wind generation are concentrated in one area of a market, the variance of the wind generation will increase as the quantity of wind generation grows. Figure 3.9 demonstrates the effect of doubling or halving the variance of the wind generation, keeping the mean constant, on the value of additional generation and storage capacity. At the low levels of wind generation in the current market, the variance of wind generation has little effect on the value of additional storage or generation capacity. As the quantity of wind generation grows, the variance of the output grows, holding steady the variance of the capacity factor, as is done in the scenarios shown in the figures. Therefore, changing the variance has a somewhat larger effect on the value of additional generation or storage capacity as the quantity of wind capacity available in the market grows although relative to the effect of changing fuel prices or changing the efficiency of the technologies, the effect of the variance of wind generation is small. Increasing the variance, increases the value of additional generation capacity and vice versa. However, at the highest level of wind capacity, the value of additional storage capacity decreases with the increase in the variance of wind generation and also with a decrease in the variance of wind generation.

3.6 Conclusion

In conclusion, it is apparent that unless natural gas costs rise significantly in the next few years, which is unlikely given the recent boom in production, it is unlikely that new storage capacity will be a more profitable investment than new combined cycle capacity. Given the assumptions of the model utilized in this paper, extremely high levels of wind generation are required to slide the net lifetime benefit of new storage capacity above the net lifetime benefit of new combined cycle capacity. However, the simplicity of the model seems to be overlooking many of the areas in which storage excels. For instance, in a model that takes into account the operating constraints of various technologies, e.g. coal-fired steam turbines that take 12 or more hours to start up, prices at night may fall close to zero more often without significantly affecting day prices, thereby increasing the value of storage. Further the cost of unexpectedly high demand in the day period would be higher if the long start-up times of low cost generation were taken into account. By contrast, most storage technologies can respond to changes in demand almost instantaneously and would therefore have a higher value in the day period.

Regardless of the accuracy of the exact values assigned to the net lifetime benefit of generation or storage capacity, these figures make it clear that combined cycle generation has much more to lose from large increases in wind generation than does storage. Thus, if wind and solar (which shares many of the same characteristics as wind) were to become the predominant sources of power in the U.S., storage would clearly have an important role to play.

Appendices

3.A Demand Parameters

25th Percentile of Night Prices

	Night Demand Parameters			
Season	P_0^N	Q_0^N	a^N	b^N
Winter	$29.62	79,141	177.75	-0.0019
Spring/Fall	$29.42	65,501	176.50	-0.0022
Summer	$28.33	74,834	169.95	-0.0019

	Shoulder Demand Parameters			
Season	P_0^{Sh}	Q_0^{Sh}	a^{Sh}	b^{Sh}
Winter	$32.43	86,397	194.57	-0.0019
Spring/Fall	$36.43	78,530	218.59	-0.0023
Summer	$38.77	94,731	232.62	-0.0020

	Day Demand Parameters			
Season	P_0^D	Q_0^D	a^D	b^D
Winter	$35.80	90,794	214.82	-0.0020
Spring/Fall	$37.74	80,708	226.45	-0.0023
Summer	$45.33	102,341	271.96	-0.0022

	Day Price Percentiles			
Season	**25th**	**50th**	**75th**	**90th**
Winter	$35.80	$39.37	$42.39	$44.28
Spring/Fall	$37.74	$41.59	$44.86	$51.99
Summer	$45.33	$52.97	$55.08	$59.95

50th Percentile of Night Prices

	Night Demand Parameters			
Season	P_0^N	Q_0^N	a^N	b^N
Winter	$33.62	80,011	201.72	-0.0021
Spring/Fall	$31.04	67,805	186.23	-0.0023
Summer	$30.44	80,200	182.64	-0.0019

	Shoulder Demand Parameters			
Season	P_0^{Sh}	Q_0^{Sh}	a^{Sh}	b^{Sh}
Winter	$39.32	85,527	235.93	-0.0023
Spring/Fall	$39.02	82,214	234.12	-0.0024
Summer	$42.09	100,670	252.53	-0.0021

	Day Demand Parameters			
Season	P_0^D	Q_0^D	a^D	b^D
Winter	$46.14	91,692	276.85	-0.0025
Spring/Fall	$43.44	83,240	260.62	-0.0026
Summer	$55.98	112,653	335.91	-0.0025

	Day Price Percentiles			
Season	**25th**	**50th**	**75th**	**90th**
Winter	$46.14	$49.12	$50.88	$54.62
Spring/Fall	$43.44	$44.91	$49.45	$60.94
Summer	$55.98	$59.49	$62.92	$72.88

75th Percentile of Night Prices

	Night Demand Parameters			
Season	P_0^N	Q_0^N	a^N	b^N
Winter	$39.70	83,247	238.23	-0.0024
Spring/Fall	$32.78	70,757	196.70	-0.0023
Summer	$33.42	85,741	200.51	-0.0019

	Shoulder Demand Parameters			
Season	P_0^{Sh}	Q_0^{Sh}	a^{Sh}	b^{Sh}
Winter	$45.99	91,818	275.9286	-0.0025
Spring/Fall	$42.97	81,719	257.82	-0.0026
Summer	$49.45	111,916	296.69	-0.0022

	Day Demand Parameters			
Season	P_0^D	Q_0^D	a^D	b^D
Winter	$55.39	94,893	332.35	-0.0029
Spring/Fall	$47.04	84,490	282.23	-0.0028
Summer	$65.23	116,020	391.36	-0.0028

	Day Price Percentiles			
Season	**25th**	**50th**	**75th**	**90th**
Winter	$55.39	$57.27	$62.51	$66.67
Spring/Fall	$47.04	$49.34	$56.39	$62.53
Summer	$65.23	$67.10	$83.82	$100.76

90th Percentile of Night Prices

	Night Demand Parameters			
Season	P_0^N	Q_0^N	a^N	b^N
Winter	$42.98	87,159	257.89	-0.0025
Spring/Fall	$34.44	72,646	206.64	-0.0024
Summer	$36.30	91,617	217.77	-0.0020

	Shoulder Demand Parameters			
Season	P_0^{Sh}	Q_0^{Sh}	a^{Sh}	b^{Sh}
Winter	$56.39	95,908	338.355	-0.0029
Spring/Fall	$42.86	82,161	257.145	-0.0026
Summer	$63.40	119,680	380.415	-0.0026

	Day Demand Parameters			
Season	P_0^D	Q_0^D	a^D	b^D
Winter	$67.89	99,023	407.35	-0.0034
Spring/Fall	$48.63	83,847	291.79	-0.0029
Summer	$83.22	129,056	499.32	-0.0032

	Day Price Percentiles			
Season	**25th**	**50th**	**75th**	**90th**
Winter	$67.89	$73.89	$91.19	$119.04
Spring/Fall	$48.63	$53.21	$56.59	$62.41
Summer	$83.22	$107.79	$145.55	$245.12

Note: The intercepts for day, a^D, vary with the day price percentiles while the slope, b^D, stays constant within a season-night bin.

Elasticity (ε) -0.2

Table 3.A.1: Parameters of the Demand Functions for Each Period

References

Alfonsi, A., A. Fruth, and A. Schied (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance 10*(2), 143–157.

Bernstein, M. and J. Griffin (2005). *Regional Differences in the Price-Elasticity of Demand For Energy.* RAND Corporation. Prepared for the National Renewable Energy Laboratory.

Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance 50*(5), 1655–1689.

Bulkeley, H. (2010). Cities and the governing of climate change. *Annual Review of Environment and Resources 35*, 229–253.

Burtraw, D. and W. Shobe (2009). State and local climate policy under a national emissions floor. *Climate Change Policy: Insights from the US and Europe*.

Bushnell, J., C. Peterman, and C. Wolfram (2008). Local solutions to global problems: Climate change policies and regulatory jurisdiction. *Review of Environmental Economics and Policy 2*(2), 175–193.

CBO (2011, July). *The Effects of Renewable and Clean Energy Standards.* Congressional Budget Office.

Crampes, C. and M. Moreaux (2010). Pumped storage and cost saving. *Energy Economics 32*(2), 325–333.

Denholm, P., E. Ela, B. Kirby, and M. Milligan (2010, January). *The Role of Energy Storage with Renewable Electricity Generation.* National Renewable Energy Laboratory.

Diringer, E. (2013). Climate change: A patchwork of emissions cuts. *Nature 501*, 307–309.

Drury, E., P. Denholm, and R. Sioshansi (2011). The value of compressed air energy storage in energy and reserve markets. *Energy 36*(8), 4959–4973.

EIA (2010, November). *Updated Capital Cost Estimates for Electricity Generation Plants.* U.S. Energy Information Administration.

EIA (2011, April). *Annual Energy Outlook 2011 with Projections to 2035.* U.S. Energy Information Administration.

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 251–276.

Engle, R. F. and A. J. Patton (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets 7*(1), 1–25.

Escribano, A. and R. Pascual (2006). Asymmetries in bid and ask responses to innovations in the trading process. *Empirical Economics 30*(4), 913–946.

Figueiredo, F., P. Flynn, and E. Cabral (2006). The economics of energy storage in 14 deregulated power markets. *Energy Studies Review 14*(2), Article 2.

Fischer, C. and L. Preonas (2010). Combining policies for renewable energy: Is the whole less than the sum of its parts? *International Review of Environmental and Resource Economics 4*, 51–92.

Fleming, M. J. and B. Mizrach (2009). The microstructure of a U.S. Treasury ECN: The BrokerTec platform. Working Paper.

Frey, S. and J. Grammig (2006). Liquidity supply and adverse selection in a pure limit order book market. *Empirical Economics 30*(4), 1007–1033.

Goulder, L. H. and R. N. Stavins (2011a). Challenges from state-federal interactions in U.S. climate change policy. *The American Economic Review 101*(3), 253–257.

Goulder, L. H. and R. N. Stavins (2011b). Interactions between state and federal climate change policies. In *The Design and Implementation of US Climate Policy*, pp. 109–121. University of Chicago Press.

Graves, F., T. Jenkin, and D. Murphy (1999, October). Opportunities for electricity storage in deregulating markets. *The Electricity Journal*.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics 22*(2), 229–252.

Hasbrouck, J. (1991a). Measuring the information content of stock trades. *The Journal of Finance 46*(1), 179–207.

Hasbrouck, J. (1991b). The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies 4*(3), 571–595.

Hautsch, N. and R. Huang (2012). The market impact of a limit order. *Journal of Economic Dynamics and Control 36*(4), 501–522.

Jaffe, A. B., R. G. Newell, and R. N. Stavins (2003). Technological change and the environment. *Handbook of environmental economics 1*, 461–516.

Jang, H. and P. Venkatesh (1991). Consistency between predicted and actual bid-ask quote-revisions. *The Journal of Finance 46*(1), 433–446.

Kumar, P. and D. Seppi (1992). Futures manipulation with "cash settlement". *The Journal of Finance 47*(4), 1485–1502.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica*, 1315–1335.

Kyle, A. (2009). Cash settlement, price manipulation, and the Modigliani-Miller Theorem. Working Paper.

Ledgerwood, S. and P. Carpenter (2012). A framework for the analysis of market manipulation. *Review of Law and Economics 8*(1), 253–295.

Ledgerwood, S. and P. Pfeifenberger (2013). Using virtual bids to manipulate the value of financial transmission rights. *The Electricity Journal 26*(9), 9–25.

Lo Prete, C. and W. Hogan (2014). Manipulation of day-ahead electricity prices through virtual bidding in the U.S. *IAEE Energy Forum 3rd Quarter*, 31–32.

Lütkepohl, H. and H.-E. Reimers (1992). Impulse response analysis of cointegrated systems. *Journal of economic dynamics and control 16*(1), 53–78.

McGuinness, M. and A. D. Ellerman (2008). The effects of interactions between federal and state climate policies. Technical report, Massachusetts Institute of Technology, Center for Energy and Environmental Policy Research.

Mizrach, B. (2008). The next tick on Nasdaq. *Quantitative Finance 8*(1).

Mokrian, P. and M. Stephen (2006). A stochastic programming framework of the valuations of electricity storage. working paper.

Oates, W. E. (2001). A reconsideration of environmental federalism. Technical report, Resources For the Future, Washington D.C., USA, Discussion Paper 01-54.

Obizhaeva, A. A. and J. Wang (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets 16*(1), 1–32.

Pascual, R., B. Pascual-Fuster, and F. Climent (2006). Cross-listing, price discovery and the informativeness of the trading process. *Journal of Financial Markets 9*(2), 144–161.

Rabe, B. G. (2008). States on steroids: the intergovernmental odyssey of american climate policy. *Review of Policy Research 25*(2), 105–128.

Rabe, B. G., M. Roman, and A. N. Dobelis (2005). State competition as a source driving climate change mitigation. *NYU Environmental Law Journal 14*, 1.

Rastler, D. (2010, December). *Electricity Energy Storage Technology Options: A White Paper Primer on Applications, Costs, and Benefits.* Electric Power Research Institute.

Sandås, P. (2001). Adverse selection and competitive market making: Empirical evidence from a limit order market. *Review of Financial Studies 14*(3), 705–734.

Sioshansi, R., P. Denholm, and T. Jenkin (2011). A comparative analysis of the value of pure and hybrid electricity storage. *Energy Economics 33*(1), 56–66.

Sioshansi, R., P. Denholm, T. Jenkin, and J. Weiss (2009). Estimating the value of electricity storage in pjm: Arbitrage and some welfare effects. *Energy Economics 31*(2), 269–277.

Walawalkar, R., J. Apt, and R. Mancini (2007). Economics of electric energy storage for energy arbitrage and regulation in new york. *Energy Policy 35*(4), 2558–2568.

Williams, R. C. (2012). Growing state-federal conflicts in environmental policy: The role of market-based regulation. *Journal of Public Economics 96*, 1092–1099.

Yergin, D. and R. Inesin (2009, 2 November). America's natural gas revolution. *Wall Street Journal Op-Ed*.