UNIVERSITY OF CALGARY

THREE ESSAYS ON ENERGY ECONOMICS

by

Xiaoli Zheng

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ECONOMICS

CALGARY, ALBERTA
JUNE, 2018

Abstract

This dissertation investigates the effect of a technology innovation, hydraulic fracturing, on the North American energy market and social welfare from both theoretical and empirical perspectives. Chapter 1 provides a theoretical framework to explain why the optimal path of oil and gas production by hydraulic fracturing can be different from conventional production. I argued that Hydraulic fracturing has substantially reduced production uncertainty since the shale boom. I model production uncertainty as a random shock that continuously shifts realized productions from proved reserves. With this new feature incorporated into a standard extraction model, I conclude that the reduced production uncertainty caused by hydraulic fracturing affects the optimal path of expected production through two channels. First, the expected production is tilted from the future towards the present as producers become less patient. Second, it limits the increases in marginal costs in response to production shocks. Both of these two channels would accelerate expected extractions and hence proved reserves would deplete more rapidly by hydraulic fracturing.

Chapter 2 examines the impact of the recent boom in U.S. oil production, caused by hydraulic fracturing, on the downstream U.S. petroleum refining industry. By using a panel dataset of 16 U.S. independent refinery firms from 2005 to 2015, the empirical evidence suggests that the shale boom caused by hydraulic fracturing was associated with an increased refinery profitability in U.S. Gulf Coast region, but leaving few benefit for U.S. gasoline and diesel consumers.

Chapter 3 explores the effect of shale boom on three important measures of the natural gas market: the rig count of active drilling rigs, production, and storage. By using a large panel dataset including 48 U.S. states and 10 Canadian provinces, I demonstrate that the expansion of hydraulic fracturing has resulted in a significant change in the natural gas spot and storage markets. In the spot market, production from existing wells has become more elastic to spot prices with the expansion of hydraulic fracturing. Moreover, with drilling also being more elastic to future prices and more productive with the expansion of hydraulic fracturing, the production from newly estab-

lished wells has become more elastic to price. Simultaneously, in the storage market, I document adjusting natural gas storage has become more flexible with the expansion of hydraulic fracturing. Since storage is primarily used to meet the demand shocks from temperature variations, hydraulic fracturing has significantly decreased the impact of temperature shocks. As a result, consumers can benefit from shale boom by avoiding paying abruptly raised prices for natural gas in the short run.

Acknowledgements

I am deeply indebted to my supervisor, John Boyce, for his guidance and patience that lead me to complete this dissertation. His passion on natural resource economics inspired me to continue to work in this field. From the very first day I worked with him, he was very generous with his knowledge, time and encouragement. I cannot complete this dissertation without his advice and support.

My appreciation also extends to W.D. Walls and Lucija Muehlenbachs, who followed my performance and provided me important suggestions and comments. I am thankful for Herb Emery who allows me to hone my skills in empirical analysis by handing administrative data.

At the end, I would like to thank my fiancée and parents, for their tolerance and understanding.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1

Production Uncertainty and Shale Boom

1.1 Introduction

Studies on exhaustible resource markets based on Hotelling (1931)'s seminal paper have appeared since 1970s. The is due to an increasing public concern about the 1973 oil embargo and the resultant energy crisis. Research in this field has thrived in the mid-2000s since an unexpected demand shock from developing countries pushed the oil price more than \$100/bbl. Until mid-2014, the fear of "peak oil" still prevailed. Nevertheless, the recent shale oil boom due to technology advances has changed this narrative and has dramatically impacted U.S. and world energy markets. By 2015, tight oil and shale gas accounted for nearly 50% of U.S. crude oil and natural gas productions. During 2010-2015, U.S. monthly crude oil production averaged 219 million barrels per month, compared to the 2004-2009 average of 158 million barrels per month. Similarly, in 2010-2015, U.S. monthly natural gas production averaged 2020 billion cubic feet per month, compared to the 2004-2009 average of 1600 billion cubic feet per month. Figure 1.1 and 1.2 illustrated that the increases in U.S. oil and gas production was mainly driven by hydraulic fracturing from U.S. shale formations.

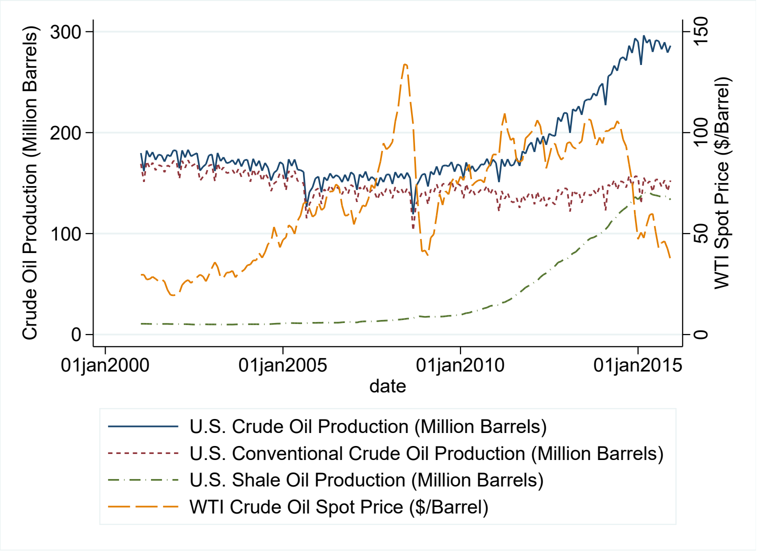


Figure 1.1: U.S. crude oil productions and prices

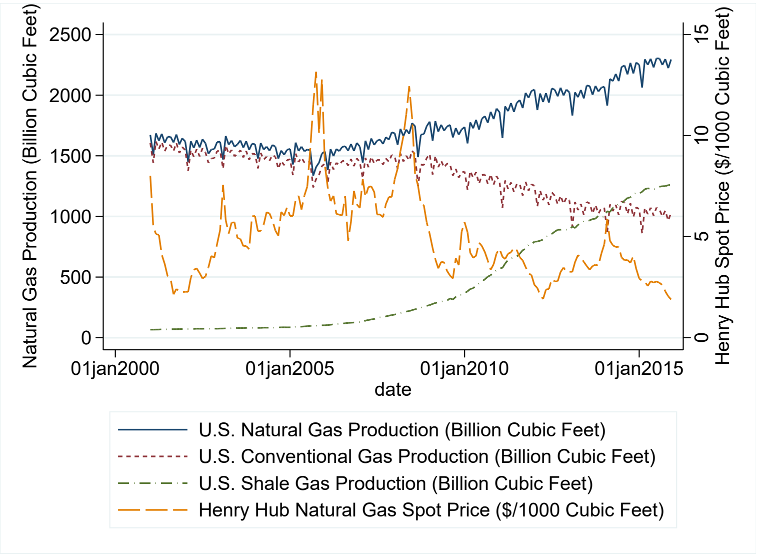


Figure 1.2: U.S. natural gas productions and prices

Despite the fast growth and spread of shale production, research on the behaviors of shale producers remain scarce. A few empirical studies recently have documented that unconventional oil and gas production by hydraulic fracturing is considerably different from conventional production. For example, Sandrea and Sandrea (2014) report that shale production usually depletes at a faster rate than conventional production. The declining production in shale plays could occur as early as

3-5 years since the initial production, in comparison to 9-12 years for conventional production with similar amounts of proved reserves. Newell et al. (2016) analyze a large dataset of conventional and unconventional natural gas wells in Texas. They find over the 2010-2014 period, the average initial production (i.e., the production in the first calendar month) for unconventional wells was 2.7 times as large as the average for conventional well.

Although previously stated empirical results offer useful sights, there is a gap in understanding as to why shale producers behave differently from conventional producers. In the context of the Hotelling model, this question is related to how hydraulic fracturing alters the optimal path of crude oil and natural gas production. This chapter attempts to shed light on this question. To accomplish it, I incorporate a new feature, an uncertainty in production, into the classic extraction model. We define the production uncertainty as a random variable that shifts the realized extraction at each period. It is important to focus on this aspect because uncertainty in U.S. oil and gas production is now substantially lower due to the expansion of hydraulic fracturing. Evidence that supports this claim is discussed in section 2. I therefore am interested in investigating how this feature affects the equilibrium path of oil and gas extractions and the implications of the current shale boom.

This chapter builds on literature on uncertainty in the market of exhaustable resources. For example, Kemp and Long (1980) assume that the amount of proved reserves is unknown at each period of extraction. Hence, similar to a driver driving without a gas gauge, producers tend to extract less resources than in the deterministic case in order to prevent reserves from suddenly depleting over a certain period. Arrow and Chang (1982) discuss a jump process of resource discovery. They find the shadow price of proved reserves move in a random cycle, thereby casting doubts on the Hotelling r-percent rule. Pindyck (1980) approaches uncertainties in a different manner. In his analysis, a random variables follows a specific stochastic process that is multiplicative to the demand or discovery of exhaustable resources. By this specification, he concludes that if the marginal extraction cost is constant, volatilities on demand or discovery have no effect on the price dynamics and thus the Hotelling r-percent rule is still applicable. However, volatilities do have an impact on the optimal path of expected production and discovery. By using a similar stochastic process, Mason (2012) explains how demand volatility causes holding of crude oil inventories.

This chapter offers new insights regrading uncertainty by modeling uncertainty in production within a simple stochastic framework. I consider a random shock similar to the specification of Pindyck (1980) and Mason (2012). I allow an uncertainty to occur at the beginning of each period. This uncertainty accounts for differences between planned and realized production at the end of each period. There are two major characteristics that distinguish my approach from previous works. First, I assume that a random variable shifts production rather than reserves or discovery. Second, unlike the pervious works where the interest revenue of contemporaneous profit is solely determined by an interest rate, I find that the production uncertainty alters the interest revenue of contemporaneous profit. This in turn affects the optimal path of extraction. The results indicate that a lower production volatility accelerates extractions through two channels. On the one hand, the representative firm becomes less patient and thus shift the productions from the future towards the present. On the other hand, a lower production volatility limits the increase in the marginal cost of extraction in response to production shocks. As a result, the representative firm spends less in adjusting its production. I argue the production uncertainty in U.S. oil and gas industry has been lower since the shale boom. Therefore, the results suggest that with the same amounts of proved reserves, oil and gas extractions from shale formation would be at a higher level initially and then deplete at a faster pace than conventional extractions.

The chapter unfolds as following: section 1.2 briefly describes the production uncertainty in U.S. oil and gas industry, with a particular focus on its relation to hydraulic fracturing; section 1.3 presents the theoretical model; section 1.4 is the conclusion.

1.2 Production Uncertainty in U.S. Oil and Gas Industry

Oil and gas extractions have been considered a high-risk business since its inception. The estimation of field production is subjected to various levels of uncertainty, such as reservoir pressure and permeability and rock compressibility (Suslick et al., 2009). These uncertainties occur because in practice the reservoir geology is known only in a probabilistic sense. For shale resources, geologist and engineers have known since the 1950s that there are considerable amounts of carbohydrate trapped in the shale formation. However, substantial uncertainties in extractions prevented producers from the commercializing productions.[1] For example, before hydraulic fracturing was introduced in the early 2000s,[2] many wells in the Bakken field, North Dakota turned out to be dry holes and produced no more than 200 barrels a day. With hydraulic fracturing, an energy company, Brigham Exploration in Bakken field, tried a multi-stage, two-mile-long horizontal well in September 2008, which no one had ever attempted to do before. This important innovation allowed the producers to access the resources underground in an unprecedented scope and hence substantially reduced the production uncertainty. As many companies used this innovation as a template, the Bakken has turned into a giant oil field since then. Bud Brigham, the founder of Brigham Exploration described this phenomenon, saying:

> *"We used to drill wildcat wells, with one-in-ten chance of success. When three-dimensional seismic mapping came along in the late 1990s, we drilled hundreds of wells with a 70%-80% success rate (i.e., 20% - 30% of wells were dry holes). Now with this technology of hydraulic fracturing, we drilled over a hundred consecutive successful wells in North Dakota. These horizontal, long, lateral, and multi-stage hydraulic fracturing wells are averaging 2,800 barrels a day, without a dry one and all commercial. I mean, in our business historically if you said you will average 2,800 barrels a day, they would say you are crazy and that would never happen. But it has happened."*
>
> —Gold (2014c), pp. 59-60

Similarly, Newell et al. (2016) argue conventional oil and gas investments resemble "trophy

---

[1] An oil and gas well or field that produces enough oil or gas to cover essential costs is considered as commercially viable.

[2] About 99% of the fracturing fluid consists of water and the rest is chemical in the modern hydraulic fracturing. However, before 2000, the industry applied a mixed gel of crude oil, sand, and chemicals to crack shale that was not only expensive but would also probably gum up the shale cracks, leaving no room for oil and gas molecules to escape out of the low-permeability shale.

hunting", with high risks compensated by high rewards; however, modern unconventional extraction from shale resembles a "manufacturing process" in which firms are much more certain over their productions.

1.3 The Model

We consider an intertemporal dynamic model in which the representative firm maximizes the lift-time profit by choosing the production $Q(t)$. The objective function $J(t)$ is specified as equation (1.1) below,

$$\begin{aligned} J = \max_{\{Q(t)\}} \quad & E_t \int_0^T \pi(\tau) d\tau \\ = \max_{\{Q(t)\}} \quad & E_t \int_0^T e^{-r\tau} \{P(\tau)Q(\tau) - C\big[Q(\tau)\big]\} d\tau, \end{aligned} \tag{1.1}$$

where $\pi(t)$ is the present-value profit of extraction at time t; $P(t)$ is the oil or gas price in a competitive market; $Q(t)$ is the production level the firm chooses at the beginning of time t; and $C\big[Q(\tau)\big]$ is twice-differentiable production cost such that $C'(Q) > 0$ and $C''(Q) > 0$. The objective function is subjected to three constraints as equation (1.2), (1.3) and (1.4) noted below,

$$\frac{dR(t)}{dt} = -\theta(t)Q(t), \tag{1.2}$$

$$\frac{d\theta(t)}{\theta(t)} = \sigma dz(t). \tag{1.3}$$

$$\frac{dP(t)}{P(t)} = \alpha dt + \eta d\nu(t). \tag{1.4}$$

Equation (1.2) states that proved reserves $R(t)$ deplete at the rate of realized production $\theta(t)Q(t)$ which is the product of the production shock $\theta(t)$ and the planned production $Q(t)$. $\theta(t)$ follows a stochastic process as equation (1.3). The law of motion of $\theta(t)$ is specified by a geometric Brownian motion without a trend, which has a constant volatility σ measuring the degree of uncertainty.

$z(t)$ is a Wiener process such that $dz(t) \sim N(0,1)$. If $\sigma = 0$ and thus $\theta(t) = 1$, there is no uncertainty on productions.[3] $\theta(t) > 1$ implies a positive shock because the realized production is more than the planned one. In contrast, $\theta(t) < 1$ denotes a negative shock. Equation (1.4) describes that the oil or gas price follows a geometric Brownian motion with a trend, where $\nu(t)$ is another Wiener process such that $d\nu(t) \sim N(0,1)$. Pindyck (1999) shows that oil prices over long run (more than 120 years) can be considered by a geometric Brownian motion with a trend by equation (1.4). For simplicity, the two Wiener processes are assumed to be uncorrelated such that $dz(t)d\nu(t) = 0$ and hence $dP(t)d\theta(t) = 0$.

Writing (1.1) in terms of its optimized value as a function of all state variables $J(t,R,\theta,P)$, and breaking the integral into a part where profits are constant over a small interval of time Δt and where the optimized profits from time $t+\Delta t$ yields:

$$\begin{aligned}
J(t,R,\theta,P) &\approx \max_{\{Q(t)\}} E_t\Big[\pi(t)\Delta t + J(t+\Delta t, R+\Delta R, \theta+\Delta\theta, P+\Delta P)\Big] \\
&\approx \max_{\{Q(t)\}} E_t\Big[\pi(t)\Delta t + J(t,R,\theta,P) + J_t(t,R,\theta,P)\Delta t + J_R(t,R,\theta,P)\Delta R \\
&\quad + J_\theta(t,R,\theta,P)\Delta\theta + J_P(t,R,\theta,P)\Delta P + \frac{1}{2}J_{\theta\theta}(t,R,\theta,P)(\Delta\theta)^2 + \frac{1}{2}J_{PP}(t,R,\theta,P)(\Delta P)^2\Big] \\
0 &= \max_{\{Q(t)\}} \Big\{\pi(t)\Delta t + J_t(t,R,\theta,P)\Delta t + J_R(t,R,\theta,P)\Delta R + J_P(t,R,\theta,P)\alpha\Delta t \\
&\quad + \frac{\sigma^2}{2}\theta(t)^2 J_{\theta\theta}(t,R,\theta,P)\Delta t + \frac{\eta^2}{2}P(t)^2 J_{PP}(t,R,\theta,P)\Delta t\Big\},
\end{aligned}$$

where the second line writes $J(t+\Delta t, R+\Delta R, \theta+\Delta\theta, P+\Delta P)$ as a second-order Taylor's series approximation about the point (t,R,θ,P). Note that J_i is the partial derivative of J with respect to i where $i \in \{t,R,\theta,P\}$ and J_{ii} is the second derivative.[4] The third line cancels $J(t,R,\theta,P)$ from each side and uses $E_t[\Delta\theta] = 0$, $E_t[\Delta P] = \alpha\Delta t$, $(\Delta\theta)^2 = \sigma^2\theta(t)^2 dt$, and $(\Delta P)^2 = \eta^2 P(t)^2 dt$.[5] Dividing the third line through by Δt and then taking the limit as $\Delta t \to 0$ and substituting from (1.2), the

---

[3] This is because the solution to $\theta(t)$ is $\theta(t) = e^{-1/2\sigma^2 t + \sigma z(t)}$ so that $\theta(t) = 1$ when $\sigma = 0$. See Øksendal (2003), page 64.

[4] In the second line, we also use the fact that $\left[dR(t)\right]^2 \approx (\Delta R)^2 = 0$, $dR(t)\cdot dP(t) \approx \Delta R \cdot \Delta P = 0$, $dR(t)\cdot d\theta(t) \approx \Delta R \cdot \Delta\theta = 0$, and $dP(t)\cdot d\theta(t) \approx \Delta P \cdot \Delta\theta = 0$. This is due to the properties of Winner processes such that $(dt)^2 = 0$, $dt \cdot d\nu(t) = 0$, $dt \cdot dz(t) = 0$, and $d\nu(t)\cdot dz(t) = 0$.

[5] We apply Ito's relationships that $\left[dz(t)\right]^2 = dt$ and $\left[d\nu(t)\right]^2 = dt$.

fundamental equation of stochastic optimality in this problem is shown as noted below:[6]

$$0 = \max_{\{Q(t)\}} \Big\{ e^{-rt}\Big[P(t)Q(t) - C\big[Q(t)\big]\Big] + J_t(t,R,\theta) - \theta(t)Q(t)J_R(t,R,\theta,P) + \alpha J_P(t,R,\theta,P)$$
$$+ \frac{\sigma^2}{2}\theta(t)^2 J_{\theta\theta}(t,R,\theta,P) + \frac{\eta^2}{2}P(t)^2 J_{PP}(t,R,\theta,P)\Big\}. \tag{1.5}$$

Differentiating equation (1.5) with respect to $Q(t)$ yields,

$$e^{-rt}\Big\{P(t) - C'\big[Q(t)\big]\Big\} = \theta(t)J_R(t,R,\theta,P). \tag{1.6}$$

Equation (1.6) states in equilibrium the present values of the extraction rent is equal to the shadow value of proved reserves. Note that this shadow value $J_R(t,R,\theta,P)$ is subjected to random fluctuations caused by the production shock $\theta(t)$. To derive the optimal path of production, we need to eliminate $J(t,R,\theta,P)$ in equation (1.6). First, we notice that the rate of changes in the expected shadow value of the proved reserves is zero according to the proposition 1.1.

Proposition 1.1 *The rate of changes in the expected shadow value of the proved reserves is zero. Therefore,*

$$\frac{1}{dt}E_t d\big[J_R(t,R,\theta,P)\big] = 0. \tag{1.7}$$

Proof *Proof of Proposition 1.1 is provided in Appendix A.*

Dividing through $\theta(t)$ on equation (1.6) and then applying Ito's differential operator $1/dtE_td(\bullet)$ yields,[7]

$$\frac{1}{dt}E_t d\Big\{\frac{e^{-rt}\{P(t) - C'\big[Q(t)\big]\}}{\theta(t)}\Big\} = \frac{1}{dt}E_t d\big[J_R(t,R,\theta)\big]. \tag{1.8}$$

---

[6] Kamien and Schwartz (1991) provide a general derivation of the fundamental equation of stochastic optimality. See Kamien and Schwartz (1991), pp. 267-268.

[7] Ito's differential generator is analogous to the time derivatives in the deterministic case. For its mathematical discussion, see Chow (1979).

However, Proposition 1.1 implies the right-hand side of equation (1.8) is zero. Thus, equation (1.8) becomes,

$$\frac{1}{dt}E_t d\left\{\frac{e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)}\right\}=0. \tag{1.9}$$

Based equation (1.9), Proposition 1.2 below describes the optimal path of expected productions.

Proposition 1.2 *The optimal path of expected production $E_t[Q(t)]$ is:*

$$\frac{1}{dt}E_t d\left[Q(t)\right]=\frac{1}{C''\left[Q(t)\right]}\left\{-(r-\sigma^2)\left\{P(t)-C'\left[Q(t)\right]\right\}+\sigma^2(C'\left[Q(t)\right]\xi_{mc\theta})+\alpha P(t)\right\}, \tag{1.10}$$

where $\xi_{mc\theta}$ is the elasticity of the marginal cost of extraction $C'\left[Q(t)\right]$ with respect to the production shock $\theta(t)$, defined by $\xi_{mc\theta}=\frac{\theta(t)}{C'\left[Q(t)\right]}\frac{\partial C'\left[Q(t)\right]}{\partial\theta(t)}$.

Proof *Proof of Proposition 1.2 is provided in Appendix B.*

With a proper functional form of $C'\left[Q(t)\right]$, equation (1.10) can be solved for the expected productions $E[Q(t)]$ and then terminal period T by using $R(0)=\int_0^T E[Q(t)]dt$. To see the effect of production uncertainty, let us consider the optimal path of expected productions without production uncertainty (i.e., $\sigma^2=0$ and thus $\theta(t)=1$) as noted below,

$$\frac{1}{dt}E_t d\left[Q(t)\right]=\frac{1}{C''\left[Q(t)\right]}\left\{-r\left\{P(t)-C'\left[Q(t)\right]\right\}+\alpha P(t)\right\}, \tag{1.11}$$

which is a simplified version of the optimal path of expected production derived by Pindyck (1980).[8] Comparing equation (1.10) to (1.11) reveals the effect of production uncertainty. The first terms in the bracket of equation (1.10) and (1.11) suggest that the presence of production volatility σ^2 changes the interest revenue of instantaneous profit from $\left\{-r\left\{P(t)-C'\left[Q(t)\right]\right\}\right\}$ to $\left\{-(r-\sigma^2)\left\{P(t)-C'\left[Q(t)\right]\right\}\right\}$. The second terms in the bracket of equation (1.10) and (1.11)

---

[8] See equation (18) at pp. 1211 of Pindyck (1980).

describe the effect of the production volatility σ^2 on marginal cost of extraction. As shown by the context between equation (B.8) and (B.10) in Appendix B, the elasticity of marginal cost with respect to production shocks, $\xi_{mc\theta}$, is more likely to be positive. As a result, $\Big\{C'\big[Q(t)\big]\xi_{mc\theta}\Big\}$ in the second term of equation (1.10) represents the extra amount of marginal costs the firm has to pay if the production shock rises by 1%. However, this amount of extra costs is weighted by the volatility σ^2.

Consider the emerging hydraulic fracturing lowers production volatility σ^2. The reduced production uncertainty affects the optimal path of production through two channels. First, the interest revenue of contemporaneous profit, $\Big\{-(r-\sigma^2)\big\{P(t)-C'\big[Q(t)\big]\big\}\Big\}$, becomes larger. As a result, it becomes more desirable to extract an additional unit of resources currently. Therefore, the firm tends to be less patient and thus tilt productions from the future towards the present. Second, the lower volatility σ^2 is, the less amount of extra costs the firm has to pay in order to adjust productions. This is because $\Big\{C'\big[Q(t)\big]\xi_{mc\theta}\Big\}$ in the second term of equation (1.10) becomes smaller. In other words, a lower production uncertainty limits the increases in the marginal cost in response to production shocks. Consequently, all else equal, the expected production would accelerate and decrease more rapidly with a lower σ^2.

1.4 Conclusion

This chapter provides a theoretical framework to analyze the effect of production uncertainty on the optimal path of crude oil and natural gas productions. I focus on production uncertainty because the recent expansion of hydraulic fracturing has substantially lowered the production uncertainty in U.S. oil and gas industry. I model the production shocks by a Geometric Brown Motion that continuously shifts realized productions from proved reserves. With this feature incorporated into a standard extraction model, I find the reduced production uncertainty affects the optimal path of expected production through two channels. First, the firm becomes less patient and tilt productions towards the present. Second, a lower production uncertainty limits increases in marginal cost in response to production shocks. Both channels imply that a lower production volatility causes the the representative firm to speed up productions. Consequently, my result suggests that crude oil and natural gas production from shale would be at a higher level initially and then depletes more rapidly than unconventional extraction.

Chapter 2

Shale Oil Boom and the Profitability of U.S. Petroleum Refiners

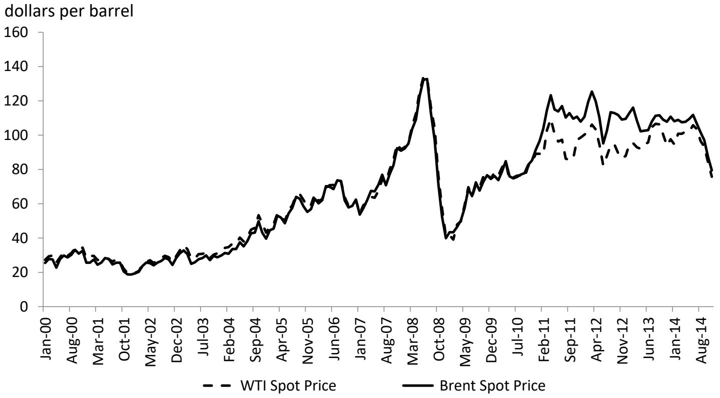
2.1 Introduction

Increased shale oil production since 2011 has had several important effects on refinery markets in North America. First, due to U.S. crude oil export ban and pipeline constraints, the increasing production from shale oil in Midwest has resulted in excess supply at the Midwest and Mid-Continent market locations. Since 2011 the domestic oversupply has resulted in the WTI price, the North American benchmark price which is tied to light sweet crude oil, being depressed relative to Brent Blend, the international benchmark crude oil price (see Figure 2.1 below). The increasing abundance in domestic supply has rekindled policymakers interest in whether or not the U.S. should lift the ban on crude oil exports.[1] Some stakeholders are strongly against lifting ban, with refiners claiming that lifting ban on crude oil export would be expected to increase domestic crude oil price and then harm their incentive to produce gasoline and diesel (GAO, 2014). According to this argument, the reduced incentives to refine may result in rising finished product prices.

---

[1] In 1975, Congress prohibited exporting crude oil without a license in response to rising prices and fears of shortages following an oil embargo by several OPEC member countries. Licenses are granted each year for very small quantities, mostly to Canada. Until December 2015, President Obama singed the legislation to finally lift the ban on crude oil export.

Figure 2.1: Brent and WTI prices (Unit: dollar per barrel)



Rising domestic oil supply resulting from U.S. shale production has stimulated much interest in the potential removal of the ban on crude oil exports.[2] Brown et al. (2014) argue that lifting the ban on crude oil exports may lower domestic gasoline prices because foreign refiners are better suited to refine light sweet crude oil than are U.S. refiners. Borenstein and Kellogg (2014) find that depressed crude oil prices in Midwest did not affect gasoline and diesel prices in the region during the period 2006–2011. These studies suggest that since the U.S does permit exports of refined products—and because refiners will charge the same fuel price in domestic as in world markets adjusted for transportation costs—rents from lower crude prices in the Midwest would have accrued mainly to refiners in those areas. U.S. gasoline and diesel fuel markets have remained integrated with the global markets, even if the global market for crude oil has fragmented (Kilian, 2014; Zavaleta et al., 2015). In some ways the removal of the Alaska North Slope (ANS) oil export ban in 1995 offers lessons for the potential effects of removing the oil export ban nationally: the effect of removing the ban on ANS exports in 1995 led to an increase in ANS crude oil prices relative to other comparable crude prices; however, gasoline and diesel prices in West Coast markets did not increase correspondingly (Bausell et al., 2001).

---

[2] See, for example, the recent stories in the popular business press (Gold, 2014b) or the recent U.S.GAO (2014) study that reviews related empirical studies and interviews different groups of stakeholders regarding the issue of crude oil export. This study points out whether the fuel prices would increase or decrease following the removal of crude export ban might differ across regions. Some market participants think that allowing crude export would lower international crude prices which would pass through into lower refinery product prices imported to U.S.

Has domestic refinery profitability been impacted by domestic crude prices as a result of the boom in shale oil production? Has lower domestic crude prices relative to international crude oil prices passed through to domestic refined product prices? In this chapter, I provide answers to these questions through an analysis of regional and firm-level data. I confirm that U.S. gasoline and diesel fuel prices are determined by international rather than domestic crude prices and that there is limited pass-through from domestic crude oil prices to finished products price. These results are consistent across most regions except Gulf Coast before and after 2011. Therefore, lifting the ban on crude oil exports would not lead to dramatic increases in domestic gasoline and diesel prices. In fact, permitting exports of U.S. crude oil could actually result in *lower* domestic finished product prices. The results also suggests that independent refiners are impacted by domestic crude oil prices after 2011 but their operating incomes (before depreciation) did not correlate with domestic crude oil prices before 2011. Consequently, there is little evidence to support the argument that refiners would suffer a great loss if domestic crude oil prices were integrated with international benchmark oil prices as was the case before 2011.

2.2 Analysis of Refiners' Financial Performance

I examine in this section how changes in U.S. domestic crude oil prices are associated with changes in U.S. refiners' operating incomes. The regression model used to quantify this relationship is

$$\Delta OIB_{it} = CONS + \eta \Delta WTI_t + \varepsilon_{it}. \tag{2.1}$$

where OIB_{it} represents the operating income before depreciation of refiner i at time t and WTI_t represents the U.S. domestic crude oil prices; ΔOIB_{it} and ΔWTI_t are their first differences. Since all the variables are first-differentiated, equation (2.1) is similar to a fixed effect model, while time-invariant variables such as firm-specific and region-specific dummies cannot be identified. The advantage of first differences is equation (2.1) can be parsimoniously estimated since firm-idiosyncratic and regional idiosyncratic variables are eliminated out. I expect the coefficient η to be negative if the reduction in domestic crude oil prices could significantly raised the independent refiners' incomes.

2.2.1 Data Description

The data are quarterly firm-level observations on 16 independent refiners operating exclusively in the U.S. from 2005–Q1 to 2014–Q3. The list of refinery companies is from an U.S. Energy Information Administration (EIA) refinery report issued in January 2014 (EIA, 2014a), which identifies each firm's operating capacity and location. The firms in the sample are independent refiners to ensure that their operating incomes are not confounded with the upstream crude oil business. Therefore, refiners such as BP and Chevron are not included. I include publicly-listed companies traded on North American financial markets to have access to quarterly data on operating incomes before depreciation. To ensure no independent refiner is omitted, I screen all the listed firms with SIC code as 2911, petroleum refining industry, from the COMPUSTAT database. Combining these data with that EIA refinery report, I am able to identify 16 independent refinery companies as reported in Table 2.1. The aggregate capacity of those refiners accounts for 45.3% of total operable capacities in U.S. as of January 2014. Half of the firms listed in Table 2.1 operated refineries in more than one region during the sample period. In this chapter, I divide the U.S. by 5 major regions in terms of Petroleum Administration for Defense Districts (PADD) [3], where PADD1 denotes East Coast, PADD2 denotes Midwest, PADD3 denotes Gulf Coast, PADD4 denotes Rocky Mountain, and PADD5 denotes West Coast.

---

[3] During World War II the Petroleum Administration for War, established by an Executive order in 1942, used these five districts to ration gasoline. U.S. Congress passed the Defense Production Act of 1950, which created the Petroleum Administration for Defense and used the same five districts, called the Petroleum Administration for Defense Districts (PADD).

Table 2.1: Petroleum Refiners included in the Sample

Refiners	Ticker Symbol	Operating PADDs
United Refining	0123A	1
Alon Israel Oil Company Ltd	ALJ	3, 5
Calumet Specialty Products Partner	CLMT (NASDQ)	2, 3, 4
CVR Refining LP	CVRR	2
Delek US	DK	3
Genesis Energy LP	GEL	4
Hollyfrontier Corporation	HFC	2, 3, 4
Kinder Morgan Energy Partners LP	KMP	3
Marathon Petroleum Corp	MPC	2 and 3
Nustar Energy LP	NS	1
Northern Tier Energy	NTI	2
Pbf Energy Co LLC	PBF	1, 2
Phillips 66	PSX	1, 2, 3, 4, 5
Tesoro Corp	TSO	2, 4, 5
Valero Energy Corp	VLO	2, 3, 5
Western Refining Inc.	WNR	3

WTI prices were obtained from the EIA. Data on operating incomes (before depreciation), defined as operating sales revenues minus operating expense by excluding extraordinary items, were extracted from COMPUSTAT.[4] Table 2.2 provides the summary statistics.

---

[4] In the COMPUSTAT database, the operation income before depreciation is equal to quarterly earnings before interest, taxes, depreciation and amortization (EBITDA), an approximate measure of a company's operating cash flow during a quarter.

Table 2.2: Summary Statistics on Operating Income before Depreciation (unit: million dollar)

Variable		Mean	Std Dev	Min	Max	Obs
OIB_{it}	overall	252.99	434.29	-448.00	3193.00	N=505
	between		342.37	11.92	1118.66	n=16
	within		282.00	-1002.40	2327.33	$\bar{T}$=31.56

2.2.2 Estimation Results

The estimation results for equation (2.1) are presented in Table 2.3 for sample periods 2005-Q1–2010-Q4 and 2011-Q1–2014-Q3, where the year 2011 is considered as a structural break of U.S. Crude oil productions (see Figure 2.2 below). These results suggest the changes in WTI are not associated with the changes in firms' financial performance before 2011. However, changes in WTI prices significantly impact refiners' operating incomes since 2011. This is strong evidence that refiners are able to capture rent from lower domestic crude oil prices.

Figure 2.2: U.S. Crude Oil productions (Unit: million barrels)

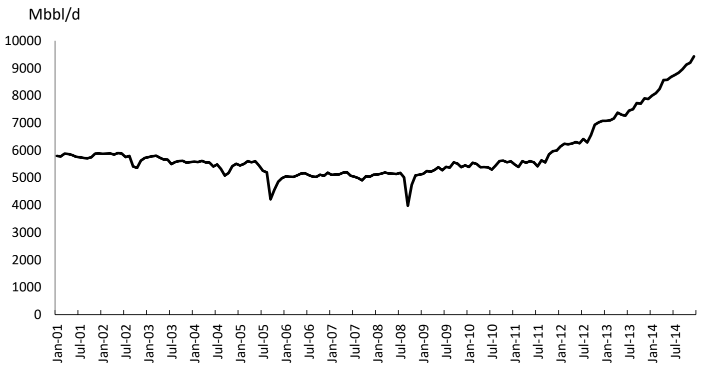


Table 2.3: Estimation Results for equation (2.1)

	$\Delta(OIB_{it})$ (2005Q1-2010Q4)	$\Delta(OIB_{it})$ (2011Q1 -2014Q3)
ΔWTI	0.893	-11.555***
	(1.102)	(3.114)
Constant	-3.225	25.541
	(17.302)	(22.776)
N	257	231
R^2	0.003	0.057
Durbin-Watson statistic	2.610	2.435
P value of Breusch-Pagan test for heteroskedasticity	0.509	0.351

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$.

2.2.3 Level Equation

Although the estimated result of equation (2.1) provides some evidences on how the correlation between domestic crude oil prices and independent refiners' incomes changes before and since 2011, it omits the impact of the regional differences due to the model specification. Regional differences in the impacts of shale oil boom on refinery markets are evident. For example, Wall Street Journal (Gold, 2014a) reports the weekly average numbers of oil trains from North Dakota to each county around U.S. It shows the various intensities of shale oil shipments to refiners in different regions. Considering these regional differences, it appears prudent that I examine these effects with some control for region. Therefore, to quantify the extent that refiners benefit from the changes of domestic crude oil prices, I estimate the following level equation:

$$ln(OIB_{it}) = CONS + \lambda ln(WTI_t) + \sum_{i=1}^{5} \gamma_i PADD_i + \varepsilon_{it} \tag{2.2}$$

The coefficient λ is interpreted as the elasticity of refinery operating income (before depreciation) with respect to WTI prices. $PADD_i$ is a dummy variable which equals 1 if a refiner operates business in PADD i and zero otherwise. Equation (2.2) is estimated by a random effect model with Newey-West standard errors that is robust to general forms of autocorrelation and heteroscedasticity over panels. A robust Hausaman test [5] is conducted to confirm that the random effect model cannot be rejected at 5% for both two sub-samples.

The results in Table 2.4 reject the hypothesis that domestic crude prices can affect refinery operating income before 2011 at 5%. Nevertheless, the results in Table 2.4 suggests that since 2011 1% drop in domestic crude prices is significantly associated with 3% increase in independent refinery operating income before depreciation. This provides a remarkable evidence that independent refiners benefit from cheaper domestic crude oil since 2011. Moreover, the estimated coefficients of *PADD* dummies indicates how the independent refiners' profitability varies before and since the shale oil boom. At first place, I find that the coefficient of $PADD_2$ in the period of 2005 to 2011 is smaller than it is in the period of 2011 to 2014. This suggests independent refiners operating in Midwest are less profitable after 2011 than they used to be, given that more and more shale oil can be shipped from the Midwest to the Gulf Coast by railway and the direction of pipeline flow between Midwest and Gulf Coast reverses after 2011.[6] This finding is also consistent with Borenstein and Kellogg (2014) who argue that the primary beneficiaries of depressed WTI prices are Midwest refiners during the years 2006–2011. In contrast, the estimated coefficients of $PADD_4$ indicates the refiners' profitability in this region is better since the shale oil boom. This is because PADD4 is connected to the Gulf Coast refiners only via PADD2, though PADD4 has also experienced a

---

[5] The robust Hausman test is developed by Wooldridge (2002) in order to include the likelihood that the random effect estimator may not be fully efficient in the standard Hausman test.

[6] In May 2012, Enbridge announced to reverse the direction of the major oil pipeline between Midwest and Gulf Coast: Seaway pipeline. The Seaway pipeline began carrying about 400,000 barrels per day of crude oil from Cushing to the Gulf Coast since then.

relative decrease in crude oil prices. Therefore, PADD4 crude oil prices is significantly lower than any other regions in U.S. since 2011. Moreover, the estimated coefficients of $PADD_1$, $PADD_3$, and $PADD_5$ are insignificant before 2011. However, they become highly significant and positive since 2011. This indicates refiners' profitability in these three regions have been improved, thank to expanded railway and pipeline capacity carrying shale oil from Midwest to these three regions. The most significant improvement of refiners' profitability appears on $PADD_3$, Gulf Coast, where 60% of U.S. refinery capacities cluster due to the waterborne access to the Atlantic. The estimated coefficient of $PADD_3$ is 1.028, which implies other things equal, an independent refiner in Gulf Coast earns \$2.8 millions more than its counterpart earns in other regions since 2011[7].

---

[7] This is calculated by $e^{1.028} = 2.8$.

Table 2.4: Estimated Results of Equation (2.2)

Variable	ln(operating income before depreciation)	
	2005-Q1–2010-Q4	2011-Q1–2014-Q4
ln(WTI)	0.463*	-3.029**
	(0.267)	(1.093)
Intercept	2.871*	17.950***
	(1.422)	(4.988)
$PADD1$	-0.888	0.316***
	(0.746)	(0.027)
$PADD2$	2.308***	1.116***
	(0.455)	(0.106)
$PADD3$	-0.615	1.028***
	(0.621)	(0.068)
$PADD4$	-2.516***	-0.584***
	(0.666)	(0.038)
$PADD5$	0.498	0.623***
	(0.381)	(0.093)
N	244	229
R^2	0.479	0.421

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$

2.3 How are U.S. Refined Product Prices Related to Crude Oil Prices?

The results in previous section suggest that U.S. independent refiners operating incomes increase by 3% when the domestic crude oil price WTI drops by 1% since 2011, but their financial performance are not associated with WTI before 2011. These results might be closely related to the pass-through effect from domestic crude prices to refined product prices. Other things equal, a refiner's financial performance is largely affected by the crack spread, the difference between refined product price and crude oil price. If the domestic crude oil price fully pass through to the refined product price, that is, one unit change in the domestic crude oil price significantly leads to one unit change in refined product prices (EIA, 2014b), it is possible the domestic crude oil prices are not associated with refiner's financial performance since the crack spread is unchanged. Another possibility is that the changes in refined product prices are completely not driven by the changes in the domestic crude oil price, but by something else, (e.g., the changes in an international crude oil price). In this case, I might conclude there is no pass-through from domestic crude oil prices to refined product prices. The last possibility is that the changes in domestic crude oil prices passed through to the changes in refined product prices to some degree (but not fully). For example, the extent to which refined product prices decrease is smaller than the extent to which domestic crude oil prices decrease. In this case, it is possible that the refiners' financial performance would have been improved.

In this section I verify that changes in U.S. refined product prices are not affected by the changes in domestic crude oil prices, but by the changes in an international crude oil price before 2011. However, since 2011 the domestic crude oil prices partially passed through to refined product prices to some degrees in the region of Gulf Coast, which allowed refiners in the region improved their financial performances.

My model to study the pass-through effects in region i is

$$\Delta G_{ist} = \alpha + \beta_1 \Delta Brent_t + \beta_2 \Delta(DOM_{it} - Brent_t) + \varepsilon_{it}. \tag{2.3}$$

where ΔG_{ist} is the first difference of gasoline or diesel prices in state s of region i at time t; ΔDOM_{it} denotes the first difference of domestic crude prices in region i and $Brent_t$ is the international crude oil price. All the price series are expressed as US dollar per gallon. I subtract $Brent_t$ from DOM_{it} to avoid multi-collinearity between domestic and international crude oil prices, so that $\Delta(DOM_{it} - Brent_t)$ solely represents the impact of domestic crude oil prices. A full pass through effect implies β_1 and β_2 are close to one (EIA, 2014b). The dataset is in a monthly frequency and available from the U.S. EIA. I estimate equation (2.3) by a state-level panel dataset for each region. The regions are East Coast, Midwest, Rocky Mountain, Gulf Coast, and West Coast.[8] The number of observation available for each estimation depends on the number of states in a region. For example, in East Coast, there are 18 states so that I have a panel dataset with more than 1000 observations. At the first step, I restrict the sample to an interval from January 2006 to December 2010, where the proportion of shale oil productions in North America was not significant. Estimated results for this period are presented in Tables 2.5 and 2.6. The standard errors are clustered on the month-of-sample to consider the possibility that refinery product prices are correlated among states within the same region.

---

[8] The domestic crude prices applied in estimation are WTI, Wyoming Sweet (WYS), Louisianan Light Sweet (LLS), Alaska North Slope (ANS) for East Coast, Midwest, Rocky Mountain, Gulf Coast, and West Coast respectively.

Table 2.5: Estimation results of equation (2.3) for the period of 2006 to 2010 with the dependent variable as the first differences on gasoline prices

Variables/ Region	East Coast	Midwest	Rocky Mountain	West Coast	Gulf Coast
$\Delta Brent_t$	1.097***	1.099***	1.095***	1.064***	1.088***
	(0.113)	(0.143)	(0.147)	(0.124)	(0.144)
$\Delta(DOM_{it} - Brent_t)$	0.210	0.319	0.415	0.810***	0.252
	(0.319)	(0.338)	(0.318)	(0.284)	(0.254)
Constant	-0.002	-0.001	-0.001	0.001	-0.002
	(0.014)	(0.018)	(0.019)	(0.015)	(0.016)
N	986	883	295	411	354
R^2	0.763	0.698	0.647	0.683	0.740
Durbin-Watson statistics	1.634	1.761	1.407	1.59	1.652
p value of Breusch-Pagan test for heteroscedasticity	0.00	0.00	0.00	0.00	0.00

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$;

DW statistics in panel data are calculated by Bhargava *et al.*'s (1982) method.

The null hypothesis of Breusch-Pagan test is constant variances.

Table 2.6: Estimation results of equation (2.3) for the period of 2006 to 2010 with the dependent variable as the first differences on diesel prices

Variables/ Region	East Coast	Midwest	Rocky Mountain	West Coast	Gulf Coast
$\Delta Brent_t$	1.037***	1.062***	1.107***	1.106***	1.066***
	(0.057)	(0.072)	(0.100)	(0.078)	(0.075)
$\Delta(DOM_{it} - Brent_t)$	0.324	0.466	0.186	0.201	0.234
	(0.260)	(0.312)	(0.276)	(0.217)	(0.209)
Constant	0.000	0.001	0.001	0.001	-0.001
	(0.009)	(0.012)	(0.016)	(0.011)	(0.010)
N	933	885	295	375	354
R^2	0.868	0.826	0.695	0.772	0.857
Durbin-Watson statistics	2.013	2.047	2.05	2.205	2.088
p value of Breusch-Pagan test for heteroscedasticity	0.00	0.00	0.00	0.00	0.00

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$;

DW statistics in panel data are calculated by Bhargava *et al.*'s (1982) method.

The null hypothesis of Breusch-Pagan test is constant variances.

Tables 2.5 and 2.6 show that changes in the international crude price, Brent, have much more power than domestic crude prices to explain the changes in gasoline and diesel prices in all regions before 2011. All the coefficients on Brent are highly significant and around unity. This indicates one unit changes in Brent completely pass through to the changes in gasoline and diesel across all the regions. With the single exception of gasoline prices in West Coast, domestic crude prices do not affect finished product prices in all the other regions before 2011.

Next, I estimate equation (2.3) from the closed interval January 2011 to October 2014, during which the shale oil boom started impacting refinery markets. The estimated results are displayed in Tables 2.7 and 2.8.

Table 2.7: Estimation results of equation (2.3) for the period of 2011 to 2014 with the dependent variable as the first differences on gasoline prices

Variables/ Region	East Coast	Midwest	Rocky Mountain	West Coast	Gulf Coast
$\Delta Brent_t$	0.957***	0.908***	0.670***	1.095***	0.960***
	(0.093)	(0.139)	(0.167)	(0.133)	(0.091)
$\Delta(DOM_{it} - Brent_t)$	0.227	0.234	0.213	0.862***	0.181*
	(0.175)	(0.252)	(0.329)	(0.300)	(0.110)
Constant	0.001	0.002	0.007	0.001	0.000
	(0.013)	(0.020)	(0.026)	(0.019)	(0.014)
N	691	690	228	230	276
R^2	0.650	0.398	0.161	0.462	0.589
Durbin-Watson statistics	1.559	1.709	1.185	1.597	1.607
p value of Breusch-Pagan test for heteroscedasticity	0.0898	0.871	0.908	0.091	0.311

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$;

DW statistics in panel data are calculated by Bhargava *et al.*'s (1982) method.

The null hypothesis of Breusch-Pagan test is constant variances.

Table 2.8: Estimation results of equation (2.3) for the period of 2011 to 2014 with the dependent variable as the first differences on diesel prices

Variables/ Region	East Coast	Midwest	Rocky Mountain	West Coast	Gulf Coast
$\Delta Brent_t$	0.919***	0.855***	0.920***	1.129***	0.912***
	(0.045)	(0.081)	(0.164)	(0.082)	(0.047)
$\Delta(DOM_{it} - Brent_t)$	0.067	0.055	0.118	0.229	0.125**
	(0.099)	(0.168)	(0.243)	(0.159)	(0.056)
Constant	0.003	0.006	0.008	0.006	0.005
	(0.008)	(0.012)	(0.020)	(0.010)	(0.007)
N	655	686	228	229	274
R^2	0.805	0.622	0.391	0.737	0.820
Durbin-Watson statistics	1.98	1.981	2	2.212	1.931
p value of Breusch-Pagan test for heteroscedasticity	0.8871	0.903	0.286	0.116	0.9

*** $P < 0.01$; ** $P < 0.05$; * $P < 0.1$;

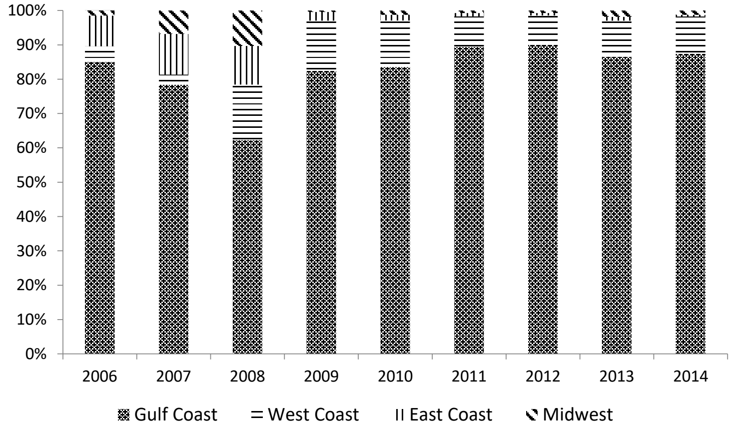
DW statistics in panel data are calculated by Bhargava *et al.*'s (1982) method.

The null hypothesis of Breusch-Pagan test is constant variances.

Comparing Tables 2.7 and 2.8 with Tables 2.5 and 2.6, several observations stand out. First, the changes in international crude price benchmark, fully passed through to the changes in domestic gasoline and diesel prices across different regions before and since 2011. Second, there is no pass-through from the domestic crude prices to gasoline and diesel prices in the East Coast, Midwest, and Rocky Mountain regions before and after 2011. More importantly, domestic crude oil prices to some degree have passed through to gasoline and diesel prices in the Gulf Coast region since 2011 because the estimated coefficients of $\Delta(DOM_{it} - Brent_t)$ are significant (Tables 2.7 and 2.8), while they are insignificant before 2011 (Tables 2.5 and 2.6). This result is further verified by the fact that the exports of gasoline and diesel in Gulf Coast have risen sharply since 2011. Because U.S. gasoline and diesel markets are integrated with world markets but U.S crude oil market is not, refiners in the Gulf Coast have taken the advantage of discounted domestic crude oil to produce refinery products and sold them with international prices. This enhances U.S refiners' competitiveness relative to their international counterparts. For this reason, the export of refinery products from the Gulf Coast has risen dramatically from that time as shown in Figures 2.3 and 2.4. Figure 2.3 graphically depicts that the share of gasoline and diesel exports from the Gulf Coast increased rapidly since 2011, while these shares from the other regions shrank. Figure 2.4 below shows the share of diesel exports from the Gulf Coast rose more sharply than the share of gasoline exports from that region. This is consistent with our finding that the pass-through effect on diesel prices is more significant than gasoline prices in Gulf Coast.[9]

On the other hand, these results also explain what I find in the analysis of last section. The fact that U.S gasoline and diesel prices are closely related to Brent rather than domestic crude oil prices explains why the domestic crude oil price has no impact on independent refiners' operating income. Furthermore, the partial pass-through effect from domestic crude oil prices to gasoline and diesel prices in Gulf Coast since 2011 explains why independent refiners benefited from cheaper domestic crude oil during that period.

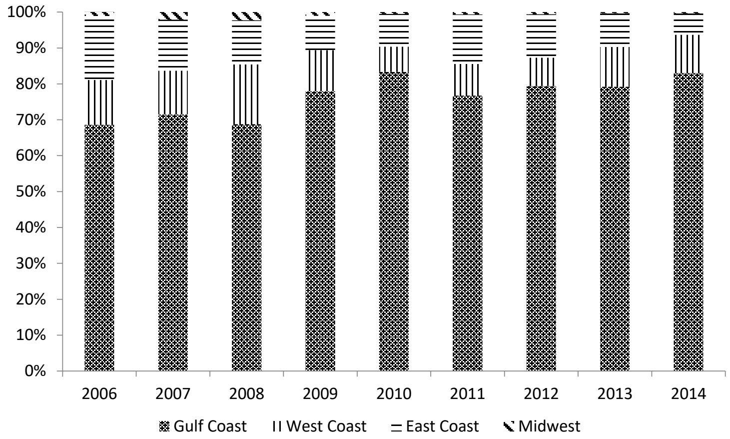
Figure 2.3: Composition of U.S. Gasoline Exports across Regions



---

[9] The coefficient of $\Delta(DOM_{it} - Brent_t)$ in the gasoline equation is at 10% significant level (Table 2.7, last column), while it is at 5% significant level in the diesel equation (Table 2.8, last column).

Figure 2.4: Composition of U.S. Diesel Exports across Regions



2.4 Conclusion and Policy Implications

This chapter has examined how the dramatic boom in U.S. domestic crude oil production has impacted the domestic petroleum refining industry. Specifically, I have quantified how independent refiner operating income has changed in response to changing domestic crude oil prices. The estimates indicate that independent refiner operating income increases by 3% for a domestic crude oil price drop of 1% since 2011. The implication is that independent refiners may have been able to capture the gains associated with lower crude oil prices. Empirical evidence confirms that U.S. gasoline and diesel prices are determined by international rather than domestic crude oil prices. Because of this, there is almost no pass-through from domestic crude oil prices to refined product prices in all U.S. regions except Gulf Coast. There was partial pass-through from the changes in domestic crude oil prices to refiner products' prices in Gulf Coast since 2011, resulting in increased refinery profitability in the region. Nevertheless, independent refiners' operating incomes were not found to be associated with domestic crude prices before 2011, a period during which U.S. domestic crude oil prices were nearly coincident with international benchmark prices.

Chapter 3

The Impact of Shale Boom on the North American Natural Gas Market

3.1 Introduction

Between 2010 and 2015, hydraulic fracturing and horizontal drilling revolutionized the U.S. and Canada petroleum and natural gas industries. According to the U.S. Energy Information Administration (EIA), the U.S. and Canada are the only two countries in the world that accomplished commercially viable natural gas from shale formations. By 2015, tight oil and shale gas accounted for nearly 50% of U.S. crude oil and natural gas production. In 2015–2016, U.S. crude production averaged 278 thousand barrels per month, almost a 70% increase over the 2000–2009 average of 164 thousand barrels per month. As well, the West Texas Intermediate (WTI) crude oil price averaged $65 per barrel during 2015–2016, compared to the 2008–2014 average of $87 per barrel. Similarly, natural gas production in 2015–2016 averaged 2,720 billion cubic feet per month, a 34% increase from 2,030 billion cubic feet per month in 2000–2009. As well, in 2015-2016, natural gas prices were $4.08 per thousand cubic feet, only 60% of their 2000–2009 average of $6.79 per thousand cubic feet. Shale oil and gas, which can be distinguished from conventional oil and gas, have played a significant role in driving the increase in total U.S. oil and gas production. These effects are shown in Figure 3.1.

In Canada, hydraulic fracturing has led to the increased production of shale gas across the country, from British Columbia to New Brunswick. By 2015, shale gas production reached 4.1 billion cubic feet per day (BCF/day), and accounted for more than 20% of total Canadian natural gas production. The National Energy Board of Canada expects this development to continue to increase and account for almost 70% of Canadian total natural gas production by 2025.[1]

These changes have reverberated throughout the industry. Natural gas consumption in the U.S. in 2015-2016 averaged 2.30 trillion cubic feet per month, which is 22% higher than its 2000-2009 average of 1.87 trillion cubic feet per month. Similarly, monthly average underground storage of natural gas has risen by 11% to 7.33 trillion cubic feet from 6.33 trillion cubic feet.[2] Storage of natural gas is of particular importance to that market because storage is primarily used to meet the large seasonal swings in demand (See Figure 3.2).[3]

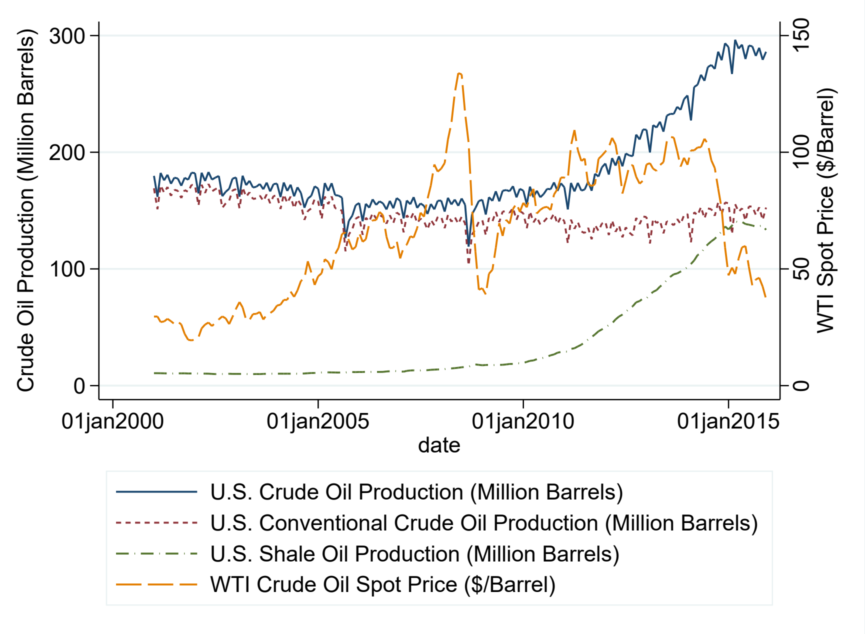
---

[1] However, due to the lack of yearly shale gas productions in Canada, I cannot plot Figure 3.1 for the Canadian version.

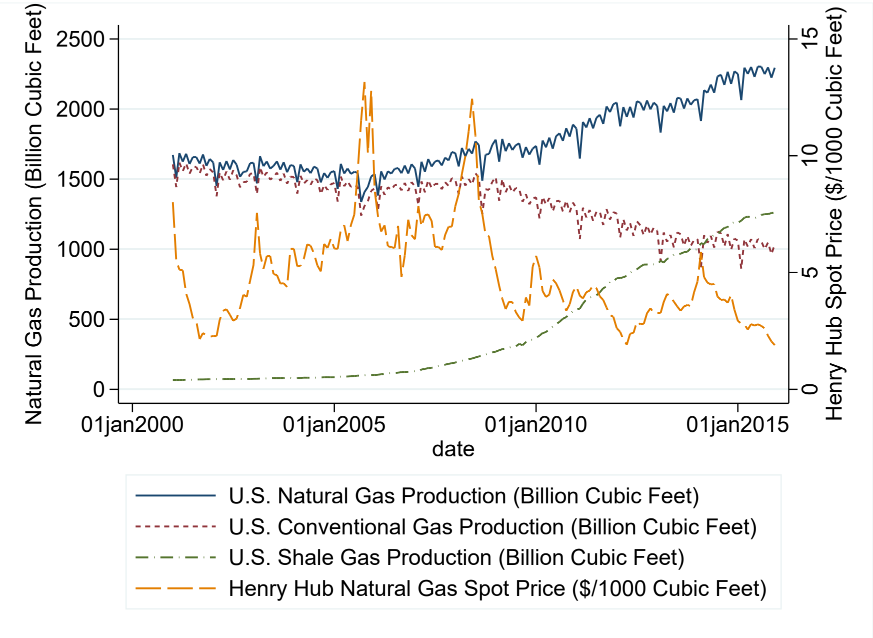
[2] Crude oil storages also rose dramatically to an average of 482 thousand barrels in 2015–16 from an average of 310 thousand barrels in 2000–2009 (stores excluding strategic reserves). But unlike natural gas, stores in petroleum show no annual cycles. Also in contrast to natural gas, consumption of finished petroleum products averaged only 518 thousand barrels per month in 2015–16, down from the average of 543 thousand barrels per month in 2000–2009.

[3] Besides consumption, natural gas pipeline transportation requires pressurization from the storage facility.

Figure 3.1: Effect of Hydraulic Fracturing on the U.S. Oil and Gas Industries

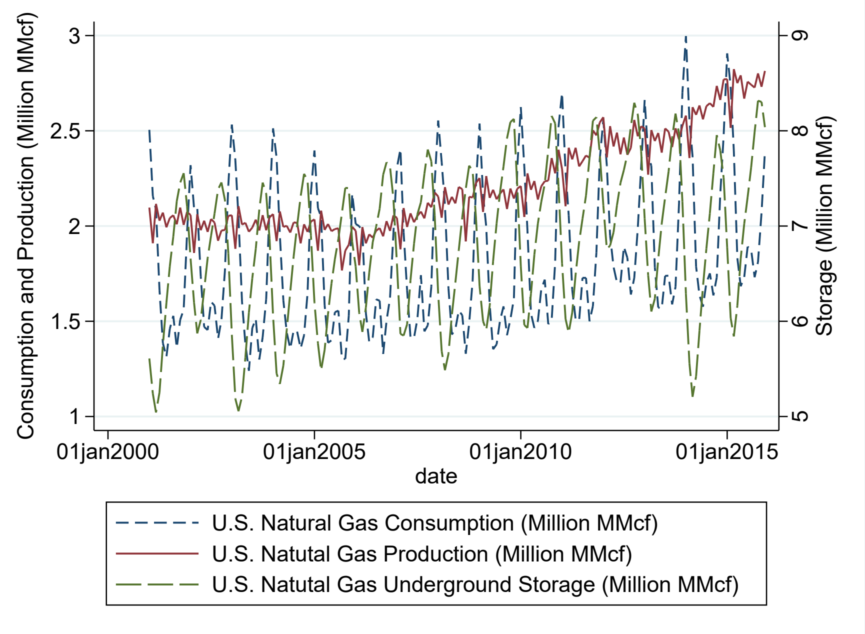


(a) Effect on the Oil Industry



(b) Effect on the Natural Gas Industry

Figure 3.2: Natural Gas Consumption, Storage, and Production



Although hydraulic fracturing has become the major and growing source of energy production, research regarding its impact on the North American and world markets is still in its infancy. Mason et al. (2015) describe the economic benefits of the shale gas boom, providing back-of-the-envelope estimates of the changes in consumer and producer surpluses resulting from hydraulic fracturing. Hausman and Kellogg (2015) estimate U.S. supply and demand elasticities of natural gas and then derive the parallel shifting of the natural gas supply curve before and after the shale boom. They calculate the counterfactual equilibrium that would have prevailed in 2013 in the absence of hydraulic fracturing. They conclude a net social benefit of $48 billion per year by subtracting a loss of producer surplus from the consumer surplus. Kilian (2017), using a structural vector-autoregressive model, shows that the global price of crude oil was lower by $10 per barrel than it would have been in the absence of hydraulic fracturing.

In this chapter, I argue that hydraulic fracturing has not only resulted in access to unconventional resources but also significantly changed how natural gas in North America is produced and stored. I find that a more elastic supply in the natural gas spot and storage markets has appeared with the expansion of hydraulic fracturing. In the North American natural gas market, storage plays a critical role in mitigating market uncertainties in the short run, such as unexpected temperature shocks during winters or summers. This is because adjusting storage is less costly than suddenly changing a production plan in the short run. Thus, hydraulic fracturing should have significantly decreased the impact of temperature shocks if it caused a more flexible adjustment in storage. As a result, consumers can benefit from shale boom by avoiding paying abruptly raised prices for natural gas in the short run. This study is most related to a recent work conducted by Newell et al. (2016), who analyze the price responsiveness of natural gas production for both conventional and unconventional (shale) wells in Texas. They find that over the 2010–2014 period, because unconventional wells produced on average 2.7 times more gas per well than conventional ones, the price responsiveness of supply for unconventional wells was almost three times larger than conventional wells. Nevertheless, our study differs from Newell et al. (2016) in two aspects. First, our analysis incorporates a more comprehensive dataset including 48 U.S states and 10 Canadian provinces. Second, our focus consists of not only the more elastic production in the spot market but also the more elastic adjustment of storage in the natural gas storage market.

Recognizing the role of natural gas storage in the natural gas market, I incorporate natural gas underground storage into a simple dynamic model by following the theory of inventory in commodity markets (Pindyck, 1994, 2001; Considine, 1997). I derive equilibrium conditions of natural gas production and storage. On the one hand, the shadow value of natural gas storage, the convenience yield, is the marginal benefit of holding an additional unit of storage. This value occurs because holding storage allows producers to avoid abrupt and costly adjustments of natural gas production in the short run (Pindyck, 1994; Mason, 2011). On the other hand, the full marginal cost of holding an extra unit of storage includes the unit cost of storage facility and an opportunity cost from interest foregone. The equilibrium storage level is governed by the condition that the convenience yield is equal to the full marginal cost of one extra unit of storage. Based on these equilibrium conditions, I develop two testable empirical hypotheses regarding hydraulic fracturing.

First, natural gas production would be more elastic with the expansion of hydraulic fracturing. Second, because of this, the demand for holding storage also should be more elastic with the expansion of hydraulic fracturing, which implies that adjusting storage has become more flexible.

Incorporating the theory of storage, I estimate a system of equations which define the equilibrium rig counts, production, convenience yield, net imports, and demand. I account for the endogenous natural gas price or storage in the right-hand side of each equation. I show that the estimated results are robust by using different sets of instruments. To test those empirical hypotheses, I allow the slopes of the rig counts, production, and convenience yield equations to change with the expansion of hydraulic fracturing. As a result, I can measure how hydraulic fracturing has affected the natural gas market in a dynamic framework.

The analysis reveals how the expansion of hydraulic fracturing has resulted in a more elastic supply in the North American natural gas spot market through two channels. First, the production from existing wells has become more elastic with the expansion of hydraulic fracturing. Second, with drilling also being more elastic to future prices and more productive with the expansion of hydraulic fracturing, production from newly established wells has also become more elastic. Furthermore, I conclude that the demand for holding storage also should be more elastic with the expansion of hydraulic fracturing. This implies that natural gas producers and distributors can adjust their storage levels more flexibly when facing unexpected shocks on the market. I document that with shale gas production at 2.3 BCF/day in 2001, one additional unit increase in convenience yields causes average storage to withdraw by 25 billion cubic feet. However, with emerging shale gas productions reaching 42 BCF/day in 2015, one additional unit increase in convenience yields causes average storage to withdraw by 273 billion cubic feet.

This chapter is organized as follows: section 2 proposes a simple dynamic model that describes how the natural gas spot and storage markets relate to each other; section 3 states the empirical strategy; section 4 describes the data sample; section 5 presents the estimated results; and section 6 gives concluding remarks.

3.2 The Model

3.2.1 A Simple Dynamic Model

I present a simple dynamic model that describes how natural gas spot and storage markets relate to each other. Consider a representative firm that produces, stores and distributes natural gas at state *i* maximizes its life-time profits as equation (3.1) subjected to an accounting identity (3.2) on natural gas storages:

$$\max_{y_{it},N_{it}} \sum_{t=0}^{\infty} E_t \beta_t \left\{ P_{it} Q_{it} - C_{it}(Y_{it}, N_{it}) - K N_{it} \right\}. \tag{3.1}$$

s.t.

$$N_{it} = N_{it-1} + Y_{it} - Q_{it} - NI_{it}, \tag{3.2}$$

where P_{it} Y_{it}, Q_{it}, N_{it} and NI_{it} are price, production, demand, storage and net import of natural gas at time t in state *i*; E_t and β_t denote the expectation and discounter factor respectively; $C_{it}(Y_{it}, N_{it})$ represents the production cost which is increasing in Y_{it} but decreasing in N_{it}; and K_i denotes the constant marginal cost paid for storage facility. The storage N_{it} enters into the production cost $C_{it}(Y_{it}, N_{it})$ because holding storage provides the "convenience" to smooth production, which avoids abrupt and costly adjustments of natural gas production in the short run. The marginal reduction of the production cost by holding an additional unit of storage $(i.e., -\partial C_{it}/\partial N_{it})$ reflects the extent of "convenience" to which this unit of storage yields for the firm. For simplicity, I assume the production cost function to be separable between production Y_{it} and storage N_{it}. Thus I write down the production cost as,

$$C_{it}(Y_{it}, N_{it}) = \frac{1}{2}\rho Y_{it}^2 + \alpha_i Y_{it} + \frac{1}{2}\tau N_{it}^2 - b_i N_{it}. \tag{3.3}$$

The cost function is specified as a quadratic form such that $\rho > 0$ denotes the slope of the supply curve and $\tau > 0$ represents absolute value of the slope of the convenience yield curve;[4] α_i and b_i are unobserved state-specific parameters respectively. The Lagrangian for the problem above is,

$$L = \sum_{t=0}^{\infty} E_t \beta_t \Big[P_{it} Q_{it} - C_{it}(Y_{it}, N_{it}) - K N_{it} + \lambda_{it}(N_{it-1} + Y_{it} - Q_{it} - NI_{it} - N_{it}) \Big], \tag{3.4}$$

where λ_{it} is the lagrangian multiplier.

Holding N_{it} and NI_{it} constant, differentiating (3.4) with respect to Y_{it} yields,

$$-\frac{\partial C_{it}(Y_{it}, N_{it})}{\partial Y_{it}} + \lambda_{it} = 0. \tag{3.5}$$

Since N_{it} and NI_{it} are held as constant, $Y_{it} = Q_{it}$. Thus, differentiating (3.4) with respect to Q_{it} yields,

$$P_{it} - \lambda_{it} = 0. \tag{3.6}$$

Next, differentiating (3.4) with respect to N_{it} yields,

$$-\frac{\partial C_{it}(Y_{it}, N_{it})}{\partial N_{it}} - K_i + \frac{E_t(\beta_{t+1})}{\beta_t}\lambda_{it+1} - \lambda_{it} = 0 \tag{3.7}$$

In the short run, $E_t(\beta_{t+1})/\beta_t \approx 1/(1+r_t)$ where r_t is risk-free interest rate. Substitute (3.6) into the equation above and it yields,

$$-\frac{\partial C_{it}(Y_{it}, N_{it})}{\partial N_{it}} = K_i - \left[\frac{1}{1+r_t} E_t(P_{it+1}) - P_{it} \right]. \tag{3.8}$$

The equation (3.8) governs the equilibrium amount of natural gas storage held by the representative firm in state i. The left-hand side is the convenience yield while the right-hand side denotes the full marginal cost of holding one extra unit of storage, which includes the unit cost of storage

---

[4] The slope of convenience yield curve is negative. This is because the marginal benefit of holding an extra unit of storage is considerably large when the level of storage approaches zero. In contrast, this marginal benefit is extremely small when storage reaches a very large level.

K_i and the opportunity cost due to expected price changes in the next period. Using (3.3), equation (3.8) becomes,

$$\left[P_{it} - \frac{1}{1+r_t}E_t(P_{it+1})\right] = -\tau N_{it} + (b_i - K_i). \tag{3.9}$$

where the left hand side denotes the difference between spot price P_{it} and the present value of future price $1/(1+r_t)E_t(P_{it+1})$. It is equal to subtracting storage cost K_i from the convenience yield CY_{it}. This refers to "net convenience yield", which represents the net benefit of holding an extra unit of storage (Pindyck, 1994).

3.2.2 The effect of hydraulic fracturing

The effect that hydraulic fracturing has had upon the natural gas spot and storage market may be illustrated by Figure 3.3. Since storage is primarily used to meet the large seasonal swings in demand, I consider a simplified scenario in which I only have two seasons, summer and winter in a closed market (i.e., no import and export). The demand curve shifts between summer demand (D_s) and winter demand (D_W) each year. The upper-right quadrant shows the equilibrium in the spot market. The demand in winter is $p = D_W(q)$ and demand in summer is $p = D_S(q)$. The initial supply function is $p = S_0(q)$. Absent storage, the winter price and quantity would occur at the intersection of $D_W(q)$ and $S_0(q)$ and the summer price and quantity would occur at the intersection of $D_S(q)$ and $S_0(q)$, respectively. With storage, however, the equilibrium smooths production and prices. With zero storage costs ($K_i = 0$), the no-arbitrage equilibrium will have the same price in both summer and winter, P_0^*. At this price, quantity Q_0^* is produced in each period, but demand in summer is Q_0^S and demand in winter is Q_0^W. The price P_0^* is such that the excess supply in summer, $\Delta N_0 = Q_0^* - Q_0^S$, is equal to the negative of excess demand in winter, $-\Delta N_0 = Q_0^W - Q_0^*$. Thus, quantity ΔN_0 is added to storage in summer and quantity ΔN_0 is removed from storage in winter to supplement winter demand.

The identity

$$N_t \equiv N_{t-1} + \Delta N_t,$$

defines changes in storages, N_t. Since I observe that end of winter storages are positive, I denote this value as $\underline{N}$. Hence, storages at the end of summer are $N_S^0 = \underline{N} + \Delta N_0$ and storages at the end of winter are $N_W^0 = \underline{N} + \Delta N_0 - \Delta N_0 = \underline{N}$. This is shown in the lower-right quadrant of Figure 3.3.

The amount of ΔN_0 is governed by the convenience yield curve CY_0, which is illustrated at the lower-left quadrant of Figure 3.3. The convenience yield curve is decreasing on storages. This is because the demand for holding storage becomes lower as the shadow price of storage, convenience yield, rises. In equilibrium, the convenience yield should be equal to the marginal cost of storage, which is the sum of the unit storage cost K_i and the difference in the expected future price $E_t(P_{t+1})$ and today's spot price P_t absence of interest rate (i.e., $r_t = 0$):

$$CY_t = K_i - \Delta P_{t+1},$$

where

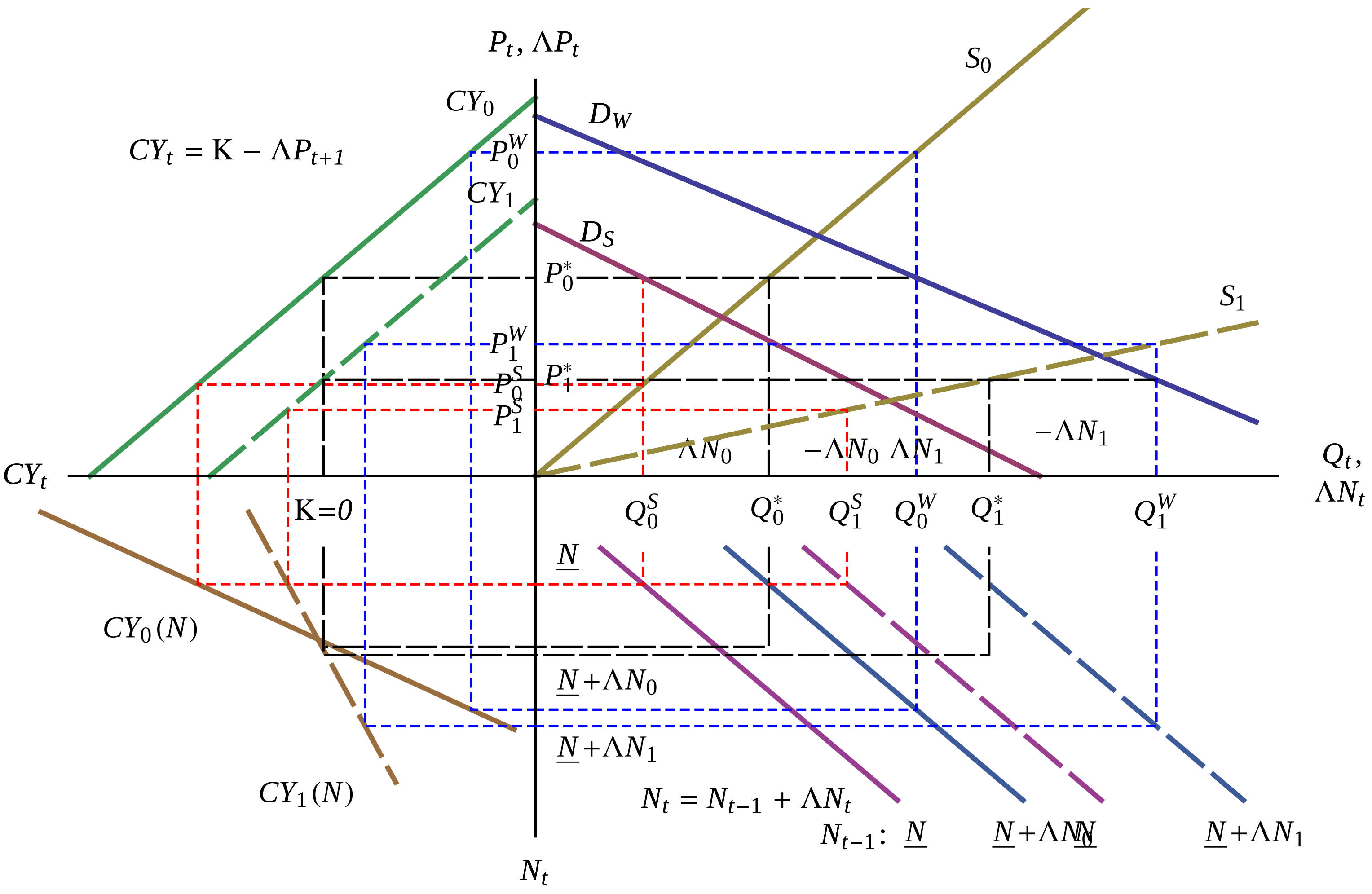
$$\Delta P_{t+1} = \left[E_t\left(P_{t+1}\right) - P_t\right].$$

This is illustrated the in upper-left quadrant.

Now consider the adoption of hydraulic fracturing enables the natural gas production in North America to become more elastic. Thus, the supply curve in the spot market (the upper-right quadrant) becomes less steep and then switches from $S_0(q)$ to $S_1(q)$. In this case, the equilibrium spot price is P_1^*, which is lower than P_0^*. Simultaneously, in the storage market the convenience yield curve is also less steep and switches from CY_0 to CY_1 if the supply curve becomes less steep.

In sum, the comparative static analysis above illustrates a critical observation in the natural gas market. If natural gas production has become more elastic with the expansion of hydraulic fracturing, the demand for holding storage has also become more elastic.

Figure 3.3: The Effect of Hydraulic Fracturing Upon the Natural Gas Market



3.3 Empirical Specification

As illustrated in Figure 3.3, I have two empirical hypotheses. First, natural gas production would be more elastic due to hydraulic fracturing. Second, because of this, the demand for holding storage should also be more elastic. In other words, the convenience yield curve would be less steep with the expansion of hydraulic fracturing. To test these hypotheses, I estimate a system of equations that consists of natural gas rig counts, production, demand, convenience yield, and net import. Next, I allow the slopes of rig counts, production, and convenience yield functions to change with the expansion of hydraulic fracturing. This is achieved by interacting the slopes of these equations with a variable describing the extent to which hydraulic fracturing has spread in North America. The specification of varying slopes allows us to understand whether the natural gas production and the demand for holding storage have become more elastic due to hydraulic fracturing.

The system of equations is identified because each equation has a unique exogenous shifter on the right-hand side that does not appear in the other equations. Therefore, I estimate equations (3.10), (3.11), (3.13), (3.18), and (3.20) in the next five subsections individually by the Generalized Method of Moment. The following five subsections describe the model specification of each equation and the last subsection discusses instrumental variables used for each equation.

3.3.1 Demand

The demand equation is specified by equation (3.10) as noted below,

$$Q_{it} = -\eta P_{it} + \kappa T_{it} + \sum_{j=1}^{11} \lambda_j M_t + u_{it} + \alpha_i, \tag{3.10}$$

where Q_{it} is the natural gas consumption in billion cubic feet in state or province i; P_{it} is the real price deflated by CPI; and T_{it} are the temperatures including heating degree days (HDD) and cooling degree days (CDD) seasonality and shocks. Seasonality represent the predictable movement of temperatures and shocks denote the unexpected temperature changes beyond the seasonalities; M_t is monthly dummy. I decompose HDD and CDD into seasonality and shocks by the approach discussed in Appendix C.

3.3.2 Rig Counts

The rig count equation is specified by equation (3.11) as noted below,

$$R_{it} = \alpha F_t + \gamma(F_t \times SHALE_t) + \beta UR_{it} + \sum_{j=1}^{11} \lambda_j M_t + \delta_i + \varepsilon_{it}, \tag{3.11}$$

where R_{it} denotes the rig counts at state or province i; F_t is the natural gas future price 3 months forward; $SHALE_t$ represents the monthly U.S. aggregate shale gas production in billion cubic feet per day; UR_{it} is the unemployment rate at state or province i; M_t is monthly dummy; and δ_i and ε_{it} are the fixed effect and error terms.

I use the future price F_t that would be delivered on contract 3-month later because it takes at least 2 or 3 months to rent and establish a rig and then to commence drilling (Kellogg, 2014; EIA, 2016). The variable $SHALE_t$ measures the extent to which hydraulic fracturing has spread in North America.[5] The future price F_t interacts with $SHALE_t$ because it is related to our first testable hypothesis. I anticipate with the increasing adoption of hydraulic fracturing, rig counts would be more responsive to price changes, which would make natural gas production more elastic. UR_{it} is an exogenous shifter that represents the effect of the business cycle, which may affect investments in establishing rigs. Both F_t and the interaction term $F_t \times SHALE_t$ are endogenous variables.

The parameters of interest are α and γ, which describe the marginal responsiveness of rig counts to future prices. Differentiating rig counts with respect to future prices yields equation (3.12) as noted below, which suggests that the marginal responsiveness of rig counts to future prices is positive if both α and γ are positive. Furthermore, this responsiveness is an increasing function of the expansion of hydraulic fracturing if γ is positive.

$$\frac{\partial R_{it}}{\partial F_t} = \alpha + \gamma SHALE_t, \tag{3.12}$$

---

[5] Unfortunately, I were unable to obtain Canadian shale gas productions on monthly frequency.

3.3.3 Production

I estimate the natural gas production function as equation (3.13) noted below:

$$\begin{aligned} Y_{it} = &\theta P_{it} + \pi(P_{it} \times SHALE_t) + \psi R_{it-3} + \delta(R_{it-3} \times SHALE_{t-3}) \\ &+ \sum_{j=1}^{11} \lambda_j M_t + \mu_i + \xi_{it}, \end{aligned} \tag{3.13}$$

where Y_{it} denotes the natural gas production in billion cubic feet; P_{it} is the natural gas real price; and R_{it-3} is the lagged 3-month rig counts. Rig counts are lagged by j month because it takes at least 2 or 3 months to rent and establish a rig and then to commence drilling. Therefore, initial productions for new wells at time t are attributed to rigs established 3 months ago.

The parameters of interest are θ, π, ψ, and δ. The first two parameters are related to the effect of hydraulic fracturing on the marginal responsiveness of production with respect to real prices, presented by equation (3.14) as noted below. Since P_{it} is the real price at the current time, this marginal responsiveness largely reflects decision making on existing productions, such as adjusting the production flow from existing wells. In contrast, the last two parameters are relevant to the effect of hydraulic fracturing on rig productivity which is measured by the incremental production of one extra rig established 3 months ago. Since this incremental production refers to the initial productions from new wells, equation (3.15) below represents the effect of hydraulic fracturing on the initial production from new wells.[6] As a result, by interacting $SHALE_t$ with both real prices and lagged rig counts, I would test the first hypothesis through two channels. In the first channel, as indicated by equation (3.14), I would test whether the expansion of hydraulic fracturing affects the marginal decision making on the existing productions. Moreover, in the second channel, as presented by equation (3.15), I would test whether the expansion of hydraulic fracturing affects the production from newly established wells. P_{it} and ($P_{it} \times SHALE_t$) are endogenous variables while

---

[6] Technically, in North America, one rig usually spuds one well on average during its rental period. See (EIA, 2016)

R_{it-3} and $(R_{it-3} \times SHALE_{t-3})$ are predetermined variables.

$$\frac{\partial Y_{it}}{\partial P_{it}} = \theta + \pi SHALE_t \tag{3.14}$$

$$\frac{\partial Y_{it}}{\partial R_{it-3}} = \psi + \delta SHALE_{t-3} \tag{3.15}$$

3.3.4 Convenience Yields

In the North American natural gas market, storage plays a more important role than production to mitigate short-run shocks. According to equation (3.8), in an equilibrium convenience yield is equal to the marginal cost of holding one extra unit of storage. This implies the following equation:

$$CY_{it} = K_i - \left[\frac{1}{1+r_t} FP_t - P_{it}\right], \tag{3.16}$$

where K_i is the unit storage cost; r_t is the risk-free interest rate; and FP_t is the one-month-forward future price of natural gas delivered from the Henry hub at time t. I use FP_t because it is an approximation to the expected price for time $t+1$, which is variable $E_t(P_{it+1})$ in equation (3.8).

On the other hand, convenience yield refers to the demand for storage, which is a downward slopping curve of storage specified as noted below,

$$CY_{it} = b_i - \tau N_{it} + \sigma(N_{it} \times SHALE_t) + \phi V_t + \sum_{j=1}^{11} \lambda_j M_t + \omega_{it}, \tag{3.17}$$

where N_{it} is the natural gas storage in billion cubic feet and V_t is the moving average Henry hub spot price volatility for the past 12 months ending at time $t-1$;[7] V_t enters into the convenience yield curve because increased price volatility implies an increase in the demand for holding storage. Market participants would want to hold a greater amount of storage in order to buffer fluctuations in production or consumption if the price volatility becomes larger and can sustain for a long time.

---

[7] V_t is equal to $1/12\sum_{j=t-13}^{j=t-1} VOL_j$ where VOL_j is the standard deviation of the Henry hub daily spot prices at month j.

The result would be an upward shift in the demand for storage as represented by the convenience yield curve (Pindyck, 2001). Therefore, the construction of V_t can well represent the extent to which the price volatility has been sustained. By interacting $SHALE_t$ with storage level N_{it}, I allow the slope of convenience yield to change with the expansion of hydraulic fracturing. This is relevant to the second testable hypothesis: a less steep convenience yield curve caused by hydraulic fracturing. As a result, I expect σ to be positive. This suggests that natural gas storage would be more responsive to the changes in convenience yield with the expansion of hydraulic fracturing.

Combining equation (3.16) and (3.17) yields an estimable equation as noted below,

$$P_{it} - \frac{1}{1+r_t} FP_t = (b_i - K_i) - \tau N_{it} + \sigma (N_{it} \times SHALE_t) + \phi V_t + \sum_{j=1}^{11} \lambda_j M_t + \omega_{it}, \tag{3.18}$$

where the left-hand side is equal to subtracting storage cost K_i from convenience yield CY_{it}. Since the storage cost K_i is unobservable, it becomes a part of the fixed effect in the right-hand side of equation (3.18). Pindyck (1994) refers to the left-hand side of equation (3.18) as "net convenience yield" which represents the net marginal benefit of holding an extra unit of storage, computed by the difference between spot price P_{it} and the present value of future price $1/(1+r_t)FP_t$. Thus I rewrite (3.18) as the following equation,

$$NCY_{it} = (b_i - K_i) - \tau N_{it} + \sigma (N_{it} \times SHALE_t) + \phi V_t + \sum_{j=1}^{11} \lambda_j M_t + \omega_{it}, \tag{3.19}$$

where NCY_{it} is net convenience yield; N_{it} and $(N_{it} \times SHALE_t)$ are endogenous variables; and V_t is predetermined.

3.3.5 Net Import

An estimable net import function is presented as noted below:

$$NI_{it} = \sum_{h=1}^{3} \omega_h (NP_{hit} - P_{it}) + \vartheta HS_{it} + \sum_{j=1}^{11} \lambda_j M_t + \omega_i + \varepsilon_{it}, \tag{3.20}$$

where NP_{1it}, NP_{2it} and NP_{3it} are the prices of first, second, and third largest neighboring states or provinces adjacent to the own state or province i;[8] HS_{it} denotes the housing starts at state or province i, which represents an exogenous shift on net imports. The ranking of the top 3 largest neighbors are based on the absolute values of the average trading volumes from 2001 to 2015 between state i and its neighboring states. Therefore, equation (3.20) describes that the net imports to state or province i depends on price differentials between state or province i and its top 3 largest neighbors. The reason for this dependence is that in a well-integrated natural gas market, local distributors at a location are able to identify arbitrage opportunities and then ship natural gas in or out of this location, given sufficiently large price differentials (De Vany and Walls, 1994; Fattouh, 2010). ω_h is a parameter denoting the marginal propensity to net import, which could be positive or negative, depending on whether the neighboring state or province h is a natural gas supplier or buyer to state or province i. For most states or provinces, their top three largest neighbors are adjacent states or provinces since natural gas is mainly shipped by pipelines. The top three neighbors account for more than 75% of the trading volume by state or province i, so the top 3 largest neighbors considerably capture the impact of neighboring markets. These price differentials are endogenous variables in equation (3.20).

3.3.6 Instrumental variables

In order to account for the endogenous variables on the right-hand side of each equation such as the contemporaneous price and storage, I apply different sets of instrumental variables to show that the estimated results are robust across various specifications of instrumental variables.

The first set of instruments are "benchmark" instruments that include exogenous shifters in the other equations and the own state price or storage lagged by more than 12 months. The second set of instruments is formed by adding the lagged prices of the first, second and third largest

---

[8] An alternative specification of equation (3.20) is using the weighted average prices of the three largest neighboring states as the independent variable. However, there are two reasons that prevent us from doing so. First, I lack the data of monthly natural gas trading between the sample state i and its adjacent states since EIA only reports the trading data annually. Therefore, I are unable to develop the weights for each adjacent state based monthly trading volumes. Second, using the neighbouring prices as the independent variables separately adds the flexibility of the model.

neighboring states or provinces to the benchmark instruments. The first, second and third largest neighbors are ranked by the average trading volumes of natural gas with the own state or province. The third set of instruments add the lagged heating degree days (HDDs) and cooling degree days (CDDs) of the top 3 largest neighboring states or provinces into the benchmark instruments. The last set of instruments add lagged prices of neighbors' neighboring' states or provinces (NNS) to the benchmark instruments. An NNS is a state or province that is not adjacent to an own state, but is directly adjacent to the own state's neighboring states. For example, the neighboring states of Alabama are Mississippi, Georgia, Florida, and the neighbor states of Alabama's neighboring states are Tennessee, Louisiana, Arkansas, South Caroline, Virginia, Kentucky, and Texas.

The first set of instruments (i.e., benchmark instruments) use the information from the system of equations. To identify one equation, exogenous shifters from the other equations can serve as instruments. In the second set of instruments, the lagged prices of neighboring states are related to the own state's contemporaneous price or storage, but may not be correlated to contemporaneous error terms in the rig counts, production, demand, convenience yield, and net import equations. For example, a sudden rise in the tax rate in an own state may affect the decision making for the own state's rig establishment, production, convenience yield, and even demand, but it is possible that this sudden shock may be orthogonal to a neighboring state's price that happened previously. In the third set of instruments, the lagged HDDs and CDDs of neighboring states are weaker instruments than lagged prices of neighboring states used in the second set of instruments, but are more likely to be orthogonal to contemporaneous error terms of each equation. Hausman and Kellogg (2015) use the third type of instruments to estimate the elasticities of U.S. natural gas supply and demand.[9] They report weak first-stage results but the Sargent's test does not reject the validity of instruments. In the last set of instruments, I add the lagged prices of NNS to the benchmark instruments because the markets of NNS should not directly affect an own state's market, but it may indirectly affect the own state only via its neighboring state. Therefore, the exclusion restriction could be satisfied.

---

[9] For each state, instead of using the HDDs and CDDs of top 3 largest neighboring states, Hausman and Kellogg (2015) instrument real prices by lagged HDDs and CDDs of all the other states.

Conceptually, lagged prices of NNS are weaker instruments than lagged prices of neighboring states, but are much more likely to be independent of the contemporaneous error terms of each equation.

In each set of instruments, I attempt to pass the Stock-Yogo weak identification test and the Hansen J test simultaneously by adjusting the months of lag for each instrumental variable. The first test is conducted to determine whether the instruments are so weak that the instrumental variable (IV) approach results in a very large bias relative to OLS estimates.[10] The second test examines whether the instruments are appropriately independent of the error terms.[11]

3.4 Data Description

A summary of all variables over the sample period is presented in Table 3.1. The analysis includes all U.S. states and 10 Canadian provinces from January 2001 to December 2015. The reason for incorporating both U.S. and Canadian data is twofold. First, U.S. and Canada are the only two countries that have commercialized hydraulic fracturing so far. Second, natural gas markets in North America are well integrated. During the sample period, more than 90% of U.S. imports of natural gas were from Canada, while more than half of U.S. exports were shipped to Canada.

Rig counts at the state and provincial level are downloaded from Baker Hughes. State and provincial level natural gas production, storage, demand (consumption) and city-gate prices are from U.S. Energy Information Agency (EIA) and Statistics Canada. However, Statistics Canada did not report the provincial-level natural gas storage over the sample period from 2001 to 2015, thus I treat storage is missing for all Canadian provinces. Since Canadian provincial-level natural gas storages are missing, I calculate provincial net imports based on domestic transfers, imports, and exports for each province, which are available in the Cansim Table 129-0002 from Statistics Canada.

---

[10] In the presence of weak instruments, the loss of precision will be severe, and IV estimates may be no improvement over OLS (Staiger and Stock, 1997).

[11] Under the assumption of *i.i.d* errors, this is known as a Sargan's test. In IV-GMM estimation, the test of overidentifying restrictions becomes Hansen J test where J is number of the GMM criterion functions.

The net imports to each U.S. state are calculated by an accounting identity as noted below,

$$NI_{it} \equiv Q_{it} - Y_{it} + (N_{it} - N_{it-1}), \tag{3.21}$$

where NI_{it} is the net imports of natural gas to state i, which is equal to subtracting production Y_{it} from the sum of consumption Q_{it} and changes in natural gas storage $(N_{it} - N_{it-1})$. This identity works for those U.S. states where both production and storage exist over the sample period. For those states having storage but no production, Y_{it} is treated as zero;[12] for those states without production and storage, their net imports reduce to consumption Q_{it}.

The future prices, obtained from Bloomberg, are based on delivery at the Henry Hub in Louisiana. I compute real prices of natural gas for each state or province by dividing state or provincial-level city-gate prices by CPI. Since city gate is a point at which local distributors deliver or receive natural gas from pipeline systems, it represents the full marginal cost of natural gas distribution at a local area, including the costs of production, shipping, and storage (Davis and Muehlegger, 2010).

U.S. state-level HDD and CDD for each state is reported by the U.S. National Oceanic and Atmospheric Administration (NOAA). I compute HDD and CDD for Canadian provinces by population-weighted averaging HDD and CDD for all stations within a province, which is consistent to the method used by NOAA. The unemployment rate and housing starts for each state or province are from the U.S. Bureau of labor and Statistics Canada. Price volatility V_t is calculated by the moving average of the standard deviations of the daily Henry hub prices from $t-13$ to $t-1$.

---

[12] However, all states with productions also own storage facilities over the sample period.

Table 3.1: Summary Statistics of All Variables

variable name	No. of Observations	Mean	Std. Dev.	Min	Max
Endogenous Variables					
Consumption, Q_{it} (Billion cubic feet)	10440	36.2	49.3	0.0	379.9
Real Price, P_{it} (U.S. Dollar per thousand cubic feet)	10651	3.8	1.9	0.4	13.5
Future Price, 3-month forward, F_{it} (U.S. Dollar per thousand cubic feet)	10800	5.3	2.3	2.1	13.9
Future Price, 1-month forward, FP_{it} (U.S. Dollar per thousand cubic feet)	10800	5.0	2.3	2.0	13.2
Storage, N_{it} (Billion cubic feet)	5400	226.7	245.4	0.3	1047.1
Net Import, NI_{it} (Billion cubic feet)	10770	-3.7	78.9	-480.1	212.6
Rig Counts, R_{it} (units)	7018	54.6	128.4	0	1208
Production, Y_{it} (Billion cubic feet)	7020	67.5	118.7	0.0	709.7
Shale Gas Production, $SHALE_t$ (Billion cubic feet per day)	10800	14.4	13.5	2.2	41.8
Exogenous or Predetermined Variables					
Heating Degree Days, HDD_{it} (Degree days per day)	10440	13.9	13.5	0.0	64.0
Cooling Degree Days, CDD_{it} (Degree days per day)	10440	2.8	4.7	0.0	25.4
Average Price Volatility for the Past 12 Months, V_t (%)	10620	4.0	1.1	2.0	10.0
Unemployment Rate, UR_{it} (%)	10260	6.1	2.1	2.0	15.4
Housing Starts, HS_{it} (1000 units)	10260	2.7	3.9	0.0	26.8
Instrumental Variables					
The First Largest Neighbor's Prices, $NP1_{it}$ (U.S. dollar per thousand cubic feet)	10187	5.9	2.3	0.6	17.4
The Second Largest Neighbor's Prices, $NP2_{it}$ (U.S. dollar per thousand cubic feet)	10008	6.3	2.4	0.6	21.7
The Third Largest Neighbor's Prices, $NP3_{it}$ (U.S. dollar per thousand cubic feet)	9720	5.6	2.9	0.0	17.4
The First Largest Neighbor's HDD, $NHDD1_{it}$ (Degree days)	10260	404.1	429.4	0.0	1919.0
The second largest neighbor's HDD, $NHDD2_{it}$ (Degree days)	10080	416.8	421.5	0.0	1919.0
The third largest neighbor's HDD, $NHDD3_{it}$ (Degree days)	9720	348.7	405.5	0.0	1919.0
The largest neighbor's CDD, $NCDD1_{it}$ (Degree days)	10260	87.9	144.5	0.0	761.0
The second largest neighbor's CDD, $NCDD2_{it}$ (Degree days)	10080	67.6	120.7	0.0	750.0
The third largest neighbor's CDD, $NCDD3_{it}$ (Degree days)	9720	76.7	136.5	0.0	761.0
The largest NNS' prices, $NNP1_{it}$ (U.S. dollar per thousand cubic feet)	10726	4.3	1.9	0.6	15.8
The second largest NNS' prices, $NNP2_{it}$ (U.S. dollar per thousand cubic feet)	10440	4.7	2.3	0.8	14.6
The third largest NNS' prices, $NNP3_{it}$ (U.S. dollar per thousand cubic feet)	10180	4.5	2.1	0.6	15.2

Data includes U.S. states and Canadian provinces from January 2001 to December 2015. The ranking of neighboring states or provinces and NNS are based on the average absolute values of the trading volume to an own state or province from 2001 to 2015; NNS refers to the neighbors' neighboring states or provinces of an own state or province.

3.5 Estimated Results

3.5.1 Demand Equation

The estimated results of demand equation (3.10) are presented in Table 3.2 as noted below. Column (1) reports the results by using the lagged 12 to 13 month real prices as the "benchmark" instruments; columns (2)-(4) report the results by adding lagged 12 and 13 month prices of the first to third largest neighbors to the benchmark instruments; columns (5)-(7) present the results by adding lagged 12-month HDDs and CDDs of the first to third largest neighbors into the benchmark instruments; finally, columns (8)-(10) show the results by adding lagged 13 month prices of the first to third largest NNS into the benchmark instruments.

All columns have the Kleibergen-Paap (KP) F statistics larger than the Stock-Yogo (SY) weak identification critical values based on 5% bias relative to OLS, rejecting the null of weak instruments. However, only column (1) and (8) do not reject the validity of instruments at 5% based on the p-values of the Hansen J tests. I try adding exogenous shifters from the other equations, such as lagged unemployment rate and average price volatility for the past 12 months to benchmark instruments; however, the Hansen J test keeps rejecting the null that they are valid at 5%.

Therefore I use estimated results by column (1) for interpretation which yields the largest KP statistics and Hansen J test's p-value. In column (1), the estimated coefficient of real price is -0.091 but statistically insignificant. This may indicate in the short run such as within one month, the real price is not a significant factor in determining natural gas consumption. In contrast, the estimated coefficients of HDD and CDD seasonalities and shocks are statistically significant. Moreover, the estimated coefficient of HDD and CDD shocks are larger than the estimated coefficients of HDD and CDD seasonalities, suggesting unexpected weather shocks are more important than predictable seasonal movements in driving natural gas consumption.

Table 3.2: Estimated results of demand equation

Dependent Variable: Consumption, Q_{it}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Real Price	-0.091	0.045	0.044	-0.147	-0.326	-0.328	-0.411	-0.096	0.086	0.014
	(0.405)	(0.409)	(0.418)	(0.451)	(0.402)	(0.407)	(0.437)	(0.407)	(0.427)	(0.425)
HDD Seasonality	0.490***	0.456***	0.413***	0.333***	0.478***	0.415***	0.421***	0.523***	0.426***	0.463***
	(0.066)	(0.066)	(0.068)	(0.070)	(0.066)	(0.067)	(0.065)	(0.066)	(0.071)	(0.071)
CDD Seasonality	1.042***	0.999***	1.079***	1.237***	0.950***	1.015***	1.139***	1.051***	1.153***	1.195***
	(0.101)	(0.101)	(0.104)	(0.109)	(0.100)	(0.093)	(0.090)	(0.101)	(0.106)	(0.107)
CDD Shock	1.077***	1.067***	1.067***	1.119***	1.102***	1.136***	1.137***	1.073***	1.129***	1.127***
	(0.149)	(0.148)	(0.149)	(0.150)	(0.149)	(0.147)	(0.147)	(0.149)	(0.149)	(0.150)
HDD Shock	0.831***	0.817***	0.806***	0.800***	0.773***	0.774***	0.788***	0.821***	0.818***	0.835***
	(0.061)	(0.061)	(0.061)	(0.062)	(0.061)	(0.061)	(0.062)	(0.061)	(0.062)	(0.063)
F stats (monthly dummies)	146.66	144.58	150.77	160.08	136.91	147.97	151.11	143.62	156.99	157.96
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Instrumental Variables (monthly lags, l)										
Own States Real Price, P_{it-l}	12-13	12-13	12-13	12-13	12-13	12-13	12-13	12-13	12-13	12-13
The First Largest Neighbor's Prices, $NP1_{it-l}$		12-13	13	13						
The Second Largest Neighbor's Prices, $NP2_{it-l}$			13	13						
The Third Largest Neighbor's Prices, $NP3_{it-l}$				13						
The First Largest Neighbor's HDD/CDD, $NHDD1_{it-l}$ /$NCDD1_{it-l}$					12	12	12			
The Second Largest Neighbor's HDD/CDD, $NHDD2_{it-l}$ /$NCDD2_{it-l}$						12	12			
The Third Largest Neighbor's HDD/CDD, $NHDD3_{it-l}$ /$NCDD3_{it-l}$							12			
The First Largest NNS' Prices, $NNP1_{it-l}$								13	13	13
The Second Largest NNS' Prices, $NNP2_{it-l}$									13	13
The Third Largest NNS' Prices, $NNP3_{it-l}$										13
Test Statistics										
Kleibergen-Paap F stats	195.57	117.85	113.31	79.63	100.88	68.75	46.82	130.49	101.21	79.65
Stock-Yogo Weak Identification Critical Values (20% bias)	8.75	6.71	6.71	6.77	6.71	6.76	6.69	6.46	6.71	6.77
Stock-Yogo Weak Identification Critical Values (10% bias)	19.93	10.27	10.27	10.83	10.27	11.12	11.39	9.08	10.27	10.83
Stock-Yogo Weak Identification Critical Values (5% bias)	20.32	16.85	16.85	18.37	16.85	19.28	20.25	13.91	16.85	18.37
Hansen J Test p-value	0.876	0.000	0.000	0.000	0.000	0.000	0.000	0.118	0.001	0.002
F stats of Testing Fixed Effect	103.15	101.23	102.27	102.85	97.99	99.39	102.21	104.01	104.24	105.88
p-value of Testing Fixed Effect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. of Observations	9224	9222	9128	8867	9224	9129	8867.000	9150	8961	8714
No. of States and Provinces Included	58	58	58	58	58	58	58	58	58	58

Robust Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

P_{it-l} is lagged l-month real price of the own state or province; NP_{it-l} is lagged l-month prices of the first, second, or third largest neighboring states; $NHDD_{it-l}$ and $NCDD_{it-l}$ are lagged l-month heating and cooling degree days of the first, second, or third largest neighbors; NNP_{it-l} is lagged l-month prices of the first, second, or third largest NNS; NNS refers to the neighbors' neighboring states or provinces of state or province i.

3.5.2 Rig Equation

The estimated results of rig equation (3.11) are presented in Table 3.3 as noted below. Column (1) reports estimates based on the "benchmark" instruments consisting of exogenous shifters from the other equations and the 3-month-forward future prices lagged by more than 12 months. The benchmark instruments thus include the lagged 4 and 5 month average price volatility, the lagged 4 and 5 month housing starts of the own state or province, the lagged 4 and 5 month HDDs and CDDs of the own state or province, and the lagged 12 to 14 month future prices; columns (2)-(4) add the lagged prices of the three largest neighbors to the benchmark instruments; columns (5)-(7) add the lagged HDDs and CDDs of the three largest neighbors to the benchmark instruments; finally, columns (8)-(10) adds the lagged prices of the three largest NNSs to the benchmark instruments .

Comparing the KP statistics among columns suggests that the lagged prices of NNS and the lagged neighbors' HDDs and CDDs are weaker instruments than the lagged neighbors' prices. The KP statistics in columns (2)-(4) can reject the null of weak identification based on 5% bias relative to OLS, while the KP stats in columns (5)-(10) only can reject weak identification based on 10%. All columns have p-values of Hansen J tests larger than 0.1, which do not reject the null of valid instruments. Therefore, the preferred specification is column (2), which yields the largest KP statistics by adding the lagged 2 to 5 month prices of the first largest neighbor to the benchmark instruments. Although I only use column (2) for interpretation, it is important to notice the estimated coefficients of F_t and $F_t \times SHALE_t$ among columns (2)-(10) are similar in magnitude.

The results from column (2) suggest the marginal response of rig counts to (future) price changes can be found as the equation below:

$$\frac{\partial R_{it}}{\partial F_t} = 2.453 + 0.289 \times SHALE_t, \tag{3.22}$$

As a result, the marginal response of rig counts to price changes is illustrated in Figure 3.4 below. With shale gas production as 2.3 BCF/day in early 2001, one additional unit increase in future prices on average leads producers to increase the establishment of additional 3 rigs. In contrast, with shale gas production reaching 42 BCF/day in late 2015, one additional unit increase in future prices on average leads to the establishment of additional 15 rigs.

Table 3.3: Estimated Results of Rig Equation

Dependent Variable: Rig Counts, R_{it}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Futures Price, F_t	6.225***	2.453***	4.025***	3.049***	5.197***	4.843***	4.864***	5.044***	4.690***	3.685***
	(1.572)	(0.809)	(1.034)	(0.766)	(1.359)	(1.381)	(1.424)	(1.389)	(1.327)	(1.236)
$F_t \times SHALE_t$	0.499***	0.289***	0.368***	0.314***	0.426***	0.401***	0.401***	0.439***	0.434***	0.367***
	(0.093)	(0.055)	(0.064)	(0.054)	(0.081)	(0.082)	(0.084)	(0.085)	(0.080)	(0.076)
Unemployment Rate, UR_{it}	0.003	-0.983**	-0.166	-0.323	-0.046	-0.043	-0.183	-0.427	-0.611	-0.858
	(0.655)	(0.482)	(0.533)	(0.446)	(0.607)	(0.607)	(0.638)	(0.594)	(0.589)	(0.564)
F stats (monthly dummies)	33.84	29.79	23.16	22.63	29.12	27.98	30.74	31.90	30.02	25.77
p-value	0.000	0.000	0.02	0.02	0.000	0.000	0.000	0.000	0.000	0.01
Instrumental Variables (monthly lags, l)										
Average Price Volatility, V_{t-l}	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
Own State's Housing Starts. HS_{it-l}	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
Own State's $HDD_{i,t-l}$ and $CDD_{i,t-l}$	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5	4-5
Future Prices, F_{t-l}	12	12-14	12-14	12-14	12-14	12-14	12-14	12	12	12
The First Largest Neighbor's Price, $NP1_{i,t-l}$		2-5	4-9	3-8						
The Second Largest Neighbor's Prices, $NP2_{i,t-l}$			4-9	3-8						
The Third Largest Neighbor's Prices, $NP3_{i,t-l}$				3-8						
The First Largest Neighbor's HDD/CDD, $NHDD1_{it-l}/NCDD1_{it-l}$					2	2	2			
The Second Largest Neighbor's HDD/CDD, $NHDD2_{it-l}/NCDD2_{it-l}$						2	2			
The Third Largest Neighbor's HDD/CDD, $NHDD3_{it-l}/NCDD3_{it-l}$							2			
The First Largest NNS' Prices, $NNP1_{i,t-l}$								4-5	4-5	4-5
The Second Largest NNS' Prices, $NNP2_{i,t-l}$									4-5	4-5
The Third Largest NNS' Prices, $NNP3_{i,t-l}$										4-5
Test Statistics										
Kleibergen-Paap stats	20.81	35.09	19.88	25.51	15.67	16.83	17.50	19.77	20.25	19.37
Stock-Yogo Weak Identification Critical Values (20% bias)	6.22	6.07	5.98	5.92	6.16	6.19	6.21	6.23	6.21	6.19
Stock-Yogo Weak Identification Critical Values (10% bias)	10.43	11.05	11.05	11.03	10.99	10.93	10.84	10.69	10.84	10.93
Stock-Yogo Weak Identification Critical Values (5% bias)	18.30	20.65	20.88	20.96	20.23	19.98	19.64	19.12	19.64	19.98
p-value of Hansen J test	0.66	0.16	0.15	0.10	0.40	0.15	0.39	0.54	0.52	0.23
F stats of testing Fixed Effect	4.11	3.93	3.94	3.74	4.03	3.74	3.84	3.93	3.97	3.46
p-value of testing Fixed Effect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. of Observations	5964	5849	5689	5596	5904	5750	5596	5962	5656	5651
No. of States and Provinces	39	39	39	39	39	39	39	39	39	39

Robust Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

F_{t-l} is the Henry hub 3-month-forward future price lagged by l months; V_{t-l} is the lagged l-month average price volatility; HDD_{it-l} and CDD_{it-l} are the lagged l-month heating and cooling degree days of the own state or province; NP_{it-l} is the lagged l-month prices of the first, second, or third largest neighbor; $NHDD_{it-l}$ and $NCDD_{it-l}$ are the lagged l-month heating and cooling degree days of the first, second, or third largest neighbor; NNP_{it-l} is lagged l-month prices of the first, second, or third largest NNS; NNS refers to the neighbors' neighboring states or provinces of an own state or province.

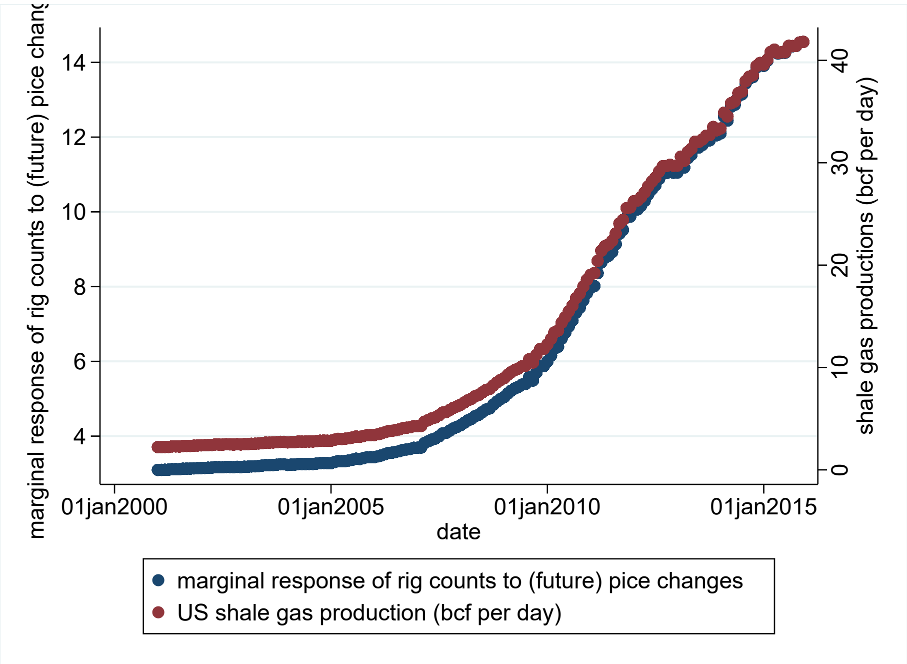


Figure 3.4: Shale gas production and marginal response of rig counts to changes in future prices

3.5.3 Production Equation

The estimated results of production equation (3.13) are presented in Table 3.4 as noted below. The "benchmark" instruments used in column (1) include the lagged 1 and 2 month average price volatility, the lagged 1 and 2 month unemployment rate of the own state or province, the lagged 1 and 2 month CDDs and HDDs of the own state or province, the lagged 1 and 2 month housing starts, and the lagged 12 month real prices of the own state or province; columns (2)-(4) add the lagged 12 month prices of the three largest neighbors to the benchmark instruments; columns (5)-(7) add the lagged 6 month HDDs and CDDs of the three largest neighbors to the benchmark instruments; finally, columns (8)-(10) add the lagged 16 month prices of the three largest NNSs to the benchmark instruments.

All the columns have the KP statistics larger than the SY weak identification critical values based on 5% bias relative to OLS, indicating these instruments are not weak. However, the Hansen J tests in columns (8)-(10) reject the null that the lagged prices of the three largest NNSs are valid instruments since the p-values in these columns are less than 0.01.

I choose column (2) for interpretation in which I only add to the benchmark instruments the lagged 12 and 13 month prices of the first largest neighbor. This is because this column yields the largest KP statistic and p-value of the Hansen J test. Although I only use column (2) for interpretation, it is important to notice the estimated coefficients among all columns are very similar in magnitude.

According to equation (3.14), the estimates in column (2) suggests that with only 2.3 BCF/day production from hydraulic fracturing in 2001, an extra unit increase in real prices on average causes production to increase by 4.6 billion cubic feet. However, with the emerging productions from hydraulic fracturing reaching 42 BCF/day in 2015, an extra unit increase in real prices on average causes production to increase by 10 billion cubic feet (Figure 3.5).

According to equation (3.15), column (2) also indicates that rig productivity has significantly improved with the expansion of hydraulic fracturing. With only 2.3 BCF/day production from hydraulic fracturing in 2001, one extra rig established 3 months ago causes the average production to increase by 0.1 billion cubic feet. However, with the emerging productions from hydraulic fracturing reaching 42 BCF/day in 2015, one extra rig established 3 month ago causes the average production to increase by 0.3 billion cubic feet (Figure 3.6).

The estimated results from rig and production equations demonstrate how the expansion of hydraulic fracturing leads to the more elastic supply in the North American natural gas market. First, the production from existing wells has become more elastic with the expansion of hydraulic fracturing. This is because the current production Y_{it} is more responsive to the current real price P_{it} with the expansion of hydraulic fracturing, as illustrated by Figure 3.5. Second, since the rig counts have become more responsive to changes in future prices and the rig productivity has improved with the expansion of hydraulic fracturing (Figure 3.6), the production from newly established wells has been more elastic. To see it, I combine equation (3.12) and (3.15) to derive the marginal response of the production from newly established wells with respect to changes in future price (3 months earlier) as the equation below:

$$\begin{aligned}\frac{\partial Y_{it}}{\partial F_{t-3}} &= \frac{\partial Y_{it}}{\partial R_{it-3}} \cdot \frac{\partial R_{it-3}}{\partial F_{t-3}} \\ &= (\psi + \delta SHALE_{t-3}) \cdot (\alpha + \gamma SHALE_{t-3}),\end{aligned} \tag{3.23}$$

where the second line uses equation (3.12) and (3.15). By applying estimated coefficients in columns (2) from Table 3.3 and 3.4, equation (3.23) shows with only 2.3 BCF/day production from hydraulic fracturing in 2001, an extra unit increase in future prices causes the average production to increase by 0.3 billion cubic feet. However, with the hydraulic fracturing production reaching 42 BCF/day in 2015, one extra unit rise in future prices causes the average production to increase by 4.5 billion cubic feet.

Table 3.4: Estimated Results of Production Equation

Dependent Variable: Production, Y_{it}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Real Price, P_{it}	4.448***	4.336***	2.910***	2.840***	4.052***	3.705***	3.537***	2.578***	2.361**	2.040**
	(1.011)	(0.987)	(0.913)	(0.930)	(0.988)	(0.992)	(1.007)	(0.902)	(0.956)	(0.941)
$P_{it} \times SHALE_t$	0.139***	0.137***	0.107***	0.105***	0.139***	0.131***	0.131***	0.088***	0.082***	0.071***
	(0.025)	(0.023)	(0.021)	(0.021)	(0.025)	(0.024)	(0.024)	(0.022)	(0.021)	(0.020)
R_{it-3}	0.085***	0.085***	0.085***	0.086***	0.080***	0.082***	0.083***	0.066**	0.062**	0.062**
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.029)
$R_{it-3} \times SHALE_{t-3}$	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
F stats (monthly dummies)	80.15	79.09	77.63	76.43	78.25	76.46	76.82	73.35	70.16	70.17
p-value	0.000	0.000	0.000	0.000	0.00	0.000	0.000	0.000	0.000	0.000
Instrumental Variables (monthly lags, l)										
Own State's Unemployment Rate, UR_{it-l}	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2
Price Volatility, V_{t-l}	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2
Own State's Housing Starts, HS_{it-l}	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2
Own State's $HDD_{i,t-l}$ and $CDD_{i,t-l}$	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2	1-2
Real Prices, P_{it-l}	12	12	12	12	12	12	12	12	12	12
The First Largest Neighbor's Prices, $NP1_{i,t-l}$		12-13	12-13	12-13						
The Second Largest Neighbor's Prices, $NP2_{i,t-l}$			12-13	12-13						
The Third Largest Neighbor's Prices, $NP3_{i,t-l}$				12-13						
The First Largest Neighbor's HDD/CDD, $NHDD1_{i,t-l}/NCDD1_{i,t-1}$					6	6	6			
The Second Largest Neighbor's HDD/CDD,$NHDD2_{i,t-l}/NCDD2_{i,t-1}$						6	6			
The Third Largest Neighbor's HDD/CDD,$NHDD3_{i,t-l}/NCDD3_{i,t-1}$							6			
The First Largest NNS' Prices, $NNP1_{i,t-l}$								16	16	16
The Second Largest NNS' Prices, $NNP2_{i,t-l}$									16	16
The Third Largest NNS' Prices, $NNP3_{i,t-l}$										16
Test Statistics										
Kleibergen-Paap F statistic	36.36	36.86	31.62	26.11	30.89	28.70	25.21	34.00	35.06	32.19
Stock-Yogo Weak Identification Critical Values (20% bias)	6.23	6.21	6.19	6.16	6.21	6.19	6.16	6.22	6.21	6.20
Stock-Yogo Weak Identification Critical Values (10% bias)	10.69	10.84	10.93	10.99	10.84	10.93	10.99	10.78	10.84	10.89
Stock-Yogo Weak Identification Critical Values (5% bias)	19.12	19.64	20.23	20.96	19.64	19.98	20.23	19.40	19.64	19.83
p-value of Hansen J Stats	0.37	0.54	0.095	0.071	0.17	0.16	0.25	0.000	0.000	0.000
F stats of Testing Fixed Effect	27.95	28.32	28.24	29.0	28.94	29.05	29.20	27.25	27.0	27.29
p-value of Testing Fixed Effect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. of Observations	5828	5795	5699	5606	5828	5732	5636	5827	5516	5512
No. of States and Provinces Included	38	38	38	38	38	38	38	38	38	38

Robust Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

UR_{it-l} is the lagged l-month unemployment rate of the own state or province; HS_{it-l} is the lagged l-month housing starts of the own state or province; V_{t-l} is the lagged l-month average price volatility; HDD_{it-l} and CDD_{it-l} are the lagged l-month heating and cooling degree days of the own state or province; NP_{it-l} is the lagged l-month prices of the

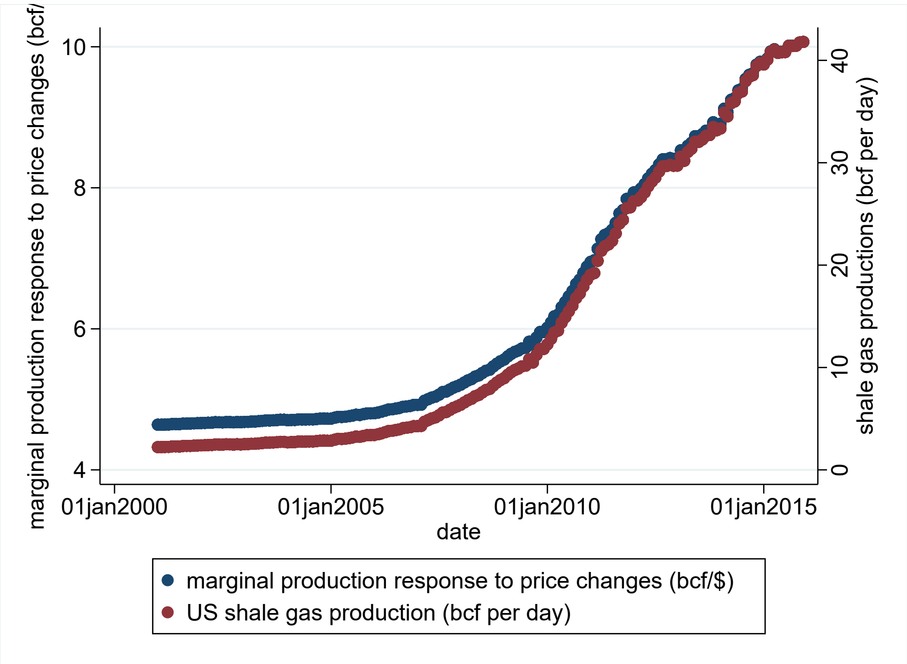


Figure 3.5: Shale gas production and marginal response of production to changes in real prices

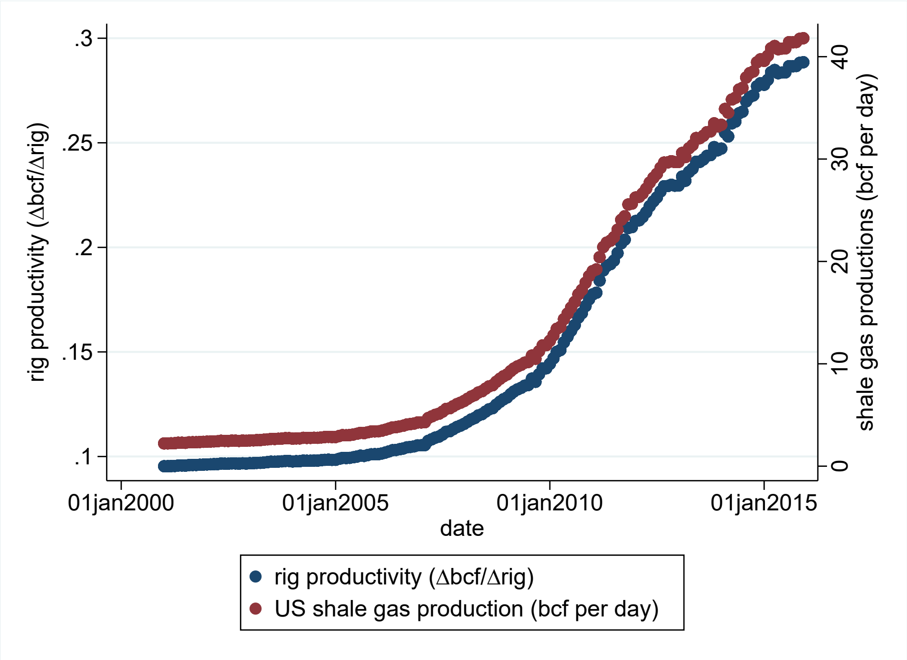


Figure 3.6: Shale gas production and rig productivity

3.5.4 Convenience Yield Equation

The estimated results of convenience yield equation (3.19) are presented in Table 3.5 as noted below. Column (1) uses the "benchmark" instruments consisting of the lagged 12 month storage of the own state, lagged 10 month unemployment rate and HDDs and CDDs of the own state; columns (2)-(4) add the lagged 8 month prices of the three largest neighbors to the benchmark instruments; columns (5)-(7) add the lagged 8 month HDDs and CDDs of the three largest neighbors to the benchmark instruments; finally, columns (8)-(10) add the lagged 7 month prices of NNSs to the benchmark instruments. I do not include the lagged housing starts of the own state to the benchmark instruments because the Hansen J test always rejects the null of valid instruments for all columns if I do so.

All the columns have the KP statistics larger than the SY weak identification critical values based on 5% bias relative to OLS, rejecting the null of weak identification. Moreover, all the columns have the p-values of the Hansen J test more than 0.1, indicating that the validity of instruments cannot be rejected.

Similar to the estimated results of rig and production equations, the preferred specification for interpretation is column (2) in which I add the lagged prices of the first largest neighbors to the benchmark instruments. This column has the largest KP statistic and p-value of the Hansen J test. Therefore, the estimates from column (2) are best suited for interpretation.

The estimates in column (2) suggest with the production from hydraulic fracturing as 2.3 BCF/day in early 2001, one additional billion cubic feet withdrawal in storage would on average cause convenience yields to rise by \$0.035 per million cubic feet. However, with emerging productions from hydraulic fracturing reaching 42 BCF/day in 2015, one additional billion cubic feet withdrawal in storage would on average cause convenience yields to increase by \$0.024 per million cubic feet (Figure 3.7). The convenience yield curve thus becomes less steep with the expansion of hydraulic fracturing. From the perspective of producers and distributors, a less steep convenience yield curve implies a higher demand for holding storage given a price of storage. As a result, producers and distributors can adjust the storage level more flexibly than ever before.

To see this, I reinterpret the convenience yield curve as following: consider in a competitive natural gas market, a temperature shock raises the spot price P_{it} by one dollar. All else equal, the convenience yield at state i would rise by one dollar as well.[13] This implies that the opportunity cost of holding an extra unit of storage is higher; therefore, frims are more likely to release their storages into the spot market. A related question under this circumstance is that how many storages are firms willing to release to the market? Following equation (3.24) as below, the estimates in column (2) suggests in 2001, one additional unit increase in convenience yields on average leads storages to withdraw by 25 billion cubic feet. However, in 2015, one additional unit increase in convenience yields on average leads storages to withdraw by 273 billion cubic feet. As a result, firms can adjust natural gas storage more flexibly when facing temperature shocks with the expansion of hydraulic fracturing.

$$\frac{\partial N_{it}}{\partial CY_{it}} = \frac{1}{-\tau + \sigma \times SHALE_t}. \tag{3.24}$$

---

[13] In practice, the future price in the next period would also change when the spot price change. Therefore the convenience yield may not change as much as the spot price.

Table 3.5: Estimated Results of Convenience Yield Equation

Dependent Variable: Net Convenience Yield, NCY_{it}	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
N_{it}	-36.911***	-37.618***	-37.913***	-36.725***	-34.273***	-34.058***	-33.505***	-33.933***	-33.187***	-33.425***
	(6.323)	(5.278)	(5.264)	(5.025)	(5.862)	(5.531)	(5.386)	(5.588)	(5.606)	(5.501)
$N_{it} \times SHALE_t$	0.831***	0.842***	0.844***	0.830***	0.792***	0.785***	0.777***	0.783***	0.780***	0.785***
	(0.076)	(0.052)	(0.052)	(0.050)	(0.072)	(0.069)	(0.068)	(0.067)	(0.068)	(0.064)
V_t	0.546***	0.562***	0.572***	0.538***	0.475***	0.459***	0.450***	0.447***	0.454***	0.459***
	(0.132)	(0.105)	(0.105)	(0.100)	(0.124)	(0.120)	(0.118)	(0.113)	(0.113)	(0.112)
F stats (monthly dummies)	41.97	43.59	46.32	49.11	42.25	43.46	43.68	44.38	43.23	43.27
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Instrumental Variables (monthly lags, l)										
Own State's Unemployment Rate, UR_{it-l}	10	10	10	10	10	10	10	10	10	10
Own State's HDD_{it-l} and CDD_{it-l}	10	10	10	10	10	10	10	10	10	10
Own State's Storage, N_{it-l}	12	12	12	12	12	12	12	12	12	12
The First Largest Neighbor's Prices, $NP1_{i,t-l}$		8	8	8						
The Second Largest Neighbor's Prices, $NP2_{i,t-l}$			8	8						
The Third Largest Neighbor's Prices, $NP3_{i,t-l}$				8						
The First Neighbor's HDD/CDD, $NHDD1_{i,t-l}/NCDD1_{i,t-l}$					8	8	8			
The Second Neighbor's HDD/CDD, $NHDD2_{i,t-l}/NCDD2_{i,t-l}$						8	8			
The Third Neighbor's HDD/CDD, $NHDD3_{i,t-l}/NCDD3_{i,t-l}$							8			
The First Largest NNS' Prices, $NNP1_{i,t-l}$								7	7	7
The Second Largest NNS' Prices, $NNP2_{i,t-l}$									7	7
The Third Largest NNS' Prices, $NNP3_{i,t-l}$										7
Test Statistics										
Kleibergen-Paap F statistic	28.56	58.68	49.02	43.40	21.74	16.64	14.09	27.13	22.70	21.82
Stock-Yogo Weak Identification Critical Values (20% bias)	5.57	5.91	6.08	6.16	6.08	6.2	6.23	5.91	6.76	6.16
Stock-Yogo weak Identification Critical Values (10% bias)	7.56	8.78	9.48	9.92	9.48	10.22	10.58	8.78	9.48	9.92
Stock-Yogo weak Identification Critical Values (5% bias)	11.04	13.97	15.72	16.88	15.72	17.7	18.76	13.97	15.72	16.88
p-value of Hansen J tests	0.79	0.92	0.28	0.13	0.25	0.12	0.21	0.26	0.23	0.35
F stats of Testing Fixed Effect	18.52	27.96	27.12	28.58	19.37	20.50	20.71	20.55	20.19	22.16
p-value of Testing Fixed Effect	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. of Observations	5036	5036	5036	5036	5036	5036	5036	5036	5036	5036
No. of States Included	30	30	30	30	30	30	30	30	30	30

Robust Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

UR_{it-l} is the lagged l-month unemployment rate of the own state or province; HDD_{it-l} and CDD_{it-l} are the lagged l-month heating and cooling degree days of the own state or province; NP_{it-l} is the lagged l-month prices of the first, second, or third largest neighbor; $NHDD_{it-l}$ and $NCDD_{it-l}$ are the lagged l-month heating and cooling degree days of the first, second, or third largest neighbor; NNP_{it-l} is the lagged l-month prices of the first, second, or third largest NNS; NNS refers to the neighbors' neighboring states or provinces of an own state or province.

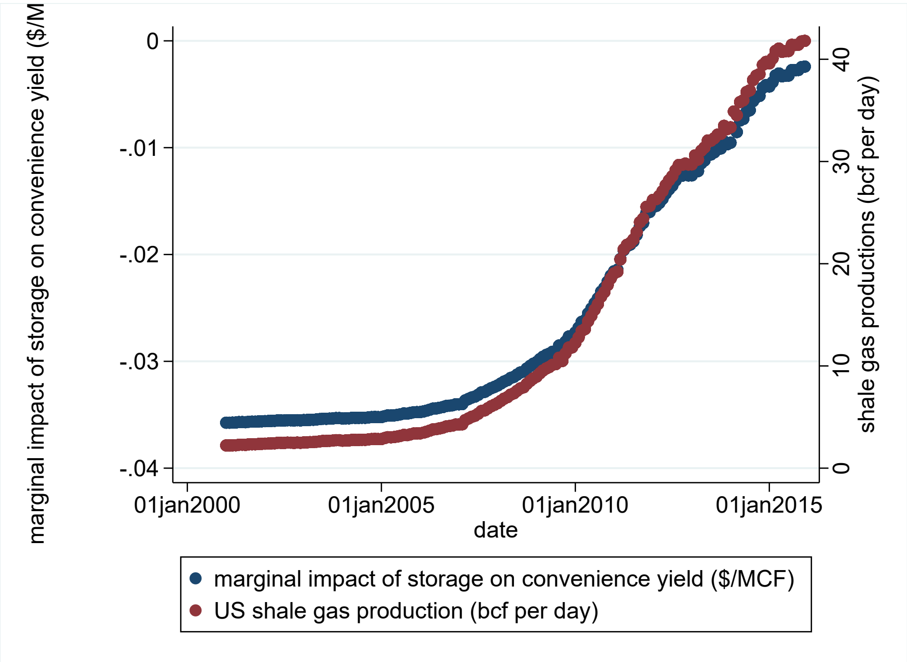


Figure 3.7: Shale gas production and marginal impact of storage on convenience yield

3.5.5 Net Import Equation

The estimated results of the net import equation (3.20) are presented in Table 3.6. The "benchmark" instruments in column (1) are the lagged 13 month unemployment rate of the own state or province, the lagged 13 month average price volatility, and the lagged 12 and 13 month price differentials between neighbors and the own state or province *i*; columns (2)-(4) add lagged 10 month HDDs and CDDs of the three largest neighbors to the "benchmark" instruments; finally, columns (5)-(7) add lagged 13 to 16 month prices of the three largest NNSs to the benchmark instruments. Since the price differentials between the three largest neighbors and the own state or province *i* (i.e., $NP1_{it}$-$NP3_{it}$) already appear on the right-hand side of equation (3.20), their lagged values are not included in the benchmark instruments as an additional set of instruments.

All columns have the KP statistics larger than the SY weak identification critical values based on 5% bias relative to OLS, rejecting the null of weak instruments. Also, the Hansen J tests do not reject the validity of instruments at 5% except column (3), (4), and (7). Therefore I use estimates by the benchmark instruments of column (1) which yields the largest KP statistics. However, it should be noted that all columns yield similar magnitudes in estimated coefficients.

The estimates in column (1) suggest the net imports to state or province *i* on average increase by 7.36 billion cubic feet if the price differential between the first largest neighbor and state *i* increases by one unit. In contrast, the net imports to state or province *i* on average decrease by 6 billion cubic feet if the price differential between the second largest neighbor and state or province *i* increases by one unit. These estimates reflect the trading pattern of natural gas in North America. For most states and provinces in the sample, they are net importers of natural gas from their first largest neighbors, while they are net exporters of natural gas to their second largest neighbors. The positive coefficient of $(NP1_{it} - P_{it})$ indicates state or province *i* would net import more natural gas if its primary supplier, the first largest neighbor, has relatively more expensive natural gas. This may be because the state or province *i* would increase its net imports from the other states or provinces to substitute the supply from the first largest neighbor. The negative coefficient of $(NP2_{it} - P_{it})$ suggests state *i* would export more (import less) natural gas out of state or province *i* if the second largest neighbor, the primary receiver of natural gas from state or province *i*, has relatively more expensive natural gas.

Table 3.6: Estimated results of net import equation

Dependent Variable: Net Import, NI_{it}	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$NP1_{it} - P_{it}$	7.356***	7.145***	9.068***	8.808***	7.023***	7.067***	7.077***
	(2.031)	(1.983)	(2.025)	(1.952)	(2.033)	(1.979)	(1.994)
$NP2_{it} - P_{it}$	-5.953***	-5.774***	-7.254***	-7.446***	-5.834***	-6.186***	-5.891***
	(1.845)	(1.805)	(1.810)	(1.786)	(1.851)	(1.797)	(1.809)
$NP3_{it} - P_{it}$	0.239	0.219	0.291	0.548	0.162	0.466	-0.117
	(1.259)	(1.254)	(1.266)	(1.245)	(1.265)	(1.257)	(1.250)
House Starts, HS_{it}	0.162	0.101	0.028	-0.104	0.169	0.060	0.054
	(0.324)	(0.317)	(0.320)	(0.314)	(0.325)	(0.318)	(0.319)
F stats (monthly dummies)	163.53	160.80	139.43	130.26	162.87	154.69	156.97
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Instrumental Variables (monthly lags, l)							
Own States Unemployment Rate, UR_{it-l}	13	13	13	13	13	13	13
Average Price Volatility, V_{t-l}	13	13	13	13	13	13	13
Price Differentials between Neighbors and Own State, $(NP_{it-l} - P_{it-l})$	12-13	12-13	12-13	12-13	12-13	12-13	12-13
The First Largest Neighbors HDD/CDD, $NHDD1_{it-l}/NCDD1_{it-l}$		10	10	10			
The Second Largest Neighbors HDD/CDD, $NHDD2_{it-l}/NCDD2_{it-l}$			10	10			
The Third Largest Neighbors HDD/CDD, $NHDD3_{it-l}/NCDD3_{it-l}$				10			
The First Largest NNS' Price, $NPP1_{it-1}$					13	16	16
The Second Largest NNS' Price, $NPP2_{it-1}$						16	16
The Third Largest NNS' Price, $NPP3_{it-1}$							16
Test Statistics							
Kleibergen-Paap F statistic	26.64	22.70	22.79	16.43	24.53	21.10	19.68
Stock-Yogo Weak Identification Critical Values (20% bias)	5.69	5.83	5.83	5.93	5.78	5.83	5.87
Stock-Yogo Weak Identification Critical Values (10% bias)	9.01	9.64	9.64	10.25	9.37	9.64	9.85
Stock-Yogo Weak Identification Critical Values (5% bias)	15.18	16.8	16.8	18.47	16.1	16.8	17.35
Hansen J Test p-value	0.173	0.143	0.000	0.000	0.091	0.055	0.008
F stats of Testing Fixed Effect	12.41	12.26	11.17	10.63	12.24	12.25	12.22
p-value of Testing Fixed Effect	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. of Observations	7814	7814	7814	7814	7813	7672	7592
No. of States and Provinces Included	57	57	57	57	57	57	57

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

UR_{it-l} is the lagged l-month unemployment rate of the own state or province; V_{t-l} is the lagged l-month average price volatility; $NHDD_{it-l}$ and $NCDD_{it-l}$ are the lagged l-month heating and cooling degree days of the first, second or third largest neighbor; NNP_{it-l} is the lagged l-month prices of the first, second, or third largest NNS; NNS refers to the neighbors' neighboring states or provinces of an own state or province.

3.6 Conclusion

This chapter provides a comprehensive analysis on the effect of hydraulic fracturing on North American natural gas market. I estimate a set of equations regarding natural gas rig counts, production, convenience yield, demand and net import by using U.S. state-level and Canadian provincial-level data. In order to account for the endogenous variables on the right-hand side of each equation, I apply different sets of instrument variables and show the estimated results are robust across different sets of instruments.

Our estimated results show that there are significant changes in how natural gas is produced and stored in North America since the shale revolution. I find that both production from existing wells and from newly established wells have become more elastic with the expansion of hydraulic fracturing. As a result, the natural gas supply in the spot market has become more elastic with the expansion of hydraulic fracturing.

Furthermore, I demonstrate that the convenience yield curve has become less steep, and thus adjusting natural gas storage is more flexible with the expansion of hydraulic fracturing. This is important because holding storage is the primary channel for producers and distributors to mitigate short-run risks (e.g., temperature shocks) in the North American natural gas market. I document that with the shale gas production at 2.3 BCF/day in 2001, one additional unit increase in convenience yields causes average storages to withdraw by 25 billion cubic feet. However, with emerging shale gas productions reaching 42 BCF/day in 2015, one additional unit increase in convenience yields causes average storages to withdraw by 273 billion cubic feet.

Bibliography

Arrow, K. and Chang, S. (1982). “Optimal Pricing, Use, and Exploration of Uncertain Natural Resource Stock”. *Journal of Environmental Economics and Management*, 9(1):1–10.

Bausell, C., Rusco, F., and Walls, W. D. (2001). “Lifting the Alaskan Oil Export Ban: An Intervention Analysis”. *The Energy Journal*, 0(4):81–94.

Benth, F. E. and Benth, J.-S. (2007). “The Volatility of Temperature and Pricing of Weather Derivatives”. *Quantitative Finance*, 7:553–561.

Bhargava, A., Franzini, L. and W. Narendranathan (1982). “Serial Correlation and the Fixed Effects Model”. *Review of Economic Studies*, 49(4):533549.

Borenstein, R. and Kellogg, R. (2014). “The Incidence of an Oil Glut: Who Benefits from Cheap Crude Oil in the Midwest?”. *The Energy Journal*, 0(1).

Brown, S., Mason, C., Krupnick, A., and Mares, J. (2014). “Crude Behavior: How Lifting the Export Ban Reduces Gasoline Prices in the United States”. *Issue Brief*, 3. Resource For Future, Washington, D.C., U.S.A.

Campbell, S. and Diebold, F. (2005). “Weather Forecasting for Weather Derivatives”. *Journal of the American Statistical Association*, 100:6–16.

Chow, G. (1979). “Optimal Control of Stochastic Differential Equation Systems”. *Journal of Economics Dynamic and Control*, 1(2):143–175.

Considine, T. (1997). “Under Joint Production: An Empirical Analysis of Petroleum Refining”. *The Review of Economics and Statistics*, 79:493–502.

Davis, L. W. and Muehlegger, E. (2010). “Do Americans consume too little natural gas? An empirical test of marginal cost pricing”. *RAND Journal of Economics*, 41(4):791–810.

De Vany, A. and Walls, W. D. (1994). “Pipeline Access and Market Integration in the Natural Gas Industry: Evidence from Cointegration Tests”. *The Energy Journal*, 14:1–19.

EIA (2014a). “Refinery Capacity Report”. Website. http://www.eia.gov/petroleum/refinerycapacity/.

EIA (2014b). “What Drives U.S. Gasoline Price?”. Website. http://www.eia.gov/analysis/studies/gasoline/pdf/gasolinepricestudy.pdf.

EIA (2016). “Drilling Productivity Report, Report Background and Methodological Overview”. `https://www.eia.gov/petroleum/drilling/pdf/dpr_methodology.pdf`.

Fattouh, B. (2010). “The dynamics of crude oil price differentials”. *Energy Economics*, 32(3):334–342.

GAO (2014). Allowing exports could reduce comsumer fuel prices, and the size of the strategic reserve should be reexamined. Technical Report GAO-14-807, U.S. General Accounting Office, Washington, D.C., U.S.A.

Gold, R. (2014a). “Oil Trains Hide in Plain Sight”. *Wall Street Journal*.

Gold, R. (2014b). “Shale-Oil Boom Puts Spotlight on Crude Export Ban”. *Wall Street Journal*.

Gold, R. (2014c). *“The Boom: How Fracking Ignited the American Energy Revolation and Change the World”*. Simon & Schuster, New York, U.S.A, 1st edition.

Hausman, C. and Kellogg, R. (2015). “Welfare and Distributional Implications of Shale Gas”. *Brookings Papers on Economic Activity*, pages 72–139.

Hotelling, H. (1931). “The Economics of Exhaustible Resources”. *Journal of Political Economy*, 39(2):137–175.

Kamien, M. and Schwartz, N. (1991). *“Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management”*. North Holland, Amsterdam, 2nd edition.

Kellogg, R. (2014). “The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling”. *American Economic Review*, 104:1698–1734.

Kemp, M. and Long, N. V. (1980). “On Two Folk Theorems Concerning the Extraction of Exhaustible Resources”. *Econometrica*, 48(3):663–74.

Kilian, L. (2014). “The impact of the shale oil revolution on U.S. oil and gasoline prices”. *CFS Working Paper*, 499.

Kilian, L. (2017). “The Impact of the Fracking Boom on Arab Oil Producers”. *The Energy Journal*, 38:137–159.

Mason, C., Muehlenbachs, L. A., and Olmstead, S. (2015). “The Economics of Shale Gas Development”. *Annual Review of Resource Economics*, 7:269–89.

Mason, C. F. (2011). “On Stockpiling Natural Resources”. *Resource and Energy Economy*, 33(2):398–409.

Mason, C. F. (2012). “On Equilibrium in Resource Markets with Scale Economies and Stochastic Prices”. *Journal of Environmental Economics and Management*, 64(3):288–300.

Newell, R., Prest, B. C., and Vissing, A. (2016). “Trophy Hunting vs. Manufacturing Energy: The Price-Responsiveness of Shale Gas”. *Resource for Future Discussion Paper*, pages 16–32.

Øksendal, B. (2003). *“Stochastic Differnetial Equations: An Introduction with Application”*. Springer, U.S.A, 6th edition.

Pindyck, R. (1980). “Uncertainty and Exhaustible Resource Markets”. *Journal of Political Economy*, 88(6):1203–1225.

Pindyck, R. (1994). “Inventories and the Short-Run Dynamics of Commodity Prices”. *The Rand Journal of Economics*, 25(1):141–159.

Pindyck, R. (1999). “The Long-run Evolution of Energy Prices”. *The Energy Journal*, 20(2):1–27.

Pindyck, R. (2001). “The Dynamic of Commodity Spot and Future Price: A Primer”. *The Energy Journal*, 22(3):pp.

Sandrea, R. and Sandrea, I. (2014). “New Well Productivity Data Provides US Shale Potential Insights”. *Oil and Gas Journal*, November, 03:Print Pages.

Staiger, D. and Stock, J. (1997). “Instrumental Variables Regression with Weak Instruments”. *Econometrica*, 65(3):557–586.

Suslick, S., Schiozer, D., and Rodriguez, M. (2009). “Uncertainty and Risk Analysis in Petroleum Exploration and Production”. *Thematic Contribution*, 6(1):30–41.

Wooldridge, J. (2002). *“Econometric Analysis of Cross Section and Panel Data”*. MIT Press, Cambridge, MA, U.S.A.

Zavaleta, A., Walls, W., and Rusco, F. W. (2015). Refining for export and the convergence of petroleum product prices. *Energy Economics*, 47:206–214.

Zhou, Q. (2010). “Dynamic Moment Analysis Of Non-Stationary Temperature Data in Alberta”. Master’s thesis, University of Lethbridge, Alberta, Canada.

Appendix A

Proof of Proposition 1.1

Differentiating the fundamental equation of stochastic optimality (1.5) with respect to $R(t)$ yields,

$$\begin{aligned}0 =& J_{tR}(t,R,\theta,P) - \theta(t)Q(t)J_{RR}(t,R,\theta,P) + \alpha P(t)J_{PR}(t,R,\theta,P)\\&+\frac{\sigma^2}{2}\theta(t)^2J_{\theta\theta R}(t,R,\theta,P)+\frac{\eta^2}{2}P(t)^2J_{PPR}(t,R,\theta,P).\end{aligned} \tag{A.1}$$

where J_{iii} is the third partial derivative of J with respect to i where $i \in \{t,R,\theta,P\}$.

Expanding $J_R(t,R,\theta,P)$ by the Ito's lemma yields,

$$\begin{aligned}dJ_R(t,R,\theta,P) =& J_{tR}(t,R,\theta,P)dt + J_{RR}(t,R,\theta,P)dR(t) + J_{R\theta}(t,R,\theta,P)d\theta(t)\\&+J_{RP}(t,R,\theta,P)dP(t)+\frac{1}{2}J_{RRR}(t,R,\theta,P)[dR(t)]^2+\frac{1}{2}J_{R\theta\theta}(t,R,\theta,P)[d\theta(t)]^2\\&+\frac{1}{2}J_{RPP}(t,R,\theta,P)[dP(t)]^2+J_{RR\theta}(t,R,\theta,P)dR(t)d\theta(t)\\&+J_{RRP}(t,R,\theta,P)dR(t)dP(t)+J_{R\theta P}(t,R,\theta,P)d\theta(t)dP(t)\\=&\Big[J_{tR}(t,R,\theta)-\theta(t)Q(t)J_{RR}(t,R,\theta)+J_{RP}(t,R,\theta,P)\alpha P(t)\\&+\frac{\sigma^2}{2}\theta(t)^2J_{R\theta\theta}(t,R,\theta,P)+\frac{\eta^2}{2}P(t)^2J_{RPP}(t,R,\theta,P)\Big]dt\\&+J_{RR\theta}(t,R,\theta,P)dR(t)d\theta(t)+J_{RP}(t,R,\theta,P)\eta P(t)d\nu(t),\end{aligned} \tag{A.2}$$

where the results following the second equality uses $[dR(t)]^2 = 0$, $[d\theta(t)]^2 = \sigma^2\theta(t)^2dz(t)$, $[dP(t)]^2 = \eta^2P(t)^2d\nu(t)$, $d\theta(t)dR(t) = 0$, $dR(t)dP(t) = 0$, $d\theta(t)dP(t) = 0$ and substitution from equation (1.2) and (1.4). Applying Ito's differential operator on equation (A.2) yields,

$$\begin{aligned}\frac{1}{dt}E_tdJ_R(t,R,\theta,P) =& J_{tR}(t,R,\theta,P) - \theta(t)Q(t)J_{RR}(t,R,\theta,P) + J_{RP}(t,R,\theta,P)\alpha P(t)\\&+\frac{\sigma^2}{2}\theta(t)^2J_{R\theta\theta}(t,R,\theta,P)+\frac{\eta^2}{2}P(t)^2J_{RPP}(t,R,\theta,P),\end{aligned} \tag{A.3}$$

where we use $E_t d\big[z(t)\big] = 0$ and $E_t d\big[\nu(t)\big] = 0$. Combing equation (A.3) with (A.1) yields,

$$\frac{1}{dt} E_t d\big[J_R(t, R, \theta, P)\big] = 0. \tag{A.4}$$

Appendix B

Proof of Proposition 1.2

This section derives the optimal path of expected production under uncertainties. Define $X(t)$ as the left-hand side of equation (1.9), thus $X(t)$ is a function of (t,P,Q,θ) such that $X(t)=X(t,Q,\theta,P)$. Using Ito's lemma on $X(t)$ yields,

$$\begin{aligned} dX(t,P,Q,\theta) =& X_t dt + X_P dP(t) + X_Q dQ(t) + X_\theta d\theta(t) + \frac{1}{2}X_{PP}\left[dP(t)\right]^2 \\ &+ \frac{1}{2}X_{QQ}\left[dQ(t)\right]^2 + \frac{1}{2}X_{\theta\theta}\left[d\theta(t)\right]^2 + X_{PQ}dP(t)dQ(t) \\ &+ X_{P\theta}dP(t)d\theta(t) + X_{Q\theta}dQ(t)d\theta(t). \end{aligned} \tag{B.1}$$

where X_i is the partial derivative of $X(t,P,Q,\theta)$ with respect to i where $i \in \{t,P,Q,\theta\}$ and X_{ii} is its second derivative. These partial deactivates are computed as following:

$$\begin{aligned} &X_t = \frac{-re^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)}, \qquad X_P = \frac{e^{-rt}}{\theta(t)}, \qquad X_Q = \frac{-e^{-rt}C''\left[Q(t)\right]}{\theta(t)}, \\ &X_\theta = \frac{-e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)^2}, \qquad X_{PP}=0, \qquad X_{QQ} = \frac{-e^{-rt}C'''\left[Q(t)\right]}{\theta(t)} = 0, \\ &X_{\theta\theta} = \frac{2e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)^3}, \qquad X_{PQ}=0, \qquad X_{P\theta} = \frac{-e^{-rt}}{\theta(t)^2}, \\ &X_{Q\theta} = \frac{e^{-rt}C''\left[Q(t)\right]}{\theta(t)^2}, \end{aligned} \tag{B.2}$$

where $X_{QQ}=0$ if we assume $C'''(\cdot)\approx 0$.

Expanding $dQ(t)$ along with the state variables (R,P,θ) yields the following equation,

$$\begin{aligned}dQ(t) =&Q_R(t)dR(t)+Q_P(t)dP(t)+Q_\theta(t)d\theta(t)+\frac{1}{2}Q_{RR}(t)\left[dR(t)\right]^2\\&+\frac{1}{2}Q_{PP}(t)\left[dP(t)\right]^2+\frac{1}{2}Q_{\theta\theta}(t)\left[d\theta(t)\right]^2+Q_{RP}(t)dR(t)dP(t)\\&+Q_{R\theta}(t)dR(t)d\theta(t)+Q_{P\theta}(t)dP(t)d\theta(t).\\=&Q_R(t)dR(t)+Q_P(t)dP(t)+Q_\theta(t)d\theta(t)+\frac{1}{2}Q_{PP}(t)\eta^2P(t)^2dt\\&+\frac{1}{2}Q_{\theta\theta}(t)\sigma^2\theta(t)^2dt\end{aligned}\tag{B.3}$$

Therefore,

$$\begin{aligned}dQ(t)d\theta(t)&=Q_\theta(t)\left[d\theta(t)\right]^2\\&=\sigma^2\theta(t)^2Q_\theta(t)dt.\end{aligned}\tag{B.4}$$

Also, from equation (1.3), we can compute the following equality as noted below,

$$\left[d\theta(t)\right]^2=\sigma^2\theta(t)^2dt.\tag{B.5}$$

Substituting those partial derivatives in (B.2), equation (B.5) and (B.4) into (B.1) yields,

$$\begin{aligned}dX(t,P,Q,\theta)=&\frac{-re^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)}dt+\frac{e^{-rt}}{\theta(t)}dP(t)+\frac{-e^{-rt}C''\left[Q(t)\right]}{\theta(t)}dQ(t)\\&+\frac{-e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)^2}d\theta(t)+\frac{e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)^3}\sigma^2\theta(t)^2dt\\&+\frac{e^{-rt}C''\left[Q(t)\right]}{\theta(t)^2}\left[\sigma^2\theta(t)^2Q_\theta(t)\right]dt.\end{aligned}\tag{B.6}$$

Applying Ito's differential operator on equation (B.6) yields,

$$\begin{aligned}\frac{1}{dt}E_t\left[dX(t,P,Q,\theta)\right]=&\frac{-re^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)}+\frac{e^{-rt}}{\theta(t)}\alpha P(t)+\frac{-e^{-rt}C''\left[Q(t)\right]}{\theta(t)}\frac{1}{dt}E_t\left[dQ(t)\right]\\&+\frac{e^{-rt}\left\{P(t)-C'\left[Q(t)\right]\right\}}{\theta(t)}\sigma^2+\frac{e^{-rt}C''\left[Q(t)\right]}{\theta(t)}\left[\sigma^2\theta(t)Q_\theta(t)\right],\end{aligned}\tag{B.7}$$

where we use $E\left[dP(t)\right] = \alpha P(t)$ and $E\left[d\theta(t)\right] = 0$. The left-hand side of (B.7) is equivalent to equation (1.9), thus it is zero. By using this fact, we solve for $1/dtE_t\left[dQ(t)\right]$ as following:

$$\begin{aligned}\frac{1}{dt}E_t\left[dQ(t)\right] =& \frac{1}{C''\left[Q(t)\right]}\Big\{-(r-\sigma^2)\{P(t)-C'\left[Q(t)\right]\} \\ &+C''\left[Q(t)\right]\left[\sigma^2\theta(t)Q_\theta(t)\right]+\alpha P(t)\Big\},\end{aligned} \tag{B.8}$$

where $Q_\theta(t)$ in the second term represents the marginal responsiveness of productions with respect to production shocks $\theta(t)$. $Q_\theta(t)$ is expected to be positive because the firm may plan more extractions if the production shocks are unexpectedly large. This is because large production shocks imply the existing well or field is very productive. As a result, the firm may update its belief regarding the productivity of the current well or field and then raise its incentive to extract.

Next, we define the elasticity of the marginal cost of extraction with respect to production shocks as below,

$$\begin{aligned}\xi_{mc\theta} &= \frac{\theta(t)}{C'\left[Q(t)\right]}\frac{\partial C'\left[Q(t)\right]}{\partial\theta(t)} \\ &= \frac{\theta(t)}{C'\left[Q(t)\right]}C''\left[Q(t)\right]Q_\theta(t),\end{aligned} \tag{B.9}$$

This elasticity describes how the firm responds to the production shocks via the adjustment of its production plan. It is positive since $Q_\theta(t)$ is positive. Using equation (B.9) to eliminate $Q_\theta(t)$ in equation (B.8), the optimal production path becomes as following,

$$\frac{1}{dt}E_t d\left[Q(t)\right] =\frac{1}{C''\left[Q(t)\right]}\Big\{-(r-\sigma^2)\{P(t)-C'\left[Q(t)\right]\}+C'\left[Q(t)\right](\sigma^2\xi_{mc\theta})+\alpha P(t)\Big\}. \tag{B.10}$$

Appendix C

Decomposition of Heating and Cooling Degree Days into Seasonalities and Shocks

The temperature fluctuations are not necessarily equivalent to temperature shocks. Campbell and Diebold (2005) argue temperature shocks are associated with risks, surprises or noises that are unpredictable components of temperature fluctuations. Thus, even very large temperature fluctuations would create little weather risk if they were highly predictable. In this sense, temperature fluctuation can be decomposed into two parts: the predictable variation of seasonality and unpredicted non-seasonal shocks. The former can be estimated based on historical data, while the latter represents unexpected temperature shocks. Specifically, I can write down temperatures as,

$$DD_t = S_t + h_t, \tag{C.1}$$

where DD_t is the temperature measured by heating or cooling degree days at time t; S_t denotes the seasonally mean temperature at time t and h_t is the unexpected temperature shocks. To obtain temperature shocks, the first step is to estimate seasonality S_t, which is considered as historical mean temperatures to which the actual temperatures would finally revert. The seasonality can be parsimoniously estimated by a sine function (Benth and Benth, 2007; Zhou, 2010). However, in terms of degree days, this method is infeasible because the symmetric property of a sine function may impose the equality between the highest degree days and the lowest degree days within a year. However, this is not reasonable since degree days are just truncated functions defined by equation (C.2). As a result, consider the heating and cooling degree days for a month as the equations below.

$$\begin{aligned} HDD_t &= \sum_{n=1}^{30} \max(65 - T_i, 0), \\ CDD_t &= \sum_{n=1}^{30} \max(T_i - 65, 0). \end{aligned} \tag{C.2}$$

where T_i is the temperature for day i measured by Fahrenheit. Based on this definition, HDD or CDD is cumulative degrees that the temperature deviates from a reference level within a month. $65^\circ F$ is the reference level because the temperature below that level would trigger more energy use for the purpose of heating, whereas temperature above $65^\circ F$ would induce using air conditioners for cooling. I estimate the seasonality by a regression below,

$$\begin{aligned} ln(HDD_t) &= c_h + \sum_{j=1}^{11} \eta_{hj} D_{jt} + \gamma_h t + \varepsilon_{ht}, \\ ln(CDD_t) &= c_c + \sum_{j=1}^{11} \eta_{cj} D_{jt} + \gamma_c t + \varepsilon_{ct}, \end{aligned} \tag{C.3}$$

where D_{jt} is a set of monthly dummy variables; t is the time trend variable to capture the potential effect of temperature rise or drop over time; and η_{hj}, η_{cj}, γ_h, and γ_c are the parameters. The log specification is applied in order to avoid negative fitted values of HDD or CDD. The exponential of fitted value of equation (C.3) is considered as seasonally mean HDD S_{ht} or CDD S_{ct} (i.e., seasonality), while the residuals ε_{ht} and ε_{ct} are considered as unexpected colder-than-normal and hotter-than-normal degree-day shocks. Because HDD and CDD have many zero in certain months, the estimated coefficients of some monthly dummies or time trend could be insignificant. I thus drop the insignificant independent variables and re-estimate (C.3) to improve the precisions. I then calculate seasonality and temperature shocks for each month in each state based on fitted values and residuals.