Capacity and Utilization Choice in the US Oil Refining Industry*

Matthew Chesnes

Department of Economics
University of Maryland, College Park
chesnes@econ.umd.edu
http://www.econ.umd.edu/~chesnes

Job Market Paper

November 2008

*I thank John Rust, Ginger Jin and Peter Cramton for their invaluable guidance and the Economic Club of Washington for their financial support. I am also grateful to John Shea, Mark Duggan, Adam Copeland, David Givens, Keith Kranker, Ariel BenYishay and seminar participants at the University of Maryland, the Federal Reserve Board of Governors, and La Pietra-Mondragone Workshop in Economics for their suggestions and comments. All remaining errors are my own.

Abstract

This paper presents a new dynamic model of the operating and investment decisions of US oil refiners. The model enables me to predict how shocks to crude oil prices and refinery shutdowns (e.g., in response to hurricanes) affect the price of gasoline, refinery profits, and overall welfare. There have been no new refineries built in the last 32 years, and although existing refineries have expanded their capacity by almost 13% since 1995, the demand for refinery products has grown even faster. As a result, capacity utilization rates are now near their maximum sustainable levels, and when combined with record high crude oil prices, this creates a volatile environment for energy markets. Shocks to the price of crude oil and even minor disruptions to refining capacity can have a large effect on the downstream prices of refined products. Due to the extraordinary dependence by other industries on petroleum products, this can have a large effect on the US economy as a whole.

I use the generalized method of moments to estimate a dynamic model of capacity and utilization choice by oil refiners. Plants make short-run utilization rate choices to maximize their expected discounted profits and may make costly long-term investments in capacity to meet the growing demand and reduce the potential for breaking down. I show that the model fits the data well, in both in-sample and out-of-sample predictive tests, and I use the model to conduct a number of counterfactual experiments. My model predicts that a 20% increase in the price of crude oil is only partially passed on to consumers, resulting in higher gasoline prices, lower profits for the refinery, and a 45% decrease in total welfare. A disruption to refining capacity, such as the one caused by Hurricane Katrina in 2005, raises gasoline prices by almost 16% and has a small negative effect on overall welfare: the higher profits of refineries partially offsets the large reduction in consumer surplus. As the theory predicts, these shocks have a smaller effect on downstream prices when consumer demand is more elastic, resulting in a larger share of total welfare going to the consumer.

1 Introduction

The United States is the largest consumer of crude oil in the world and this resource accounts for 40% of the country's total energy needs.[1] Although a majority of this oil comes from foreign sources, almost all is refined domestically. Refineries distill crude oil into a large number of products such as gasoline, distillate (heating oil), and jet fuel. While much attention has been paid to the upstream crude oil production industry (see Hamilton (1983) and Hubbard (1986)), and the downstream retail sector (see Borenstein (1991 & 1997)), very little research has focused on the role of the refining industry. Two important dynamic decisions faced by refiners are their investment in capacity and the utilization rate at which they run their plant. These choices are defined over different time horizons.[2] The optimal choice of capacity accumulation, i.e., the increased ability to distill crude oil into higher valued products, is a long-term decision. Capacity is expensive to build and may take time to come online so forecasts of future market conditions are crucial. A shorter-term problem involves a refiner's choice of capacity utilization. This rate measures the intensity with which a firm uses its capital, which for a refinery may include the use of boilers, distillation columns, and downstream cracking units.[3]

The refiner's problem is further complicated by changing market conditions, geopolitical tensions, and unexpected events, such as hurricanes. The largest component of refiners' output is gasoline. New alternative technologies, such as hybrid cars, and changing perceptions on the environmental impact of gas-powered vehicles has affected the sensitivity of consumer demand to the price of gasoline.[4] This affects the ability of refiners to pass through shocks to the price of crude oil resulting from, for example, reduced production from OPEC countries or a war in the Middle East. With about one-half of US refining capacity located along the Gulf of Mexico, the potential for hurricanes can also dramatically affect the ability of the industry to supply a consistent flow of gasoline and other products to the rest of the country.

This paper develops and estimates a new dynamic model of the operating and invest-

[1]Source: *2007 Annual Energy Review*, Energy Information Administration (EIA).

[2]In addition, they must solve a complicated linear programming problem because their relative output prices are constantly changing and they have the choice of utilizing different types of crude oil, some of which are better adapted to producing certain products.

[3]More details on the refining process can be found in section 2 and in appendix A.

[4]Knittle et al. (2008) and Espey (1996) both study the recent changes in consumers' price elasticity of demand for gasoline.

ment decisions of US oil refiners. These refiners face the possibility of breaking down if they run their plant too intensively, so they make costly investments in capacity to reduce this potential and to meet the growing demand for their products. My model assumes that firms are Cournot competitors in the refined product market. With many small firms, each is approximately a price-taker in the market, so the model of Kreps and Scheinkman (1983), with quantity pre-commitment (capacity choice) and Bertrand price competition, is similar to my approach. The model enables me to predict how shocks to crude oil prices and refinery shutdowns (e.g., in response to hurricanes) affect the price of gasoline, refinery profits, and overall welfare.[5] I also estimate how a change in the price sensitivity of consumers may affect the results of these shocks, particularly in regards to the division of welfare between the refiner and the consumer.

I estimate a fully dynamic model of the oil refining industry incorporating key decisions made by plants which affect both contemporaneous and future profitability. The refining industry is inherently forward-looking and decisions made today rely heavily on forecasts of future market conditions. A static model would not, for example, account for the increased breakdown potential of a plant from high utilization rates or the appropriate long-term investments of a refiner facing rising crude oil costs and uncertain demand. My estimation algorithm involves classic policy function iteration nested inside a GMM optimization, which allows me to compute the equilibrium value and policy functions.[6] This approach allows me to run various counterfactual experiments and determine the optimal policy and future discounted profits of each firm. Several recent papers, including Bajari et al. (2007) and Ryan (forthcoming), estimate dynamic models of firm behavior using a 2-step method that reduces the computational complexity of finding the structural parameters, but does not allow one to compute the equilibrium under counterfactual environments.

My model predicts that a 20% increase in the price of crude oil is only partially passed on to consumers, resulting in a 13% increase in gasoline prices, lower profits for the refinery, and a 45% decrease in total welfare. The pass-through result is fairly close to the historic rate of about 50%.[7] Consumer surplus falls following the shock, but the change in the overall distribution of welfare depends on the sensitivity of consumer

[5] I define total welfare to be the sum of consumer surplus and refiner profit.

[6] See Rust (2008).

[7] See Borenstein and Shepard (1996) and Goldberg and Hellerstein (2008) for related literature on price pass-through.

demand to the prices of refined products. More sensitive consumers sacrifice less and receive a larger share of the (smaller) surplus. I also show that a disruption to refining capacity, such as the one caused by Hurricane Katrina in 2005, raises gasoline prices by almost 16% and has a small negative effect on overall welfare: the higher profits of operating refineries partially offset the large reduction in consumer surplus. When Hurricane Katrina hit the Gulf Coast in August 2005, the actual wholesale gasoline price rose by 14% the following month.

Much of the literature on retail gasoline markets has focused on the asymmetric response of gasoline prices to crude oil shocks, the so-called *rockets and feathers* phenomenon (for example, see Borenstein (1997), Bacon (1991), and Noel (2007)).[8] Recent research on the wholesale gasoline market includes Hastings et al. (2008), which analyzes wholesale prices and the effects of new environmental regulations, and studies by The Government Accountability Office (2006), the Federal Trade Commission (2006), and the Energy Information Administration (2007).

To my knowledge, this is the first dynamic model of the US oil refining industry. Refiners play an important role as an intermediary between upstream crude suppliers and downstream retail markets. A complete analysis of the oil industry must account for the important effects of the refiners' dynamic decisions. I show that the model fits the data well and can be used to generate insights into the pass-through of crude oil shocks and the impacts of refinery shutdowns on consumers. The model's main features include a dynamic decision process, long-term investment choices, and the possibility of plant break-down. The framework could be applied to other energy markets as well as industries, such as shipping, that make large investments in capacity based on expectations of future market conditions.

The remainder of this paper is organized as follows. In section 2, I provide an overview of the oil refining industry to better understand the complicated problem facing the refiner. I describe my data in section 3 and lay out a dynamic model of the industry in section 4. Section 5 provides the details of my empirical strategy and I summarize the fit and results of the model in section 6. Finally, in section 7, I use my estimated parameters to run several counterfactual experiments involving shocks to the price of crude oil,

[8]The market power gained by the refining industry due to a tight capacity environment is one potential explanation. Others include search costs in the retail market, inventory management by consumers who may fill their tank more frequently as prices rise, but are less eager to "top-off" when prices are falling, and adjustment costs at the refinery.

refining capacity, and consumers' price elasticity of demand. Section 8 concludes and provides a discussion of potential extensions.

2 The US Oil Refining Industry

The oil industry is broadly comprised of several vertically oriented segments. They include crude oil exploration and extraction, refineries which distill crude oil into other products, pipeline distribution networks, terminals which store the finished product near major cities, and tanker trucks which transport products to retail outlets.[9] The largest refined product, gasoline, accounts for about 50% of total production, while distillate makes up another quarter. A full 68% of output from the oil refining industry is used in the transportation industry. Figures 1 and 2 provide a description of the production process and average product yields. The main distillation process produces some final products like gasoline, but it is complemented by other units that extract more of the highest valued products. Technical details of the refining process and background on the types of crude oil available can be found in the appendix.

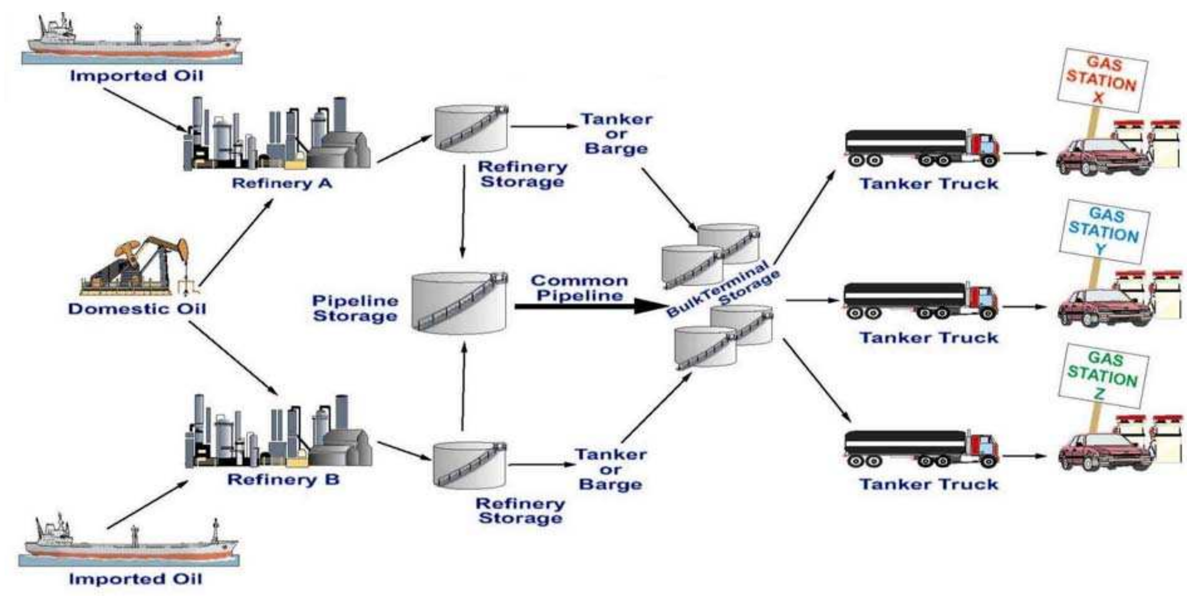


Figure 1: Production Process

[9]75% of terminals in the US are owned by companies not involved in the upstream exploration and refining.

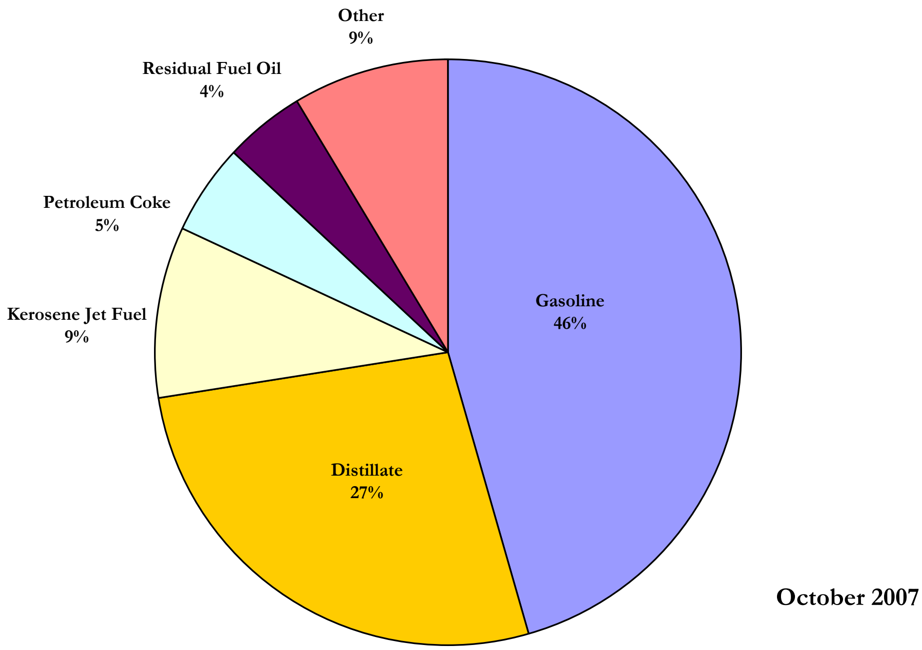


Figure 2: Average Yields

The market for refined oil products is large and growing, with the US consuming 388 million gallons of gasoline each day and one quarter of the world's crude oil.[10] Aside from refining crude oil into gasoline, refineries produce many products that are important inputs into other industries. Retail gasoline prices have recently experienced increased variability in the US and in summer 2008 hit an all time high of $4.11 per gallon. Wholesale prices peaked around $3.40 a gallon in the same period.[11] Many justify the high prices as a result of the growing demand for gasoline and supply limitations, including the scarcity of crude oil, Middle East uncertainty, hurricanes, and the OPEC cartel. Others claim the high prices result from coordinated anticompetitive behavior by big oil companies. It may be that the strategic capacity investment and utilization choices by oil refineries play a significant role in affecting downstream prices, profits, and consumer welfare.

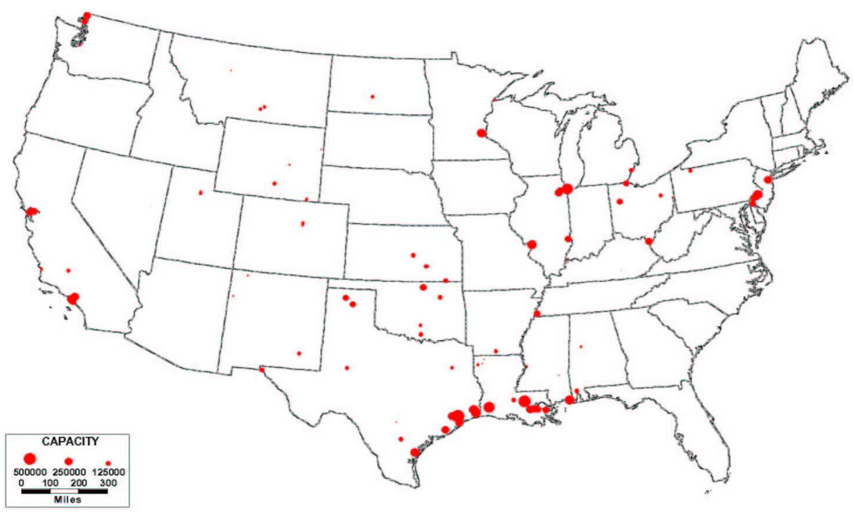
[10] Annual world consumption of crude oil totals 30 billion barrels, of which 7.5 billion barrels comes from the US. About 60% of crude oil used by refineries is imported and US consumption of refined gasoline represents 40% of world consumption.

[11] US regular gasoline, source: EIA.

2.1 Competition

Concentration

The refining industry is fairly competitive, with 144 refineries owned by 54 refining companies in January 2006. About one-half of US production occurs near the Gulf of Mexico in Texas and Louisiana, though there are significant operations in the Northeast, the Midwest, and California. During World War II, the country was divided into Petroleum Administration for Defense Districts (PADDs) to aid in the allocation of petroleum products. Figure 3 displays a map of refinery locations along with delineations of PADDs and PADD districts. PADDs are often used by regulators such as antitrust authorities when assessing market concentration. See table D.1 in appendix D for concentration ratios and Herfindahl-Hirschman Indices (HHIs) for various PADDs and regions at the refiner level. The degree of market concentration is clearly dependent upon how one defines the relevant geographic market.[12]



PADD	District	Area
I	1	East Coast
II	2	Midwest
	3	Upper Midwest
	4	Central Plains
III	5	Louisiana
	6	Texas
IV	7	New Mexico
	8	Rockies
V	9	West Coast

Figure 3: Refinery Locations (Scaled by Capacity)

Market Definition

While retail markets for gasoline tend to be very small, markets for wholesale gasoline are relatively large due to the extensive pipeline network use to transport most refined

[12]At the national level, the top four refiners (who each own multiple refineries) controlled 44.1% of the market in 2007. The HHI for refiners on the Gulf Coast was about 1,100, which would be classified as *moderately concentrated* according to the Horizontal Merger Guidelines.

products. While a PADD may have roughly approximated a market in 1945, these delineations were made before the pipeline network had been fully developed, so they are now just a convenient way to report statistics on the industry.[13] A map of major crude oil and production piplines is shown in figure 4. With important pipelines connecting the Gulf Coast production center to the population centers in the Northeast and the Midwest, I combine PADDs 1, 2, and 3 into one large market for wholesale gasoline. I denote the Rocky Mountain region, PADD 4, as another market, because it is isolated from the rest of the country and imports only limited refined product from other regions. Finally, my third market is the West Coast, PADD 5, which includes California, a state that, due to strict environmental regulations, is limited in its ability to use products that are refined in other states.

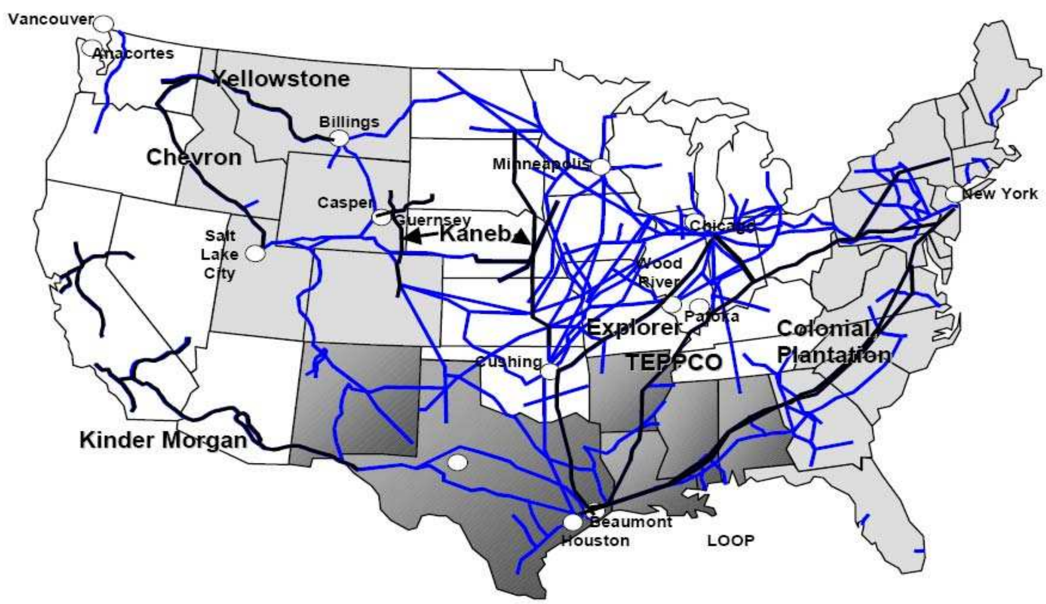


Figure 4: Major Refined Product Pipelines

Aside from the domestic refining industry, US refiners face limited competition from abroad. While the US is very dependent on foreign oil, domestic production accounts for about 90% of US gasoline consumption, though the import share has grown since the mid 1990s. These imports come primarily into the Northeast, which receives 45%

[13]For instance, the Colonial pipeline, which runs from the Gulf Coast up to the Northeast, was built in 1968. Pipelines now carry 70% of all refined products shipped between PADDs.

of its supply from sources, such as the US Virgin Islands, the United Kingdom, the Netherlands, and Canada. Recent US regulations limiting certain types of fuel additives combined with increased European dependence on diesel fuel has limited the ability of US markets to rely on foreign imports.

2.2 Capacity and Utilization

Capacity utilization rates at US refineries have been steadily rising and are now at their maximum sustainable levels. From 2000 to 2008, the average utilization rate in US manufacturing industries was 77%, while in the refining industry it was 91%.[14] At the same time, no new refineries have been built in the US since 1976. In fact, many plants have closed and the number of refineries has fallen from 223 in 1985 to just 144 today. However, most of these closures were small and inefficient plants, and those that remain have expanded, so total operable capacity has grown from 15.6 million barrels per day (bbl/day) in 1985 to almost 17 million bbl/day today. However, this figure is lower than in 1981, when capacity was 18.6 million bbl/day. The overall number of refineries along with their production capacity are displayed in figure 5. The average plant size has increased from 74,000 bbl/day in 1985 to almost 124,000 bbl/day in 2007.

Building a new refinery is very expensive, and environmental requirements and permits create significant hurdles.[15] Evidence from a 2002 US Senate hearing estimated the cost of building a 250,000 bbl/day refinery at around 2.5 billion dollars, with a completion time of 5-7 years (Senate (2002)). This assumes the various environmental hurdles and community objections are satisfied. No one wants a dirty refinery operating near them.[16] In May 2007, the chief economist at Tesoro, Bruce Smith, was quoted as saying that the investment costs in building a new refinery are so high that "you'd need 10 to 15 years of today's margins [at the time, around 20%] to pay it back."[17] Even without new refineries, existing refineries have invested to expand capacity. The distribution of

[14]See http://www.federalreserve.gov/releases/G17/caputl.htm.

[15]One of the few new plants in development is in Yuma, Arizona. The builder of the 150,000 bbl/day refinery has spent 30 million dollars over 6 years to acquire all the permits. If not blocked, construction on the new refinery will begin in 2009.

[16]Commonly referred to as "NIMBY," an acronym for Not In My Back Yard.

[17]The National Petrochemical & Refiners Association estimates that the average return on investment in the refining industy between 1993-2002 was 5.5%. The S&P 500 averaged over 12% for the same period. See "Lack of Capacity Fuels Oil Refining Profits" available online at http://www.npr.org/templates/story/story.php?storyId=10554471 (downloaded: 09/13/2008).

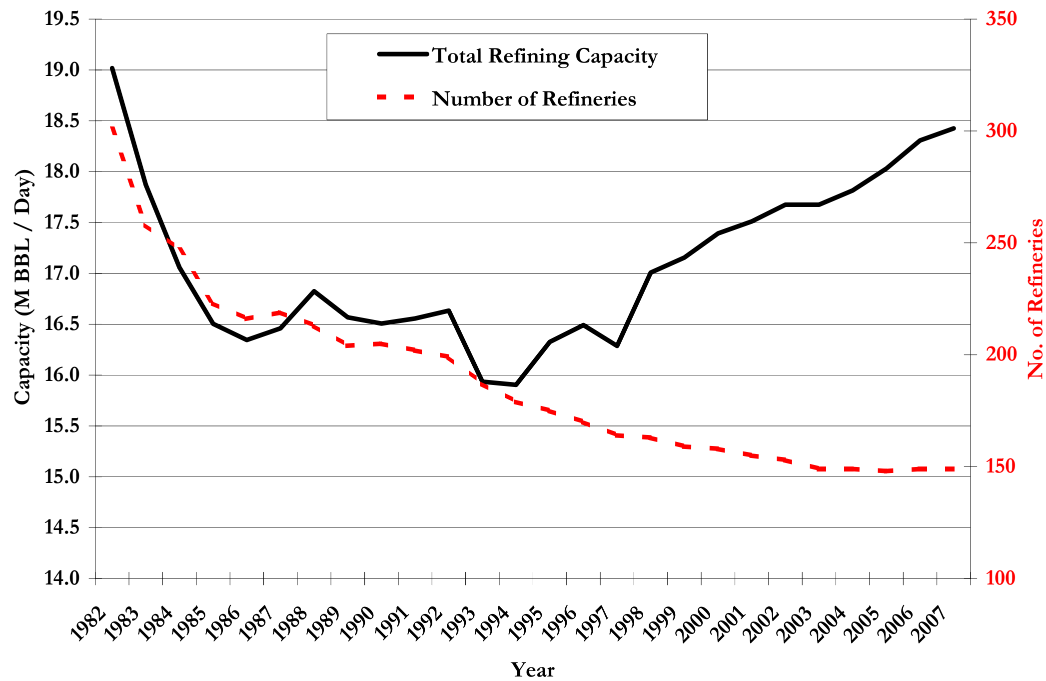


Figure 5: Capacity and Number of Refineries

historical investment rates is shown in figure 6. While the mean investment has been 1.3% per year, the median is zero as plants tend to make very infrequent investments. Even restricting the sample to non-zero changes as shown in the graph, investments tend to be small, with almost 85% of the non-zero changes less than 10%.

Although oil refining has historically been an industry plagued by thin profit margins, oil producers are now starting to make higher profits from their refining business. One simple measure of the profit margin at a refinery is the "crack spread." For every barrel of crude oil the refinery uses, technological constraints require that about half of it goes into gasoline production and about a quarter into distillate. So the crack spread, expressed in dollars per barrel, is calculated as:

$$Crack = \frac{1 * Price(distillate) + 2 * Price(gasoline) - 3 * Price(crude\ oil)}{3}.$$

The crack spread along with the utilization rates of refineries are shown in figure 7. The crack spread hit a record high of nearly $30 per barrel in July 2006. Some argue

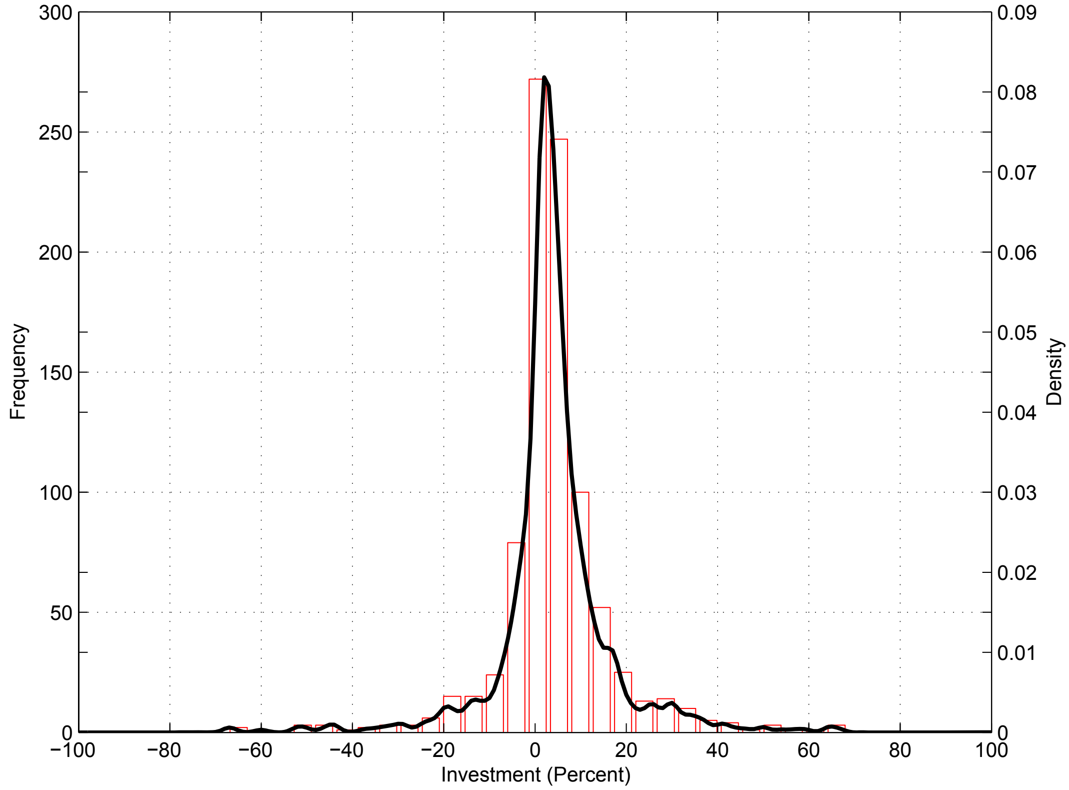


Figure 6: Non-Zero Changes in Capacity, All Plants, 1986-2007

that based on this measure of profitability, it is surprising that more refiners have not overcome the setup costs and entered this industry. The increase in the crack spread after 2000 occurred after the utilization rate had already been at a very high level. This may imply that a refiner's ability to pass through their crude oil cost has changed since 2000, perhaps due to the scarcity of crude oil, an increase in industry concentration, or an increase in the demand for gasoline.

While total refining capacity has risen in the past 10 years, it has not kept up with demand growth. Capacity of oil refiners has increased by 10% in the past 10 years, while demand for gasoline has increased about 17%. The gap has been filled by higher utilization rates and, to a lesser degree, growing imports. New regulations requiring the shift from MTBE[18] oxygenates to ethanol poses a problem for this segment of supply because foreign refiners have not invested in the facilities to produce ethanol blended gasoline. With capacity tight and supply alternatives limited, even a minor supply

[18] Methyl Tertiary Butyl Ether.

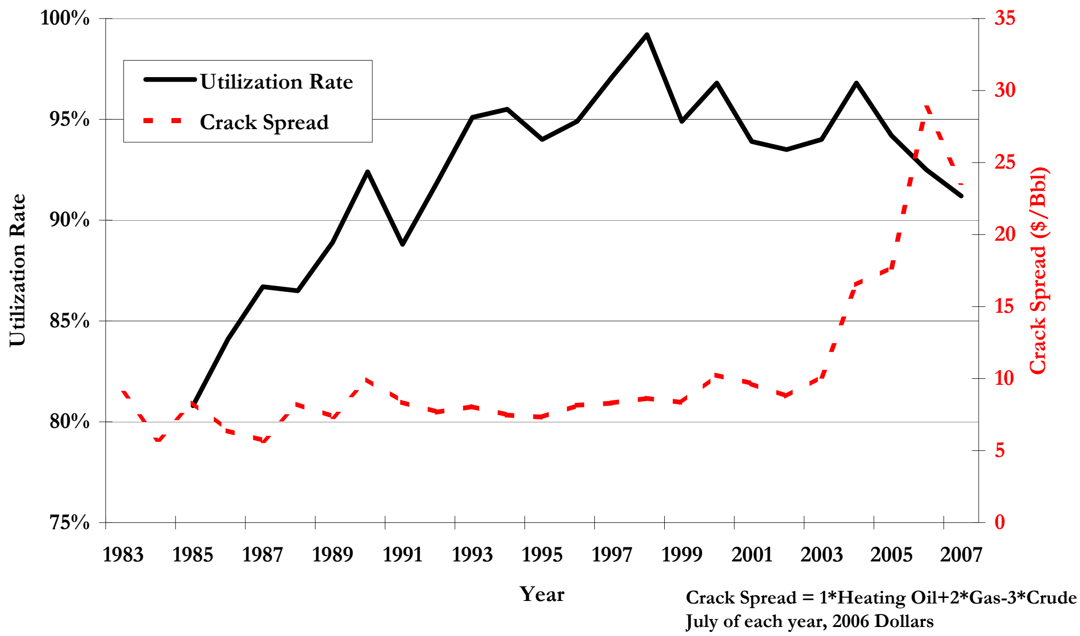


Figure 7: Capacity Utilization Rate and Crack Spread

disruption (or a major one like Hurricane Katrina) can have a large price impact.[19]

2.3 Refinery Maintenance and Outages

An oil refinery is a complex operation that requires frequent maintenance, ranging from small repairs to major overhauls.[20] The regular maintenance episodes tend to be short and have minimal impact on production as they are strategically scheduled for low demand periods. Unplanned major outages, by definition, can take place at any time and can have a major impact on production capability. The EIA divides refinery outages into four classes, summarized in table 1.

Planned turnarounds are major refinery overhauls, while planned shutdowns bridge the gap between turnarounds. Unplanned shutdowns involve unexpected issues that may allow for some strategic planning of the downtime, but often may force a refinery

[19]Following Hurricane Katrina on 9/23/05, capacity fell by 5 MBbl/Day. This represented a full one third of US refining capacity. Inventories are also limited as there is only about 20-25 days worth of gasoline in storage at any time.

[20]Refinery maintenance is crucial not only for production sustainability, but also for the safety of the plant. A 2005 fire at BP's Texas City refinery killed 15 workers and injured over 100 more.

Table 1: Refinery Downtime

Type	Typical Length of Outage	Frequency
Planned Shutdowns	1-2 Weeks	Every year
Unplanned Shutdowns	2-4 Weeks	-
Planned Turnarounds	3-9 Weeks	Every 3-5 years
Emergency Shutdowns	Varies	-

Source: EIA.

to reduce production sub-optimally. Finally, emergency shutdowns are those that cause an immediate plant breakdown like a refinery fire.

Organization for planned turnarounds typically start years in advance, and cost millions of dollars to implement, in addition to the revenue lost from suspending production. Due to the hiring of outside personnel, major refineries often have to plan these turnarounds at different times because of the shortage of skilled labor to implement them. Given the typical seasonal variation in product demand, the ideal periods for maintenance are the first and third quarter of the year, though in some northern refineries, cold winter weather forces shifts in planned downtimes.

Even though refineries consist of several components, such as distillation columns, reformers and cracking units, these components are dependent on one another so a breakdown of any one component can affect the production capability of the entire refinery. Downstream units include hydrocrackers, reformers, fluid catalytic cracking (FCC) units, alkylation units, and coking units. They are responsible for breaking down hydrocarbons into more valuable products and removing impurities such as sulfur. For example, in a typical refinery, only 5% of gasoline is produced from the primary distillation process; the rest comes hydrocrackers (5%), reformers (30%), FCC and alkylation units (50%), and coking units (10%). Not all refineries have all of these components, so such refineries are even more affected when one component goes down (EIA (2007)).

At the PADD level, EIA reports that in the 1999-2005 period, refineries experienced reductions in monthly gasoline and distillate production of up to 35% due to outages. At the monthly frequency, there is little effect of outages on product prices. This is primarily because most (planned) outages occur during the low-demand months when markets are not tight; most outages last less than a month; and the availability of imports, increased production from other refineries, and inventories provide a cushion to supply. However, major outages, like those caused by a hurricane, still affect the downstream prices and profitability of all refineries.

Overall, the oil refining industry features several economic *puzzles*, some of which I explore in this paper. While the industry is relatively competitive, refiners have recently been earning significant profits, as measured by the growing crack-spread. However, entrants have yet to overcome the regulations and costs of setting up a new plant and existing firms have been cautious in their expansion. As a result, plants run at high rates of utilization, which leads to instability in the face of unexpected capacity disruptions.

3 Data

The EIA publishes data on the oil refining industry at various frequencies and levels of aggregation.[21] I observe monthly district level data, which is publicly available on EIA's website.[22] For every month in the years from 1995 to 2006, and for each of the 9 refining districts, I have the following data:

- Wholesale gasoline production, sales, and prices.
- Wholesale distillate production, sales, and prices.
- Crude oil first purchase price and inputs into refineries.
- The capacity utilization rate.

This provides 1,296 observations. I also have annual firm level data for the same years on the capacity to distill crude oil. The reported capacity, called the *atmospheric crude oil distillation capacity*, measures the number of barrels of crude oil that a refinery can

[21] Although monthly plant level data is collected from individual refineries on EIA form 810, this data remains proprietary and unavailable to academic researchers. A new program, joint with the National Institute for Statistical Sciences (NISS), called the NISS-EIA Energy Micro Data Research Program, may allow access to this data (http://www.niss.org/eia/niss-eia-microdata.html). The dataset includes monthly observations for all refineries in the US on production, capacity, utilization, and inputs into production. The program is currently on hold.

[22] See http://tonto.eia.doe.gov/dnav/pet/pet_pnp_top.asp. There are 9 refining districts, including the East Coast, the Midwest, the upper Midwest, the Central Plains, Louisiana, Texas, New Mexico, the Rockies, and the West Coast.

process through the initial distillation process. This measure is calculated on a stream-day basis.[23]

There are 246 unique plants in the dataset, with 179 active in 1995 and 144 active in 2006. Overall, I observe a total of 1,959 plant-year observations. Table 2 summarizes the data by district and indicates the market definitions I use in my estimation. The number of plants and aggregate capacity are for January 2006.

Table 2: Industry Summary

Market	District	States	No. Plants	Ref. Cap. (Mbbl)
1	1	CT, DE, DC, FL, GA, ME, MD, MA, NH, NJ, NY, NC, PA, RI, SC, VT, VA, WV	14	659
1	2	IL, IN, KY, MI, OH, TN	14	913
1	3	MN ND, SD, WI	4	171
1	4	IA, KS, MO, NE, OK	8	306
1	5	TX	23	1,812
1	6	AL, AR, LA, MS	27	1,353
2	7	NM	3	42
2	8	CO, ID, MT, UT, WY	16	232
3	9	AK, AZ, CA, HI, NV, OR, WA	35	1,220
			144	6,709

Proceeding with the district level data on production and utilization combined with capacity at the firm level requires some discussion. Implicitly, I must make the strong assumption that all firms within a district are identical and respond the same way to shocks. When aggregating to the district, one firm that increases production may be cancelled out by another that breaks down. Thus, results from this approach will be meaningful only in terms of assessing the "average" behavior of a firm within a district. However, there is significant variation in district production levels as well as in the break-down episodes described below. Also, aggregating to the district level when I estimate my model avoids having to account for the complicated linear programming problem

[23]Capacity reported in barrels per stream-day equals the maximum number of barrels of oil that a refinery can process on a given day under optimal operating conditions. Calendar-day capacities assume *usual* rather than optimal operating conditions, though these two numbers are frequently reported as identical.

faced by an individual refinery. These idiosyncratic differences should be smoothed out in the higher level data.

4 Model

Firms make annual investments to increase or decrease their available capacity. I assume these investments increase or decrease capacity immediately and that firms then choose their utilization rates each month. While empirically, some plants make major investments in capacity that take years to complete, the average investment is small and can be completely quickly.[24] Though plants require a certain minimum level of maintenance each year (usually carried out just before the summer driving season), running a plant at a high utilization rate in one month increases the probability of a plant breakdown or an extended maintenance episode in the next month. Thus, faced with relatively high product prices or low crude oil input prices (a high refining margin or *crack spread*), firms may want to run their plants at a high rate of utilization to maximize profits. However, this intensive use of capital may increase the possibility of a breakdown next month when prices may be even higher.

I model the competitive environment by assuming that plants are price-takers in the market for crude oil but are Cournot competitors with some (small) market power in the downstream refined products market. Since I do not observe plant level production choices, the model is best described as a representative-agent Cournot model. In each period, a firm optimally chooses its utilization rate in response to its estimate of the aggregate production of its competitors.

With the development of a network of pipelines across the US after World War II, markets tend to be large and feature many firms producing a homogeneous product. Firms are differentiated not only by their capacity to turn crude oil into gasoline and other products, but also by their technical capabilities to utilize varying types of crude oil in their production. I focus on the capacity differentiation and average firm behavior to smooth over the technical production heterogeneity.

[24]These small investments, known as *capacity creep*, include both additional infrastructure and improved through-put of existing capital.

4.1 A Firm's Problem

Consider the problem of firm i in month m.[25] I will focus only on gasoline and distillate production by refineries, since these account for about three-quarters of the production of an average refinery. Denote production of gasoline and distillate as q^g_{im} and q^d_{im}, and the capacity of the refinery as $\overline{q}_{iy}$, where y indexes the current year. Given the investment behavior of firms, I assume that investments in capacity are made only once per year and the resulting capacity is fixed for the entire year. Let r_{iy} denote the investment of the firm, expressed as the proportional increase or decrease in capacity.

A firm's problem can be written as:

$$Max_{\{r_{iy}\}_{y=0}^{\infty}} \;\; E\left[\sum_{y=0}^{\infty} \delta^y \Pi_{iy}(r_{iy}; x_{iy})\right], \tag{1}$$

$$\Pi_{iy} = Max_{\{u_{im}\}_{m=1}^{12}} \;\; E\left[\sum_{m=1}^{12} \mu^{m-1} \pi_{im}(u_{im}; x_{im}, \overline{q}_{iy})\right]. \tag{2}$$

I assume capacity evolves according to:

$$\overline{q}_{iy} \;\; = \;\; \overline{q}_{i,y-1}(1 + r_{iy}), \tag{3}$$

where r_{iy} is net of any depreciation of existing capital. The utilization rate can be expressed as:

$$u_{im} \;\; = \;\; \frac{q_{im}}{\overline{q}_{iy}}, \tag{4}$$

where $q_{im} = q^g_{im} + q^d_{im}$. While this is not a classic utilization rate, in that it does not assess the proportion of available inputs that are actively being used, technical constraints on the proportion of total capacity that can be used to produce gasoline and distillate makes this ratio approximately a scaled down version of the actual rate. $\pi_{im}(\cdot)$ is the per-period profit function, x_{im} and x_{iy} are vectors of state variables, and δ and μ are the discount rates, with $\delta = \mu^{12}$. Note that $\overline{q}_{iy}$ appears as a state variable in equation 2 and equals last year's capacity plus or minus the investment made at the beginning of the current year.

[25] I assume that firms are individual plants and use the two terms interchangeably.

Throughout a given year, state variables observable to the firm include the following:

P^c_{jm}	The price of crude oil
B_{im}	An indicator equal to 1 if the firm is in a breakdown episode
$Q_{-i,m}$	The estimated aggregate competing production by other firms in the market
$\overline{q}_{iy}$	A firm's capacity
$Time$	Month & year

I explicitly include a district j index on the crude oil price because, while I assume this price is exogenous, there are differences in the quality and price of oil in different districts. The competing production state is needed to calculate the price of a firm's output. With the large number of firms in the industry, each firm has only a small impact on the prices of gasoline and distillate.[26] Firms form a statistical forecast of competing production as follows:

$$E[Q_{-i,m}] = Q_{-i,m-1}(1+g_m), \tag{5}$$

where g_m is the historical growth rate of production in the market between months $m-1$ and m. The month of the year is included to capture the obvious and important seasonal effects. For example, a refinery operator may forgo preventative maintenance measures during the summer high-demand period to capitalize on the high prices and profit margins. The expectation operator is taken over the future profile of the state variables, some of which are deterministic (month and year), others of which evolve according to the firm's choices (capacity and breakdown), and still others are stochastic, for which firms base their expectations on historical values (the crude price and competing production).

Due to breakdowns, only a portion of $\overline{q}_{iy}$ will be available in a given month. I denote the *available* capacity as $\overline{q}^*_{iy}$. Because the numerator in equation 4 is the volume of downstream products and the denominator is the number of barrels of crude oil that a refinery can distill, the utilization rate may be greater than 1 in some cases. This occurs because chemicals called blending components are added in the distillation process (such

[26] With plant-level production data, I could explicitly solve for the (asymmetric) Cournot equilibrium in each period. I plan to adopt this approach in future research.

as oxygenates like MTBE and ethanol).

Note that the firm's objective function can be written recursively. Denote $V(\cdot)$ to be the present discounted value of the stream of refiner's profits with optimal choices. Then, after dropping subscripts and discretizing the state space, the Bellman equation can be written:

$$V(x) \;=\; Max_r\Big\{\Pi(r;x) + \delta\sum_{x'} V(x')P(x'|x,r)\Big\}. \tag{6}$$

Here $P(\cdot)$ is the annual probability transition matrix and it reflects the transition between average annual values of the state variables. To solve for $\Pi(r;x)$, I apply backward induction from December back to January. For example, the expected value of a refiner's aggregate discounted profit from July onward is:

$$W_6 \;=\; Max_{u_6}\Big\{\pi_6(u_6;x_6,\overline{q}) + \mu\sum_{x_7} W_7(x_7)P^*(x_7|u_6,x_6,\overline{q})\Big\}. \tag{7}$$

Here, $P^*(\cdot)$ is conditional on u and $\overline{q}$ because plants that do not invest in new capacity and choose to operate more intensively increase their probability of breaking down.

4.2 Per-Period Profit

Prices are determined at the market level, which I index by k. Per-period profit is defined as gasoline and distillate revenue less production costs and investment costs. Thus, in month m, profits of firm i are:

$$\begin{aligned} \pi_{im}(u_{im};P^c_{jm},,B_{im},Q_{-i,m},\overline{q}_{iy},m,y) \;&=\; u_{im}\overline{q}^*_{iy}[(yield^g)P^g_{km}(Q^g_{km};m,y) \\ &+\; (1-yield^g)P^d_{km}(Q^d_{km};m,y)] \\ &-\; C_{im}(u_{im};P^c_{jm},\overline{q}^*_{iy}) \\ &-\; \frac{1}{12}C^r_{iy}(r_{iy}), \end{aligned} \tag{8}$$

where,

$$\overline{q}^*_{iy} = \begin{cases} \overline{q}_{iy} & \text{if } B_{im}=0 \\ \phi\overline{q}_{iy} & \text{if } B_{im}=1. \end{cases} \tag{9}$$

The term $yield^g$ represents the proportion of available capacity that can be distilled into gasoline. It is fixed over time and across firms. Functional forms for the demand and cost functions will be specified below. The last term in the profit function is the investment cost, which is spread equally across the 12 months of a year. Note that $\phi \in [0, 1)$ reflects the percentage reduction in capacity that a refinery experiences during a breakdown. While I allow this term to vary stochastically, the data suggest this value averages around 0.9 and can fall as low as 0.7. In other words, district level breakdowns occur that result in a 30% reduction in capacity relative to normal levels. It should be noted that a 25% capacity reduction in a given month could result from one week of complete breakdown and three weeks of optimal operation.

4.3 Demand

The prices of gasoline and distillate are determined at the "market" level. The three markets defined earlier are: the East Coast, Midwest and Gulf Coast; the Rocky Mountain region; and the West Coast. The first is by far the largest, with several large pipelines connecting the major production area near the Gulf of Mexico with the population centers on the East Coast and in the Midwest. I estimate the demand for wholesale gasoline (and similarly for distillate) according to:

$$log\ Q^g_{km}(P^g_{km}) \quad = \quad \alpha^g_0 + \alpha^g_1(Month) + \alpha^g_2\ log\ P^g_{km} + \epsilon^g_{km}. \tag{10}$$

P^g and Q^g are the price and quantity demanded of wholesale gasoline. Here I specify a log-linear demand equation with month fixed effects to account for the strong seasonal variation in demand. I estimate the demand separately for each of the 12 years to account for not only the growing demand for refined products, but also changes in the sensitivity of consumers to prices.

Note that the East Coast receives a significant amount of their refined product from abroad (mostly from Europe and the Caribbean). Imports increase in periods of high demand or tight supply, as the price must be high enough to justify the transportation costs. Thus the demand for refined products from US refineries may be affected by the availability of imports, though robustness checks reveal that the effect is small relative to the size of the East Coast's overall market (which includes the Midwest and Gulf Coast).

4.4 Probability of Breakdown

Consider the following specification for the likelihood of a plant breakdown or extended period of maintenance beyond the regular minimum level:

$$Pr(\text{breakdown in month } m) = F(\beta u_{i,m-1}) \;=\; \frac{exp(\beta_0 + \beta_1 u_{i,m-1})}{1 + exp(\beta_0 + \beta_1 u_{i,m-1})}, \qquad (11)$$

which assumes the probability follows the logistic distribution. With more detailed firm-level data, an ordered probit may be the ideal specification, as it would account for both the magnitude and length of the breakdown episode. Modeling the breakdown dynamics based solely upon last month's utilization rate, and not, say, the average rate over the last six months, is primarily a computational simplification. The results using only last month's utilization rate are robust to other specifications.[27] See below for how I define a breakdown using district-level production data.

4.5 Production and Investment Costs

I assume the following production cost specification:

$$C_{im}(u_{im}; P^c_{jm}, \overline{q}^*_{iy}) \;=\; \gamma_0 * q_{im} + \gamma_1 * q^2_{im} + \gamma_2 * q_{im} * P^c_{jm}, \qquad (12)$$

where $q_{im} = u_{im}\overline{q}^*_{iy}$, the firm's actual production of gasoline and distillate in the current month.

I assume firms face increasing costs as they near their capacity constraint. To model this, I suppose firms have a quadratic production cost function and also include a term, γ_2, reflecting the major input of the refiner, crude oil. Refiners take this crude oil price as exogenous since the price is determined on the world market. As firms produce near their capacity, they may face increasing costs due to less time for maintenance, excess wear on their capital, and other effects that raise their marginal costs.

Investments in capacity are available immediately, and capacity is fixed within the year. This is a strong assumption since firms likely make investment decisions far in advance and spread the costs over a long time period. In future work, I will relax this

[27]Specifications involving the prior 3-month average rate or last month's deviation from historical rates yielded similar results. With firm-level data on production, one could also include the age of the refinery and perhaps the length of time since the last significant maintenance period.

assumption, allowing for a one-year "time-to-build." Investments come at a cost:

$$C^r_{iy}(r_{iy}) \quad = \quad \gamma_3(\overline{q}_{i,y-1}r_{iy}) + \gamma_4(\overline{q}_{i,y-1}r_{iy})^2. \tag{13}$$

The parameters, γ_3 and γ_4, reflect the cost of capacity expansion. They embody both the cost of physical expansion and any regulatory costs faced by the plant. Unfortunately, I will not be able to differentiate these two components with currently available data. Note that the investment cost parameters reflect the cost of a change in the number of barrels of a capacity that is created or destroyed. Large plants may benefit from economies of scale in capacity expansion as compared with smaller plants, but since I am estimating my model for an average capacity firm, this consideration is not necessary.

5 Empirical Estimation Strategy

In general, I split the estimation into two stages. I first estimate the demand parameters, $(\alpha^g_0, \alpha^g_1, \alpha^g_2, \alpha^d_0, \alpha^d_1, \alpha^d_2)$, via GMM. This is a static relationship between the market price and quantity. I also estimate the logit parameters governing the probability of breakdown, (β_0, β_1), via maximum likelihood.

In the second stage, I take the demand and breakdown coefficients as given and solve the firms' dynamic utilization and investment choice problem using a *nested fixed-point GMM algorithm* to recover the cost parameters $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ for each market. I allow for the cost parameters to vary each year to reflect changes in technology over time. I assume an annual discount rate of $\delta = 0.95$, implying a monthly rate of $\mu = 0.996$. When a firm enters a breakdown episode, I assume their capacity is reduced by a random amount, ϕ, which follows a beta distribution with mean 0.9.[28]

The firms' dynamic problem can be thought of as a finite-horizon monthly utilization choice problem nested inside an infinite-horizon annual investment choice problem. The annual investments in capacity can raise or lower the optimal utilization rate throughout the year, (e.g., a larger investment allows for the same level of output with a lower level

[28] Formally, $\phi \sim \mathcal{B}(9, 1)$.

of utilization). Recall that the problem can be written:

$$Max_{\{r_{iy}\}_{y=0}^{\infty}} \quad E\left[\sum_{y=0}^{\infty} \delta^y \Pi_{iy}(r_{iy}; x_{iy})\right], \tag{14}$$

$$\Pi_{iy} = Max_{\{u_{im}\}_{m=1}^{12}} \quad E\left[\sum_{m=1}^{12} \mu^{m-1} \pi_{im}(u_{im}; x_{im}, \overline{q}_{iy})\right]. \tag{15}$$

The aggregate discounted profits of the firm over the course of the year becomes the per-period (annual) payoff of the investment choice problem. Given the frequency with which refiners adjust their capacity and their utilization rate, this modeling strategy is not only realistic, but it is computationally appealing. Solving the finite horizon problem in equation 15 is simply a matter of backward induction.

The state variables available to the firm are the same in both sub-problems, aside from the month of the year, which is only relevant in the utilization choice problem. For the annual investment choice, the firm considers the average values of last year's crude oil price and market production, the proportion of time the refinery was broken down in the last 12 months, and the current level of capacity.

5.1 Demand

The demand parameters, the α's, are estimated in the first stage using 2-stage least squares with appropriate instruments. Given the endogeneity of P and Q, I need to find instruments, Z_{km}, that are correlated with the price, $Cov(P_{km}, Z_{km}) \neq 0$, and unrelated to error term, $Cov(\epsilon_{km}, Z_{km}) = 0$.[29] An obvious cost shifter in the oil refining industry is the price of crude oil, which should be exogenous as it's determined in the world market. However, it is likely that the market for crude oil and the market for refined products are both subject to the same demand shocks, which invalidates the contemporaneous crude oil price as a good instrument.

Therefore, I instrument for the price of wholesale products with the inventories of gasoline, distillate, and crude oil. These are industry-wide inventories, not just at the refinery. I have also included an indicator of major hurricanes and the lagged crude oil price, though the resulting estimates are largely unaffected. These should all be related to the price of the refiner's products though unrelated to the downstream demand. I can

[29]Essentially, I need cost shifters that move around the supply curve to trace out a demand curve.

use the R^2 from the first stage to test for the correlation between my instruments and the endogenous price. Since I've instrumented for price in the first stage, in the second stage I then regress the log of Q_{km} on $\hat{P}_{km}$ and month dummies.

5.2 Breakdown Probability

The parameters of the breakdown logit, β_0 and β_1, are estimated by maximum likelihood. I define a "breakdown" in district j as a month when the observed utilization rate u_{jm} (published by EIA, reflecting gross inputs of crude oil divided by the capacity to distill crude oil) drops below $\underline{u}_{jm}$, defined as:

$$\underline{u}_{jm} = min\left\{\frac{1}{9}\sum_{i=1}^{9} u_{im}, \frac{1}{4}\sum_{i=1}^{4} u_{j,m-12i}\right\}.$$

So the threshold is the smaller of the contemporaneous average across all districts and the average of the selected district's production in the *same* month for the last 4 years. So a breakdown is only triggered when 1) a district is producing relatively less than all other districts in the current month, and 2) the district is producing relatively less than it has historically in the same month. Figure 8 displays the breakdown dynamics for districts that experience a breakdown. The plots show that districts that run their plants more intensively in one month are more likely to break down the following month.

Once a breakdown episode is started, a district may stay below the threshold for a period of months. The data show that median episode length is 1 month, the mean is 2.3 months, and the maximum is 15 months.[30]

5.3 Production Cost Parameters

The cost parameters, $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$, are estimated by GMM in the second stage dynamic optimization. In order to solve for the production and investment cost parameters, I need to solve a dynamic optimization problem. To achieve this, I first discretize the

[30] The 15 month episode occurred in district 9 (the West Coast) from February 1999 - May 2000. It resulted from two California refinery fires at the Tosco Refinery in Avon on 02/23/99 and at the Chevron Refinery in Richmond on 03/25/99. The fall in gasoline production from these two fires was only 7% but due to California's strict environmental standards for gasoline, shipments from other (less regulated) districts were impossible so prices rose by about 25%. This implies a demand elasticity for retail gasoline of -0.28.

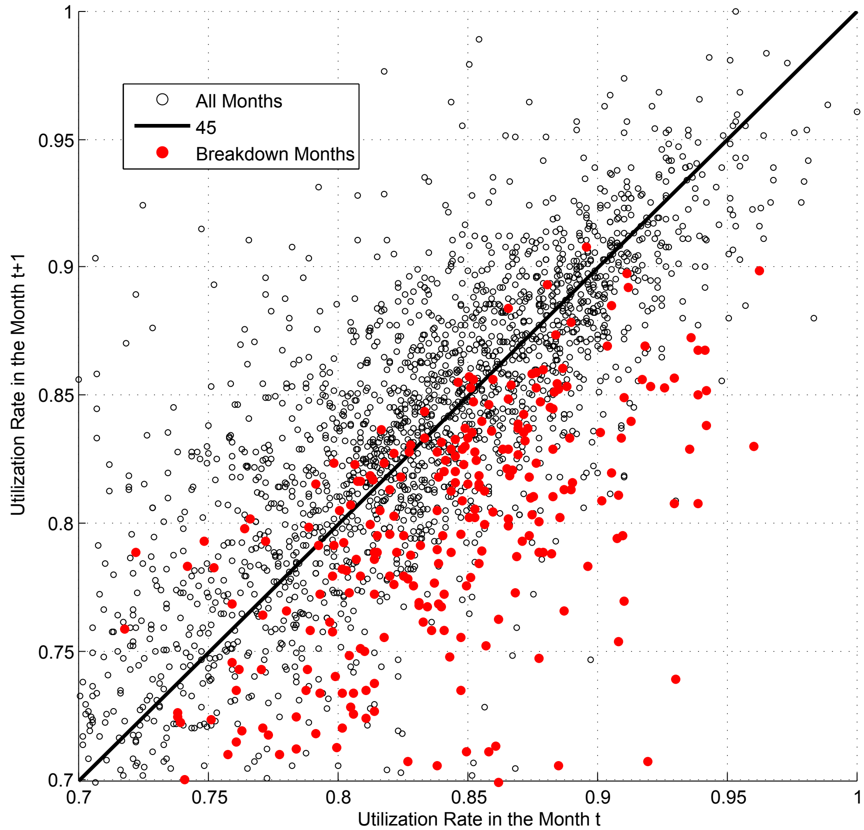


Figure 8: District Breakdowns

state space, which includes deterministic time states. The transition probability for the crude price is found using the empirical distribution of its historical series. The transition probabilities between breakdown states depend on the choice variable in the previous period according to the logit estimation done in the first stage. In a given year, the transition matrix for months reflects moving from one month to the next with certainty. Therefore, I can simplify the analysis by taking advantage of the cyclic nature of the month state. This dramatically reduces the computational time; see Rust (forthcoming). Further details of the estimation algorithm can be found in appendix C.

For a candidate parameter vector, I iterate on the policy function until convergence. I then interpolate the policy function on the actual states in my data and estimate the utilization rate for each district-month observation. Since the optimization is preformed at the firm level, I aggregate to the market level and form the following moments:

$$\begin{aligned} M_1 &= J^{-1}\sum_j (u_{mj} - \hat{u}_{mj}) \\ M_2 &= N_j^{-1}\sum_i (r_{ijy} - \hat{r}_{ijy}) \end{aligned}$$

where $\hat{u}_{mj}$ is the *average* utilization rate in district j and month m and $\hat{r}_{ijy}$ is the estimated investment rate by firm i located in district j in year y. I average the utilization rate moments over districts and the investment rate moments over firms and then stack them to form a moment vector: $M(\gamma) = (M_1, M_2)'$. I then numerically solve the following problem:

$$Min_\gamma \ \left\{ M(\gamma)'\Psi^{-1}M(\gamma) \right\}, \tag{16}$$

where Ψ is the variance-covariance matrix of the moment vector. With estimated parameters in hand, I estimate the standard errors of the cost estimates using Hansen's GMM estimator of the VC matrix. Given the matrix G of numerical derivatives, where (for parameter k and moment l)[31],

$$G_{lk} = \frac{M_l(\overline{\gamma}_k) - M_l(\underline{\gamma}_k)}{\gamma_k * 1\%}, \tag{17}$$

I can then compute:

$$VC(\gamma) = \frac{1}{N}(G'\Psi^{-1}G)^{-1}. \tag{18}$$

6 Results

6.1 Model Fit

I first assess the fit of the dynamic model by plotting actual and estimated values of key variables in figure 9. This is an in-sample analysis and shows that, on average, the estimated values match the data fairly well. Prices are estimated very precisely due to the flexibility gained by including monthly fixed effects. The estimated utilization

[31] For a 1% window, I perturb the parameter by 0.5% above and below the estimate.

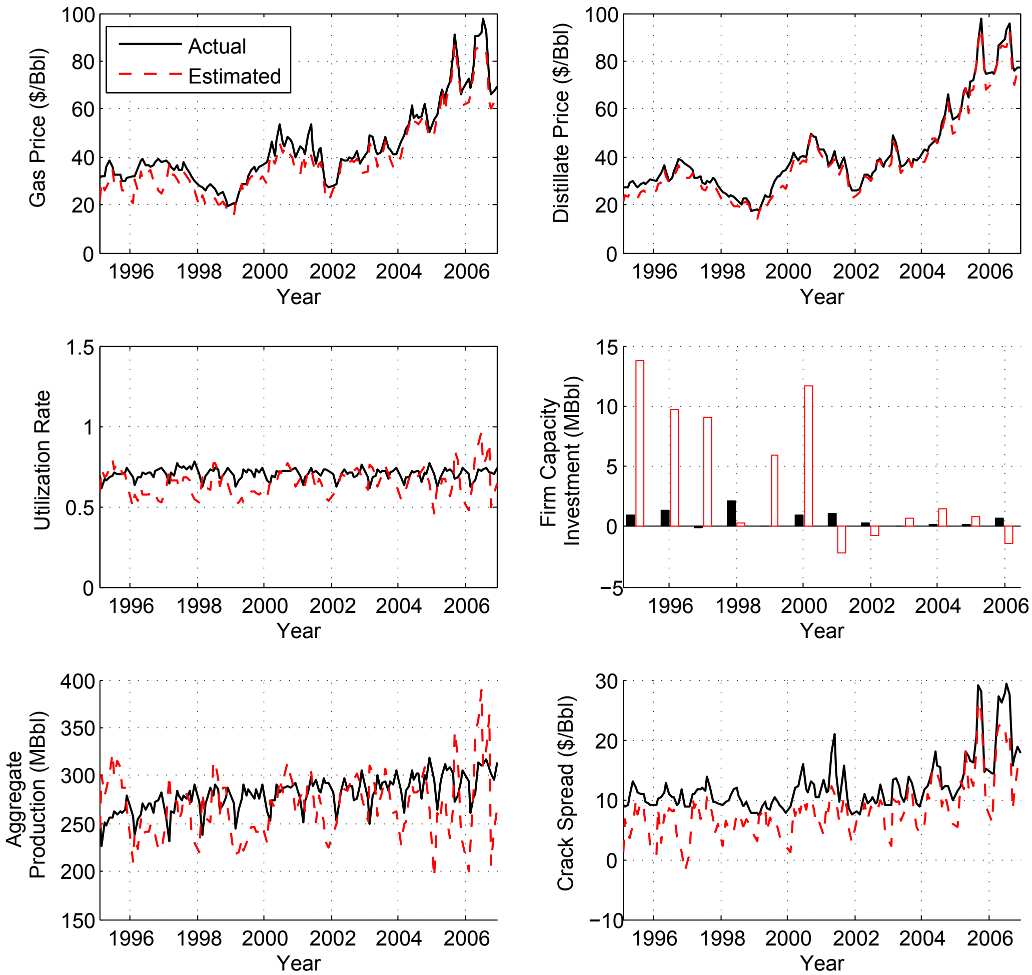


Figure 9: Model Fit (In Sample)

rate is more variable than the actual rate though the month-to-month fluctuations are approximated well. The model does not do as well at predicting the level of investment because firms tend to make lumpy investments every few years instead of updating their plant continuously. This means the median investment in any given year is zero and the reduced variation makes identification more difficult.

Finally, though the model tracks the movements in the crack spread very well, it tends to predict a value that is below the actual spread. This occurs because the estimated prices of gasoline and distillate are also biased down, because I do not include the small share of imports in the market. Excluding this amount means that domestic refineries are predicted to be producing slightly more than they actually are, which pushes down

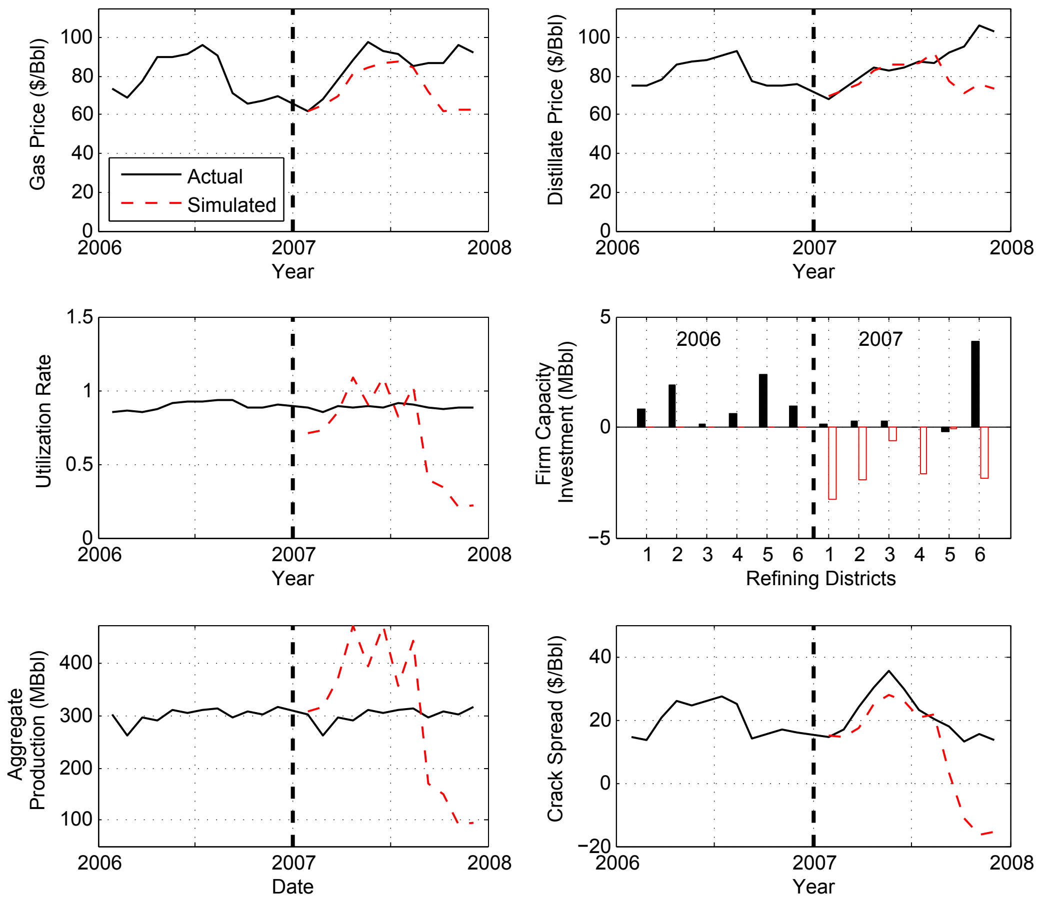


Figure 10: Model Fit (Out of Sample)

the estimated price.[32]

In figure 10, I do an out-of-sample test of the model, where I use the parameter estimates based on data through 2006 and simulate the investment and utilization policy of firms in 2007. The predicted prices of gasoline and distillate are close to the data for the beginning of 2007 but then begin to deviate. This pattern, also shown in the crack spread plot, is partially a result of unprecedented levels of the price of crude oil in 2007. The model predicts that refineries should optimally respond to these high input prices by cutting their utilization rate to drive up their product prices and maintain their profit margin.

[32]The estimated demand equations are based on sales of wholesale product, which includes imports of gasoline and distillate.

6.2 First Stage Estimates: Demand and Breakdown

Tables 3 and 4 present the results of the first stage demand and breakdown estimations. Most of the demand coefficients are significant at the 1% or 5% level and have the expected signs. I omit monthly fixed effects estimates, but they show the peak in gasoline demand during the summer months and distillate toward the fall. The elasticity estimates show a growing sensitivity to wholesale gasoline prices over the years. The R^2 from the first stage regression of price on stocks is 0.87. The logit estimation of breakdown reveals an increasing probability of breakdown as a refiner runs the plant more intensively. Estimating the probability of breakdown next period conditional on being broken down this period reveals that refiners with more severe breakdowns are less likely to recover in the next period.

6.3 Second Stage Estimates: Costs

The cost coefficients are generally significant and reflect a production cost function that is increasing and convex. I display the cost functions at the average values of the estimates in figure 11 and report all estimates in appendix D, table D.2. The cost functions show that firms in market 2, the isolated Rocky Mountain region, are the most sensitive to production changes and have the highest overall production costs. Market 1 enjoys relatively easy access to crude supplies in the Gulf region and has the lowest production costs. The curvature of the production cost functions shows that refiners face increasing marginal costs as they approach the limitations of their capacity. I use a constant crude oil price of $50/bbl in my estimated production cost function.

The estimates of investment cost functions reflect an almost linear relationship, with the quadratic term often insignificant. While the figure shows the average investment costs over time, table D.2 displays the increase in expansion costs that refiners have faced in recent years. The Senate's (2002) estimated cost of building a new 2,700 barrel/day refinery was about $27 million. I estimate the cost of the same size *expansion* at around $10 million, further evidence that expanding existing sites is more cost-effective than building a new plant.

Table 3: Demand Estimates

Year	Parameter	Gasoline Coefficient	Gasoline Std. Err.	Distillate Coefficient	Distillate Std. Err.
1995	constant (α_0)	1.55***	0.40	3.60**	1.82
	log P_{km} (α_2)	-0.55***	0.11	-1.48***	0.54
1996	constant (α_0)	1.26***	0.42	6.31**	2.71
	log P_{km} (α_2)	-0.46***	0.12	-2.21***	0.78
1997	constant (α_0)	2.50***	0.48	6.32***	2.22
	log P_{km} (α_2)	-0.78***	0.13	-2.10***	0.61
1998	constant (α_0)	1.53***	0.32	4.16**	1.79
	log P_{km} (α_2)	-0.55***	0.09	-1.65***	0.54
1999	constant (α_0)	1.95***	0.27	2.40	1.92
	log P_{km} (α_2)	-0.75***	0.08	-1.24**	0.64
2000	constant (α_0)	3.20***	0.57	12.41***	4.80
	log P_{km} (α_2)	-1.00***	0.15	-3.79***	1.31
2001	constant (α_0)	2.64***	0.38	9.32**	5.30
	log P_{km} (α_2)	-0.80***	0.10	-2.79**	1.40
2002	constant (α_0)	4.92***	0.59	12.31**	6.65
	log P_{km} (α_2)	-1.54***	0.17	-4.15**	2.03
2003	constant (α_0)	3.96***	0.56	13.56**	6.90
	log P_{km} (α_2)	-1.16***	0.15	-4.01**	1.86
2004	constant (α_0)	6.24***	1.00	8.57***	3.65
	log P_{km} (α_2)	-1.72***	0.26	-2.61***	0.97
2005	constant (α_0)	7.90***	1.70	11.90**	6.74
	log P_{km} (α_2)	-2.06***	0.42	-3.26**	1.67
2006	constant (α_0)	6.78***	1.94	14.29**	8.24
	log P_{km} (α_2)	-1.67***	0.45	-3.58**	1.90

****, **, * Significant at the 1%, 5%, and 10% level respectively. Month fixed effects omitted. Dependent variables: log of gasoline and distillate sales. First stage regression of price on stocks of crude oil, gasoline and distillate.*

Table 4: Breakdown Probability Estimates

Parameter	Conditional on No Breakdown Coefficient	Conditional on No Breakdown Std. Err.	Conditional on Breakdown Coefficient	Conditional on Breakdown Std. Err.
Constant (β_0)	-2.40***	0.44	0.91**	0.45
Utilization_{t-1}(β_1)	0.74	0.62	-4.03***	0.67

*Maximum likelihood estimates. ***, **, * Significant at the 1%, 5%, and 10% level respectively. Dependent variable = breakdown indicator.*

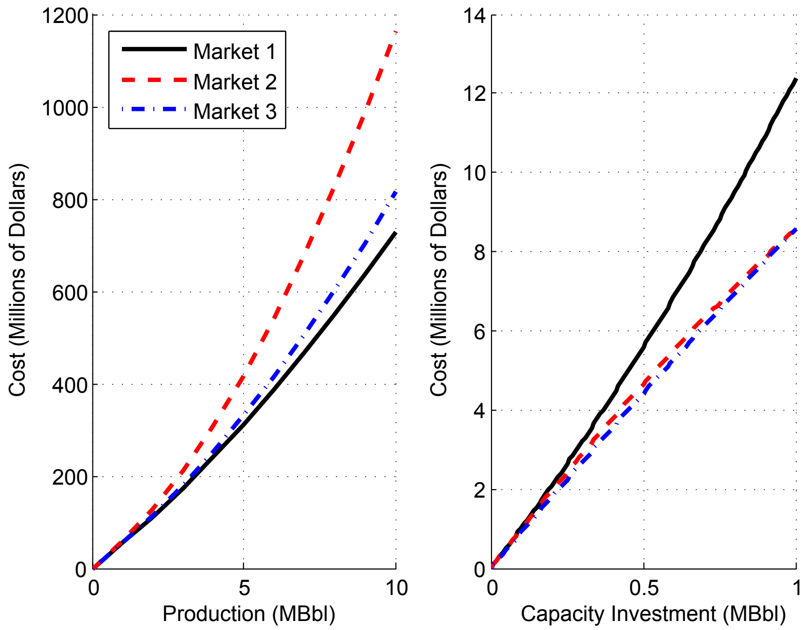


Figure 11: Estimated Production and Investment Cost Functions

6.4 Policy Function

In figure 12, I plot the optimal policy function over the course of a year at the average values of the other state variables. The optimal utilization rate increases during the late winter and early spring but then falls off around April and May, before rising again to a peak in August. A likely explanation is that refiners, anticipating the high demand summer driving season in July and August, scale back operations in the late spring to prevent the possibility of a breakdown occurring during the peak. This pattern is replicated in most markets and years. Figure 13 displays the optimal policy function in 3-dimensional space, varying by both the month of the year and the crude oil price. It shows that refiners cut back production when the oil price rises, a competitive response to a rising input price. The pattern across months is replicated at each crude oil price.

7 Counterfactuals

With a fully estimated dynamic model of the US oil refining industry, I can now use the model to determine the effects of various shocks that may occur. There are many interesting questions that could be examined with my model given the importance of oil

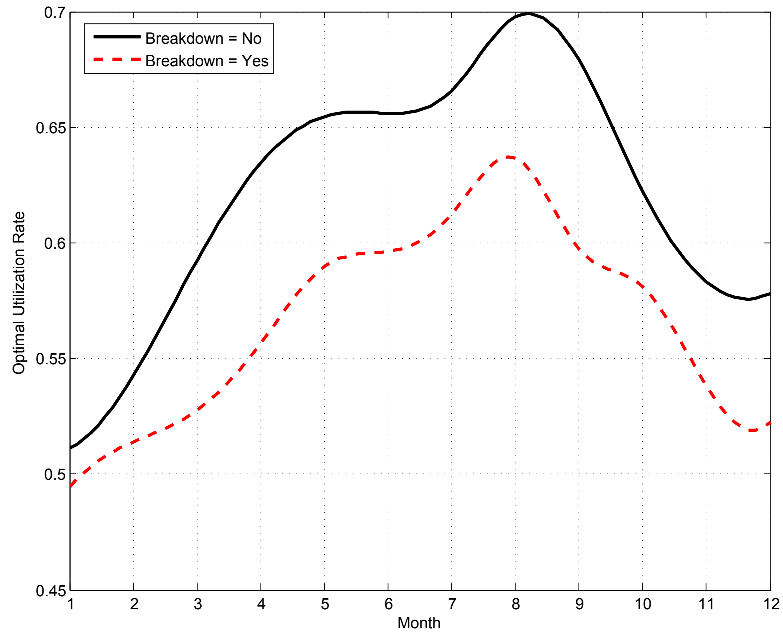


Figure 12: Optimal Utilization Rate Versus Month

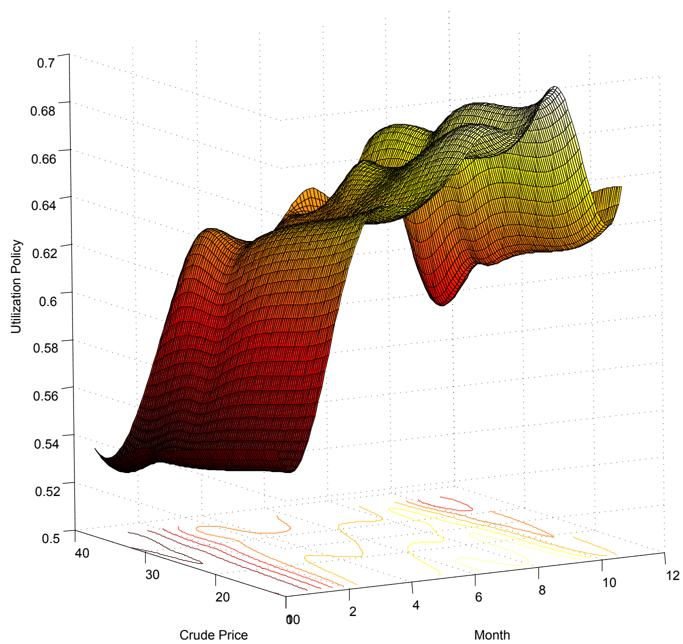


Figure 13: Optimal Utilization Rate Versus Month and Crude Price

refining in US and global energy markets. I focus on three stylized facts that I believe to be particularly important in the following analysis: crude oil prices are rising to unprecedented levels; there is little to no excess capacity in the oil refining industry; and end-use consumers of refined products are becoming increasingly sensitive to the prices they face (See Knittel et al. (2008)). Elasticities may be changing due to the availability of other fuels or because of changing perceptions of the environmental impact of oil usage (see figure 14). As a result, I will consider 2 experiments:

1. What are the effects of an increase in the crude oil price and how do the results change when the demand for refined products is more elastic?

2. What are the effects of a fall in available capacity and how do the results change when the demand for refined products is more elastic?

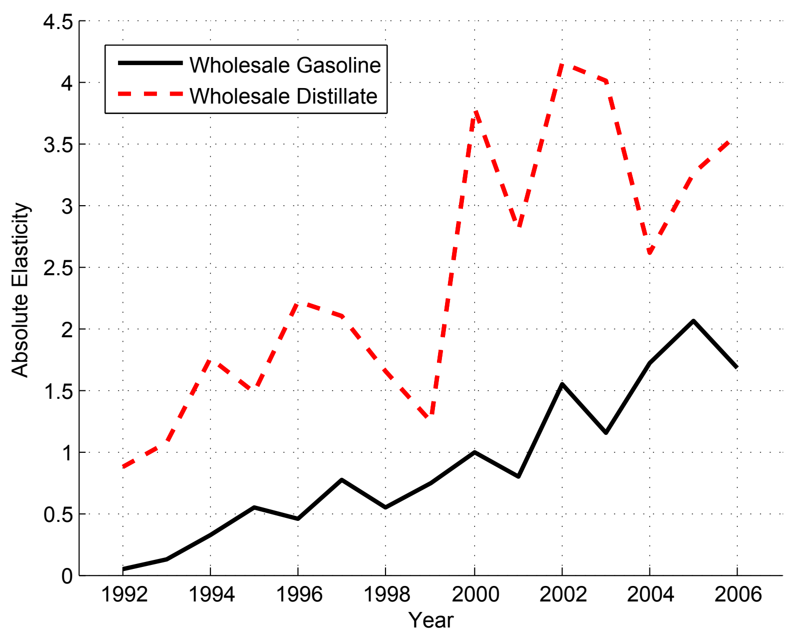


Figure 14: Price Elasticity of Demand

7.1 Methodology

Both counterfactuals are based on the coefficients and policy functions from 2006, the most recent year in my data. I shock the crude oil price in May to determine the effects throughout the peak demand summer months. The shock is permanent and I compute the average effects throughout the remainder of the year. I shock capacity in August to approximate the effects of a late summer hurricane hitting the Gulf of Mexico. I compute impacts assuming both the actual estimated elasticity in 2006 and an elasticity that is higher by 2.5% (in absolute terms) for both gasoline and distillate. Even this small increase in the sensitivity of consumers is enough to induce a dramatic response.

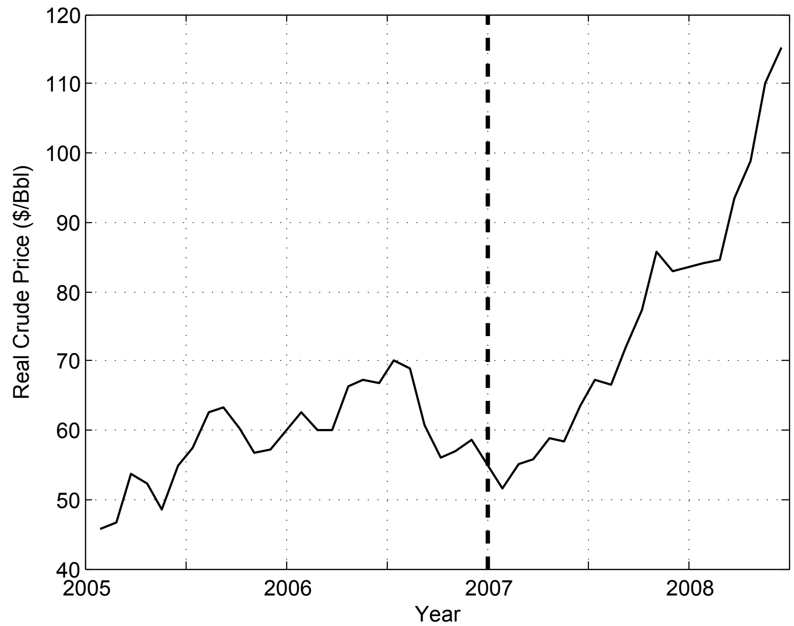


Figure 15: Crude Oil Price

In my sample, the maximum observed real crude oil price is around $70/bbl. However, as shown in figure 15, crude oil prices have been driven to record levels more recently, exceeding $115/bbl (in real 2006 dollars). Thus, I simulate the effects of a 20% increase in the price of crude oil to determine the impact on prices of gasoline and distillate and the resulting crack spread. Since the price elasticity of demand is one of the parameters estimated in the first stage and it influences the per-period payoff of the firm, I must solve my model at each new elasticity estimate. The optimal policy functions change as a result. Since the crude oil price is a state variable, I extrapolate my policy functions to the new crude prices.

About one-half of the US refining capacity is located on the Gulf of Mexico. Major hurricanes like Katrina and Rita in 2005, and more recently, Gustav and Ike in 2008,

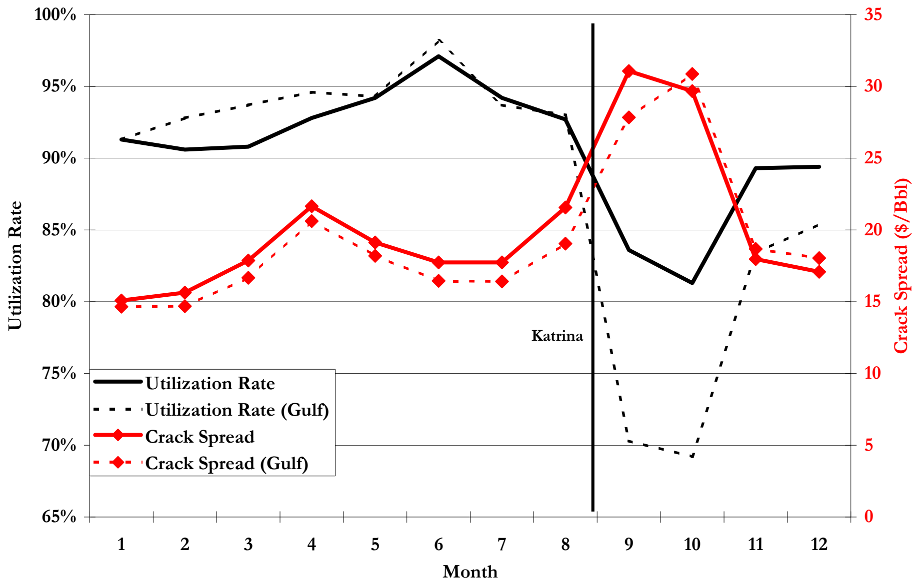


Figure 16: Loss in Capacity: Hurricane Katrina

reduced US oil refining capacity by 25% to 35% and had a major impact on downstream prices and refiners' profit margins (see figure 16). Therefore, in my second counterfactual experiment, I simulate the effects of a 25% reduction in capacity on downstream prices, the crack spread, refiner profits, and consumer welfare.

7.2 Results of Experiments

The effect of a 20% increase in the price of crude oil (from 2006 prices) is shown in figure 17 and summarized in table 5. Note, the price and crack spread changes in the table are the average changes relative to the baseline prediction following the shock for the remainder of the year. The changes in surplus, profit and welfare are based on totals for the remainder of the year following the shock. The graphs in figure 17 show the future path of product prices, the utilization rate, and the crack spread through the remainder of the year.

The first column of graphs corresponds to the actual estimated elasticity (in 2006) and the second column of graphs assumes more sensitive demand estimates. The price of gasoline and distillate both rise following the crude oil price shock, though the price

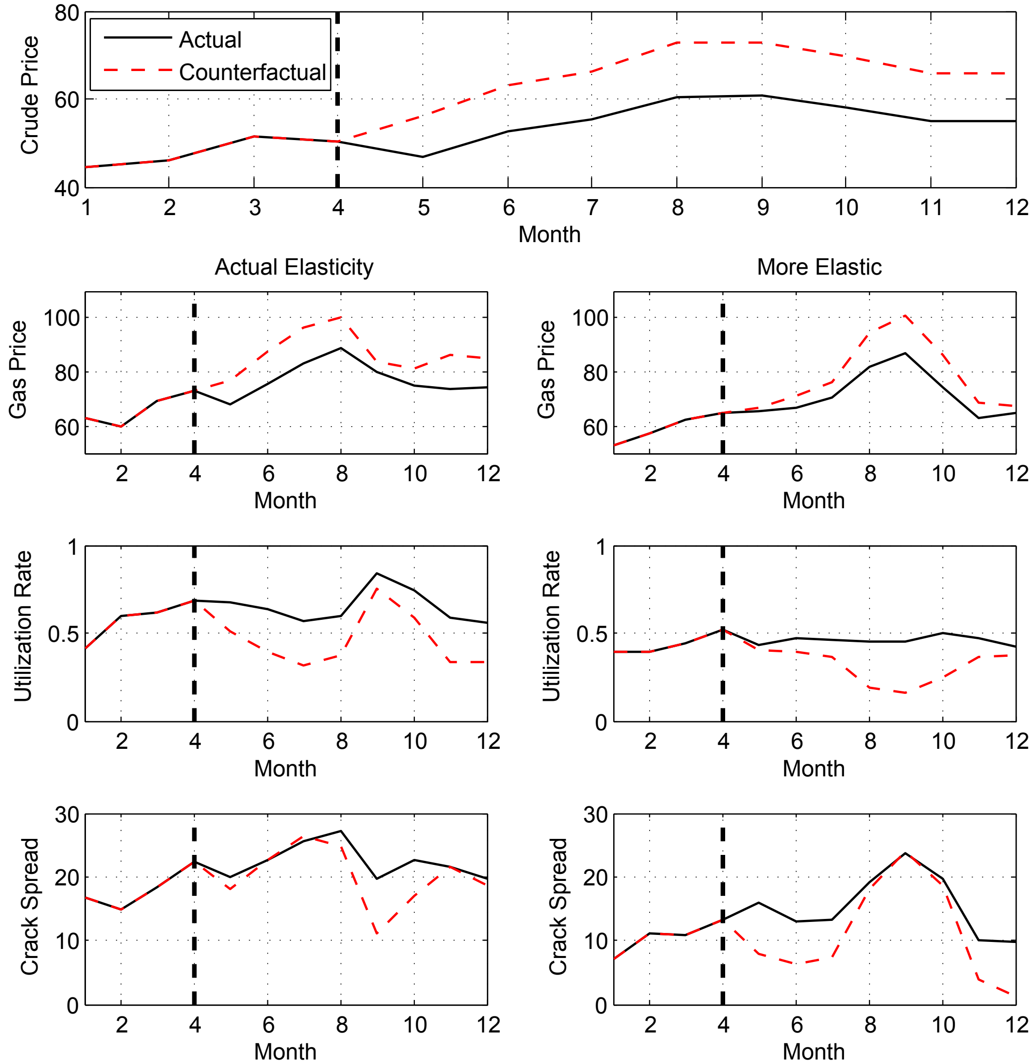


Figure 17: Crude Oil Counterfactual: Simulation

increases do not cover the entire cost increase as refiner profits fall after the shock. The amount of the increase that can be "passed through" to consumers appears to vary over the year. The crack spread graph reflects this, as it shows that although refiners are immediately hurt by the crude oil shock, they recover during the summer months by reducing their utilization rates before the spread falls again in September with weaker product demand.

Comparing the two levels of demand sensitivity, we see that refiners are less able to pass on the crude price increase to more sensitive consumers, and thus their crack

Table 5: The Effect of a 20% Increase in the Crude Oil Price

Percent Change	Actual Elasticity	More Elastic
Gasoline Price	12.7	10.2
Distillate Price	8.1	6.7
Crack Spread	-10.8	-30.1
Consumer Surplus	-58.3	-34.1
Refiner Profit	-37.1	-70.8
Total Welfare	-45.2	-49.7

Table 6: The Effect of a 25% Loss in Capacity

Percent Change	Actual Elasticity	More Elastic
Gasoline Price	15.9	3.0
Distillate Price	9.8	2.0
Crack Spread	47.9	11.9
Consumer Surplus	-69.0	-17.6
Refiner Profit	15.4	-4.8
Total Welfare	-11.1	-11.3

spread is dramatically reduced immediately following the shock. In addition to analyzing the effects on prices and profit margins, it is interesting to calculate the distribution of welfare between consumers and refiners. Total welfare declines by 45% in the months following the shock. According to table 5, overall welfare falls for both the actual and more sensitive elasticity estimates, although more sensitive consumers end up with a larger share of the surplus following the shock.

Figure 18 and table 6 display the results of my second counterfactual experiment, in which I reduce the size of the average refinery by 25%. Again, the table shows the average response to the shocks and figure 18 shows the longer-term effects for different levels of demand sensitivity. My counterfactual assumes that all refiners are hit equally hard by the shock, though in reality, some plants close completely while others operate

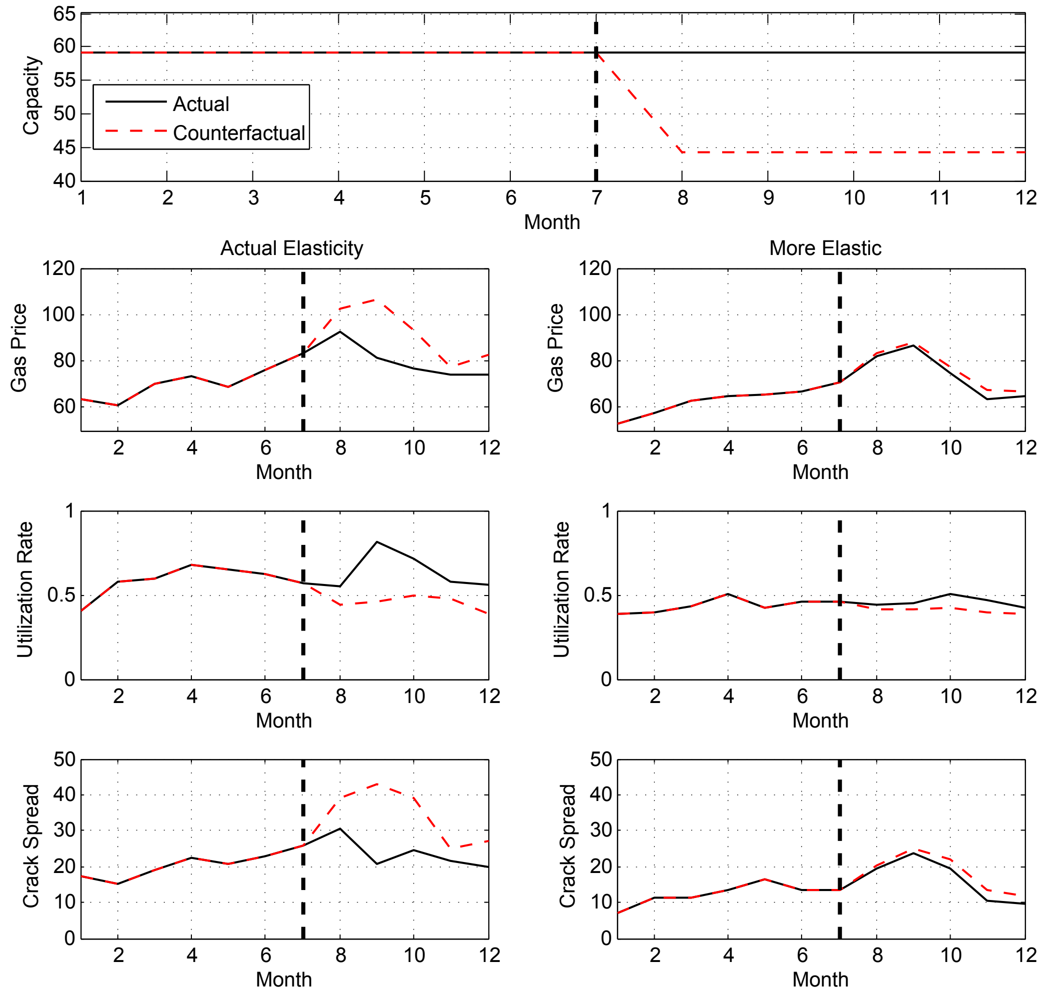


Figure 18: Capacity Counterfactual: Simulation

even more intensively following events like Katrina.

The impact of the shock on the crack spread depends strongly on the demand elasticity. With the crude oil price the same in both cases and the percentage increases in the prices of gasoline and distillate about five times higher at the actual elasticity, the refiners facing more sensitive consumers benefit immediately following the shock, though the longer-term crack spread is higher for the less sensitive consumer group. Utilization rates change only slightly following the shock and the real cost is borne by consumers in the form of gasoline prices, which rise by almost 16%, reducing consumer surplus by 69%.

In terms of the distribution of welfare, the overall pie decreases by about the same amount in both cases, but at the actual elasticity, the increase in profits at operating refineries partially offsets the loss in consumer surplus. However, the more sensitive consumers retain a larger proportion of welfare following the shock. It's important to note that my measure of total welfare puts equal weight on consumer surplus and refiner profit and makes no consideration for the variability of prices faced by consumers. Given the economy's extraordinary reliance on gasoline, an extra dollar per gallon paid at the pump may hurt consumers more than it helps refiners.

8 Conclusion

In this paper, I have developed and estimated a new dynamic model of the US oil refining industry. Energy markets, and in particular, the production and distribution of gasoline, are a hot topic in both academic research and the popular media. While the focus has tended to be on the upstream supply of crude oil (from both foreign and domestic sources) and the downstream retail stations, relatively little attention has been given to the role that oil refiners play in the industry. My analysis helps clarify and quantify the crucial role of the refiners in the transmission of crude oil and capacity shocks into downstream product prices, refiner profits, and consumer surplus.

The model matches the historical data and provides reasonably good out-of-sample predictions of key variables. I show that refiners are only partially able to pass through crude oil shocks to consumers and this ability varies across months of the year. As consumers have become more sensitive to changes in the price of gasoline, refiners face an even tougher competitive environment. Capacity disruptions, such as those caused by hurricanes, increase industry profits because the resulting price increase outweighs the loss in profits caused by reduced production. The effect on overall welfare is negative, though fairly small because the large loss in consumer surplus is partially offset by a rise in refiner profits.

My analysis not only models the behavior of refiners and the role they play in an important energy market, it also may have policy implications regarding optimal environmental regulations. In conversations with refiners, I found that current regulatory policies regarding both the building of new plants and the expansion of existing sites is the main hurdle that managers face when making their investment decisions. Regulatory policies have, at the very least, contributed to the current situation where capacity is tight and small shocks can have large effects. Realizing the importance of production flexibility in the refining industry means that new policies must balance responsible environmental concerns with incentives for capacity investment to meet the growing demand for refined products.

There are many extensions to this work that could provide further insights into the industry, though some require access to plant-level data which the EIA is considering making available. While this paper only addresses the production and investment decisions of active firms, including the possibility of exit may improve the model. Firms would likely follow a cut-off rule, exiting if the expected discounted stream of future profits fell below some critical level. Another potentially important determinant of firm behavior in this industry is a refiner's relationship with upstream crude oil producers. Currently, 60% of refiners are part of an integrated oil company, and although they benefit from a consistent supply of their major input, they are also constrained by having to exhaust their partner's stream of crude oil before seeking other, potentially more cost-effective sources. Independent refiners tend to invest in technologies that allow them to utilize different types of crude oil more flexibly, though may suffer relatively more when there is a supply disruption. Modeling the decisions of each type of refiner and the interaction between the two could help clarify the role of these vertical relationships. I leave these extensions for future work.

References

[1] Aguirregabiria, V. P. Mira, (2006). "Sequential estimation of dynamic discrete games." *Econometrica,* 75(1), 2006.

[2] Attanasio, Orazio, (2000). "Consumer Durables and Inertial Behavior: Estimation and Aggregation of Ss Rules for Automobiles." *Review of Economic Studies,* October 2000.

[3] Bacon, Robert W., (1991). "Rockets and Feathers: The Asymmetric Speed of Adjustment of UK Retail Gasoline Prices to Cost Changes." *Energy Economics,* 13 July 1991.

[4] Bajari, Patrick, Lanier Benkard, and Jonathan Levin, (2007). "Estimating Dynamic Models of Imperfect Competition." *Econometrica,* 75(5), 2007.

[5] Benkard, Lanier, (2004). "A dynamic analysis of the market for wide-bodied commercial aircraft." *Review of Economic Studies,* 71(3), 2004.

[6] Besanko, David and Ulrich Doraszelski, (2004). "Capacity Dynamics and Endogenous Asymmetries in Firm Size." *The RAND Journal of Economics,* Vol. 35, No. 1. Spring 2004.

[7] Besanko, David A., Ulrich Doraszelski, Lauren Xiaoyuan Lu, and Mark A. Satterthwaite, (2008). "Lumpy Capacity Investment and Disinvestment Dynamics." Harvard Institute of Economic Research Discussion Paper No. 2154 Available at SSRN: http://ssrn.com/abstract=1117991.

[8] Borenstein, S., C. A. Cameron and R. Gilbert, (1997). "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *Quarterly Journal of Economics,* 112(1), 1997.

[9] Borenstein, S., (1991). "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins." *The RAND Journal of Economics,* 22(3), 1991.

[10] Borenstein, S., Andrea Shepard (1996). "Dynamic Pricing in Retail Gasoline Markets." *The RAND Journal of Economics,* 27(3), 1996.

[11] Energy Information Administration, US Department of Energy, (2007). "Refinery Outages: Description and Potential Impact on Petroleum Product Prices." March 2007.

[12] Energy Information Administration, US Department of Energy, (2008). "A Primer on Gasoline Prices." Online: `http://www.eia.doe.gov/bookshelf/brochures/gasolinepricesprimer/index.html` [Downloaded: 09/11/2008], May 2008.

[13] Ericson, Richard, and Ariel Pakes, (1995). "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies,* 62:1, 53-83, 1995.

[14] Espey, Molly, (1996). "Explaining Variation in Elasticity of Gasoline Demand in the United States: A Meta Analysis." *The Energy Journal,* 17, 1996.

[15] The Federal Trade Commission, (2006). "Investigation of Gasoline Price Manipulation and Post-Katrina Gasoline Price Increases." Spring 2006.

[16] Goldberg, Pinelopi K. and Rebecca Hellerstein, (2008). "A Structural Approach to Explaining Incomplete Exchange-Rate Pass-Through and Pricing-to-Market." *The American Economic Review,* 98(2), 2008.

[17] The Government Accountability Office, (2006). "Energy Markets: Factors Contributing to Higher Gasoline Prices." GAO-06-412T. February 2006.

[18] Gron, Anne, Deborah Swenson, (2000). "Cost Pass-Through in the U.S. Automobile Market." *The Review of Economics and Statistics,* 82(2), 2000.

[19] Hamilton, James D., (1983). "Oil and the Macroeconomy since World War II." *The Journal of Political Economy,* 91(2), 1983.

[20] Hastings, Justine, Jennifer Brown, Erin Mansur, and Sofia Villas-Boas, (2008). "Reformulating Competition? Gasoline Content Regulation and Wholesale Gasoline Prices." *Journal of Environmental Economics and Management,* January 2008.

[21] Hotz, V. J., and R. A. Miller, (1993). "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies,* 60:3, 497-529, 1993.

[22] Hubbard, Glenn, (1986). "Supply Shocks and Price Adjustment in the World Oil Market." *The Quarterly Journal of Economics,* 101(1), 1986.

[23] ICF Consulting, (2005). "The Emerging Oil Refinery Capacity Crunch: A Global Clean Products Outlook." 2005.

[24] Knittel, Christopher, Jonathan E. Hughes, and Daniel Sperling, (2008). "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *The Energy Journal,* 29(1), January 2008.

[25] Kreps, David M., Jose A. Scheinkman, (1983). "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes." *The Bell Journal of Economics,* 14(2), Autumn 1983.

[26] Lidderdale, T.C.M. (United States Energy Information Administration), (1999). "Environmental Regulations and Changes in Petroleum Refining Operations." Online: `http://www.eia.doe.gov/emeu/steo/pub/special/enviro.html` [Downloaded: 12/07/2007], 1999.

[27] Noel, Michael D., (2007). "Edgeworth Price Cycles, Cost-Based Pricing, and Sticky Pricing in Retail Gasoline Markets." *Review of Economics and Statistics,* Vol. 89, 2007.

[28] Pakes, Ariel, (2000). "A framework for applied dynamic analysis in I.O." Working paper no. 8024, NBER, Cambridge, 2000.

[29] Pakes, Ariel, Michael Ostrovsky, and Steven T. Berry, (2004). "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)." Harvard Institute. Economic Research Discussion Paper No. 2036, May 2004.

[30] Pakes, Ariel and P. McGuire, (1994). "Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model." *Rand Journal of Economics,* 25(4), 1994.

[31] Peterson, D. J. and Sergej Mahnovski, (2003). "New Forces at Work in Refining: Industry Views of Critical Business and Operations Trends." Santa Monica, CA : RAND, 2003.

[32] Rust, John and Harry Paarsch, (forthcoming). "Valuing Programs with Deterministic and Stochastic Cycles." Forthcoming in the *Journal of Economic Dynamics and Control*.

[33] Rust, John, (2008). "Dynamic Programming." The New Palgrave Dictionary of Economics. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008.

[34] Rust, John, (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica,* 55:5, 999-1033, 1987.

[35] Ryan, Stephen, (forthcoming). "The Costs of Environmental Regulation in a Concentrated Industry." Forthcoming in *Econometrica*.

[36] Tirole, Jean, (1988). The Theory of Industrial Organization. Cambridge, MA: M.I.T. Press. 1988.

[37] The United States Senate, (2002). "Gas Prices: How are they Really Set?" Online: `http://www.senate.gov/~gov_affairs/042902gasreport.htm` [Downloaded 10/01/2007], May 2002.

A The Distillation Process

Since the various components of crude oil have different boiling points, a refinery's essential task is to boil the crude oil and separate it into the more valuable components. Figure A.1 displays a simplified diagram of a typical refinery's operations. The first and most important step in the refining process is called fractional distillation. The steps of fractional distillation are as follows:

1. Heat the crude oil with high pressure steam to 1,112 degrees fahrenheit.
2. As the mixture boils, vapor forms which rises through the fractional distillation column passing through trays which have holes that allow the vapor to pass through.
3. As the vapor rises, it cools and eventually reaches its boiling point at which time it condenses on one of the trays.
4. The substances with the lowest boiling point (such as gasoline) will condense near the top of the distillation column.

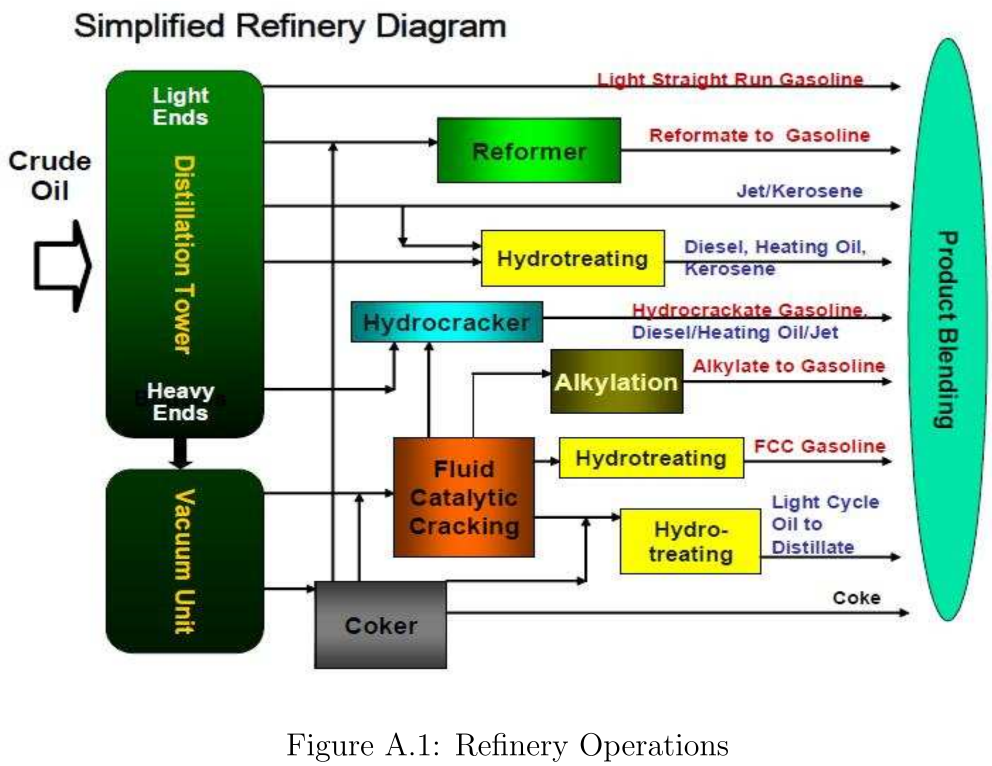


Figure A.1: Refinery Operations

While some gasoline is produced from pure distillation, refineries normally employ several downstream processes to increase the yield of high valued products by removing impurities such as sulfur. Cracking is the process of breaking down large hydrocarbons into smaller molecules through heating and/or adding a catalyst. Cracking was first used in 1913 and thus changed the problem of the refiner from choosing how much crude oil to distill into choosing an appropriate mix of products (within some range). Refineries practice two main types of cracking:

- Catalytic cracking: a medium conversion process which increases the gasoline yield to 45% (and the total yield to 104%).
- Coking/residual construction - a high conversion process which increases the gasoline yield to 55% (and the total yield 108%).

The challenge of choosing the right input and output mix given the available technology creates a massive linear programming problem.

B Crude Oil Quality

Crude oil is a flammable black liquid comprised primarily of hydrocarbons and other organic compounds. The three largest oil producing countries are Saudi Arabia, Russia and the United States.[33] Crude oil is the most important input into refineries and this raw material can vary in its ability to produce refined products like gasoline. The two main characteristics of crude that determine its quality are American Petroleum Institute (API) gravity and sulfur content. The former is a measure (on an arbitrary scale) of the density of a petroleum liquid relative to water.[34] Table B.1 summarizes these characteristics and includes some common crude types and their gasoline yield from the initial distillation process.

Worldwide, light/sweet crude is the most expensive and accounts for 35% of consumption. Medium/sour is less expensive and accounts for 50% of consumption while heavy/sour is the least costly and accounts for 15%. Figure B.1 show how the average crude oil used by US refiners is becoming heavier and more sour over time. This means

[33] Production in this sense refers to the quantity extracted from a country's endowment.

[34] Technically, API gravity = (141.5/ specific gravity of crude at 60° F) −131.5. Water has an API gravity of 10°.

Table B.1: Crude Qualities

API Gravity	Sulfur Content < 0.7%	Sulfur Content > 0.7%
$< 22°$	Heavy Sweet	Heavy Sour - 14% yield *(Maya, Western Canadian)*
$22° - 38°$	Medium Sweet	Medium Sour - 21% yield *(Mars, Arab light)*
$> 38°$	Light Sweet - 30% yield *(WTI, Brent)*	Light Sour

Source: EIA.

that the production costs of a gallon of gasoline are changing as refineries must invest in more sophisticated technology in order to process lower quality crude oil.

Since crude oil by itself has very little value to any industry, the price of a barrel of oil reflects the net value of the downstream products that can be created from it. The two major sources of movements in the crude oil price are upstream supply shocks (due to OPEC's quotas and hurricanes affecting oil rigs in the Gulf of Mexico) and downstream demand shocks (due to consumer's demand for refined products). The other source often sited by industry experts are refinery inventories of crude oil. Maintaining stocks of crude oil allow the refinery to respond quickly to downstream shocks like an unexpectedly cold winter increasing the demand for heating oil.

Within the various types of crude oil, the prices of each quality respond differently to shocks. The "light/heavy" differential is one measure that indicates the benefit a refiner can achieve by investing in sophisticated equipment to process heavier crude oil into highly-valued refined products. The differential has varied significantly over the last 10 years from 3 dollars per barrel to almost 20 dollars per barrel. An oil refinery faces a unique decision when making its production choice, one that provides for both flexibility and complexity. One one hand, consumers do not care about the type of crude oil, oxygenates, or distillation process used to make, for example, the gasoline they put in their cars. They just want their car to run well. While this would appear to make a refiner's problem easier, choosing their heterogeneous inputs, such as crude oil, satisfying federal, state and city environmental regulations, and all while maximizing profits, makes for an enormously complex optimization.

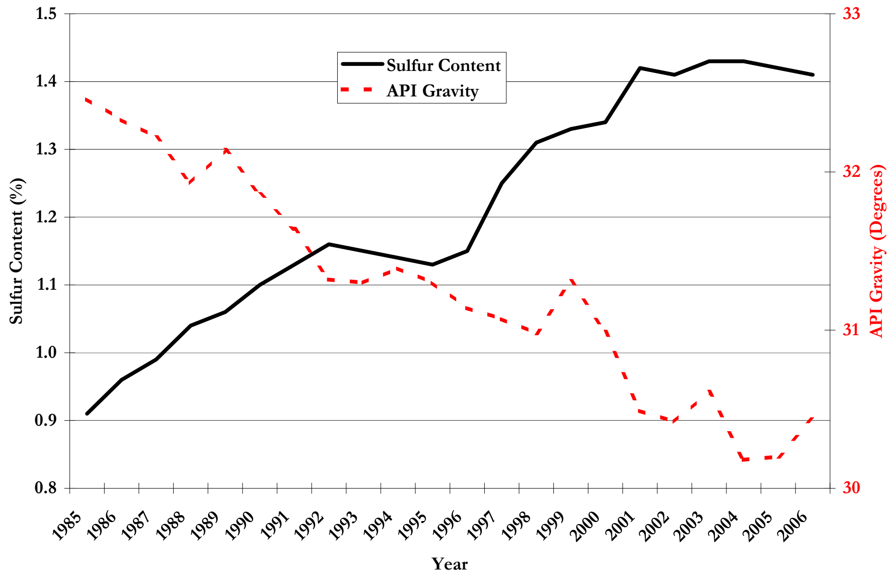


Figure B.1: Average Crude Oil Quality: Heavier and More Sour

C Estimation Algorithm

My estimation strategy involves matching utilization and investment moments. This requires that I solve for a policy function for each of these decisions and interpolate the functions to the realizations of the state variables in the data. The monthly utilization choice problem is a simple finite horizon dynamic program that I am able to solve by backward induction. So, for a given level of investment which induces a capacity for the plant, I can write the problem as:

$$\Pi_{iy} = Max_{\{u_{im}\}_{m=1}^{12}} \quad E\left[\sum_{m=1}^{12} \mu^{m-1} \pi_{im}(u_{im}; x_{im}, \overline{q}_{iy})\right]. \tag{19}$$

Then, Π_{iy}, the aggregate discounted annual profit of the plant, becomes the payoff function for the infinite horizon problem. The Bellman equation for that problem is:

$$V(x) \quad = \quad Max_r \Big\{ \Pi_{iy}(r; x) + \delta V(x') P(x'|x, r) \Big\}. \tag{20}$$

To solve this equation, I could have used several different methods including successive approximations or collocation, but I chose policy function iteration, also known as the

Howard Policy Improvement Algorithm. The first step is to guess a candidate policy function, which I call, $\sigma_t(x)$, where t indexes the iteration. Since this policy governs investment which effects optimal utilization, which in turn effects the probability of breakdown, I have to calculate the transition matrix given the policy: $P(x'|x,\sigma_t(x))$. Then comes the "policy evaluation step" which is to solve 20, i.e.:

$$V_t(x) = [I - \delta P(x'|x,\sigma_t(x))]^{-1}\Pi_{iy}(\sigma_t(x);x). \tag{21}$$

For a size K state space, this involves the inversion of a KxK matrix which makes it difficult to estimate the with too fine of a discretization. With the value function in hand, I move to the "policy improvement step" which updates the policy function:

$$\sigma_{t+1}(x) = argmax_r \left\{\Pi_{iy}(r;x) + \delta V_t(x')P(x'|x,r)\right\}. \tag{22}$$

Finally, I compare $\sigma_{t+1}(x)$ to $\sigma_t(x)$ and repeat the process until convergence.

D Additional Tables

Table D.1: Industry Concentration

	1970	1980	1991	2001	2004	2005	2006	2007	2008
US									
4-Firm (%)			31.4	40.2	44.4	43.0	45.8	44.1	41.2
8-Firm (%)			52.2	61.6	69.4	68.4	72.0	69.5	63.7
HHI			437.0	611.0	728.0	727.0	776.4	730.3	644.2
PADD 1									
4-Firm (%)			59.2	80.7	76.7	85.8	87.3	87.3	87.0
8-Firm (%)			88.7	99.0	97.9	99.4	99.4	99.4	99.4
HHI			1,225.0	2,158.0	1,943.0	2,505.0	2,537.5	2,540.2	2,524.7
PADD 2									
4-Firm (%)	38.3	37.4	39.3	50.9	57.1	57.1	59.6	55.5	50.5
8-Firm (%)	59.7	60.0	65.0	75.6	82.6	82.6	85.0	80.9	75.9
HHI			675.0	961.0	1,063.0	1,059.0	1,114.0	1,031.3	950.8
PADD 3									
4-Firm (%)	44.0	36.2	36.3	48.4	56.3	56.0	57.8	56.0	50.9
8-Firm (%)	64.8	54.5	58.5	66.5	78.8	78.2	81.2	77.6	73.2
HHI			578.0	851.0	1,018.0	1,005.0	1,052.2	976.7	909.2
PADD 4									
4-Firm (%)	53.5	48.0	55.8	58.1	46.1	45.7	50.9	50.7	58.7
8-Firm (%)	81.7	75.3	83.6	86.9	81.2	80.4	85.5	85.2	84.3
HHI			1,080.0	1,179.0	944.0	935.0	1,047.7	1,031.5	1,405.5
PADD 5									
4-Firm (%)	66.5	54.4	53.8	60.2	62.4	62.4	59.1	59.2	61.8
8-Firm (%)	95.2	76.5	74.2	86.9	92.7	92.8	89.5	89.6	89.4
HHI			965.0	1,148.0	1,246.0	1,247.0	1,162.2	1,168.7	1,195.7
California									
4-Firm (%)			58.9	68.7	66.2	66.5	62.3	62.5	63.0
8-Firm (%)			82.5	95.1	96.3	96.3	92.1	93.2	93.2
HHI			1,184.0	1,481.0	1,475.0	1,475.0	1,354.9	1,367.2	1,368.8
Gulf Coast									
4-Firm (%)							59.1	60.1	53.7
8-Firm (%)							83.5	83.1	76.7
HHI							1,107.9	1,110.5	995.0
PADDs 1 & 3									
4-Firm (%)	40.9	35.0	36.7	44.6	54.6	52.5	55.4	54.0	50.2
8-Firm (%)	62.3	55.0	57.2	65.3	76.1	75.5	79.5	76.6	72.8
HHI			561.0	741.0	919.0	890.0	967.9	991.1	861.2
PADDs 2 & 3									
4-Firm (%)			30.7	42.5	46.2	45.9	50.0	47.5	44.4
8-Firm (%)			56.5	64.9	75.6	75.2	79.9	76.2	70.3
HHI			455.0	681.0	826.0	818.0	894.6	822.7	742.9
PADDs 1, 2, & 3									
4-Firm (%)	35.2	30.7	30.2	39.4	45.9	44.5	49.2	47.1	43.9
8-Firm (%)	58.0	49.2	53.6	63.5	73.1	72.6	78.3	75.1	69.6
HHI			460.0	638.0	789.0	783.0	872.7	807.9	731.4

Source: EIA. Concentration based on operating capacity of crude oil distillation measured per calendar day on January 1st of the given year. The FTC generated the table through 2004 and I extended it through 2008. Upper Midwest: Illinois, Indiana, Kentucky, Michigan, and Ohio. Increase from 2004 to 2005 HHI's in PADDs I and III primarily due to the merger between Valero and Premcor. Capacities used in this table are at the corporate level (multiple refineries owned by the same corporation are aggregated).

Table D.2: Cost Estimates

Year	Parameter	Market 1 Coefficient	Market 1 Std. Err.	Market 2 Coefficient	Market 2 Std. Err.	Market 3 Coefficient	Market 3 Std. Err.
1995	Q (γ_0)	3.45***	0.01	0.36***	0.10	7.99***	0.75
	Q^2 (γ_1)	2.70***	0.01	10.86	11.18	5.45***	0.21
	Q*P^c (γ_2)	0.29***	0.00	0.06***	0.02	0.28***	0.04
	Investment (γ_3)	4.41***	0.14	4.56	5.36	7.80	8.70
	Investment^2 (γ_4)	-4.41***	0.07	-2.99***	0.74	-5.52	5.01
1996	Q (γ_0)	3.48***	0.00	2.62***	0.38	0.05	2.09
	Q^2 (γ_1)	6.19***	0.01	5.21***	0.31	6.02***	0.44
	Q*P^c (γ_2)	0.03***	0.00	0.03*	0.02	1.00***	0.03
	Investment (γ_3)	4.01***	0.15	5.58	51.23	3.84	11.82
	Investment^2 (γ_4)	-1.27***	0.05	-0.97	8.91	-2.09**	1.03
1997	Q (γ_0)	0.05*	0.03	0.92***	0.19	1.08	1.98
	Q^2 (γ_1)	5.14***	0.05	7.85***	0.15	7.30***	0.04
	Q*P^c (γ_2)	0.08***	0.00	0.05***	0.01	0.38***	0.00
	Investment (γ_3)	4.25***	0.03	3.60**	1.64	8.88***	0.21
	Investment^2 (γ_4)	-0.81***	0.01	1.03	1.88	-1.86***	0.04
1998	Q (γ_0)	0.17***	0.03	0.05	26.36	1.16***	0.31
	Q^2 (γ_1)	1.00***	0.04	3.68	8.20	3.40***	0.24
	Q*P^c (γ_2)	1.00***	0.01	0.02	55.93	0.86***	0.08
	Investment (γ_3)	-17.65	110.67	3.28	6.13	5.15	95.97
	Investment^2 (γ_4)	25.35	33.80	-4.30	32.06	-1.91	1.75
1999	Q (γ_0)	2.70***	0.04	0.44	51.57	6.94	35.07
	Q^2 (γ_1)	5.79***	0.18	2.13	6.43	7.35***	0.05
	Q*P^c (γ_2)	0.01***	0.00	0.27	3.96	0.12	19.64
	Investment (γ_3)	4.65	14.90	5.90	11.03	9.53***	0.73
	Investment^2 (γ_4)	-0.82	1.31	-6.05	58.11	-0.92***	0.13
2000	Q (γ_0)	6.19***	0.57	0.04	0.19	10.29***	1.57
	Q^2 (γ_1)	5.89***	0.11	11.36***	0.63	6.36***	0.41
	Q*P^c (γ_2)	0.00	0.00	0.00	0.00	0.01	0.06
	Investment (γ_3)	5.65*	4.16	4.08	4.40	11.85***	1.33
	Investment^2 (γ_4)	-2.82***	0.44	-0.99	2.13	5.26	9.43
2001	Q (γ_0)	0.32***	0.06	0.05***	0.01	0.03	2.92
	Q^2 (γ_1)	5.75***	0.06	23.84***	1.07	2.63**	1.19
	Q*P^c (γ_2)	0.02***	0.00	0.00***	0.00	1.00***	0.20
	Investment (γ_3)	4.56***	0.53	3.91***	0.35	9.74	15.17
	Investment^2 (γ_4)	1.12***	0.07	-4.79***	0.36	-5.05***	0.99
2002	Q (γ_0)	2.24***	0.74	0.12***	0.03	0.58	0.52
	Q^2 (γ_1)	4.51***	0.10	3.70***	0.74	6.90***	0.58
	Q*P^c (γ_2)	0.16***	0.03	0.98***	0.08	0.28***	0.05
	Investment (γ_3)	17.48**	9.18	5.49**	2.74	6.75	1,402.90
	Investment^2 (γ_4)	3.49	14.69	-1.09	0.86	-0.87	6.75
2003	Q (γ_0)	0.88***	0.18	13.42	394.71	0.03	0.22
	Q^2 (γ_1)	5.87***	0.11	0.56	27.99	4.50***	0.24
	Q*P^c (γ_2)	0.08***	0.01	0.32	3.94	0.79***	0.04
	Investment (γ_3)	4.32***	0.70	5.43**	3.15	4.73***	1.64
	Investment^2 (γ_4)	2.75***	0.89	-1.02	1.88	-3.08*	2.14
2004	Q (γ_0)	3.18***	0.22	0.17***	0.07	0.15	0.45
	Q^2 (γ_1)	8.04***	0.13	28.65***	8.47	11.49***	0.68
	Q*P^c (γ_2)	0.00***	0.00	0.01***	0.00	0.00	0.02
	Investment (γ_3)	7.48***	0.70	5.35***	1.19	7.07	7.23
	Investment^2 (γ_4)	2.09***	0.10	-5.14***	0.57	-2.84	4.96
2005	Q (γ_0)	0.34***	0.02	0.90***	0.11	0.04	34.02
	Q^2 (γ_1)	2.85***	0.05	8.52***	0.07	1.39	13.95
	Q*P^c (γ_2)	1.00***	0.01	1.00***	0.01	1.00***	0.05
	Investment (γ_3)	10.42	24.93	11.69***	1.40	10.74***	4.42
	Investment^2 (γ_4)	2.05	1.60	-2.97***	0.71	-1.15	2.36
2006	Q (γ_0)	2.92***	0.06	0.01	0.19	1.01***	0.35
	Q^2 (γ_1)	1.39***	0.02	4.67***	0.93	4.79***	0.34
	Q*P^c (γ_2)	1.00***	0.00	1.00***	0.03	1.00***	0.03
	Investment (γ_3)	9.44	443.89	8.42	7.81	7.43	493.22
	Investment^2 (γ_4)	2.85	134.06	0.01	130.15	0.15	157.43

****, **, * Significant at the 1%, 5%, and 10% level respectively.*