Clean Air Regulation and Heterogeneity in U.S. Gasoline Prices

By

Ujjayant Chakravorty, Celine Nauges and Alban Thomas[1]

Abstract

In order to improve public health in areas with air quality problems, the U.S. Clean Air Act imposes a variety of federal regulations on gasoline, which have led to a proliferation of fuel blends known as "boutique fuels." More than 45 fuel blends are sold nationwide. We examine the effects of this program on wholesale gasoline prices. The methodological innovation in this study is the use of a regulatory distance measure as a proxy for measuring market power that arises from product differentiation. We find that Clean Air regulation increases gasoline prices by increasing the cost of refining, but more importantly, by creating regulatory "islands," it segments the market and increases the market power of firms. Our estimation controls for the potential endogeneity of the regulatory variables. We find that OLS techniques systematically underestimate the effect of regulation on gasoline prices.

Key Words: Boutique Fuels, Clean Air Act, Environmental Regulation, Market Structure, Product Differentiation

JEL Classification: L51, L71, D43

This Version: June 2007

Forthcoming, Journal of Environmental Economics and Management

---

[1] Chakravorty: Department of Economics, University of Central Florida, College of Business Administration, PO Box 161400, Orlando FL 32816, uchakravorty@bus.ucf.edu; Nauges and Thomas: LERNA-INRA, University of Toulouse, 21 Allée de Brienne, 31000 Toulouse France, cnauges@toulouse.inra.fr and thomas@toulouse.inra.fr, respectively. Correspondence: Chakravorty, 407 823 4728, fax: 407 823 3269.

Clean Air Regulation and Heterogeneity in U.S. Gasoline Prices

1. Introduction

In order to improve public health in areas with air quality problems, the U.S. Clean Air Act of 1990 led to a variety of federal regulations that aim to reduce emissions from motor vehicles. The Act allows individual states to implement their own clean fuel programs for gasoline to address local or regional air quality concerns. These federal and state regulations have not only led to a significant improvement in air quality but also to a proliferation of clean fuel blends. Differential gasoline standards include the Reformulated Gasoline program, the Oxygenated Gasoline program, and federal or state programs that impose lower volatility requirements, caps on sulfur content, the use of fuel additives such as MTBE (methyl tertiary butyl ether) and ethanol, as well as requirements for minimum oxygen content.

At least 15 different types of fuel specifications are currently in use. Combined with the three octane grades of gasoline available at pumps - regular, mid-grade and premium[2] - over 45 different blends are used nationwide (Ryan, 2003). A new ozone rule proposed by the EPA is expected to add another 24 new blends into the mix in the near future.[3] These fuels are often called "boutique fuels."

An investigative committee of the United States Senate concluded that "the mix of state and federal standards in effect today has resulted in a situation where adjacent areas may be using gasoline with significantly different properties" (US Senate, 2002, p.74). Because requirements vary between and within states, refiners find it difficult to move product quickly from one area to another. Critics have suggested that the multiplicity of fuels in use have caused price volatility especially during periods of supply disruption, such as winter-summer transitions, periods of high demand and refinery and pipeline breakdowns, since refiners mostly specialize in producing certain fuel specifications and cannot switch between them immediately. They have also resulted

---

[2] Price levels vary by grade, but the price differential between grades is generally constant (EIA, 2003).

[3] The number of fuels may increase further because of a ban on the use of MTBE (suspected of contaminating groundwater supplies) by some states, introduction of a renewable fuel standard and use of low sulfur fuels (EIA, 2002).

in supply bottlenecks and pipeline congestion because various types of fuels must often use the same pipeline system.[4]

Boutique fuel programs have “fewer fuel producers, are less fungible and have fewer distribution system supply options” (EPA, 2001). The magnitude of the problem varies with volumes, distance from supply sources and their number, which in turn depends on the degree of product differentiation. For example, in the summer, fuel produced for the Charlotte, North Carolina area cannot be used in Norfolk, Virginia (which must use Reformulated Gasoline) or Atlanta, Georgia (which requires a lower Reid Vapor Pressure and sulfur cap). However, Atlanta and Norfolk fuels can be moved to Charlotte (Yacobucci, 2004). The United States Congress is considering enacting legislation to prevent further proliferation of boutique fuel islands through the Boutique Fuels Reduction Act of 2006 (Green, 2006). The bill proposes to establish a mechanism to start reducing the number of boutique fuels and broadens the authority of the EPA to waive boutique fuel mandates when supply shortages and price spikes occur.

This paper uses panel data techniques to explain heterogeneity in wholesale gasoline prices[5] due to federal and state Clean Air regulation during the period 1995-2002. We consider two major types of clean fuels, the Reformulated Gasoline (RFG) and Oxygenated Gasoline (OXY) programs, which aim to reduce local ozone and carbon monoxide pollution, respectively. We show that Clean Air regulation of gasoline affects prices in two ways: the “direct cost” effect: clean gasoline is more costly to produce so the larger the regulated area within a state, the higher the average price in that state, and the “market segmentation” effect: the greater the regulatory distance between a state and its neighbors, i.e., the difference in relative size between the clean gasoline markets, the higher the price in that state. The methodological innovation in this study is the use of a regulatory distance measure as a proxy for measuring market power that arises from product differentiation. Compared to an unregulated market, the number of refineries supplying a particular clean fuel may decline, leading to a potential increase in their market power in the regulated market.

---

[4] Several pipelines put refiners into an allocation system during peak periods that delays fuel transportation and increases costs (EPA, 2001). Often the same pipeline has to be washed before carrying a different fuel blend.

[5] Unless stated otherwise, “price” will mean the wholesale price of gasoline.

Other papers have examined the issue of boutique fuels.[6] In a recent paper, Muehlegger (2004) uses a structural model to isolate the effect of differential regulation on the price of gasoline during refinery outages in the states of California, Illinois and Wisconsin. These states were chosen because they exhibited significant price spikes in recent years. He finds that during supply shocks these regulations increase refining costs by 3-4.5 cents per gallon. He concludes that during refinery outages, the incompatibility between state and federal regulation may contribute about 5-7 cents/gallon to the price of gasoline, numbers that are higher than our average estimates for the country as a whole.[7] Chouinard and Perloff (2007) estimate a reduced form model of gasoline price differences across states and over time, using monthly panel data. They use dummy variables to control for the implementation of RFG and OXY gasoline programs in each state. They measure the direct cost effects of these clean fuel programs but not the effects on the market power of refiners.[8]

Brown et al (2006) have also investigated the price effect of two gasoline programs: the RFG program and the Reid Vapor Pressure (RVP) program, using weekly wholesale prices for selected metropolitan areas. They estimate a reduced-form model using a treatment and control approach. Each regulated city (i.e., each city which is under a RFG or RVP program) is matched with an unregulated city in close geographic proximity. The estimated price effect of gasoline content regulation is found to be on average about 3 cents per gallon, with significant variation (by about 8 cents per gallon) across regulated markets. Their approach focuses on the marginal impact of environmental regulation on gasoline prices by considering the resulting decrease in the number of firms that supply fuel when regulation is introduced. They measure geographical isolation caused by regulation by using two variables: the number of distribution terminals (often called

---

[6] The dynamics of gasoline prices as well as the transmission of price changes from crude to wholesale and from wholesale to retail markets has been studied by Borenstein and Shephard (1996a) and Borenstein, Cameron and Gilbert (1997). Borenstein and Shepard (1996b) examine market power in wholesale gasoline markets through the concentration of refiners supplying products at gasoline terminals.

[7] His higher estimates may at least be partly due to the choice of states with significant gasoline price spikes while our numbers are national averages.

[8] They do not explicitly model the effect of boutique fuels on market power but recognize the importance of this issue: “the requirement that stations sell only specially formulated pollution-reducing gasoline increases refining costs and may create market power for wholesalers within a state. To produce reformulated gasoline, refiners must make several costly modifications to their production equipment. If producers in surrounding states avoid incurring these large capital costs, producers in states mandating the use of reformulated or oxygenated gas do not face competition from these out-of-state suppliers” (Chouinard and Perloff, 2007).

"distribution racks" in the industry jargon) with which a city could potentially trade ('the potential partner count') and a 'proximity measure' equal to the sum of the inverse distance between a city and every city with which it could trade. The latter is found significant and negatively related to prices, which lends support to the hypothesis that geographical segmentation may have contributed to the increase in gasoline prices in regulated areas.

While they examine a similar issue, their approach is different than the one in this paper. They use city-level data from selected regions and a treatment and control approach. We use state-level data for the entire country and define regulatory distance by considering the difference in the size of markets under regulation in a given state and its adjoining states. An important difference is that we explicitly model the potential endogeneity of regulation which may arise, for example when firms lobby the state to introduce regulation beyond the minimum required under Federal guidelines or by introducing an unique fuel. These issues are not considered in their paper. In fact we show that OLS estimates systematically underestimate the effect of regulation on the wholesale price of gasoline. However, their findings with a complementary technique and disaggregated data support our conclusion that market segmentation following gasoline regulation has led to a significant increase in wholesale prices.

Section 2 provides background information on the U.S. gasoline market and regulation under the Clean Air Act. Section 3 describes the empirical model, the data used and provides estimation results and discussion. Section 4 concludes the paper.

2. Characteristics of the U.S. Gasoline Market

The U.S. gasoline market uses about a quarter of the world's crude oil and produces about 40% of the world's gasoline. Gasoline prices have been especially volatile in recent years, as shown in Fig. 1 which graphs the monthly price of gasoline and the price of crude oil between 1995 and 2002. This may be partly due to the volatility in the price of crude oil, which accounts on average, for about two-thirds of the price of a gallon of regular grade gasoline.[9] Crude oil supply

---

[9] During 1995-2002, the average wholesale price of gasoline per gallon in nominal terms was 73.9 cents, the retail price was 86.4 cents, and federal and state taxes were 41.3 cents. Thus the nominal retail price inclusive of taxes was $1.28. Since we consider wholesale prices and implicitly assume inelastic demand, state sales and excise taxes on gasoline are excluded from the analysis.

disruptions stemming from world events, natural disasters and refinery or pipeline outages have had a significant impact on both wholesale and retail gasoline prices. Even when crude oil prices are stable, gasoline prices normally fluctuate due to seasonality: prices tend to rise gradually before and during the summer driving season, and decline in the fall and winter, when people drive less.

Imported and domestically produced crude oil is distilled by refiners and converted into gasoline, kerosene (jet fuel), heating oil and several other petroleum products. Approximately 50% of all crude oil in the U.S. is imported but 96% of the gasoline consumed is refined domestically (EIA, 2000). Crude oil is transported in tankers from Europe, Asia and the Middle East and through pipelines from Mexico and Canada into major ports located in the New York Harbor, the Gulf Coast and the West Coast. It is then moved by barge or pipelines to refineries. Refined petroleum products are mostly carried by pipelines into wholesale terminals, and from there on trucks to retail outlets.

The U.S. refining industry has gone through a substantial restructuring in recent years. In 1981, a total of 189 firms owned 324 refineries; by 2001 the number of firms had reduced to 65 which together owned a total of 155 refineries, a decrease of about 65 percent in the number of firms and 52 percent in the number of refineries. During this period the market share of the ten largest refiners increased from 55 to 62 percent. This consolidation happened while gasoline demand continued to increase and consumption rose by about 30 percent (U.S. Senate, 2002). Several important mergers occurred in the 1998-2001 time frame beginning with Marathon with Ashland Oil, followed by British Petroleum with Amoco and ARCO, and Exxon-Mobil with Chevron-Texaco.[10]

Many small refiners used domestically produced crude oil and benefited from price controls on imported oil during the era of high world oil prices in the 1970s. The end of the Crude Oil Entitlements Program led to the shut down of some of these inefficient units. Conservation programs of the 70s took effect in the 80s, reducing demand and hence refining margins. By 1981, only about two-thirds of the refinery capacity was being utilized. In addition, the Clean Air

---

[10] Other mergers include Philips with Tosco and Conoco and Valero with Ultramar Diamond Shamrock.

Act of 1990 mandated higher gasoline standards, such as oxygenated and reformulated gasoline, forcing refiners to upgrade their refineries and add to capacity. Many refiners that did not make the necessary investments exited the industry. Recent capacity utilization rates in the refinery industry are routinely more than 90 percent. Increased concentration and capacity utilization has also meant reduced inventories. Average gasoline storage in 1981 was equal to 40 days consumption. By 2001, it fell to 25 days. This restructuring has resulted in a tight gasoline market characterized by frequent price spikes.

Table 1 shows state level data on population, the price of gasoline (in real 1995 dollars), total number of refineries, total refinery capacity, the refinery capacity per capita, and the share of population under RFG and OXY programs. These figures are averaged for the 1995-2002 period.[11]

The states in Table 1 have been sorted by increasing average price. States with larger refining capacities generally have lower prices. Examples include Texas with 23 refineries out of a national total of 140, Mississippi with 3 refineries, Louisiana with 17 refineries, and Oklahoma with 6 refineries. Conversely, in the East Coast region (or PADD)[12] prices in Connecticut, Massachusetts, and Maryland with no refineries are about 5 cents per gallon higher than Pennsylvania, New Jersey, and Virginia where some crude oil is refined. California is an exception with 17 refineries and prices that are among the highest nationally. This premium is due to the state requiring the use of a unique, cleaner and thus costlier, gasoline. The 4-firm concentration index (which is not shown here) is computed as the sum of the largest four market shares for states with a positive number of refineries, is lower than 0.60 in only Texas (0.40) and Louisiana (0.54).[13] All states belonging to the Rockies and West Coast regions have higher prices and therefore are located in the bottom half of the table. The states of Alaska and Hawaii, which incur high costs of transportation for imported gasoline, have the highest prices. The Gulf Coast

---

[11] See Appendix A for definitions and sources of data.

[12] Historically, crude oil allocation in the United States has been divided into five Petroleum Administration for Defense Districts (PADD). These districts were originally classified during World War II for purposes of administering an oil allocation program. The five PADDs are: West Coast, the Rockies, Midwest, the Gulf Coast and East Coast. The PADD identification of each state is shown in Table 1.

[13] Measuring market power through concentration is common in empirical studies. Kim and Singhal (1993) use the Herfindahl index to measure the degree of concentration in the airline industry. Evans and Kessides (1993), Berger and Hannan (1989) and Cotterill (1986) use similar approaches. Shepherd (1999) defines a market as a tight oligopoly when this index is above 0.60.

has the highest number of refineries and the lowest prices. The Midwest also has a significant number of refineries and relatively low prices. Because it contains the states of California, Alaska and Hawaii, the West Coast is an exception – a large number of refineries and high gasoline prices.

Clean Air Regulation of Gasoline

The Clean Air Act Amendments of 1990 established a clean fuels program to reduce harmful emissions from motor vehicles. Areas that do not meet EPA national ambient air quality standards are required to implement clean gasoline programs, the most important of which are the RFG and OXY fuel programs.

The RFG program was implemented since January 1995 in areas with major ozone problems. RFG is a gasoline blend that contains lower levels of benzene, sulfur and aromatic compounds. RFG provides the same vehicle performance as regular gasoline but does not evaporate as easily, especially in the summer. It also reduces volatile organic compounds and toxic emissions. Areas with less severe pollution were given the option of using RFG. It is now used in 17 states and the District of Columbia and accounts for nearly 30 percent of the gasoline sold in the U.S. The RFG program runs for the whole year (EIA, 1999a). RFG fuels must contain 2% oxygen by weight, since oxygen aids combustion and thus reduces emissions of certain harmful compounds. But how it is done is entirely at the discretion of the refiner. Most RFG fuels contain the chemical MTBE as an oxygenate (since oxygen cannot be added directly), but in Chicago and Milwaukee, which are closer to the grain-producing regions of the Midwest, ethanol is the required oxygenate.

The OXY program was launched in November 1992 and was mandatory in carbon monoxide non-attainment areas in order to reduce its production from gasoline in the winter months. Federal standards require oxygen content to gasoline of at least 2.7 percent by weight. OXY accounts for about 5 percent of the gasoline sold, is run only in winter months (November through February) and averages about 1.3 percent over the whole year (EIA, 1999a). Originally 39 areas qualified for this program, but only 16 of them use OXY at present, the rest having already achieved target regulatory standards.

These regulations are implemented in two steps. First, based on pollution standards for carbon monoxide and ozone, the Clean Air Act mandates the use of regulated gasoline in specific areas in the country. Next, each state that is so regulated can decide to impose more stringent regulation, e.g., by extending the use of RFG or OXY to a larger area than the one required by federal regulation. States can also impose stricter standards on the gasoline content. For example, two states (California and Minnesota) chose to produce special blends (not sold anywhere else in the country) and 11 other states decided to impose a lower RVP than what EPA imposed originally.[14]

We focus only on the more important RFG and OXY programs and do not consider state fuel programs that impose lower gasoline volatility requirements and caps on sulfur content. Because of the chemical characteristics of the pollutants, RFG and OXY programs are mutually exclusive except in the Los Angeles region, i.e., either an area is under RFG or under OXY regulation.[15]

The price of RFG gasoline is higher than the price of regular gasoline. During the study period, their observed average price differential was about 5 cents, ranging from 2.63 cents in the Gulf Coast region to 8.23 cents in the Midwest (Petroleum Marketing Annual, 1995-2002).[16] The first explanation for this price difference is that RFG and OXY blends are more costly to produce than regular gasoline as refiners have to make adjustments in their production technology. It is estimated to cost an additional 2-4 cents per gallon to produce these fuels relative to regular gasoline (EIA, 1999a). Although these numbers may seem small, an industry rule of thumb is that a 10 cent/gallon price increase translates into additional annual industry revenues of 10-12 billion dollars (US Senate, 2002, p.20).

This cost effect may not explain the entire price differential between RFG and regular gasoline. Another possible explanation is that boutique fuel regulation effectively segments the market, leading some refiners to produce these special fuels while others produce the standard gasoline

---

[14] See EPA (2001, Appendix B) for a list of these states and the specific standards imposed.

[15] Both programs are imposed in the Los Angeles area because of severe urban smog problems coupled with the especially strict regulatory standards in the state of California.

[16] Information on the price of OXY gasoline was not available.

blends. Yet other refiners may produce multiple blends and vary their output mix over time. In the short run product differentiation may lead to a smaller number of firms selling in each wholesale market, thereby reducing competition and increasing prices. Our empirical analysis below aims to measure these cost and segmentation effects.

States with neither of these programs tend to have the lowest prices (Mississippi, Louisiana, Oklahoma, South and North Carolina, Tennessee, Georgia, Alabama, Arkansas, Florida and Kansas), as seen in Table 1. On the other hand, states with both programs have higher prices (Connecticut, Massachusetts, Maryland, Arizona, California and the District of Columbia).

Table 2 shows annual descriptive statistics averaged over all states. Refinery capacity per state increased over the study period from 302,649 to 329,125 barrels per day. The refinery concentration index remained stable at about 0.60 during this period. The share of population under RFG remained constant (around 0.25) with a decrease for OXY from 0.09 in 1995 to 0.05 in 2002. Some OXY areas achieved their goals and left the program during the study period.

3. The Empirical Model

We only consider mid-grade gasoline in the estimation. The average price of gasoline and the price of crude oil are measured in 1995 dollars.

The Wholesale Price Equation

The equation for the wholesale price of gasoline (in logs) denoted by P_{it} can be specified as:

$$P_{it} = \mathbf{x}'_{\mathbf{it}}\beta + \lambda_t + \alpha_i + \upsilon_{it} \tag{1}$$

where the subscript i $(i = 1, 2, .., N)$ represents states and t $(t = 1, 2, .., T)$ is the time index in years, respectively.[17] The vector $\mathbf{x}_{\mathbf{it}}$ represents the characteristics of the gasoline market in state *i* that may affect price P_{it}. Time effects that affect all states simultaneously are captured by dummies λ_t.

---

[17] The original data for the wholesale price of gasoline and the price of crude oil is by month. When estimating the price model with monthly data, the month-specific indicators were highly significant because of the seasonal effects described earlier. The significance of the parameters of interest was generally higher in the model using annual data.

Examples include the terrorist attack of September 11, 2001 and crude oil output decisions by OPEC. To control for unobserved state heterogeneity, we specify time-invariant, state-specific effects denoted by α_i which are estimated as fixed parameters.[18] We include the usual idiosyncratic error term υ_{it} with mean zero.

Variables that describe the characteristics of the gasoline market in state *i* include the (log) price of crude oil, the refinery capacity per capita in the state,[19] the refinery concentration index in the state and average distance to refinery, defined below. We expect a positive relationship between the price of crude oil and the price of gasoline. A state with a higher refinery capacity per capita is expected to have a lower gasoline price, *ceteris paribus*. Gasoline prices are expected to be positively correlated with the refinery concentration index. The greater the market concentration, the larger the potential market power of firms in the segmented market.

An important issue is the need to distinguish transportation costs from non-competitive markups. Some regions may be geographically remote from refineries yet be served by many of them. Their prices will be high due to transportation costs. On the other hand, remote regions may be served by only a few refineries and their prices may be high both due to the cost of transportation and because of non-competitive markups. Therefore we need to define variables that allow a distinction between distance effects and market concentration. As a proxy for transportation costs, we define the variable 'average distance to refinery' for state *i* in year *t* by the ratio

$\dfrac{\sum_{j=1, j\neq i}^{J} D_{ij} CAP_{jt}}{\sum_{j=1}^{J} CAP_{jt}}$ where D_{ij} is the distance (in kilometers) between the capitals of state *i* and state *j*,

CAP_{jt} is total refinery capacity in state *j* and year *t*, and *J* represents the total number of states in the country.[20] This is the average distance from a state to all refineries in the country, weighted

---

[18] When studying an exhaustive population, Arellano (2003) suggests that a fixed-effects specification is more appropriate than a random-effects model.

[19] A better fit was obtained with refinery capacity per capita, instead of total refinery capacity.

[20] Distance measures between state capitals are great circle distances ("as the crow flies") computed using latitude and longitude coordinates available from the US Geological Survey. For computational details, see http://www.cpearson.com/excel/latlong.htm. This distance measure varies from an average of 1,346 km in the Gulf Coast region to an average of 2,321 km in the West Coast. States located in the Gulf Coast region are in general closer to refineries than states in other regions.

by the capacity of each refinery. The higher the distance to refineries, the higher the value of this index. This measure is not perfect since state capitals may not be a good proxy for the center of gravity of the gasoline market in the state. Many states may have multiple urban centers. Implicitly this measure also assumes that the refineries in other states are located in the capitals of those states. However, it may serve as a reasonable first approximation in the absence of data on gasoline inflows into states from individual refineries. In the empirical section, we check the validity of this measure as a proxy for transportation costs for each region. A higher average distance to refinery (or equivalently, higher transportation costs) is expected to induce a higher wholesale price of gasoline, *ceteris paribus*.

The impact of transportation costs on gasoline price is likely to vary depending on whether a state is an importer of gasoline or not. A state which has a higher production of gasoline should be less affected by transportation costs. We thus include in the model a cross term defined as the product of refinery capacity per capita and the average distance to refinery.

Clean Air regulation may directly affect the price of gasoline by increasing the cost of refining and distribution (the cost effect). It may also affect the gasoline price indirectly by reducing competition among refineries (the segmentation effect). We measure the cost effect by introducing the relative size S_{kit} of the RFG and OXY markets in each state as explanatory variables in equation (1). These are defined as follows:

$$S_{kit} = \frac{POP_{kit}}{POP_{it}}, \tag{2}$$

where POP_{kit} is the size of the market (population) for regulated gasoline of type k $(k = RFG, OXY)$ in state i and year t, i.e., the total population in the area covered by program k, and POP_{it} is total population in state i and year t. The ratio S_{kit} thus measures the relative size of the market covered by the clean fuel program and takes values from zero to one. A larger regulated market leads to a larger market share for clean gasoline and hence a higher price.

Variables Measuring Regulatory Distance between States

To test for the effect of segmentation on price in a state, we build a variable that measures regulation in the state relative to regulation in the adjoining states. We implicitly assume that refineries in any state compete only with refineries in adjoining states and are not impacted by markets in states located more remotely.[21] We adopt a differentiation index from the literature on international trade (see Helpman and Krugman, 1985) in which trade flows between countries is written as a function of GDP per capita and population:

$$I_{kit} = \left| \frac{POP_{kit}}{POP_{it}} - \frac{NPOP_{kit}}{NPOP_{it}} \right| \tag{3}$$

where $NPOP_{kit}$ is the market size (population) for regulated gasoline in the states neighboring state *i,* and $NPOP_{it}$ is total population in the neighboring states.[22] The index I_{kit} measures the *difference* in the market size for regulated gasoline in the state and its adjoining states. When I_{kit} takes the minimum value of zero it indicates that the relative size of the regulated market in state *i* and its neighbors is the same. When it equals the maximum value of one, we have complete differentiation between regulation in the state and in the adjoining states. A higher value of I_{kit} implies a higher price of gasoline. This measure abstracts from accounting for heterogeneity both in the types of crude oil used for refining and in the refinery product mix. Not all refineries produce the same type of gasoline, leading to greater transportation of gasoline than if all refined products were homogenous. There may be heterogeneity in the size of states. California is several times larger than some states on the East Coast. So for California, regulation in surrounding states may have less of an effect than say for a smaller state like Rhode Island. The effect of differential sizes is partly captured by the average distance to refinery index discussed earlier. We have also

---

[21] The use of variables describing market conditions in neighboring states can be found in Baltagi and Levin (1986) and Baltagi and Li (1999). These authors build a model in which cigarette consumption in one state is assumed to depend not only on the price of cigarettes in that state but also on the price of cigarettes in neighboring states, the latter reflecting the price of a substitute good. In our case, the dependent variable is price which is affected by the availability of the good in neighboring states, the latter being measured by the difference in the size of the regulated market between a state and its neighbors.

[22] We have considered other indices such as the relative size of the regulated gasoline market in the whole region (i.e., state *i* and its adjoining states) defined by $\frac{POP_{kit} + NPOP_{kit}}{POP_{it} + NPOP_{it}}$. We use the specification which was found to be the most significant in the price model.

re-estimated the model with an interaction term between I_{kit} and POP_{kit} but the estimated parameters did not change significantly, as discussed later in the paper. Coastal states may have fewer neighbors, but that also implies that each neighbor has a higher degree of influence which is captured adequately by the I_{kit} index.

Endogeneity of Regulation

The choice of a boutique fuel that is different from those used in neighboring states creates opportunities for firms to exercise market power. Even though the regulatory programs we consider in this study are primarily federal, firms may have an incentive to lobby the state to expand area coverage beyond the minimum required under Federal guidelines or to choose a unique fuel.[23] This is not an issue in a competitive market, since firms will pass on any increase in costs to the consumer. However, given the tight concentration in the energy market, this opportunity to influence the choice of regulation and the type of boutique fuel exists.

In our model, local firms may lobby the federal government to influence the area covered by these regulated fuels. These factors may imply that the S_{kit} and I_{kit} variables (which measure the size of the regulated markets) may be endogenous in the price model if local firms jointly determine gasoline price and decide on a lobbying activity to influence areas of regulation. The variable measuring market concentration in the state (the refinery concentration index) may also depend on the price of gasoline and hence be endogenous in the price model. As a result, all six variables can be considered endogenous in the model: the wholesale price of gasoline P_{it} and the refinery concentration index RC_{it} may depend on the four variables S_{kit} and I_{kit} (one each for RFG and OXY).

Moreover, dependent variables S_{kit} and I_{kit} can take on a value of zero with positive probability, when states do not impose any regulation on gasoline. Parameter estimates may be affected by a selectivity bias if OLS are used on the sub-sample of positive observations only. We therefore specify each equation for S_{kit} and I_{kit} as a Tobit model for variables censored at zero.

---

[23] There is limited evidence that this may have occurred. According to the US Senate (2002, p.74), some refiners have encouraged states to develop unique fuel requirements in order to create distinct fuel markets with limited competition while simultaneously urging federal officials to reduce the number of fuels nationally.

As the variables RC_{it}, S_{kit} and I_{kit} also vary across states and through time, we must account for time effects and unobserved state heterogeneity. Thus the full model is a simultaneous equation system which reads:

$$\begin{cases} P_{it} = \mathbf{x}_{\mathbf{it}}'\boldsymbol{\beta} + \lambda_t + \alpha_i + \upsilon_{it}, & (4) \\ RC_{it} = \mathbf{w}_{\mathbf{it}}'\boldsymbol{\theta} + \delta_t + \tau_i + \zeta_{it}, & (5) \\ S_{kit} = \max(0, \mathbf{z}_{\mathbf{kit}}'\boldsymbol{\gamma}_{\mathbf{1k}} + \varpi_t + c_{1ki} + \varepsilon_{1kit}), \quad k = RFG, OXY, & (6\text{-}7) \\ I_{kit} = \max(0, \mathbf{z}_{\mathbf{kit}}'\boldsymbol{\gamma}_{\mathbf{2k}} + o_t + c_{2ki} + \varepsilon_{2kit}), \quad k = RFG, OXY, & (8\text{-}9) \\ \qquad\qquad i = 1, 2, \ldots, N \; ; \; t = 1, 2, \ldots, T, & \end{cases}$$

where (λ_t, δ_t, ϖ_t, o_t) are year-specific effects, (α_i, τ_i, c_{1ki}, c_{2ki}) are the state-specific unobserved effects, and $\left(\upsilon_{it}, \zeta_{it}, \varepsilon_{1kit}, \varepsilon_{2kit}\right)$ are the idiosyncratic error terms of the structural equations.

The **x**-vector represents variables that describe the characteristics of the gasoline market in state *i,* including the (log) price of crude oil, the refinery capacity per capita in the state, the size of the regulated markets (S_{kit}), the difference in market size (I_{kit}), the refinery concentration index in the state (RC_{it}) and average distance to refinery.

The **w**-vector includes the (log) crude oil price, the (log) wholesale price of gasoline (P_{it}), refinery capacity per capita, average distance to refinery, the interaction term between refinery capacity per capita and average distance to refinery, state population density, MTBE consumption by state and year, and the S_{kit} and I_{kit} indices.

The **z**-vector, which is assumed to be the same in the equations (6-9), contains identifying instruments for the impact of environmental regulation (S_{kit} and I_{kit}) on the price of gasoline (P_{it}). These instruments are total state population, state population density, share of population classified in non-attainment areas for particulate matter and sulphur dioxide, average number of vehicles per capita, and the ratio of net receipts over net production of gasoline for the corresponding region or PADD.[24,25] The net receipts over net production of gasoline variable provides a measure of the dependence of each region or PADD on imports from outside the region and is measured at the PADD level and not at the state level. Thus it may be assumed exogenous in the price model. The share of population classified in non-attainment areas for particulate matter and sulphur dioxide provides a measure of the extent of air pollution in each state while not being directly related to RFG and OXY regulation - RFG and OXY regulation are designed to reduce pollution by ozone and carbon monoxide.

To allow for identification of the system parameters, we impose exclusion restrictions on some exogenous variables for appropriate equations, to satisfy the usual rank condition in simultaneous-equation systems. Estimation details for this simultaneous equation system of panel data equations are provided in Appendix B. Estimation is done for 384 observations corresponding to 48 states over a period of 8 years. The states of Alaska and Hawaii are not considered at the estimation stage since the I_{kit} index is not defined for these states, since they share no border with any other state.

Results of Estimation

Estimation results for the full six-equation system are reported in Table 3 and diagnostic tests are shown in Table 4.[26] For reasons of space, estimates of the year- and state-specific effects in the six equations are not shown but are available upon request from the authors.

We first test for endogeneity of regressors in the first two equations in Table 3, namely the price and refinery concentration equations. The Durbin-Wu-Hausman test rejects the null hypothesis that the OLS estimator would yield consistent estimates and therefore indicates that the refinery concentration index (RC_{it}) and the indices for environmental regulation (S_{kit} and I_{kit}, $k=RFG,$

---

[24] The average number of vehicles per capita in the state was not found significant in the price equation.

[25] The East Coast and Midwest PADDs are major importers of gasoline while the Gulf Coast is primarily an exporter. The West Coast and the Rockies exported small quantities of gasoline during the period studied. See Table 1 and footnote 11 for PADD identification and definition.

[26] The model was re-estimated with an interaction term between the I_{kit} index and the population in each state (one for RFG, one for OXY). These terms are endogenous since the index I_{kit} is an endogenous variable. This new model is a system of eight simultaneous equations combining price, refinery concentration index, S_{kit} ($k=RFG, OXY$), I_{kit} ($k=RFG, OXY$), and the two interactive terms. The same procedure as the one described in Appendix B was applied and in general, the estimated parameters in the price equation were not found statistically different from the estimated parameters when the interactive terms were excluded.

OXY) are endogenous in the price equation, and that the use of instruments is justified (Chi-square = 83.89, p-value = 0.000, see Table 4).[27] Similarly, the price of gasoline (in logs) and the S_{kit} and I_{kit} indices for environmental regulation are found endogenous in the equation fitting the refinery concentration index (Chi-square = 28.82, p-value = 0.000).

The Anderson canonical correlations Likelihood Ratio test for underidentification is also presented (Table 4). A rejection of the null hypothesis indicates that the model is identified and that the instruments are relevant (see Hall, Rudebusch and Wilcox, 1996). For both the price equation and the refinery concentration index equation, the null hypothesis is rejected at the 10% level of significance.

Finally, we test that the instruments are valid, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equations, using the Sargan test of overidentifying restrictions. Under the null hypothesis, the test statistic is distributed as Chi-square in the number of overidentifying restrictions. For both the price equation and the refinery concentration index equation, the p-value is above 0.10, thus confirming the validity of the selected instruments (see Table 4).

In the price equation, we find that the price of crude oil is a major determinant of gasoline prices (see Table 3). The estimated parameter measuring the elasticity of the wholesale price to the price of crude oil is obtained as 0.63. It implies that a 10% increase in the price of crude oil leads to a 6.3% increase in the wholesale price of gasoline.[28] The refinery concentration index has a significant positive coefficient, suggesting a link between concentration and the price of gasoline. Because the variables measuring refinery capacity per capita and average distance to refinery enter the price model non-linearly, their coefficients (as shown in Table 3) are not directly interpretable. The marginal effect of the two variables is obtained by computing the derivative of the right-hand-side expression with respect to refinery capacity per capita and average distance to refinery, respectively. At the sample mean, we find, as expected, that states with a larger refinery

---

[27] Under the null of exogeneity of the refinery concentration index RC_{it} and the S_{kit} and I_{kit} indices for environmental regulation, the Durbin-Wu-Hausman test-statistic follows a Chi-square distribution with *l* degrees of freedom, *l* being the number of endogenous regressors (Davidson and MacKinnon, 1993).

[28] Chouinard and Perloff (2007) also find that the price of crude oil is a significant determinant of gasoline price.

capacity per capita have a lower gasoline price. The marginal effect is estimated at -0.719, significant at the 1 percent level. States with a larger average distance to refinery have a higher gasoline price. The marginal effect is estimated at 0.635, significant at the 1 percent level. To check for the validity of the average distance to refinery measure as a proxy for transportation costs, we compute the marginal effect of this variable at the mean of each region or PADD. The strongest marginal effect of the distance to refinery measure on gasoline prices is obtained for the East Coast region (0.711, significant at the 1% level) which is highly dependent on imported gasoline. The lowest marginal effect which is not statistically different from zero is obtained for the Gulf Coast region (0.403) which is largely gasoline self-reliant.

In the price equation, three of the seven year dummies are found significant at the 5 percent level - five are significant if we consider the 10 percent level. The year 1995 is chosen as the reference. These coefficients capture effects that have had an impact on the gasoline market as a whole (see Figure 1). The largest coefficient is obtained for the year 2001, which may be explained by the combined effect of consolidation in the gasoline market and the terrorist attack on September 11 (U.S. Senate, 2002).

Table 3 shows that the four variables used to measure the impact of gasoline regulation, namely $S_{RFG}, S_{OXY}, I_{RFG}, I_{OXY}$ have positive signs, and are highly significant. The larger the relative size of the market for regulated gasoline in a state, the higher the price. This result illustrates the cost effect described earlier: RFG and OXY are more costly to produce than regular gasoline and those states in which a larger segment of the market uses RFG or OXY fuels exhibit a higher price for gasoline. Our estimates indicate that the price of gasoline would increase by 16 percent if a state with no regulation would impose either RFG or OXY regulation to the whole population (the S_{kit} ratio would increase from 0 to 1).[29]

The difference in the size of the RFG (respectively, OXY) market between a state and its neighbors is also positively related to the price in the state, suggesting that market segmentation has a positive effect on the price of gasoline. This segmentation effect is found to be highly

---

[29] An increase in the S_{kit} ratio from 0 to 1 induces an increase in the average price of gasoline measured by $e^{0.151}=1.163$ for the case of RFG and $e^{0.150}= 1.162$ for the case of OXY.

significant for both the RFG and OXY markets. Compared to a situation in which there is no differentiated regulation between a state and its neighbors (the I_{kit} index is equal to 0), a complete differentiation in the RFG market regulation (respectively OXY market regulation), i.e., when the I_{kit} index is equal to 1, induces an increase in the wholesale price of gasoline by about 14 percent (15 percent for OXY).[30]

Several explanatory variables are found significant in the refinery concentration index equation (see Table 3). In particular, the results confirm the positive correlation between gasoline prices and the refinery concentration index. The variable measuring the extent of RFG regulation in each state (S_{RFG}) has a negative and significant coefficient. This would suggest that the requirement to produce a cleaner gasoline in a state has induced a lower concentration in the refinery sector in that state, all other things equal.

In Table 5, we report OLS estimation results for the price equation (including year- and state-specific effects) as benchmark. If the signs of the coefficients that have been estimated with OLS and 3SLS are the same, the magnitudes are different, in particular for the five endogenous variables. We find evidence that the OLS strategy systematically underestimates the impact of the four variables measuring environmental regulation and the refinery concentration index on the price of gasoline, suggesting that the endogeneity of these variables needed to be addressed in the estimation.

4. Concluding Remarks

This paper examines the effect of differential Clean Air regulation on wholesale gasoline prices. We consider two major boutique fuels – reformulated gasoline and oxygenated fuels which aim to reduce ozone and carbon monoxide from automobile emissions. Using measures that include the relative size of the regulated market in the state as well as the difference in market sizes between a state and its adjoining states, we find that boutique fuels cause an increase in gasoline

---

[30] Although we consider the most important clean fuel programs RFG and OXY, the price in states with no RFG and OXY regulation may be impacted by other federal and state clean fuel programs that we do not study here. Since we are in effect pooling observations with and without non-RFG and OXY fuels, the mark-ups we estimate are lower bounds. Alternatively, our estimates are relative to prices if no RFG (or OXY) is used. This could be corrected by including a variable that measures the population under all other boutique fuels relative to total state population. However, such data were not easily available. We thank an anonymous referee for this important observation.

prices in two ways – by increasing the cost of refining and by segmenting the market and increasing the market power of firms. While the refinery concentration in the regulated market in a state leads to an increase in the price of gasoline, the price is also affected by the regulatory distance between a state and its neighbors. The larger this distance, the higher the price in that state. These conclusions support the notion that heterogeneity in gasoline regulation following the Clean Air Act may have been an important factor in the increase in domestic gasoline prices in recent years, as suggested by several industry analysts (e.g., Fesharaki, 2004).

An important issue addressed in the paper is the potential endogeneity of regulatory policies. For example, firms may lobby their states to extend regulation beyond federally mandated levels or select a unique blend of gasoline for sale in regulated markets. The econometric model is estimated by treating all the regulatory variables as endogenous. Comparison with benchmark OLS estimates suggests that the latter systematically underestimates the effect of regulation on gasoline prices.

A policy implication is that homogenizing gasoline regulation over the entire country may increase gasoline prices because of the higher cost of regulation but there may be an opposite effect from reduced segmentation in the market. At least for some states that may currently be under little or no regulation but are surrounded by states that are heavily regulated, ratcheting up regulation may lead to a decline in the price of gasoline. It may be useful to examine these state level price effects in future work especially if a uniform national standard were to be in place. Future research could also focus on extending this analysis to the state or PADD levels with disaggregated price data. One could test for price volatility and examine whether gasoline prices are more volatile in states with heterogeneity in regulation or those at a greater regulatory distance from their neighbors.

References

Anselin, L. (1988), *Spatial Econometrics:Methods and Models*, Dordrecht: Kluwer.

Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press: Advanced Texts in Econometrics.

Baltagi, B.H. and D. Levin (1986), "Estimating Dynamic Demand for Cigarettes using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered," *Review of Economics and Statistics* 68 (1), 148-155.

Baltagi, B.H. and D. Li (1999), "Prediction in the Panel Data Model with Spatial Correlation," Working Paper, Texas A&M University.

Baltagi, B. H. (2001). Econometric Analysis of Panel Data. 2nd Edition. New York: John Wiley and Sons.

Berger, A. N. and T. H. Hannan (1989), "The Price-Concentration Relationship in Banking," *Review of Economics and Statistics* 71(2), 291-299.

Borenstein, S., and A. Shepard (1996a), "Dynamic Pricing in Retail Gasoline Markets," *RAND Journal of Economics* 27(3), 429-451.

Borenstein, S., and A. Shepard (1996b), "Sticky Prices, Inventories, and Market Power in Wholesale Gasoline Markets," NBER Working Paper No. 5468.

Borenstein, S., A. C. Cameron and R. Gilbert (1997), "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes," *Quarterly Journal of Economics* 112(1), 305-339.

Brown, J., Hastings, J., Mansur, E.T., and Villas-Boas, S.B. (2006), "Reformulating competition? Gasoline Content Regulation and Wholesale Gasoline Prices," CUDARE Working Paper number 1010, University of California, Berkeley.

Chouinard, H., and J. M. Perloff (2007), "Gasoline Price Differences: Taxes, Pollution Regulations, Mergers, Market Power, and Market Conditions," *The B.E. Journal of Economic Analysis & Policy*, Volume 7(1), Article 8.

Cotterill, R. W. (1986), "Market Power in the Retail Food Industry: Evidence from Vermont," *Review of Economics and Statistics* 68(3), 379-386.

Dansby, R. E. and R. D. Willig (1979), "Industry Performance Gradient Indexes," *American Economic Review* 69(3), 249-260.

Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Energy Information Administration (1999a), *Environmental Regulations and Changes in Petroleum Refining Operations*, November.

Energy Information Administration (1999b), *Areas Participating in the Oxygenated Gasoline Program, http://www eia.doe.gov/emeu/steo/pub/special/oxy2.html.*

Energy Information Administration (2000), *Petroleum Supply Annual*, Volume 1, Table S4.

Energy Information Administration (2002), *Gasoline Type Proliferation and Price Volatility*, Office of Oil and Gas, September.

Energy Information Administration (2003), *A Primer on Gasoline Prices*, September.

Energy Information Administration (2003), *Motor Gasoline Outlook and State MTBE Bans*, April.

Environmental Protection Agency (2001), *Study of Unique Gasoline Fuel Blends ("Boutique Fuels"), Effects on Fuel Supply and Distribution and Potential Improvements,* Staff White Paper, Office of Transportation and Air Quality, October.

Evans, W. N. and I. N. Kessides (1993), "Localized Market Power in the U.S. Airline Industry," *Review of Economics and Statistics* 75(1), 66-75.

Fesharaki, F., (2004), "Crystal Ball 2004," Presented at the Asian Oil and Gas Conference, Kuala Lumpur, June 13-15.

Green, M. (2006), "Reps. Ryan and Green Advance Fuel Bill with House Leadership," http://www.house.gov/apps/list/press/wi08_green/0628grtycnvb.html.

Greenspan, A., (2001), "Impacts of Energy on the Economy," Remarks before the Economic Club of Chicago, Chicago, Illinois, June 28.

Hall A.R., G.D. Rudebusch and D.W. Wilcox (1996), "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review* 37(2), 283-298.

Hausman, J.A., and W.E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49(6), 1377-1398.

Helpman, E., and P. Krugman (1985), *Market structure and foreign trade: Increasing returns, imperfect competition and the international economy*. Cambridge: MIT Press.

Kim, E. H., and V. Singhal (1993), "Mergers and Market Power: Evidence from the Airline Industry," *American Economic Review* 83(3), 549-569.

Muehlegger, E. (2004), "Gasoline Price Spikes and Regional Gasoline Content Regulations: A Structural Approach," Unpublished manuscript, Department of Economics, MIT.

Newey, W. K. and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55(3), 703-708.

Petroleum Marketing Annual, (1995-2002), http://www.eia.doe.gov/bookshelf.html

Ryan, P., (2003), www.house.gov.ryan/press/_releases (2003), "Ryan Passes Amendment to Stabilize Gas Prices for Wisconsin over Long Term," April 11.

Shepherd, W. G. (1999), *The Economics of Industrial Organization*, 4th edition, Waveland Press.

Shonkwiler, J.S. and S.T. Yen (1999), “Two-step estimation of a censored system of equations,” *American Journal of Agricultural Economics* 81(4), 972-982.

United States Senate (2002), *Gas Prices: How Are They Really Set?* Permanent Subcommittee on Investigations, Committee on Governmental Affairs, Washington, DC.

Wooldridge, J.M. (1995), “Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions,” *Journal of Econometrics* 68(1), 115-132.

Yacobucci, B., (2004), “Boutique Fuels and Reformulated Gasoline: Harmonization of Fuel Standards,” Congressional Research Service, Report for Congress, January 9.

Appendix A: Data Description and Sources

Wholesale gasoline prices are obtained from the Petroleum Marketing Annual reports (1995-2002) prepared by the Energy Information Administration (EIA). The price of crude oil is the monthly national average price of the composite (domestic and imported) refiner acquisition cost, which is the average price of crude oil purchased by refiners (in cents per gallon). The wholesale price is the monthly average price of regular motor gasoline wholesale sales (in cents per gallon excluding taxes) within the state. We deflate prices using the Consumer Price Index (CPI) provided by the Bureau of Labor Statistics.

Refinery capacity (in million barrels per calendar day) is the aggregate capacity of all refineries operating in the state (source: EIA). Information on refineries was missing for the years 1996 and 98, so 1995 and 1997 figures were used as substitutes.[31] Total refinery capacity is a good proxy for crude oil production since the annual average refinery utilization rate regularly exceeds 90 percent of installed capacity (US Senate, 2002, p. 5). These figures are used to compute the market share (based on capacity) of each firm owning a refinery in any given state and the firm concentration index is defined as the sum of the largest four market shares.[32] These indices are computed on an annual basis. MTBE consumption by state and year were obtained from EIA (2003b). Consumption is measured in thousand barrels per day. Net production and net receipts of gasoline by PADD and year were obtained from the Petroleum Supply Annual, Volume 2 (1995-2002) prepared by the EIA.

Information on control areas under the RFG and OXY programs, as well as population in non-attainment areas for particulate matter and sulphur dioxide, was obtained from EPA. EPA also provided data on the population of the mandated and opt-in RFG control areas and the population of OXY control areas, both estimated as of July 1, 1996. The duration of the oxygenated fuel

---

[31] We estimated the price model with linear interpolation of the refinery capacity data, as suggested by a referee. Because refinery capacity does not exhibit significant variation from year to year, we find that coefficients for all variables are almost unchanged relative to those reported in the paper. The coefficient on the variable measuring refinery capacity is slightly different but not statistically different.

[32] Different concentration measures have been proposed in the literature. We follow Dansby and Willig (1979) who advocate the use of a *m*-firm concentration ratio when the largest *m* firms collude and the remaining firms are price-takers. We choose *m=4*. The US Senate report (2002, p.4) also uses the four-firm concentration ratio in its analysis of mergers in the gasoline industry.

programs is for at least four months, and typically runs from November 1 to February 29, although it may vary by state. We control for the period of implementation (number of days per month) for the OXY program (source: EIA, 1999b).

Information on population by state and year was obtained from the US Census bureau, and vehicle registration data were obtained from the US Federal Highway Administration.

Appendix B: Details of the Econometric Estimation

The year-specific effects (λ_t, δ_t, ϖ_t, o_t) and the state-specific effects (α_i, τ_i, c_{1ki}, c_{2ki}) are assumed to be fixed parameters to be estimated. Equations (6-9) for the S_{kit} and I_{kit} indices can be estimated using a Tobit procedure. To apply the Tobit approach to panel data, we need to account for two issues. First, unobserved state heterogeneity (state-specific effects (c_{1k}, c_{2k}), $k = RFG, OXY$) may be correlated with the explanatory variables z_{kit}. To control for this possible effect, we assume that such a correlation disappears when state-specific effects are projected onto a set of observed exogenous variables, denoted $\bar{z}_{ki}$. More specifically, we assume that

$$c_{1ki} \left| z_{ki} \sim \text{Normal } \left(\psi_{1k} + \bar{z}_{ki}\xi_{1k},\ \sigma^2_{c_{1k}} \right), \right.$$
$$c_{2ki} \left| z_{ki} \sim \text{Normal } \left(\psi_{2k} + \bar{z}_{ki}\xi_{2k},\ \sigma^2_{c_{2k}} \right), \right.$$

where $\sigma^2_{c_{1k}}$ (respectively, $\sigma_{c_{2k}}$) is the variance of a_{1ki} (respectively, a_{2ki}) in the equation $c_{1ki} = \psi_{1k} + \bar{z}_{ki}\xi_{1k} + a_{1ki}$ (respectively, $c_{2ki} = \psi_{2k} + \bar{z}_{ki}\xi_{2k} + a_{2ki}$). In addition, we require that $E\left(a_{1ki}\ z_{kit}\right) = E\left(a_{2ki}\ z_{kit}\right) = 0$. See Wooldridge (1995) for more on such procedures for nonlinear panel data models.

Second, extending the Tobit approach to panel data is usually done by specifying a random-effects model, in which explanatory variables are uncorrelated with idiosyncratic errors. The conditional expectation of censored dependent variables is typically computed for predicting

positive values on the associated sub-sample of observations. However, as this can result in a significant loss in the number of observations used for estimation, we consider instead the non-conditional expectation of S_{kit} and I_{kit} based on the full sample, using for example, the fact that $E(S_{kit}) = P(S_{kit} > 0) \times E(S_{kit} | S_{kit} > 0) + P(S_{kit} \leq 0) \times 0 = P(S_{kit} > 0) \times E(S_{kit} | S_{kit} > 0)$.
We thus estimate the expectation of S_{kit} $(k = RFG, OXY)$ using all observations in the sample, by extending the Shonkwiler and Yen (1999) approach to the panel data case.

In particular, we make the following assumptions on the idiosyncratic error terms:
H1: $\upsilon_{it} \; | \mathbf{x_i}, \lambda_t, \alpha_i \sim \text{Normal}(0, \sigma_\upsilon^2), \; \forall i, t$

H2: $\zeta_{it} \; | \mathbf{w_i}, \delta_t, \tau_i \sim \text{Normal}(0, \sigma_\zeta^2), \; \forall \text{i}, t$

H3: $\varepsilon_{1kit} | \mathbf{z_{ki}}, \varpi_t, c_{1ki} \sim \text{Normal}(0, \sigma_{\varepsilon_{1k}}^2), \; \forall i, t, k = 1, 2$

H4: $\varepsilon_{2kit} | \mathbf{z_{ki}}, o_t, c_{2ki} \sim \text{Normal}(0, \sigma_{\varepsilon_{2k}}^2), \; \forall i, t, k = 1, 2$.

This gives:

$$
\begin{aligned}
E(S_{kit} | \mathbf{z_{kit}}, \varpi_t, c_{1ki}) &= P(S_{kit} > 0 | \mathbf{z_{kit}}, \varpi_t, c_{1ki}) \times E(S_{kit} | \mathbf{z_{kit}}, \varpi_t, c_{1ki}, S_{kit} > 0) \\
&= \Phi\left(\frac{\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki}}{\sigma_{1k}}\right) \left(\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki} + \sigma_{1k} \left[\frac{\phi\left((\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki}) / \sigma_{1k}\right)}{\Phi\left((\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki}) / \sigma_{1k}\right)}\right]\right) \\
&= \underbrace{\Phi\left(\frac{\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki}}{\sigma_{1k}}\right)}_{\text{(term 1)}} \underbrace{(\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki})}_{\text{(term 2)}} + \sigma_{1k} \underbrace{\phi\left((\mathbf{z'_{kit}} \boldsymbol{\gamma_{1k}} + \varpi_t + c_{1ki}) / \sigma_{1k}\right)}_{\text{(term 3)}}
\end{aligned}
\tag{10-11}
$$

with $\sigma_{1k} = \sqrt{\sigma_{\varepsilon_{1k}}^2 + \sigma_{c_{1k}}^2}$, $\phi(.)$ and $\Phi(.)$ being the standard normal density and probability distribution functions respectively. Similarly for I_{kit} $(k = RFG, OXY)$:

$$\begin{aligned}E\left(I_{kit}\left|\mathbf{z_{kit}},o_t,c_{2ki}\right.\right)&=P\left(I_{kit}>0\left|\mathbf{z_{kit}},o_t,c_{2ki}\right.\right)\times E\left(I_{kit}\left|\mathbf{z_{kit}},o_t,c_{2ki},I_{kit}>0\right.\right)\\&=\underbrace{\Phi\left(\frac{\mathbf{z'_{kit}}\boldsymbol{\gamma}_{\mathbf{2k}}+o_t+c_{2ki}}{\sigma_{2k}}\right)}_{\text{(term 1)}}\underbrace{\left(\mathbf{z'_{kit}}\boldsymbol{\gamma}_{\mathbf{2k}}+o_t+c_{2ki}\right)}_{\text{(term 2)}}+\sigma_{2k}\underbrace{\phi\left(\left(\mathbf{z'_{kit}}\boldsymbol{\gamma}_{\mathbf{2k}}+o_t+c_{2ki}\right)/\sigma_{2k}\right)}_{\text{(term 3)}}\end{aligned}$$

(12-13)

with $\sigma_{2k}=\sqrt{\sigma^2_{\varepsilon_{2k}}+\sigma^2_{c_{2k}}}$.

The estimation proceeds in two stages. In the first stage, we estimate four random effects Tobit models fitting S_{kit} and I_{kit} $(k=RFG,OXY)$. Estimates for the four random-effects Tobit models are not shown here but are available upon request. Estimated parameters are then used to compute term 1 and term 3 in equations (10-13). In the second stage, the system fitting simultaneously equations (4), (5) and equations (10)-(13) is estimated by Three Stage Least Squares (3SLS). In Table 3, we report the four estimates for σ, the coefficient associated with term 3, one each for the *S* and *I* indices for *RFG* and *OXY*.

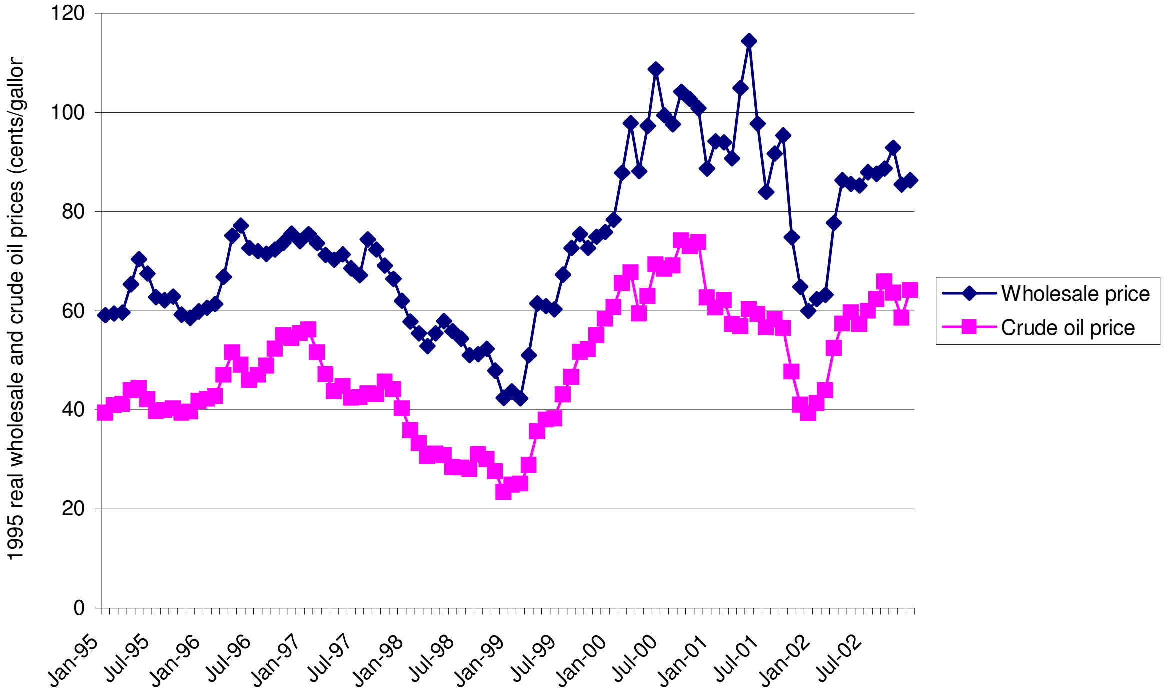

Figure 1. Monthly wholesale and crude oil prices (1995-2002), source: EIA

Table 1. Gasoline Market Statistics by State$^{(a)}$

State	Region or PADD$^{(b)}$	Population	Wholesale price (cents / gallon)	Number of refineries	Refinery capacity (barrels / day)	Refinery capacity per capita	Share of population under RFG program	Share of population under OXY program
Texas	3	20,357,661	65.36	23	4,154,651	0.204	0.43	0.02
Mississippi	3	2,807,667	65.63	3	337,050	0.120	0.00	0.00
Louisiana	3	4,440,333	65.78	17	2,539,488	0.572	0.00	0.00
Oklahoma	2	3,410,162	66.71	6	434,455	0.127	0.00	0.00
South	1	3,936,429	66.99	0	0	0.000	0.00	0.00
Tennessee	2	5,587,687	67.20	1	129,750	0.023	0.00	0.00
Georgia	1	7,953,055	67.31	2	33,470	0.004	0.00	0.00
North	1	7,858,645	67.36	0	0	0.000	0.00	0.00
Alabama	3	4,404,759	67.84	3	126,500	0.029	0.00	0.00
Arkansas	3	2,633,774	67.93	3	64,458	0.024	0.00	0.00
Florida	1	15,620,174	68.17	0	0	0.000	0.00	0.00
Kansas	2	2,662,544	69.04	3	291,694	0.110	0.00	0.00
Pennsylvania	1	12,260,058	69.48	5	713,425	0.058	0.41	0.02
New Jersey	1	8,329,169	69.59	5	641,000	0.077	0.98	0.28
Virginia	1	6,968,506	70.24	1	56,975	0.008	0.59	0.01
Indiana	2	6,016,794	70.42	2	435,275	0.072	0.11	0.00
Kentucky	2	3,996,711	70.81	2	224,875	0.056	0.26	0.00
Nebraska	2	1,697,541	70.83	0	0	0.000	0.00	0.00
Iowa	2	2,907,058	70.90	0	0	0.000	0.00	0.00
Missouri	2	5,536,208	71.03	0	0	0.000	0.16	0.00
Ohio	2	11,318,068	71.08	4	511,250	0.045	0.00	0.00
Michigan	2	9,875,224	71.50	2	102,650	0.010	0.00	0.00
Maine	1	1,266,183	71.58	0	0	0.000	0.32	0.00
Delaware	1	768,877	72.13	1	151,375	0.197	0.98	0.00
Wisconsin	2	5,316,610	72.18	1	33,800	0.006	0.34	0.00
West Virginia	1	1,812,900	72.19	1	13,188	0.007	0.00	0.00
New York	1	18,831,173	72.81	0	0	0.000	0.66	0.24
North Dakota	2	643,939	72.83	1	58,000	0.090	0.00	0.00
Illinois	2	12,311,089	72.96	5	989,443	0.080	0.63	0.00
South Dakota	2	749,500	73.02	0	0	0.000	0.00	0.00
Colorado	4	4,171,483	73.36	2	85,688	0.021	0.00	0.22
Rhode Island	1	1,039,361	73.42	0	0	0.000	0.97	0.00
New Mexico	3	1,794,611	73.67	3	94,975	0.053	0.00	0.10
Connecticut	1	3,383,624	74.33	0	0	0.000	0.98	0.13
Massachusetts	1	6,290,891	74.36	0	0	0.000	0.99	0.04
Maryland	1	5,244,375	74.43	0	0	0.000	0.88	0.03
Minnesota	2	4,845,224	75.19	2	320,900	0.066	0.00	0.51
Wyoming	4	491,486	75.35	5	133,099	0.271	0.00	0.00
New	1	1,215,568	75.87	0	0	0.000	0.59	0.00
Vermont	1	603,070	77.13	0	0	0.000	0.00	0.00
Utah	4	2,176,219	77.21	5	157,963	0.073	0.00	0.17
Oregon	5	3,363,534	77.44	0	0	0.000	0.00	0.17
Montana	4	895,108	77.68	4	154,679	0.173	0.00	0.03
Washington	5	5,788,980	78.38	6	587,869	0.102	0.00	0.07
Idaho	4	1,262,296	78.48	0	0	0.000	0.00	0.00
Arizona	5	4,949,269	79.21	0	475	0.000	0.39	0.24
Nevada	5	1,886,244	79.41	1	6250	0.003	0.00	0.35
California	5	33,301,897	80.58	17	1,944,123	0.058	0.92	0.26
District of Columbia	1	571,554	82.81	0	0	0.000	0.95	0.06
Alaska	5	621,972	92.19	4	314,816	0.505	0.00	0.18
Hawaii	5	1,215,297	100.72	2	147,500	0.121	0.00	0.00

Notes. (a) Averages for the period 1995-2002. (b) PADD: Petroleum Administration for Defense Districts. They are: East Coast (PADD 1), the Midwest (PADD 2), the Gulf Coast (PADD 3), the Rockies (PADD 4), and the West Coast (PADD 5).

Table 2. Descriptive Statistics by Year

	1995	1996	1997	1998	1999	2000	2001	2002
Refinery capacity (barrels per day)								
Mean	302,649	302,574	302,976	302,976	318,849	323,860	325,399	329,125
Std. Dev.	694,213	694,246	698,933	698,933	728,748	743,027	747,921	764,002
Min	0	0	0	0	0	0	0	0
Max	4,004,050	4,004,050	4,043,450	4,043,450	4,185,350	4,246,050	4,292,430	4,418,380
Refinery concentration index								
Mean	0.604	0.584	0.588	0.588	0.592	0.594	0.592	0.593
Std. Dev.	0.468	0.472	0.473	0.473	0.473	0.473	0.473	0.473
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Share of population under RFG								
Mean	0.245	0.244	0.245	0.251	0.245	0.246	0.246	0.246
Std. Dev.	0.372	0.370	0.369	0.372	0.368	0.369	0.369	0.369
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	0.986	0.986	0.996	0.996	0.996	0.996	0.996	0.996
Share of population under OXY								
Mean	0.086	0.076	0.060	0.057	0.055	0.054	0.054	0.051
Std. Dev.	0.131	0.130	0.119	0.114	0.113	0.113	0.113	0.110
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	0.510	0.556	0.556	0.556	0.556	0.556	0.556	0.556

Table 3. Three Stage Least Squares Estimation Results(a,b)

List of variables	Wholesale price of gasoline ($\log P_{it}$)	Refinery concentra-tion index (RC_{it})	Size of RFG market (S_{RFGit})	Size of OXY market (S_{OXYit})	Difference in RFG market size (I_{RFGit})	Difference in OXY market size (I_{OXYit})
	Coeff.(c) *Std. Err.*	Coeff. *Std. Err.*	Coeff. *Std. Err.*	Coeff. *Std. Err.*	Coeff. *Std. Err.*	Coeff. *Std. Err.*
Constant	-	-	-0.079	1.047*	-0.340	0.470
			0.218	*0.487*	*9.246*	*1.044*
Crude oil price (log)	0.625*	-0.312	-	-	-	-
	0.141	*0.225*				
Size of RFG market (S_{RFGit})	0.151*	-0.500*	-	-	-	-
	0.028	*0.063*				
Size of OXY market (S_{OXYit})	0.150*	-0.176	-	-	-	-
	0.052	*0.117*				
Index measuring difference in RFG market size (I_{RFGit})	0.135*	0.161	-	-	-	-
	0.037	*0.084*				
Index measuring difference in OXY market size (I_{OXYit})	0.137*	0.135	-	-	-	-
	0.058	*0.131*				
Refinery concentration index	0.136*	-	-	-	-	-
	0.038					
Refinery capacity per capita	3.097*	0.051	-	-	-	-
	0.942	*2.193*				
Average distance to refinery *(in thousand kms)*	0.755*	0.149	-	-	-	-
	0.202	*0.478*				
Refinery capacity per capita times average distance to refinery	-2.104*	0.318	-	-	-	-
	0.618	*1.442*				
Wholesale price (log)	-	0.481*	-	-	-	-
		0.210				
MTBE consumption	-	0.007*	-	-	-	-
		0.002				
Population density *(in thousands)*	-	0.016	-4.082*	1.761*	2.842	1.216*
		0.397	*0.940*	*0.241*	*1.520*	*0.488*
State population *(in millions)*	-	-	0.093*	0.009	-0.052	-0.048
			0.014	*0.019*	*0.088*	*0.036*
Population in non-attainment areas for particulate matter	-	-	0.722*	0.009	-0.093	-0.097
			0.273	*0.133*	*3.487*	*0.427*
Population in non-attainment areas for sulphur dioxide	-	-	0.212	-0.019	-0.118	-0.095
			0.122	*0.069*	*0.192*	*0.168*
Number of vehicles per capita	-	-	0.039	0.100	-0.033	-0.006
			0.173	*0.110*	*0.922*	*0.086*
Net receipts / net production of gasoline by PADD	-	-	0.562*	-0.301	-0.033	-0.037
			0.083	*0.166*	*1.478*	*0.231*
σ (d)	-	-	1.305*	-8.442*	-3.372	-19.941*
			0.647	*1.731*	*5.240*	*5.712*

Notes. (a) Alaska and Hawaii are not included in the sample. Total number of observations is 384, for 48 states over 8 years. (b) 8 year-specific and 48 state-specific effects have been included in all six equations but are not shown here. (c) * indicates parameter significance at the 5 percent level. (d) This is the parameter associated with the Normal density function evaluated at predicted values of *S* and *I* for *RFG* and *OXY* (see term 3 in equations (10-13) in Appendix B). Data sources: Energy Information Administration, Environmental Protection Agency, US Census Bureau, and US Federal Highway Administration (see Appendix A).

Table 4. Diagnostic Tests

Endogenous regressors (Durbin-Wu-Hausman test)		
Null: Regressors are endogenous		
Price equation		
Test of endogeneity of $RC_{it}, S_{kit}, I_{kit}$ *with k = RFG, OXY*	Chi-square = 83.89	*p-value = 0.000*
Refinery concentration index equation		
Test of endogeneity of P_{it}, S_{kit}, I_{kit} *with k = RFG, OXY*	Chi-square = 28.82	*p-value = 0.000*
Underidentification (Anderson canonical correlation Likelihood Ratio Test)		
Null: Model is underidentified		
Price equation	Chi-square = 29.15	*p-value = 0.000*
Refinery concentration index equation	Chi-square = 7.86	*p-value = 0.097*
Instrument validity$^{(a)}$ (Sargan test for overidentifying restrictions)		
Null: Instruments are valid		
Price equation	Chi-square = 4.292	*p-value = 0.508*
Refinery concentration index equation	Chi-square = 5.212	*p-value = 0.157*

Notes. (a) Instruments are price of crude oil, refinery capacity per capita, average distance to refinery, the interaction term between refinery capacity per capita and average distance to refinery, state population, state population density, MTBE consumption, share of population classified in non-attainment areas for particulate matter and sulphur dioxide, average number of vehicles per capita, ratio of net receipts over net production of gasoline for the corresponding region or PADD, and dummies for year, PADD and state.

Table 5. Wholesale Price Equation: 3SLS and OLS Estimation Results[(a,b)]

	3SLS estimation (Table 3, column 1)		**OLS estimation**	
	Coeff.[(c)]	*Std. Err.*	Coeff.	*Std. Err.*
Crude oil price (log)	0.625*	*0.141*	-	-
Size of RFG market $\left(S_{RFGit}\right)$	0.151*	*0.028*	0.087*	*0.022*
Size of OXY market $\left(S_{OXYit}\right)$	0.150*	*0.052*	0.112*	*0.039*
Index measuring difference in RFG market size $\left(I_{RFGit}\right)$	0.135*	*0.037*	0.051*	*0.023*
Index measuring difference in OXY market size $\left(I_{OXYit}\right)$	0.137*	*0.058*	0.078*	*0.039*
Refinery concentration index	0.136*	*0.038*	0.060*	*0.025*
Refinery capacity per capita	3.097*	*0.942*	3.256*	*1.022*
Average distance to refinery *(in thousands kms)*	0.755*	*0.202*	0.831*	*0.217*
Refinery capacity per capita x average distance to refinery	-2.104*	*0.618*	-2.183*	*0.669*
Year 1995 Dummy - *reference*	-	-	-	-
Year 1996 Dummy	0.023	*0.026*	0.136*	*0.005*
Year 1997 Dummy	0.059*	*0.013*	0.121*	*0.005*
Year 1998 Dummy	0.024	*0.047*	-0.178*	*0.005*
Year 1999 Dummy	0.019*	*0.006*	0.002	*0.005*
Year 2000 Dummy	0.131	*0.068*	0.440*	*0.005*
Year 2001 Dummy	0.168*	*0.038*	0.342*	*0.005*
Year 2002 Dummy	0.077	*0.044*	0.281*	*0.005*

Notes. (a) Alaska and Hawaii are not included in the sample. Total number of observations is 384, for 48 states over 8 years. (b) 48 state-specific effects have been included in the models but are not shown here. (c) * indicates parameter significance at the 5 percent level. Data sources: Energy Information Administration, Environmental Protection Agency, US Census bureau, and US Federal Highway Administration (see Appendix A).