# Consumer Search and Dynamic Price Dispersion: An Application to Gasoline Markets[*]

Ambarish Chandra  Mariano Tappata[†]

*July 7, 2010*

**Abstract**

This paper studies the role of imperfect information in explaining market price dispersion. We claim the need for an alternative to the common tests used to identify consumer search, by pointing out the ambiguous predictions from search theory. Using a new panel dataset on the U.S. retail gasoline industry, we establish the importance of consumer search with a simple and novel test of temporal price dispersion involving price-spreads between pairs of gasoline stations in a given market. We exploit the cross-sectional variation in our dataset to establish the equilibrium relationships between price dispersion and key variables from consumer search models. We find that price dispersion increases with search costs, increases with the number of firms in the market and decreases with the production cost.

## 1 Introduction

Retail markets generally exhibit price dispersion regardless of the attributes of the products transacted. Establishing conditions under which firms will choose to set a range of prices has been a central and classic question in price theory. Beginning with Stigler (1961), the literature has acknowledged the role of imperfect information in generating equilibrium price dispersion. This literature on search posits that markets consist of consumers who acquire information by actively searching for lower prices, as well as consumers who remain uninformed as they prefer to avoid search costs. This behavior is what allows some firms to set higher prices than others in equilibrium, even when all firms sell a homogenous good and have identical production costs.[1]

---

[*] We thank Dan Ackerberg, Jim Brander, Keith Head, Thomas Hellmann, Tim Hannan, Phil Leslie, Alan Sorensen, Hal White and various seminar and conference participants for their comments and suggestions.

[†] Sauder School of Business, University of British Columbia. Contact: ambarish.chandra@sauder.ubc.ca and mariano.tappata@sauder.ubc.ca

[1] See Stahl (1996) and Baye et al. (2006) for a comprehensive literature review on consumer search theory.

Establishing evidence that price dispersion results from costly consumer search has been challenging. Empirical studies have mostly relied on a single comparative static to identify the role of search in a given market. Some studies regress price dispersion on proxies of search costs or proxies of search benefits; others regress price and price dispersion on the number of firms in the market.[2] However, a drawback of these tests is that the relationships tested are not monotone, as we show in the next section, and so, in principle, any result is consistent with search. For example, the theoretical relationship between search intensity and price dispersion is non-monotonic; yet it has usually been assumed to be negative and results that find a negative relationship are taken as validation of the role of search. A better test of consumer search comes from temporal price dispersion, i.e. changes in the ranking of firms' prices over time. Many prior studies do suggest the existence of temporal price dispersion with transition matrices that show prices jumping from one quartile of the distribution to another over time; however, no formal test is provided.[3] Moreover, temporal price dispersion can also be generated in a product differentiation environment with idiosyncratic demand or cost shocks; the studies mentioned above do not have a control group or benchmark with which to compare the observed temporal price dispersion.[4]

To address these issues we propose a simple yet powerful test of information frictions in the gasoline industry, comparing the relative prices over time between stations for which imperfect information may possibly play a role, to a control group of stations where imperfect information is absent. In this way we can identify the role of consumer search without making assumptions on the relation between search and price dispersion or other comparative statics. We use a new and extremely rich panel dataset, providing daily retail prices for US gasoline stations. We exploit the cross-sectional dimension of the data to characterize the equilibrium relationships of a model of search. This model emphasizes that key variables do not necessarily have a monotonic effect on price dispersion and average posted prices; something that has not been made explicit in previous studies.[5]

There is a large empirical literature that links price dispersion with consumer search behavior in many industries. Van Hoomissen (1988) finds that — consistent with costly search and repeated purchases — price dispersion and inflation are positively related. Higher inflation reduces the incentives to search since information depreciates at a higher rate. Sorensen (2000) follows a similar approach although the identification comes from comparing

---

[2] See, respectively, Dahlby and West (1986) and Brown and Goolsbee (2002); Sorensen (2000) and Van Hoomissen (1988); and Barron et al (2004).

[3] See, for example, Lach (2002).

[4] These issues are exacerbated due to the absence in the literature of a general theoretical framework which incorporates the interaction of product differentiation and imperfect information. We return to this point in Section 4.

[5] One exception is Brown and Goolsbee (2002).

price dispersion across products rather than across time. The main finding is that the price dispersion for a prescription drug is negatively correlated with its associated frequency of use. An alternative approach is to focus on the relationship between price dispersion and consumers' search cost. Dahlby and West (1986) show that car insurance premiums are less dispersed for the class of drivers that are associated with lower search costs. Hortaçsu and Syverson (2004) use a structural model to estimate the search cost distribution in the S&P500 index fund market. They find that the increase in fees and dispersion during the late 1990s is explained by the entry of novice investors (with higher search costs) to the market.[6] Brown and Goolsbee (2002) show how the reduction in searching costs due to the introduction of the Internet made the life insurance market more competitive.[7]

The gasoline market is an appealing industry in which to study the role of search for two main reasons.[8] First, the phenomenon of gas stations prominently displaying their prices allows us to use a simple test involving price spreads that controls for imperfect information. In particular, we find that price dispersion *over time* is significantly lower for stations located at the same street intersection than for stations further apart although still in the same local market. This is consistent with the predictions of search models since, for stations in the same intersection, differences in prices are driven only by product characteristics and not by imperfect information. Second, since gas stations sell more than one fuel-type we can compare price dispersion across products that have varying search costs, but are still in the same market. This allows us to pin down the equilibrium relationship between price dispersion and search intensity which, in theory, is non-monotonic and therefore hard to estimate. There are other reasons that make the gasoline market appropriate for empirical research in this area, essentially due to the fit between industry characteristics and the assumptions of consumer search theories : *i)* demand is inelastic in the short run and similar across consumers, *ii)* stockpiling is not a feasible option, *iii)* firms face a fairly homogeneous marginal cost, and *iv)* have no capacity constraints.[9]

In addition to the contributions to the literature on search and price dispersion, our

---

[6] Both Dahlby and West (1986), and Hortaçsu and Syverson (2004) assume a model where the price dispersion is generated by the combination of production cost dispersion and costly consumer search (Carlson and McAfee, 1983). While this is a reasonable assumption for the industries analyzed, it is important to note that identifying consumer search as responsible for price dispersion is a more difficult task given that the theory predicts no temporal dispersion.

[7] Other studies of online markets include Clay et al. (2001), Smith and Brynjolfsson (2001) and Baye et al. (2004).

[8] This is also an important industry for economic reasons. Gasoline retailing is a large and growing sector within the U.S. economy. According to the Census Bureau, retail sales of gasoline reached $241.9 billion in 2002 (www.census.gov). Additionally, gasoline's share of total consumer expenditures rose by 43% to 4.3% between 2002 and 2005 (www.bls.gov/cex/).

[9] See Borenstein *et al.* (1997, p. 328) and Noel (2007b, p. 88).

paper also adds to the empirical literature on price dispersion in gasoline industries.[10] We employ new data which provide key advantages over prior datasets. First, we have daily data on gasoline prices; this is important since existing studies rely on weekly observations, which are not ideal for a study of temporal price dispersion. Second, we have data on 25,000 gas stations in four large states, thereby employing more representative data than samples from a single city or region. Our data cover all grades of gasoline and span over 18 months during which there were large changes in the wholesale price of gasoline. This allows us to test relationships, such as between price dispersion and production costs and search costs, that have not been examined thus far.

We find that fuel-types associated with higher consumer search cost exhibit higher equilibrium levels of price dispersion, suggesting relatively high levels of search intensity (low information rents). Markets with more firms show greater price dispersion and higher average markups. Interestingly, we find that consumers should search less when pump prices increase as a result of shocks to the wholesale price. A spike in the oil price that translates into a 10 cent increase in retail prices is associated with up to a 4 cent reduction in the market price dispersion for Regular Unleaded. Our results are strengthened when we use as the dependent variable the price dispersion that remains after controlling for sellers' observed and unobserved characteristics.

This paper is organized as follows. In the next section we present a simple model of consumer search and summarize its main empirical predictions. Section 3 describes the industry and the data. We then move on to our empirical exercises in two parts. In Section 4 we analyze temporal price dispersion and establish the importance of consumer search in the gasoline market. Then in Section 5, having affirmed the role of search in generating price dispersion, we analyze price dispersion across markets, estimate the equilibrium relationships between key variables and compare them to the predictions of our model of search. Section 6 concludes.

## 2 Predictions of a consumer search model

In this section we present a simple model of consumer search and establish its equilibrium properties. We do not attempt to model the gasoline industry; rather, our goal is to discuss alternative strategies for the identification of consumer search from price dispersion data. The main message is that many of the comparative statics from models with fixed search

---

[10] See, for example, Barron et al (2004), Hosken et al (2008), Lewis (2008) and Lach and Moraga-Gonzalez (2009).

intensity change and become non-monotonic once we allow for endogenous search.[11] This suggests that rather than relying on testing the usual comparative statics, the importance of consumer search should be inferred from more general results, such as temporal price dispersion, that characterize search models.

The literature on consumer search has been refined and extended in many directions since Stigler's (1961) seminal work. However, its main result still prevails: price dispersion is a natural outcome in markets where consumers have imperfect information. The intuition is that when prices are not freely observed, some consumers decide to search while others choose to remain uninformed. Given this behavior, a firm can set a high price and sell to a fraction of uninformed consumers or have a *sale* and increase the probability of higher demand by the mass of informed consumers. In equilibrium, the firms are indifferent between a broad range of prices and follow mixed strategies.

Price dispersion is the outcome of search models irrespective of the search protocol assumed for consumers.[12,13] We focus on the case of nonsequential search and discuss a dynamic model similar to Tappata (2009).[14] Its simplicity allows us to focus on the effect of heterogeneous search costs on price dispersion, and examine how the market search intensity adjusts to changes in the parameters of the model. A limitation, shared with most search models, is that it abstracts away from the interaction between imperfect information and product differentiation. As will be argued in Section 4, our only assumption on this point is that the main equilibrium property, i.e. mixed strategies, is present when products are heterogeneous. Additionally, we avoid the common assumption that product characteristics imply a shift in prices and hence that the effects of product differentiation can be removed by demeaning observed prices with firm fixed-effects. We now present the model, its main predictions, and the implications for identifying search empirically.[15]

Assume a homogeneous good market with $n$ firms that compete on prices and have the same constant unit production cost $c$. The demand side is characterized by a unit mass of

---

[11] Nearly all current models of consumer search incorporate consumers who balance the cost of search and its expected benefits. In this section we focus on the empirical predictions and challenges for applied work.

[12] Most of the search protocols analyzed in the literature are variations of sequential (Stahl, 1989) and nonsequential (Varian, 1980) models. For examples, see Dana (1994), Burdett and Judd (1983), Armstrong et al. (2009), Janssen and Moraga-Gonzalez (2004) and Janssen *et al.* (2005).

[13] The optimality of each protocol depends on the characteristics of the market analyzed and the search technology available to consumers (Morgan and Manning, 1985). In the case of retail gasoline, nonsequential and sequential search are likely to be the appropriate protocols for commuters and tourists respectively.

[14] The empirical predictions from sequential search models are very similar to the nonsequential case and we describe them in a supplementary note to this paper which is available from the authors' webpage.

[15] See Tappata (2009) for a formal analysis of the comparative statics.

consumers that have inelastic demands with valuations $v$.[16] A fraction $\lambda \in (0,1)$ of them have zero search cost and are called shoppers while the rest face positive — and different — search costs. Nonshoppers decide, before observing any price, between paying the search cost to know all the market prices or remaining ignorant and buying from a random store. Two conditions need to be satisfied in equilibrium: *i)* For any given search intensity $\mu \in [\lambda, 1]$, the pricing strategies of the firms must be a Nash Equilibrium, and *ii)* the search intensity in the market has to be consistent with the firms' pricing strategies. That is, when consumers compare the cost of search with the benefits of search, they correctly anticipate the firms' pricing strategies.

Varian (1980) showed that, given a proportion of informed consumers $\mu$, there is a unique symmetric NE that involves firms playing mixed strategies. In each period, firms simultaneously draw prices from

$$F\left(p;c,v,n\right) = 1 - \left[\frac{\left(1-\mu\right)}{\mu n}\frac{\left(v-p\right)}{\left(p-c\right)}\right]^{1/(n-1)} \tag{1}$$

where $p \in [p^* = \frac{cn\mu+(1-\mu)v}{1+(n-1)\mu}, v]$. For prices below $p^*\left(c,v,n\right)$, a firm always prefers to charge a monopoly price and sell to $\left(1-\mu\right)/n$ consumers. The amount of search is directly related to the intensity of price competition. As the number of informed consumers increases, the domain of the price distribution increases and in the limit, the entire distribution collapses to the marginal cost (competitive outcome). At the other extreme, when no consumer searches, each firm becomes a monopolist over $1/n$ consumers and the domain collapses to $p = v$ (monopoly outcome). As Varian pointed out, the main implication of imperfect information is that prices are dispersed across sellers *and* across time. The latter is due to the fact that the static game is repeated over time. Firms draw new prices every period to avoid rival firms being able to systematically undercut a fixed price (Baye, Morgan and Scholten, 2004). At the same time, 'hit and run' pricing strategies keep consumers guessing as to which firms have sales in a particular period.

The second requirement for an equilibrium is that the amount of search in the market needs to be consistent with the firms' pricing strategies. That is, each consumer calculates the gains from search ($GS$) anticipating (1) and taking the market search intensity as given:

$$GS = E\left[p - p_{\min}|\mu;c,v,n\right] = \int_{p^*}^{v} p\left[1 - n\left[1-F\left(p;c,v,n\right)\right]^{n-1}\right] dF\left(p;c,v,n\right) \tag{2}$$

---

[16] This is a simplifying assumption and the results in this section hold for a large set of downward sloping demand functions.

It is easy to see that $GS$ is a non-monotonic function of the search intensity. There is no point in searching in the monopoly and competitive cases ($\mu = 0$ and $\mu = 1$). In fact, the gains from search are low when very few or too many consumers decide to search, but are greater when the search intensity takes intermediate values. Consumers draw their search cost from a continuous distribution $G(s_i)$ with $s_i \in [0, \overline{s}]$. In equilibrium, the indifferent consumer has a search cost $\widetilde{s}$ such that $GS = \widetilde{s}$ and the search intensity becomes $\mu = \lambda + G(\widetilde{s})(1-\lambda)$. Figure 1.a shows the equilibrium when search costs are drawn from a Beta distribution.[17] The proportion of informed consumers is measured on the horizontal axis, while the search costs and gains from search are on the vertical axis. The concave curves represent the gains from search to consumers in markets with 2, 5 and 10 firms. The upward sloping curve represents the search cost of the marginal consumer who decides to search. There is a unique equilibrium represented by the intersection of the two curves. It can be seen that as the number of firms increases in the market, the search cost of the indifferent consumer is higher, implying a greater search intensity in more atomistic markets.

The traditional measures of price dispersion are plotted as a function of the search intensity in Figure 1.c. Both the sample range (SR) and the standard deviation (SD) resemble the shape of the GS and show that dispersion measures alone cannot be used to predict the search intensity or level of competition in a market.[18] To see this, define $\widehat{\mu}$ as the search intensity that maximizes GS for a given production cost and number of firms.[19] The relationship between the search intensity and price dispersion is positive when $\mu < \widehat{\mu}$ and negative if $\mu > \widehat{\mu}$. Therefore, comparing the price dispersion for products with different search costs does not answer the question of whether the dispersion is consistent with consumer search, since any relation — positive, negative or even zero — between consumers' search costs and price dispersion will be consistent with a model of consumer search. Instead, this comparison answers the following question: Given that price dispersion is generated by consumer search, is the equilibrium search intensity consistent with prices closer to the competitive level or closer to the monopoly level? That is, by examining the price dispersion for products with different search costs, we can identify whether we are at the point in the relationship where increases in the search intensity increase price dispersion ($\mu < \widehat{\mu}$) or decrease price dispersion ($\mu > \widehat{\mu}$). We carry out such an exercise in Section 5.

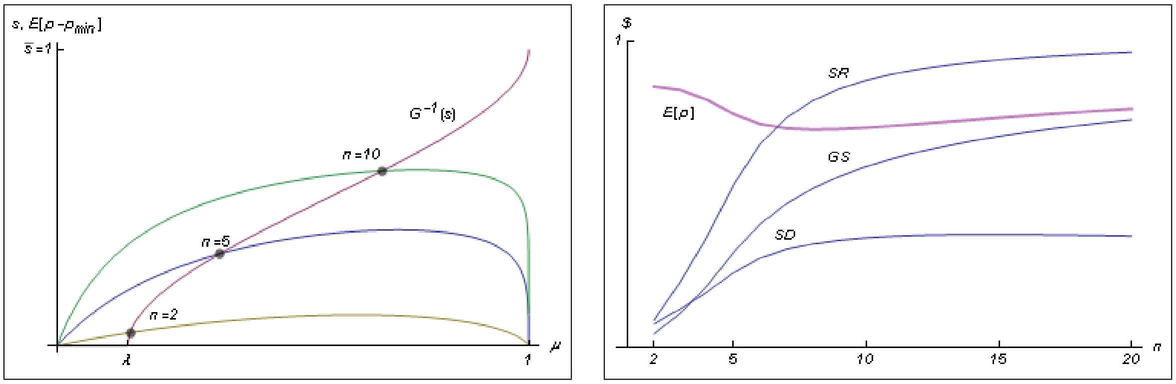
We now analyze how changes in the number of firms, production cost and consumer

---

[17] The parameter values used are $\lambda = 0.15, v = 1, c = 0, G(s) = I_s[2,2]$ where $I_s$ is the regularized incomplete beta function.

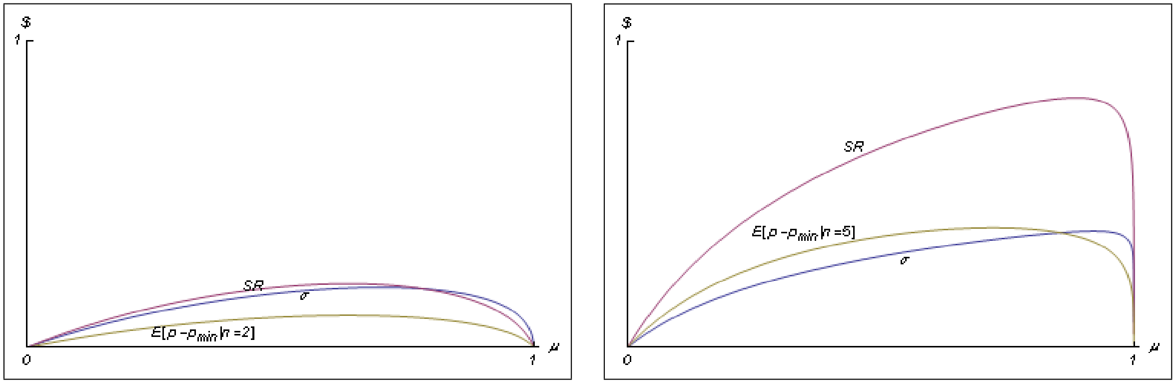
[18] We use the term price dispersion to refer to GS, SR and SD. The predictions of this section involve GS although they also hold for SR and SD.

[19] We show this visually in Figure 6 in Section 5.



a) Search intensity and the number of firms

b) Expected price and price dispersion



c) Price dispersion and search intensity ($n = 2$ and $n = 5$)

Figure 1: Expected Prices and Price Dispersion in Equilibrium

search cost affect the equilibrium price dispersion and price levels. To examine the effect of entry and exit, assume first that the search intensity is fixed ($\mu = \overline{\mu}$). Then, as the number of firms in the market increases, the expected profit of a seller changes in two ways. First, the fraction of captive uninformed consumers for each firm decreases at a rate $1/n$. Second, the probability of being the lowest price in the market decreases at an exponential rate. These two effects imply that firms become more likely to set extreme prices at the expense of middle-range prices. All else constant, the price dispersion increases with $n$ (Figure 1.a). Moreover, since the gains from setting low prices decrease at a faster rate than the gains from setting high prices, the price distribution shifts toward higher prices.

But the complete effect of the number of firms on price levels and dispersion needs to incorporate the reaction by rational consumers. Since, for a given $\mu$, price dispersion increases with $n$, the marginal consumer will have a greater search cost in markets with more firms. When $\mu < \widehat{\mu}$, the higher search intensity strengthens the partial (and positive) effect of the number of firms on the price dispersion. When $\mu > \widehat{\mu}$, the higher search intensity reduces this partial effect although it never offsets it.[20] However, the total effect on the average posted price cannot be signed since more consumers searching push prices down. As shown in Figure 1.b, the relationship between the average price and the number of firms is not monotonic. It is important to emphasize this result since previous work assumed that a unique testable implication of search models is a negative relationship between average prices and market concentration (Barron et al, 2004; and Lewis, 2008). Figure 1.b also shows that the two alternative measures of price dispersion increase at a decreasing rate with $n$.

The comparative statics with respect to the production cost are straightforward. Holding the search intensity constant, as the cost of production increases, the gap between the monopoly price and the minimum profitable price $p^*$ decreases and firms set higher but less dispersed prices (the extreme case being $c = p^* = v$). Given the search intensity, the cost pass-through is lower than 100% and increases with $\mu$. Figure 2 shows — for any given level of search intensity — the negative effect of production cost on GS and markups. With endogenous search, the response by consumers to an increase in production cost is to search less. Thus, the new equilibrium involves higher and less dispersed prices. The final effect on markups depends on the magnitude of the search intensity adjustment. In general, it is expected to be negative but it is possible that a large reduction in consumer search generates an equilibrium pass-through greater than 100%.

The last comparative static is about changes in the search cost. Higher search costs can

---

[20] Note that the number of firms in the market has no effect on consumers' search costs in a model of non-sequential search, though it may have an effect in a model of sequential search.

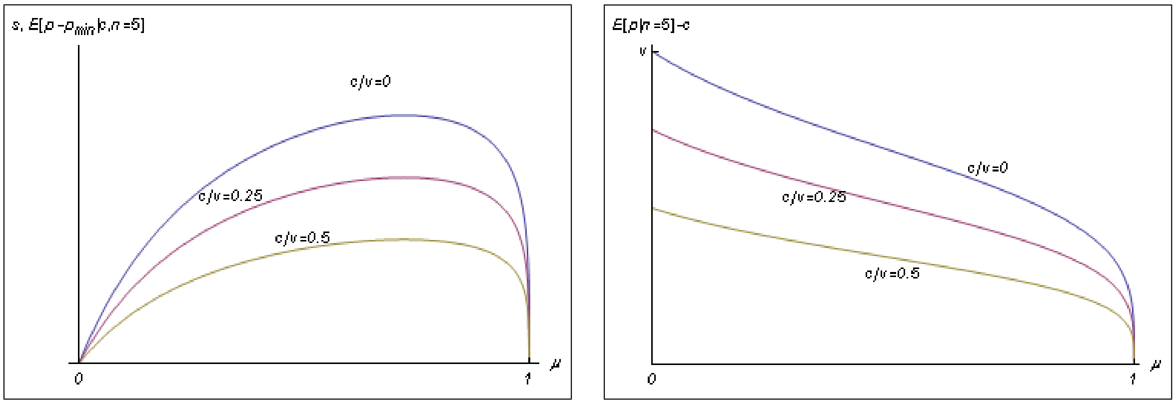


Figure 2: Production Cost, Gains from Search and Markups

be thought of as a decrease in the number of shoppers or as a change in the distribution of search costs. In terms of the curves in Fig 1a, a decrease in search costs implies either a shift or a rotation of $G^{-1}(\cdot)$ to the right, arising from either a change in its intercept or in its slope. In both cases it is easy to see that the equilibrium search intensity is reduced and therefore prices increase. But, as mentioned above, the price dispersion will increase or decrease depending on the initial equilibrium ($\mu \lesseqgtr \widehat{\mu}$). In Section 5 we use the price dispersion observed in the data for each fuel-type to pin down the equilibrium relationship between search intensity and price dispersion. Since gas stations sell three fuel-types that are associated with different search costs, we can obtain the effect of search cost on price dispersion controlling for all other possible factors that affect pricing decisions (see Figure 6).

Table 1 summarizes the comparative statics predicted by our model.[21] The entries in columns 2 and 3 show the qualitative and *partial* effect of an increase in the parameters $n$ and $c$ on the expected price, price dispersion and markup. The market search intensity is held constant so these results are driven entirely by changes in the firms' pricing strategies. As discussed earlier, some of the comparative statics become non-monotonic once we allow consumers to adjust their search strategies to changes in the number of firms or production costs (columns 4 and 5). The last two columns show the effect of an increase in the search cost (a change in $G$ or a reduction in the number of shoppers).

The stylized model in this section leads us to claim that identification of information frictions in the market cannot be based on estimating the relationships between the number of firms and prices, or between search costs/benefits and price dispersion. We propose simple and more general tests to assess the importance of search in the gasoline market.

---

[21] See the supplementary notes to this paper for a similar table with the predictions in a sequential search environment.

|  | $\overline{\mu}$ |  | $\mu$ |  |  |  |
|---|---|---|---|---|---|---|
|  | $n$ | $c$ | $n$ | $c$ | $s$ | $-\lambda$ |
| $E[p]$ | $+$ | $+$ | $-/+$ | $+$ | $+$ | $+$ |
| $PD$ | $+$ | $-$ | $+$ | $-$ | $-/+$ | $-/+$ |
| $E[p]-c$ | $+$ | $-$ | $-/+$ | $-/+$ | $+$ | $+$ |

PD=E[p-$\mathrm{p_{min}}$] (price dispersion). $\overline{\mu}$ and $\mu$ refer to fixed and endogenous search intensity, respectively.

Table 1: Comparative statics

The tests are related to temporal price dispersion and look at changes in relative prices over time since mixed strategies are a feature common to most consumer search models. We shall return, in Section 5, to analyzing the comparative static predictions of the model, when we compare them to estimated results in the gasoline industry.

## 3 Data description

In this section we describe the dataset that we use for the empirical analysis and present some descriptive statistics which help to understand the scope of the data and the nature of the retail gasoline market. Our dataset is unique in the sense that it covers more cross-sectional observations and a higher temporal frequency than the data used in other studies of this nature. We obtained daily gasoline prices for virtually every gas station in the states of California, Florida, Texas and New Jersey. Moreover, our sample time period stretches for almost 18 months (January 2006- May 2007).[22]

The data were originally collected by OPIS (Oil Price Information Service) and are widely available through various commercial and other organizations. OPIS provides daily service station level data for up to 120,000 stations across the US, which translates to more than 25,000 stations in the four states that we analyze.[23,24] The prices are obtained from "reconciled credit card transactions, direct feeds of data and other survey methods" (opisnet.com). The data are from all kinds of service stations: company owned, jobber owned, or independently owned. We have data on all three grades of unleaded gasoline — Regular, Mid-grade and Premium — as well as Diesel, although not every station sells all fuel-types or necessarily reports a price on each day for all fuel-types. Each observation is a

---

[22] There are some breaks in this period; however, we have data for over 400 individual days in all states.

[23] California, Florida and Texas have the greatest number of observations among all the states in the OPIS dataset while New Jersey provides geographic balance to our sample.

[24] The Census Bureau reports a total of 28,153 stations for the four states in 2002, a 5% decrease from the 1997 census, and 5% more stations than in our dataset for 2006-2007.

station-date-fuel-type triple. We dropped some observations which could not be geocoded; either because their addresses were ambiguous, or because the geocoding software could not find a match with a high enough degree of accuracy.

A second dataset includes weekday spot prices from the Energy Information Administration (EIA) for the ports relevant to the states that we analyze: Los Angeles Harbor, New York Harbor and the Gulf Coast. Depending on the vertical contract with the refinery, a gas station buys its gasoline in the wholesale market at the rack price or obtains it directly from the refinery at the Dealer Tank Wagon (DTW) price which is private and includes delivery to the station. Since we do not observe data on rack or DTW prices we use the spot price as a proxy for the shifts in the wholesale cost faced by stations. Figure 7 in the Appendix plots monthly spot, rack and DTW prices and shows that they are almost perfectly correlated. In particular, the spot and rack prices behave very similarly (average spread is less than 1 cent and the correlation is above 0.99).[25]

Figure 3 shows the variation in prices over time by plotting the price series for our sample of Regular Unleaded prices for California, Texas and Florida, along with a weekly price series for California obtained from the EIA.[26] As is widely known, gasoline prices peak in the summer months and reach their lowest point around January. Also, our sample of prices for California very closely tracks the official price average for that state, which provides reassurance that our sample is representative. The figures for the other states also match the official averages very closely, and are not presented here.

We now present summary statistics on our dataset. Table 2 contains means at the station level, separately for each of the four states. The first panel shows the raw price data. Texas and New Jersey are the cheapest states for gasoline, while California is the most expensive. Variation in gasoline prices across states is due to state and local taxes, varying regulatory standards, and variation in the spot prices of gasoline and diesel. On average, Mid-grade gasoline is about 8 cents more expensive per gallon than Regular, while Premium is about 12 to 18 cents more expensive. Diesel appears to be closest to Premium gasoline in its price level.

Data on raw prices are not very meaningful due to considerable variation across time in the price of crude oil. Therefore we also present data on the 'markup' which is defined here as the retail price minus the corresponding spot price on that date.[27] This measure exhibits

---

[25] EIA collects DTW and rack prices through surveys. Average values can be downloaded from their website.

[26] http://tonto.eia.doe.gov/dnav/pet/pet_pri_gnd_a_epmr_pte_cpgal_w.htm

[27] We compute this measure for Regular Unleaded and Diesel only; the spot price of the higher grade gasoline fuels is the same as for Regular. The markup is not intended to measure the actual profit per gallon for retailers but rather the variability of retail prices net of the wholesale cost volatility.

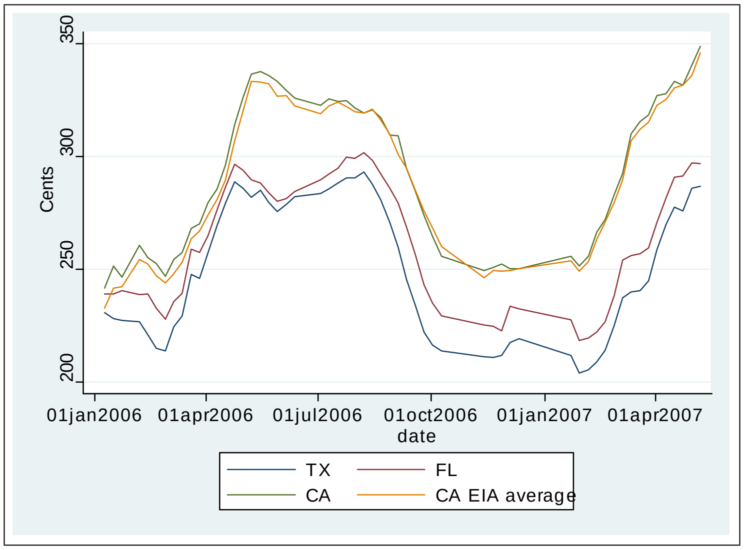


Figure 3: Weekly retail gasoline price series

| | California | | Florida | | New Jersey | | Texas | |
|---|---|---|---|---|---|---|---|---|
| Prices (cents/gal): | | | | | | | | |
| Regular | 301.9 | (33.4) | 269.2 | (26.8) | 258.1 | (33.1) | 257.6 | (29.9) |
| Mid-grade | 309.4 | (34.2) | 278.6 | (27) | 264.9 | (33.2) | 265.6 | (30.5) |
| Premium | 314.0 | (35.6) | 286.5 | (27.2) | 269.5 | (32.7) | 271.3 | (30.6) |
| Diesel | 311.9 | (22.4) | 284.2 | (18) | 270.6 | (20.9) | 272.1 | (19) |
| Markups (cents/gal): | | | | | | | | |
| Regular | 85.2 | (21) | 75.5 | (15.9) | 64.7 | (19.9) | 64.4 | (15.1) |
| Diesel | 100.5 | (16.4) | 85.3 | (14.6) | 72.0 | (16.6) | 73.7 | (14.2) |
| Number of rivals (all stations): | | | | | | | | |
| Within 1 mile | 4.62 | (3.2) | 4.59 | (3.5) | 4.25 | (3.5) | 4.74 | (3.5) |
| Within 2 miles | 13.45 | (8.4) | 13.59 | (9.9) | 13.38 | (10.1) | 14.22 | (10.4) |
| Dist. to closest rival (mi) | 0.38 | (1.13) | 0.4 | (0.94) | 0.38 | (0.62) | 0.49 | (1.38) |
| Dist. to closest same-brand rival (mi) | 2.88 | (3.79) | 2.94 | (3.6) | 2.56 | (2.75) | 3.28 | (4.51) |
| Number of Stations Observed: | | | | | | | | |
| Diesel | 3345 | | 3039 | | 928 | | 5909 | |
| Regular | 7396 | | 7004 | | 2233 | | 9856 | |

Note: Standard deviations in parentheses. Markups refer to retail price less spot price on that day.

Table 2: Station level Summary Statistics

considerably less variation. As noted above, this value includes taxes and other state and county specific price differences. Nevertheless, it provides a better picture of the variation in prices according to fuel-types and states.

Table 2 also contains data on the number of rival gas stations that surround a given station. For example, the average gas station in Texas has 4.74 other stations within a 1 mile radius. Despite the differences in the price levels, the station density in each market is similar across states. Appendix Table 10 shows the distribution of gasoline brands in the dataset. There is considerable variation across states in the shares of various brands. However the larger brands, such as Shell, Chevron and Citgo are observed across all states and unbranded stations account for between 6 and 10 percent of the observed stations.[28]

Our empirical strategy in Section 5 will rely on estimating the effect on price dispersion of varying market conditions, as well as varying levels of production and search costs. In order to accurately represent the competitive environment in this industry, we define each gas station as lying in a unique market. This comprises the station itself, plus all the stations that lie within a certain radius. This implies that each station will be counted as being part of many distinct markets. A similar approach has been taken in other work in this industry.[29,30]

To study price dispersion in a given market we calculate three statistics: The standard deviation of prices reported by these stations, the range of prices (i.e. maximum price minus minimum price), and the gains from search in these markets. The last statistic is defined in Equation 2. Table 3 contains summary statistics of price dispersion at the market level. The radius used to define markets in Table 3 was 1 mile; in Appendix Table 9 we present summary statistics on markets defined according to a 2-mile radius. Table 3 presents summary measures separately for Regular Unleaded, Mid-grade, Premium and Diesel Stations. Note that the price dispersion measures for Premium gasoline are higher than for Mid-grade, which in turn are higher than for Regular. It appears that, for all 3 grades of gasoline, markets in California have the highest price dispersion among the 4 states, while for Diesel, New Jersey appears to have the greatest dispersion. The same holds true for markets defined using a 2 mile radius. The price dispersion measures imply,

---

[28] This is consistent with other studies; Verlinda (2007), for example, reports around 9% of the stations population in Orange County as being unbranded.

[29] See Shepard (1991), Hastings (2004) and Lewis (2008).

[30] Although the frequency of our gasoline pricing data is at the daily level, we do not observe prices posted by every station on every date. To avoid making assumptions regarding missing data, we define markets in the following way: A gas station reporting a price on a given date is the center of a potential market. The entire market will consist of that station on that date, along with all the stations which fall within a certain radius, and which also report prices on that same date. Therefore, each market corresponds to a particular date. We restrict markets to contain 3 or more stations.

|  | California |  | Florida |  | New Jersey |  | Texas |  |
|---|---|---|---|---|---|---|---|---|
| Regular: |  |  |  |  |  |  |  |  |
| Number of Firms | 3.60 | (0.96) | 3.77 | (1.13) | 3.79 | (1.25) | 3.78 | (1.14) |
| Range | 10.69 | (8.27) | 7.85 | (6.73) | 8.60 | (7.33) | 7.50 | (6.84) |
| StdDev of Price | 5.23 | (4) | 3.77 | (3.18) | 4.12 | (3.45) | 3.62 | (3.23) |
| Gains from search | 5.10 | (3.92) | 3.59 | (3.12) | 4.00 | (3.7) | 3.49 | (3.13) |
| Observations | 112089 |  | 159758 |  | 36298 |  | 207231 |  |
| Mid-grade: |  |  |  |  |  |  |  |  |
| Number of Firms | 3.29 | (0.64) | 3.34 | (0.67) | 3.24 | (0.58) | 3.35 | (0.67) |
| Range | 11.55 | (9.49) | 8.48 | (7.1) | 12.36 | (7.36) | 8.24 | (7.52) |
| StdDev of Price | 5.89 | (4.83) | 4.31 | (3.57) | 6.37 | (3.77) | 4.20 | (3.78) |
| Gains from search | 5.60 | (4.74) | 4.07 | (3.42) | 6.21 | (4.05) | 3.95 | (3.52) |
| Observations | 15341 |  | 25799 |  | 593 |  | 9871 |  |
| Premium: |  |  |  |  |  |  |  |  |
| Number of Firms | 3.26 | (0.55) | 3.33 | (0.67) | 3.29 | (0.66) | 3.31 | (0.61) |
| Range | 17.31 | (11.83) | 9.01 | (7.66) | 18.52 | (11.04) | 10.50 | (9.01) |
| StdDev of Price | 8.90 | (6.04) | 4.58 | (3.83) | 9.45 | (5.63) | 5.39 | (4.62) |
| Gains from search | 8.87 | (6.44) | 4.66 | (4.32) | 10.06 | (6.56) | 5.56 | (5.21) |
| Observations | 10877 |  | 20030 |  | 1748 |  | 6995 |  |
| Diesel: |  |  |  |  |  |  |  |  |
| Number of Firms | 3.25 | (0.55) | 3.25 | (0.53) | 3.32 | (0.66) | 3.33 | (0.67) |
| Range | 15.76 | (12.22) | 10.95 | (8.47) | 17.39 | (14.6) | 10.68 | (8.45) |
| StdDev of Price | 8.12 | (6.28) | 5.65 | (4.33) | 8.98 | (7.53) | 5.46 | (4.23) |
| Gains from search | 7.45 | (5.85) | 5.28 | (4.29) | 7.73 | (7.02) | 5.23 | (4.25) |
| Observations | 9786 |  | 13626 |  | 4092 |  | 25202 |  |

Note: Standard Deviations in Parentheses. Markets are restricted to having a minimum of 3 stations.

Table 3: Market level Summary Statistics, 1-mile radius

for example, that consumers in California can save up to 13 cents per gallon on average if searching in a 2 mile radius for Regular gasoline and 19 cents if searching for Premium gasoline.

It is worth emphasizing the richness and detail that our dataset provides. By having daily station level data, we are able to examine *local* market price dispersion for all states and fuel-types *over time*, without having to rely on samples which are not always representative along these dimensions. The main empirical prediction from consumer search models is that firms use mixed strategies and have sales from time to time. Having a panel dataset with daily data allows us to test the temporal price dispersion for different fuel-types as well as the effects on price dispersion of time-varying variables such as the spot price of gasoline. Having data from 4 large states allows us to generalize our results beyond the possible idiosyncrasies of city or region specific data. Previous work on price dispersion and search in the gasoline market faced the limitations of cross-sectional datasets, weekly data from a single region, and small samples for a unique fuel-type.[31]

## 4 Temporal price dispersion

In this section we try to answer the following question: Is price dispersion in the gasoline industry consistent with a search based theory of sales? In order to do this, we look at the properties of price dispersion over time. Dispersed prices can be the outcome of both product heterogeneity and costly consumer search. However, a critical difference between the two is that prices are not expected to change in product differentiation models as long as the characteristics of the products remain constant.[32] By contrast, models of consumer search, including the one presented in Section 2, predict temporal price dispersion (sales) since firms use mixed strategies and change their prices every period to keep buyers from learning about the identity of the store with the lowest price. We analyze temporal price dispersion by looking at the variation over time in the price spreads between all pairs of stations in a given market.

A straightforward way to analyze temporal dispersion is to look at the changes in price rankings over time. Let $\mathbf{s}_{ij}$ be a vector of the price spread between two gas stations $(i, j)$ over $T_{ij}$ days, such that $p_{it} \geq p_{jt}$ is observed most of the time. Define the *rank reversals* between stations $i$ and $j$ as the proportion of observations in which $p_{jt} > p_{it}$:

$$r_{ij} = \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} \mathbf{I}_{\{p_{jt} > p_{it}\}}$$

We construct this statistic for all possible pairs of stations separated by less than 2 miles. Figure 4 shows a histogram of the rank reversals for Regular Unleaded gasoline and Table 4 presents the summary statistics for all fuel-types and distance bounds of 1 and 2 miles. Both suggest the existence of temporal price dispersion: more than 90% of the pairs of stations have positive rank reversals and the average rank reversal is around 0.15 (Regular and Mid-grade). That means that a station that usually charges the lower price has a higher price 15% of the time.[33] The table also shows that the average price spread between two gas stations is not negligible (more than 5 cents per gallon) and that this spread increases with the octane rating, suggesting that the intensity of price competition is different across

---

[31] For examples, see Barron et al (2004), Hosken et al (2008), and Lewis (2008).

[32] The equilibrium in these type of models is characterized by firms using pure strategies.

[33] By definition, a rank reversal can never be higher than 0.5.

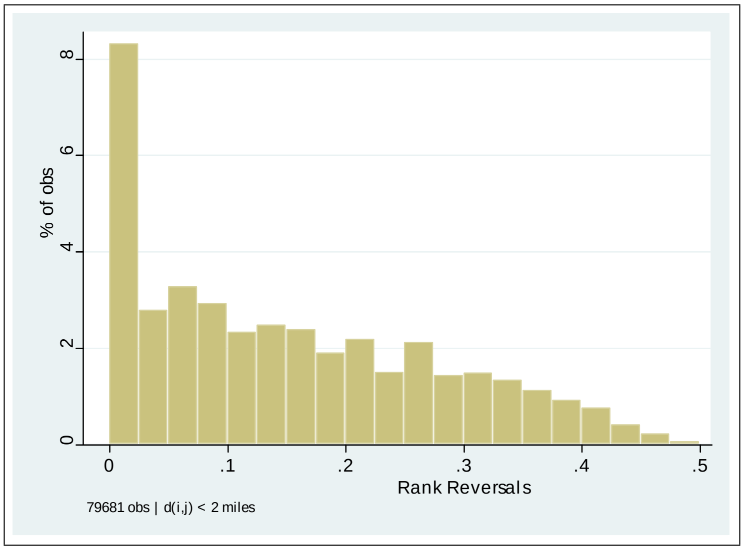


Figure 4: Temporal price dispersion, Regular unleaded gasoline

| | Regular | Mid-grade | Premium | Diesel |
|---|---|---|---|---|
| $d_{ij} < 1mi.$ | | | | |
| Number of obs. | 26,106 | 5,194 | 3,959 | 4,279 |
| Avg. rank reversal | 0.138 | 0.149 | 0.119 | 0.121 |
| Avg. spread | 4.89 | 5.94 | 8.91 | 7.52 |
| $d_{ij} < 2mi.$ | | | | |
| Number of obs. | 79,681 | 15,771 | 11,890 | 12,255 |
| Avg. rank reversal | 0.149 | 0.159 | 0.123 | 0.131 |
| Avg. spread | 5.25 | 6.37 | 9.44 | 8.52 |

Table 4: Summary statistics, rank reversals

fuel-types. We return to this point in the next Section. Appendix Figure 8 shows examples of station-pairs with varying levels of rank reversals and price spreads.

The rank reversals statistic conveys information that is similar to the transition probabilities calculated in other studies of price dispersion. For example, Lewis (2008), Hosken et al. (2008), and Lach and Moraga-Gonzalez (2009) find evidence of mixed strategies — and hence consumer search — in the gasoline market by noting that the probability that a seller's price remains in the same quartile of the distribution in the following period is very low. There are some practical differences between those transition probabilities and our rank reversals. Given the data requirements in those studies, the observed price distributions are constructed at the city level. As we discuss below, localized (market specific) shocks may be the reason for rank changes in the city-wide price distributions. Therefore, the fact that we use station-pairs that are within 1 or 2 miles of each other is an improvement over city-wide samples. Additionally, the transition probabilities in prior studies are calculated for *residual prices*, i.e. prices net of station-specific fixed effects. The logic is that even if firms use mixed strategies, their positions in the price distribution may not change over time if product differences are large, since the sets from which each firm draws prices may not overlap. However, the drawback of this procedure is that misspecification of the regression that estimates station fixed-effects may erroneously suggest evidence of mixed strategies.[34] Our findings also support the existence of mixed strategies, but the evidence of unstable rankings over time is based on actual or 'raw' prices at the local market level.

Positive rank reversals are expected when information frictions underlie the data generation process, but other factors could be present as well. First, we observe prices at any time during the day. Although there is reason to believe that stations change their price at most once a day (this is mandated in some jurisdictions including in the entire state of New Jersey),[35] positive reversals could be reflecting the fact that we observe prices for stations at different moments.[36] Second, other models with or without mixed strategies could explain positive reversals; for example, models with idiosyncratic (firm-specific) demand and cost shocks, or Edgeworth Cycles as in Maskin and Tirole (1988). Assume for now that it is plausible that some or all of these factors are behind the rank reversals in Figure 4 and Table 4. To test whether imperfect information is also responsible for the observed price dispersion, the ideal study would entail comparing price dispersion in markets where the researcher knows that search is absent (a control group), with the dispersion in markets where search could be present.

The nature of the gasoline retailing market allows for such a test. Since gas prices are prominently posted outside stations and visible to all drivers, we expect that the price spread between gas stations that are located in the same street intersection reflect only product differences. Stations in the same corner set different prices according to their heterogeneity (brand, amenities, accessibility, degrees of vertical integration) but they do not compete among each other for informed or uninformed consumers since every driver in that corner knows their prices.[37] However, these stations located in the same corner might choose their prices to compete for informed consumers with other distant stations and then distribute

---

[34] These regressions are only valid if station fixed effects are additively separable from stations' costs. See Wildenbeest (2009) for a specific example where this is the case. However, suppose the data generation process corresponds to a monopolistic competition model with firms setting deterministic prices according to $p_{it} = \beta_i c_t$. Regressing $p_{it}$ on time and firm fixed effects will generate — by construction — random residual prices that could be interpreted as evidence of mixed strategies.

[35] http://www.nytimes.com/2005/09/27/nyregion/27prices.html

[36] In this regard, the fact that we use daily data alleviates many of the problems with earlier studies that used weekly data.

[37] See Png and Reitman (1994) for evidence of product differentiation across stations with similar location.

their captive customers based on product differences.[38] In other words, rank reversals are expected to happen less frequently for stations that are close to each other than for stations that are further apart but still in the same market.

Assume that the rank reversals between stations in the same corner are drawn from the distribution $F_1(r)$ and that the rank reversals between the stations that are separated by more than a block but still in the same market (1 mile or 2 miles) are drawn from $F_2(r)$. If consumer search plays an important role in the gasoline market, we should expect $F_1(r) > F_2(r)$. We use the Kolmogorov-Smirnov (K-S) test to evaluate whether the observed rank reversals for the two groups are drawn from the same population. This is a non parametric test that evaluates the alternative hypothesis of $F_1 \gtrless F_2$, against the null hypothesis $H_0 : F_1 = F_2$. Figure 5 shows the empirical distribution of rank reversals for Regular unleaded gasoline and Table 5 presents the test results for the four fuel-types and the two market bound specifications. Except in the case of Premium gasoline, we reject the null hypothesis at the 1% significance level. That is, consistent with the theory, the temporal price dispersion is significantly lower for the control group than at the market level.[39]

Analogous results can be obtained in a regression environment. Consider a regression of the following form:

$$r_{ij} = \beta_0 + \beta_1 I[corner]_{ij} + \beta_2 X_{ij} + \epsilon_{ij} \tag{3}$$

Here, $r$ represents the rank reversals between stations $i$ and $j$, $I[corner]$ is an indicator for whether the stations are in the same corner, and $X$ contains other control variables. This regression provides a test of the equality of the means of the dependent variable for each of the two groups; i.e., station-pairs in the same corner, and those that are not. However, our hypothesis is that the entire distribution of rank reversals among pairs of stations at the same intersection is systematically different from the distribution among pairs further away. To test this hypothesis, we also employ Quantile regressions that correspond to equation 3. Before doing so, we address a possible concern regarding our reliance on rank reversals as evidence of mixed strategies by gas stations.

---

[38] Note that we are implicitly assuming some degree of coordination between firms located at the same intersection. In terms of the model of Section 2, only one station draws a price and the rest adjust their prices based on product characteristics.

[39] Note that the rank reversals for stations at the same corner are not zero (Figure 5). This may be related to measurement error since we define stations as being in the same corner when their distance is under 270 feet. The reason is that the mapping of stations' addresses to coordinates is not precise and therefore distances calculated for the stations can easily be overstated. On the other hand, stations that we assign to the same corner could actually be further apart.

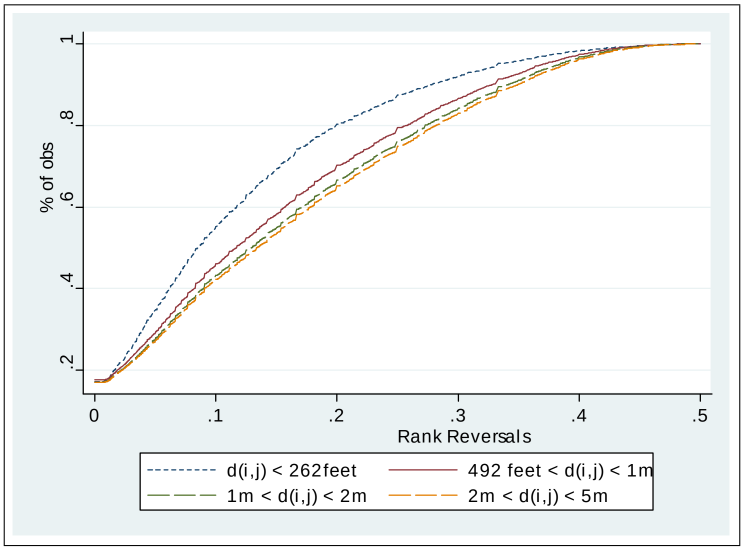


Figure 5: CDF of rank reversals and distance, Regular unleaded gasoline

| | | 1 mile | | 2 miles | |
|---|---|---|---|---|---|
| fuel-type | Ha | D | p-value | D | p-value |
| RU | 1 | 0.1144 | 0.0000 | 0.1485 | 0.0000 |
| | 2 | -0.0048 | 0.8911 | -0.0005 | 0.9985 |
| | KS | 0.1144 | 0.0000 | 0.1485 | 0.0000 |
| MU | 1 | 0.0930 | 0.0000 | 0.1096 | 0.0000 |
| | 2 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| | KS | 0.0930 | 0.0001 | 0.1096 | 0.0000 |
| PU | 1 | 0.0595 | 0.0512 | 0.0677 | 0.0147 |
| | 2 | -0.0020 | 0.9967 | -0.0066 | 0.9612 |
| | KS | 0.0595 | 0.1024 | 0.0677 | 0.0295 |
| DI | 1 | 0.1165 | 0.0000 | 0.1437 | 0.0000 |
| | 2 | -0.0197 | 0.6583 | -0.0122 | 0.8375 |
| | KS | 0.1165 | 0.0000 | 0.1437 | 0.0000 |

Note: $H_0 : F_1 (r_{ij}) = F_2 (r_{ik})$ where $d_{ij} < 80mts$ and $150mts < d_{ik} < 1/2mi$. $H_a(1) : F_1 (r) > F_2 (r)$ ; $H_a(2) : F_1 (r) < F_2 (r)$; $H_a(KS) : F_1 (r) \neq F_2 (r)$

Table 5: Equality of distributions test for rank reversals. Corners vs. Market

The rank reversal test and regression assume that the only difference between the two groups being compared is whether or not the gas stations $(i, j)$ share the same location. But stations can differ along other dimensions. In fact, characteristics are endogenous and are not expected to be randomly chosen in equilibrium. Indeed, stations at the same intersection may try to differentiate themselves more on other dimensions in order to attenuate price competition. Therefore, our preceding results may conceivably be driven by the following factors: $(i)$ greater price spreads between station-pairs at the same corner than pairs further apart, due to endogenous characteristics, and $(ii)$ firm-specific cost shocks. Together, these may cause fewer rank reversals at corners. However, Table 5 suggests that this is not the case since price spreads increase with the distance between stations.[40] That is, the fewer rank reversals for neighboring stations occur despite the fact that their price spreads are smaller.

We test this more formally by using an alternative measure of randomization as the dependent variable in equation 3 above. We define the standard deviation of price differences for each pair of stations as follows:

$$\sigma_{ij} = \sqrt{\frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} [s_{ijt} - \overline{s}_{ij}]^2}$$

where $s_{ijt} \in \mathbf{s}_{ij}$ and represents the price spread between stations $(i, j)$ in day $t$, and $\overline{s}_{ij}$ is the average of $T_{ij}$ days observed. While the rank reversals only use information on changes in the sign of the price spread, the advantage of the standard deviations is that they use all the available price information to measure the degree of volatility of stations' prices with respect to each other. This helps to characterize cases such as the one illustrated in Appendix Figure 8b, where the price spread between the pair of gas stations is highly volatile, but there are zero rank reversals.

OLS and Quantile regression results of estimating equation 3, using both dependent variables, are presented in Table 6. We only report the coefficient on the variable of interest, namely the indicator for being in the same corner. The quantile regression results are presented for the 25th, 50th, 75th and 90th quantiles.[41] The coefficient on the indicator for the corner is negative and highly significant in all specifications.[42] The Table presents results

---

[40] For comparison purposes, the average price spreads between corner stations for Regular, Mid-grade, Premium and Diesel are 3.95,4.96,7.37, and 5.20 cents respectively, which are lower than for station-pairs within 1 and 2 miles (Table 5 and Appendix Table 9).

[41] Results using $\sigma_{ij}$ hold at lower quantiles as well. However, coefficients using $r_{ij}$ are generally not identified at quantiles lower than 15% as there are a significant number of pairs of stations with zero rank reversals, regardless of their distance from each other.

[42] The results are robust to adding various controls such as the distance between stations and state-fixed

| Sample | Depvar | OLS | Quantile Regressions | | | |
|---|---|---|---|---|---|---|
| | | | 25% | 50% | 75% | 90% |
| Station pairs | $r_{ij}$ | -0.027 | -0.007 | -0.033 | -0.056 | -0.051 |
| within 1 mile | | [11.10]** | [2.52]* | [8.54]** | [13.65]** | [9.84]** |
| N=25345 | $\sigma_{ij}$ | -0.006 | -0.007 | -0.006 | -0.006 | -0.004 |
| | | [10.89]** | [20.14]** | [14.87]** | [7.95]** | [2.35]* |
| Station pairs | $r_{ij}$ | -0.036 | -0.011 | -0.045 | -0.076 | -0.064 |
| within 2 miles | | [15.09]** | [5.85]** | [10.00]** | [15.32]** | [10.71]** |
| N=78920 | $\sigma_{ij}$ | -0.007 | -0.009 | -0.008 | -0.007 | -0.005 |
| | | [15.19]** | [29.45]** | [22.80]** | [11.34]** | [3.40]** |

Note: T statistics in brackets. * significant at 5%; ** significant at 1%. Values represent coefficients from a regression of the dependent variable on an indicator for whether the pair of stations is in the same corner.

Table 6: OLS and Quantile Regressions of measures of price dispersion

for Regular Unleaded; the results for Mid-grade and Diesel are similar, and not reported here. The results for Premium using $\sigma_{ij}$ are also similar. However, results for Premium using $r_{ij}$ are weaker, as was the case with the Kolmogorov-Smirnoff test. This indicates that station characteristics play a more important role in determining prices for Premium gasoline. This is also in line with our observation in Table 3 that prices for Premium gasoline exhibit more dispersion than for other fuel-types, which explains why the rank reversals test for Premium yields weak results (due to a lower likelihood of stations switching ranks) but the price-spread test is similar to that of the other fuel-types.[43] Overall, these results strongly indicate that temporal price dispersion is lower for pairs of gas stations at the same corner than for pairs that are further apart. These parametric tests bear out the previous results using the non-parametric K-S test and support our hypothesis that search is important in the gasoline market.

We now focus on possible alternative explanations that might generate the temporal dispersion patterns found above. First, it might be argued that the rank reversals are driven by *correlated* shocks to stations' costs. While this does not seem the case for the retail gasoline market, temporal cost shocks, if any, are expected to be correlated across

---

effects, as well as robust to different definitions of corners. They also hold with the same significance level if we restrict the sample to observations where both stations have changed their price since the previous observation, or expand the sample to include stations within 5 miles of each other. These results are available from the authors upon request.

[43] This is illustrated by the pair of stations in Appendix Figure 8b, where there appears to be high variation in the price spreads of the station-pair, but with zero rank reversals.

stations carrying the same brand.[44] Given that stations of the same brand are unlikely to locate in the same corner (which can be seen from Table 2), we should expect more rank reversals in the control group than the market group ($F_1 < F_2$). This is the opposite of what we observe in Figure 5.

Second, stations could face demand shocks and hence adjust their prices relative to other firms that did not receive a demand shock. In general, a demand shock should be thought of as affecting a whole market rather than a particular gas station or corner. Thus, if demand shocks explained rank reversals we would observe that gas stations in the same market (1 or 2 miles apart) have lower reversals than those further apart. The K-S test was used (not reported) to test this prediction and we find some evidence of market demand shocks. In the case of Regular unleaded, the rank reversals for pairs of stations located in the same market (1 mile) are lower than for stations separated by more than 2 miles. However, this difference disappears once we consider a market bound of 2 miles or other fuel-types. Additionally, to explain the differences in rank reversals between the control and treatment groups we need to consider the possibility of localized demand shocks such as sport and other events. To correct for that, only weekday prices were used to calculate the rank reversals.

Third, it could be argued that some station owners set prices based on the accounting rather than opportunity cost of gasoline. If this is the case, rank reversals could simply arise because station fill their underground storage tanks at different moments. Despite the doubts about the rationality of this behavior, we carry out the same test of rank reversals but restricting the observations to those for which both prices in the station pair change with respect to the last reported prices.[45] Table 12 in the Appendix shows that the results are even stronger (including Premium gasoline) than those obtained when prices are not conditioned to change.

Finally, rank reversals may be consistent with Edgeworth Cycles, a price pattern generated by firms taking turns to change their prices.[46] We do not discard this as a possible

---

[44] Refineries buy gasoline from each other when facing disruptions in their production process. Additionally, stations that are not vertically integrated with their supplier face identical rack prices although prices can vary when delivery costs are included (Hastings, 2004). However, delivery cost differences are expected to be very small (and stable over time) since stations are located within 1 to 2 miles of each other.

[45] Evidence that current practices are to set prices according to the opportunity cost can be inferred from Connecticut Senate Bill 1136 that attempts to "mandate that retailers sell gasoline based on the actual prices . . . paid for the gasoline located in underground storage tanks located on the premises of the retail gasoline station at which gasoline is sold," (http://www.ftc.gov/opa/2007/05/fyi07241.shtm)

[46] Evidence of Edgeworth Cycles in the gasoline market has been found for some Canadian cities (Eckert, 2003; and Noel, 2007b). The evidence in the US is not as clear (Hosken et al, 2008), and research indicates that Edgeworth cycles in the US are mainly concentrated in the Midwest, which does not apply to our dataset (Lewis, 2009).

explanation for positive rank reversals at the market level; however, to explain our other results, this would require two cycles occurring in parallel (one at the intersections and one in the entire market of 1-mile radius), with the properties shown in Table 5. Moreover, Table 12 in the Appendix provides evidence that reversals in the two groups are still significantly different when firms are changing their prices simultaneously rather than in turns.

To summarize this section, our results establish the link between consumer search theory and the price dispersion observed in gasoline markets. By employing a simple test involving street corners, we are able to compare pairs of stations which may be randomizing their prices, in order to keep consumers uninformed, to pairs of stations where imperfect information is not a consideration in determining relative prices. Our results strongly indicate that, on top of other possible sources, costly consumer search plays an important role in explaining the observed temporal variation of prices. We emphasize again that it is the nature of the gasoline industry, and the nature of our panel dataset, that allow us to conduct this test.

## 5 Price Dispersion Across Markets

The previous section used temporal price dispersion to establish the importance of imperfect information in gasoline markets. We now shift the focus to the variation in price dispersion across markets. We have two goals in this section. First, we characterize the equilibrium relationships between price dispersion and key parameters that vary both cross-sectionally and temporally. We compare our estimates to the predictions of our stylized search model of Section 2; refer to Table 1 for these predictions.

Second, we attempt to pin down the equilibrium relationship between price dispersion and search intensity. Recall, from Section 2, that examining the relationship between price dispersion and measures of search benefits or costs does not lead to identification of the role of search in generating price dispersion. Rather, a test of this nature helps to pin down whether the observed equilibrium is closer to the competitive outcome or the monopoly outcome, *conditional on* imperfect information affecting pricing behavior. We will use our data on gasoline prices for different fuel-types to analyze this issue.

Our basic estimation strategy relies on estimating the relationship between price dispersion in a market and (i) measures of the input cost of gas stations, such as the spot price of gasoline or diesel, and (ii) characteristics of the market such as the average level of prices. Recall, from Section 3, that a market is defined as a central station on a given date

|  | Regular |  | Mid-grade |  | Premium |  | Diesel |  |
|---|---|---|---|---|---|---|---|---|
| Depvar: | Range | Std Dev | Range | Std Dev | Range | Std Dev | Range | Std Dev |
| Cost | -0.053 | -0.025 | -0.056 | -0.028 | -0.001 | 0.001 | -0.126 | -0.065 |
|  | [84.62]** | [83.87]** | [27.73]** | [27.34]** | [0.53] | [0.22] | [38.76]** | [38.96]** |
| Avg. Price | 0.048 | 0.023 | 0.05 | 0.025 | -0.019 | -0.011 | 0.133 | 0.068 |
|  | [78.76]** | [77.95]** | [24.60]** | [24.11]** | [7.17]** | [7.88]** | [38.78]** | [38.82]** |
| Rival Firms | 0.303 | 0.077 | 0.286 | 0.101 | 0.395 | 0.151 | 0.501 | 0.163 |
|  | [100.72]** | [53.24]** | [27.17]** | [18.98]** | [28.38]** | [21.21]** | [25.26]** | [16.06]** |
| Constant | 5.713 | 3.352 | 8.113 | 4.237 | 23.763 | 12.241 | 0.634 | 0.498 |
|  | [53.07]** | [65.06]** | [20.65]** | [21.33]** | [44.92]** | [45.61]** | [0.85] | [1.30] |
| N | 515376 |  | 51604 |  | 39650 |  | 52706 |  |

Note: T-Statistics in brackets. All regressions include state fixed effects.

Table 7: Price Dispersion Regressions

surrounded by all stations within a specified radius that report prices on the same date. Our specification is:

$$PRICEDISP_{jt} = \beta_0 + \beta_1 MC_t + \beta_2 AVPRICE_{jt} + \beta_3 N_j + \varepsilon_{jt} \tag{4}$$

where $PRICEDISP_{jt}$ denotes various measures of price dispersion in market $j$ on date $t$, $MC_t$ is a measure of the marginal cost faced by gas stations on date $t$, $AVPRICE_{jt}$ is average retail price of all gas stations in market $j$ on date $t$, and $N_j$ is the number of stations in market $j$. Note that $MC$ is not market specific whereas $AVPRICE$ and $N$ are.[47] The results of estimating Equation 4 are presented in Table 7. Results are presented separately for two dependent variables: the range and the standard deviation of prices. We have combined observations across all 4 states, for each fuel-type, with state-fixed effects included. Our measure of marginal cost is the same-day spot price of the wholesale market corresponding to the state.

The results show that $\beta_1 < 0$, $\beta_2 > 0$ and $\beta_3 > 0$, all of which are consistent with our search model of Section 2. These results are strongly significant (p-value less than 0.01) and hold across all fuel-types, except for Premium. Moreover, the results are robust to running the regression state-by-state, as well as to using the number of rival firms within either 1 or 2 miles; we do not report those results here. Below, we discuss the reasons for Premium gasoline not exhibiting the same results as for the other fuel-types. The first result implies that shocks to the spot price of gasoline or diesel are associated with a decrease in price dispersion. The range of prices for Regular Unleaded decreases by about 5% of the increase in the spot price. This is consistent with the notion that an increase in the input cost

---

[47] The spot price does vary across states, according to the nearest port, but is constant within a state.

(and, therefore, the retail price) of gasoline across all markets leads to a decrease in price dispersion and, consequently, a decrease in the search intensity of consumers.

The second result implies that markets with higher retail prices for gasoline or diesel are associated with higher price dispersion, holding all else constant (in particular, the market spot price). Two reasons could explain this effect. First, all else equal, markets with more differentiated firms are exposed to less price competition and also exhibit higher price dispersion. Second, markets where consumers face higher search costs are expected to have lower search intensity, hence higher prices.[48] As discussed in Section 2, the effect on price dispersion depends on whether the equilibrium is closer to the competitive ($\mu > \widehat{\mu}$) or monopoly ($\mu < \widehat{\mu}$) extremes. $\beta_2 > 0$ indicates that the equilibrium is closer to competitive. We return to this below.

The effect of the number of firms is positive and strongly significant, suggesting that an additional gas station is associated with an average increase in the range of prices of around 0.4 cents. The interpretation of this result is that, as the number of firms increases, there is an increasing tendency towards more extreme pricing. A firm that tries to undercut all its rivals now needs to set a lower price on average; whereas if firms choose not to set low prices to attract informed consumers then they will choose to be at the upper end of the distribution of prices in order to maximize margins. Therefore, the domain from which firms draw prices increases and so does any measure of price dispersion, holding constant the search intensity. In equilibrium, the search intensity also adjusts, as consumers find it more attractive to search, but the final result is generally higher price dispersion (Figure 2.a).

Our theoretical model had predicted that an increase in the number of firms in the market would increase price dispersion, but at a decreasing rate (see Figure 2.b). Allowing a quadratic specification for the number of stations, we find the effect is indeed concave and in fact quite similar to the effect predicted by our model. A plot (not shown) of the relation between the number of gas stations within a 1 mile radius and the range of prices, using our estimated coefficients, looks very similar to the relation predicted by our theoretical model.

The regression above uses the contemporaneous spot price as a measure of the input cost of stations. However, in reality, it may take some time for shocks to the spot price to be reflected in the input prices at retail gas stations.[49] Accordingly, we ran different

---

[48] We do not have a direct explanation of which kind of markets experience higher search costs and lower search intensity. One mechanism could be through income. However, the income of the residents of the market area does not necessarily correspond to the income of its customers, since these could be commuters (Houde, 2009).

[49] Burdette and Zyren (2003) show that lagged values of the spot price are extremely important in determining retail fuel prices.

specifications with spot price in the above regression replaced by various measures of lagged spot prices. We also replaced the contemporaneous or lagged spot prices in these regressions with the contemporaneous state-wide average retail price for the corresponding fuel-type on that date. This strategy assumes that shocks to the spot price will, ultimately, be reflected in the retail prices of all gas stations, after the appropriate time has elapsed. The results are robust to all of these methods.

We now conduct two additional robustness checks. One concern with our methodology so far may be that we assign a single price observation — that is, a station-date combination — to many different, overlapping, markets and therefore to many regressions. This implies that there is a high degree of correlation in the regressors and, potentially, the unobserved component, in each regression. To address this issue we identify, separately for each date, markets according to our original definition, but subject to the constraint that no station is assigned to more than one market.[50] Doing so leaves us with approximately 30% to 50% of the original markets depending on the fuel-type.

The second robustness check is that we use residual or 'cleaned' prices rather than actual prices. This has been the traditional approach in the empirical literature; i.e. the analysis has focused on the price dispersion that remains after product characteristics are controlled for. We construct our price dispersion measures with the estimated residuals from a regression of prices on station fixed effects. In Section 4 we had mentioned the problem of using cleaned prices for identifying consumer search. Once search is established, the problems are not so severe. However, note that this exercise still makes certain assumptions about the interaction between product differentiation and imperfect information, namely that they are additively separable.[51] If these variables interact in a different way then using residuals does not correct for product differentiation, and in fact it would be preferable to examine actual prices rather than demeaned prices. With this caveat in mind, we present estimates in Table 8 from regressions of the range in the market level residuals on the state average residual (i.e. a statewide measure of cost as explained above) and the market average residual. We also constrain the set of markets to be non-overlapping.

The estimates confirm our earlier results. Controlling for station characteristics, increases in cost decrease price dispersion, while increases in the market average price and

---

[50] Other possible corrections for this issue involve panel specifications using market fixed effects, as well as clustering standard errors at the market level. We ran all of these tests and found very similar results, with every coefficient of interest remaining significant at the 99% level.

[51] To the best of our knowledge, there is no theoretical work that analyzes the testable predictions of general models that incorporate the interaction of product differentiation and consumer search. See Hortaçsu and Syverson (2004) and Wildenbeest (2009) for particular examples of models with vertical differentiation and consumer search.

| Depvar=Range | Regular | Mid-grade | Premium | Diesel |
|---|---|---|---|---|
| Cost | -0.369 | -0.285 | -0.237 | -0.235 |
|  | [97.92]** | [30.97]** | [23.34]** | [29.63]** |
| Avg. Price | 0.315 | 0.219 | 0.152 | 0.174 |
|  | [84.80]** | [24.36]** | [15.37]** | [23.08]** |
| Rival Firms | 0.25 | 0.219 | 0.203 | 0.328 |
|  | [34.33]** | [12.05]** | [10.01]** | [17.79]** |
| Constant | 11.784 | 15.315 | 19.515 | 10.256 |
|  | [192.53]** | [97.45]** | [101.58]** | [92.49]** |
| N | 144364 | 28857 | 23114 | 32532 |

Note: T-Statistics in brackets. All regressions include state fixed effects. Each station appears in exactly one market. Prices are demeaned. Cost refers to the state-wide average residual.

Table 8: Regression of Residuals, non-overlapping markets only

the number of firms increase price dispersion. In fact, our results are stronger using residuals than actual prices: they hold across all states and fuel-types. Moreover, the coefficient on Premium gasoline is now similar, in sign and magnitude, to those for the other fuel-types, unlike our previous results. This may suggest that product differentiation plays a greater role in explaining prices and price dispersion for this grade of gasoline; but that once station characteristics are controlled for, the effect of cost and market average prices are the same as for the other grades.[52] These results also hold using the standard deviation of residuals as the dependent variable, and withstand all of our other robustness checks as well.

The average price dispersion (range) for Premium is 14% larger than for Mid-grade, which in turn is 33% larger than for Regular.[53] Except in the case of cars that require higher octane fuel, the demand for Premium is comprised of those who believe this grade will not hurt the engine, even if the car does not explicitly require it. This insurance motive can be expected to be stronger for people driving expensive cars which should be correlated with higher income and search costs.[54] If this assumption is correct, we would expect the price dispersion for higher octane grades to be lower (higher) than for Regular when the

---

[52] We had noted a similar result, using the rank reversals and price spread tests, in Section 4. As we had explained there, the average difference between two gas stations selling Premium gasoline is more than twice the difference for Regular gasoline (Table 4).

[53] These results were obtained by estimating equation 4 and combining data on all fuel-types, but with separate intercepts for each fuel-type, and using markets defined according to a 1-mile radius. See Appendix Table 13 for results. Results were similar using markets of 2-mile radius. The basic pattern of this result can also be seen by comparing price dispersion across fuel-types in Table 3.

[54] See Setiawan and Sperling (1993) for details. The authors show that Premium gasoline is a luxury good, whose demand falls as average gasoline prices rise. They also show that the propensity to buy Premium is positively correlated with consumers' income.

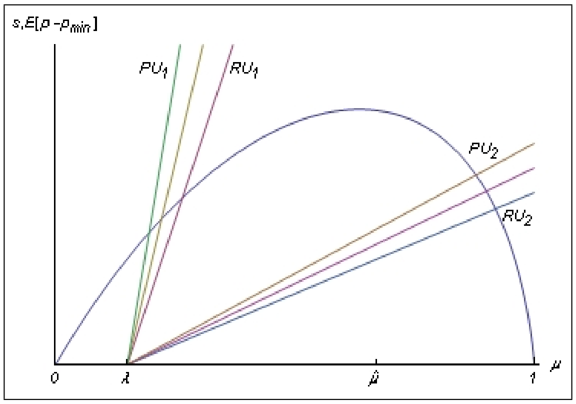


Figure 6: Consumer search cost and equilibrium search intensity

search intensity is low (high). This can be seen in Figure 6, which shows two scenarios. In both cases, search costs are highest for Premium gasoline, followed by Mid-grade and then Regular, as per our assumption. The difference is that in scenario 1 the equilibrium search intensity is low ($\mu < \widehat{\mu}$), while in scenario 2 it is high.

As noted earlier, the non-monotonic relationship between price dispersion and search intensity does not, in general, permit the identification of the local equilibrium. However, combining our assumption regarding search costs with our observed result that price dispersion is greatest for Premium, followed by Mid-grade, followed by Regular gasoline, we can pin down the location of the current equilibrium: it is on the downward sloping portion of the function in Figure 6. Thus, our results suggest that the gasoline market is closer to the competitive outcome than to the monopoly outcome and the relationship between search intensity and price dispersion is negative.[55] We can also conclude that the market for Regular gasoline is the most competitive, followed by Mid-grade and then Premium.

The theory suggests that higher production costs can have either a positive or negative effect on firms' markups, once consumer search is endogenized (see Figure 2). This is independent of the effect on price dispersion, which has an unambiguously negative relationship with production costs. In order to test this, we examined the relation between average markups and the spot price while controlling for the number of firms and other variables.[56] The results imply that firms' markups are lower when contemporaneous spot prices are higher. This suggests that the reaction by consumers to lower price dispersion — caused by the increase in firms' costs — is not enough to offset the direct negative effect on firms' markups. Our results suggest that a 10 cent increase in the spot price reduces the average

---

[55] This is consistent with the findings by Sorensen (2000) in the prescription drug market, and Brown and Goolsbee (2002) in the insurance market once the Internet became a widespread method of search.

[56] See Appendix table 14 for results using both raw prices and demeaned prices.

markup by 1.4 cents for Regular gasoline and by almost 3 cents for Diesel. This also implies that the market for Regular gasoline is considerably more competitive than for Diesel.[57]

We now summarize the results of this section. We have shown that observed price dispersion is positively associated with the number of firms in the market and negatively associated with measures of firms' marginal costs. The results relating market structure to price dispersion are counter to those found by Barron et al (2004), but similar to the results found by Lach and Moraga-Gonzalez (2009) and by Lewis (2008) for local sub markets. The results relating production cost to price dispersion are, to our knowledge, new as data used in previous studies have not spanned as long a period as ours (18 months), which permits an examination of this issue. Our results indicate that there are lower gains from search as the average price of gasoline rises. Thus consumers would be better off reducing their search intensity during periods of high oil prices. All the results — with either actual prices or residuals — are consistent with the predictions of our model of endogenous search with homogeneous products from Section 2. This may help to establish the desired properties for a more general model that incorporates product differentiation and costly consumer search.

While we cannot directly measure search intensity, we use two plausible measures to show that, at the current equilibrium, there is a negative relationship between price dispersion and search intensity, implying that the equilibrium is closer to the competitive outcome than to the monopoly outcome. This was done by showing that markets with higher average prices, suggesting lower search intensity by consumers, have greater price dispersion. Additionally, products that we believe are associated with higher search costs exhibit greater price dispersion. All of our results are robust to many different specifications.

# 6 Conclusions

In this paper we examine the role of imperfect information in explaining equilibrium price dispersion. We show that identification of the role of imperfect information cannot be done by simply testing the usual comparative statics of price dispersion on search costs or benefits, or the number of firms in the market. Price dispersion becomes a non-monotonic function of those variables once we allow consumers to adjust their equilibrium search strategies. Previous work has often assumed a particular equilibrium relationship and has accordingly inferred consumer search from positive or negative relationships among these variables. We argue that identifying the role of consumer search in explaining price dispersion requires a

---

[57] The result that higher costs lead to reduced markups is not unique to search models. This property will hold for a large family of demand functions, though not all; constant elasticity demand functions, for example, will not have this property.

careful examination of temporal dispersion, a dimension in which predictions from consumer search models and other models are orthogonal. Using a novel test of rank reversals and price spreads among station-pairs, we find that the temporal price dispersion at the market level is consistently higher than for stations at the same intersection. This is consistent with the theory of consumer search since the dispersion in the latter group is driven only by product differentiation.

We then use a unique panel dataset to examine equilibrium price dispersion in the U.S. retail gasoline industry. We verify that increases in production costs are associated with a decrease in price dispersion. In particular, a 10 cent increase in the marginal cost of Regular Unleaded gasoline decreases price dispersion by over 3 cents. The number of stations in the market is positively related to price dispersion; on average one additional gas station within a market of a 1 mile radius increases price dispersion within the market by almost 4 cents. These results are in line with the predictions of our simple model of endogenous search. Our findings are similar across all fuel-types, hold across all states, and across a range of robustness checks.

Our results imply that consumers could save as much as 5% by price shopping within a one-mile radius. The fact that search costs deter consumers from price-shopping is reinforced by the result that grades of gasoline associated with higher search costs involve greater price dispersion. To the extent that search costs act as a friction, sources that alleviate imperfect information will reduce prices and price dispersion. Centralized sources of information regarding gas prices would achieve this. Existing web sites where users periodically list stations' gas prices may be one step in this direction.[58] Moreover, our results indicate that price dispersion decreases when the aggregate level of prices rise, implying that there are less gains to searching at such times. Therefore, employing a policy of greater search during periods of peak pricing may be sub-optimal. Increased information along these lines may help consumers to make better decisions regarding their search strategies.

# References

Armstrong, M., J. Vickers, and J. Zhou (2009): "Consumer Protection and the Incentive to Become Informed," *Journal of the European Economic Association*, 7(2-3), 399–410.

---

[58] The government in Western Australia has introduced a "24-hour rule" which requires stations to submit daily their prices for the next day; this information is available to consumers by telephone and online.

Barron, J. M., B. A. Taylor, and J. R. Umbeck (2004): "Number of sellers, average prices, and price dispersion," *International Journal of Industrial Organization*, 22(8-9), 1041–1066.

Baye, M. R., J. Morgan, and P. Scholten (2004): "Price Dispersion In The Small And In The Large: Evidence From An Internet Price Comparison Site," *Journal of Industrial Economics*, 52(4), 463–496.

——— (2006): "Information, Search, and Price Dispersion," in *Handbook of Economics and Information Systems*, vol. forthcoming. Elsevier Press.

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

Brown, J. R., and A. Goolsbee (Jun., 2002): "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *The Journal of Political Economy*, 110(3), 481–507.

Burdett, K., and K. L. Judd (1983): "Equilibrium Price Dispersion," *Econometrica: Journal of the Econometric Society*, 51(4), 955–970.

Burdette, M., and J. Zyren (2003): "Gasoline Price Pass-through," Discussion paper, available at http://www.eia.doe.gov/.

Carlson, J. A., and R. P. McAfee (1983): "Discrete Equilibrium Price Dispersion," *The Journal of Political Economy*, 91(3), 480–493.

Clay, K., R. Krishnan, and E. Wolff (Dec., 2001): "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *The Journal of Industrial Economics*, 49(4), 521–539.

Dahlby, B., and D. S. West (1986): "Price Dispersion in an Automobile Insurance Market," *The Journal of Political Economy*, 94(2), 418–438.

Dana, James D., J. (1994): "Learning in an Equilibrium Search Model," *International Economic Review*, 35(3), 745–771.

Eckert, A. (2003): "Retail price cycles and the presence of small firms," *International Journal of Industrial Organization*, 21(2), 151–170.

Hastings, J. (2004): "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California," *American Economic Review*, 94(1), 317–328.

Hortaçsu, A., and C. Syverson (2004): “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds,” *Quarterly Journal of Economics*, 119(2), 403–456.

Hosken, D. S., R. S. McMillan, and C. T. Taylor (2008): “Retail gasoline pricing: What do we know?,” *International Journal of Industrial Organization*, 26(6), 1425 – 1436.

Houde, J.-F. (2009): “Spatial Differentiation and Vertical Contracts in Retail Markets for Gasoline,” Working paper, University of Wisconsin.

Janssen, M. C., and J. L. Moraga-Gonzalez (2004): “Strategic Pricing, Consumer Search and the Number of Firms,” *The Review of Economic Studies*, 71(4), 1089–1118.

Janssen, M. C., J. L. Moraga-Gonzalez, and M. R. Wildenbeest (2005): “Truly costly sequential search and oligopolistic pricing,” *International Journal of Industrial Organization*, 23(5-6), 451–466.

Lach, S. (2002): “Existence And Persistence Of Price Dispersion: An Empirical Analysis,” *The Review of Economics and Statistics*, 84(3), 433–444.

Lach, S., and J. Moraga-Gonzalez (2009): “Asymmetric Price Effects of Competition,” Discussion Paper Discussion Paper No. 7319, CEPR.

Lewis, M. (2008): “Price Dispersion and Competition with Differentiated Sellers,” *Journal of Industrial Economics*, 56(3), 654–678.

Lewis, M. S. (2009): “Temporary Wholesale Gasoline Price Spikes Have Long-Lasting Retail Effects: The Aftermath of Hurricane Rita,” *The Journal of Law and Economics*, 52(3), 581–605.

Morgan, P., and R. Manning (1985): “Optimal Search,” *Econometrica: Journal of the Econometric Society*, 53(4), 923–944.

Noel, M. (2007a): “Edgeworth Price Cycles, Cost-based Pricing and Sticky Pricing in Retail Gasoline Retail Markets,” *Review of Economics and Statistics*, 89 (2), 324–334.

——— (2007b): “Edgeworth Price Cycles: Evidence form the Toronto Retail Gasoline Market,” *Journal of Industrial Economics*, 55 (1), 69–92.

Png, I. P. L., and D. Reitman (1994): “Service Time Competition,” *The Rand Journal of Economics*, 25(4), 619–634.

SETIAWAN, W., AND D. SPERLING (1993): "Premium Gasoline Overbuying in the U.S.: Consumer-Based Choice Analysis," Discussion Paper Research Report UCD-ITS-RR-93-25, Institute of Transportation Studies, UC Davis.

SHEPARD, A. (1991): "Price Discrimination and Retail Configuration," *Journal of Political Economy*, 99(1), 30–53.

SMITH, M. D., AND E. BRYNJOLFSSON (Dec., 2001): "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters," *The Journal of Industrial Economics*, 49(4), 541–558.

SORENSEN, A. T. (2000): "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, 108(4), 833–850.

STAHL, D. O. (1989): "Oligopolistic Pricing with Sequential Consumer Search," *The American Economic Review*, 79(4), 700–712.

——— (1996): "Oligopolistic pricing with heterogeneous consumer search," *International Journal of Industrial Organization*, 14(2), 243–268.

STIGLER, G. J. (1961): "The Economics of Information," *The Journal of Political Economy*, 69(3), 213–225.

TAPPATA, M. (2009): "Rockets and feathers: Understanding asymmetric pricing," *RAND Journal of Economics*, 40(4), 673–687.

VAN HOOMISSEN, T. (Dec., 1988): "Price Dispersion and Inflation: Evidence from Israel," *The Journal of Political Economy*, 96(6), 1303–1314.

VARIAN, H. R. (1980): "A Model of Sales," *The American Economic Review*, 70(4), 651–659.

VERLINDA, J. (2005): "Price-Response Asymmetry and Spatial Differentiation in Local Retail Gasoline Markets," Working paper, University of California Irvine.

WILDENBEEST, M. R. (2009): "An empirical model of search with vertically differentiated products," mimeo, Indiana University.

# 7 Appendix

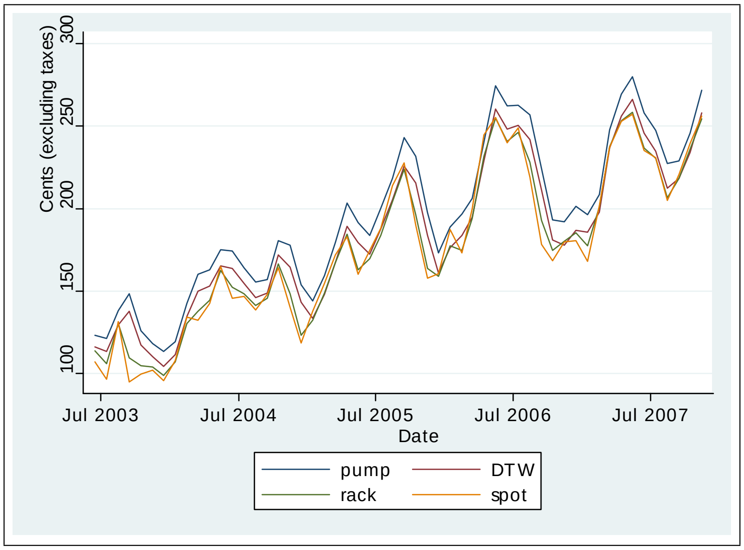
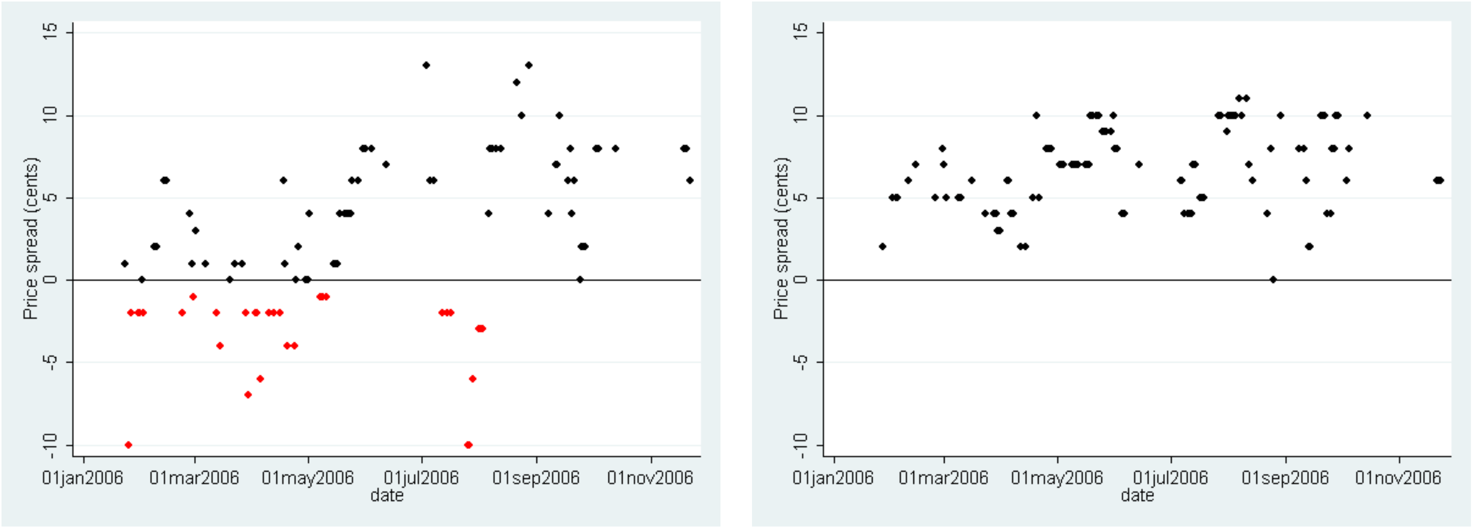
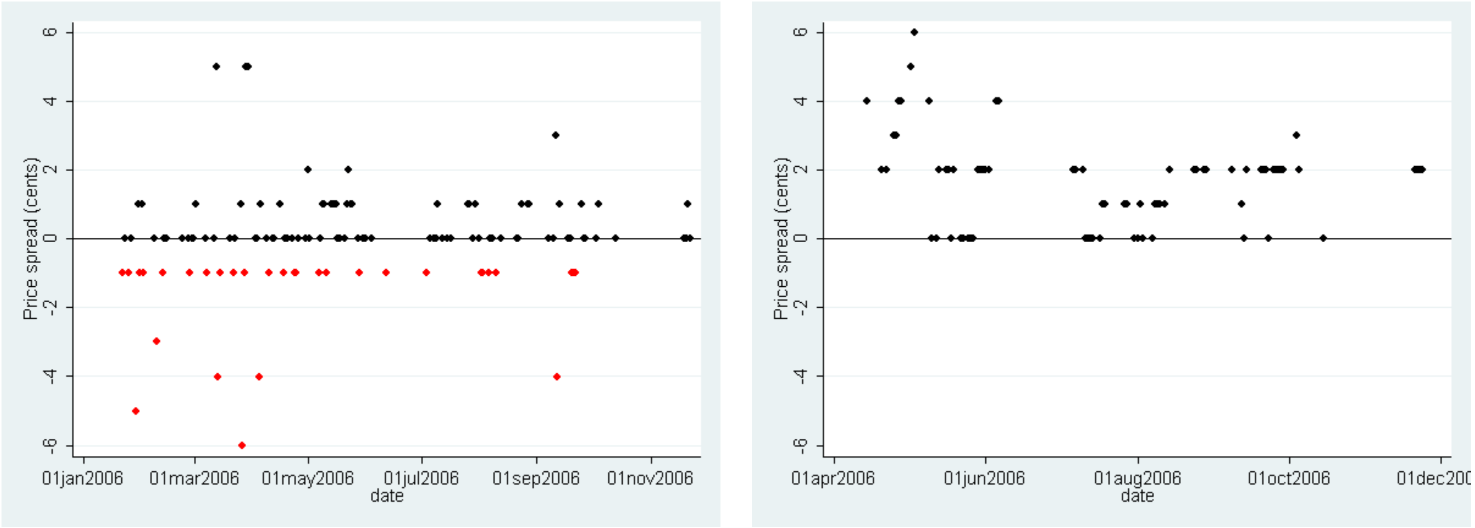


Figure 7: Regular unleaded retail and wholesale prices in California



a) High average spread, high $r$

b) High average spread, zero $r$



c) Low average spread, high $r$

d) Low average spread, zero $r$

Figure 8: Price spread patterns between pair of stations

| | California | | Florida | | New Jersey | | Texas | |
|---|---|---|---|---|---|---|---|---|
| Regular Unleaded: | | | | | | | | |
| Number of Firms | 5.23 | (2.59) | 5.72 | (3.06) | 5.56 | (2.97) | 5.87 | (3.17) |
| Range | 13.07 | (8.86) | 10.09 | (7.61) | 10.99 | (8.16) | 9.83 | (7.86) |
| StdDev of Price | 5.48 | (3.51) | 4.09 | (2.89) | 4.48 | (3.08) | 3.97 | (2.96) |
| Gains from search | 6.04 | (3.96) | 4.51 | (3.53) | 4.95 | (4.13) | 4.45 | (3.46) |
| Observations | 318314 | | 386031 | | 96275 | | 496152 | |
| | | | | | | | | |
| Mid-grade: | | | | | | | | |
| Number of Firms | 3.89 | (1.26) | 4.14 | (1.45) | 3.48 | (0.83) | 3.95 | (1.5) |
| Range | 12.82 | (9.41) | 10.46 | (7.66) | 13.26 | (7.54) | 10.08 | (7.91) |
| StdDev of Price | 6.01 | (4.34) | 4.78 | (3.39) | 6.57 | (3.66) | 4.73 | (3.58) |
| Gains from search | 6.09 | (4.56) | 5.01 | (3.75) | 6.61 | (4.13) | 4.84 | (3.72) |
| Observations | 71469 | | 94492 | | 3470 | | 44171 | |
| | | | | | | | | |
| Premium: | | | | | | | | |
| Number of Firms | 3.79 | (1.15) | 4.02 | (1.32) | 3.67 | (1.03) | 3.86 | (1.38) |
| Range | 19.79 | (12.36) | 11.29 | (8.25) | 19.81 | (11.35) | 13.44 | (9.75) |
| StdDev of Price | 9.48 | (5.82) | 5.22 | (3.7) | 9.51 | (5.29) | 6.42 | (4.62) |
| Gains from search | 10.23 | (6.85) | 5.84 | (4.69) | 10.64 | (6.76) | 7.15 | (5.64) |
| Observations | 49753 | | 73772 | | 8194 | | 30085 | |
| | | | | | | | | |
| Diesel: | | | | | | | | |
| Number of Firms | 3.73 | (1.08) | 3.82 | (1.22) | 4.02 | (1.33) | 4.02 | (1.45) |
| Range | 19.19 | (13.48) | 13.49 | (9.79) | 20.25 | (14.55) | 13.21 | (9.62) |
| StdDev of Price | 9.23 | (6.32) | 6.40 | (4.51) | 9.52 | (6.73) | 6.15 | (4.23) |
| Gains from search | 8.92 | (6.34) | 6.40 | (4.84) | 8.54 | (6.27) | 6.40 | (4.8) |
| Observations | 51605 | | 53971 | | 15168 | | 101676 | |

Note: Standard Deviations in Parentheses. Markets are restricted to having a minimum of 3 stations.

Table 9: Market level Summary Statistics, 2 mile radius

| California | | Texas | | Florida | | New Jersey | |
|---|---|---|---|---|---|---|---|
| CHEVRON | 0.16 | SHELL | 0.13 | CITGO | 0.18 | MOBIL | 0.13 |
| SHELL | 0.15 | CHEVRON | 0.12 | BP | 0.12 | EXXON | 0.12 |
| 76 | 0.14 | CITGO | 0.11 | CHEVRON | 0.11 | SUNOCO | 0.10 |
| UNBRANDED | 0.10 | DIAMOND SHAMROCK | 0.10 | SHELL | 0.11 | GETTY | 0.09 |
| ARCO | 0.08 | EXXON | 0.10 | UNBRANDED | 0.06 | GULF | 0.09 |
| VALERO | 0.08 | FINA | 0.06 | 7-ELEVEN | 0.05 | UNBRANDED | 0.09 |
| MOBIL | 0.07 | TEXACO | 0.06 | HESS | 0.05 | SHELL | 0.08 |
| DIAMOND SHAMROCK | 0.05 | UNBRANDED | 0.06 | MOBIL | 0.05 | BP | 0.07 |
| TEXACO | 0.03 | CONOCO | 0.05 | CIRCLE K | 0.04 | CITGO | 0.07 |
| 7-ELEVEN | 0.02 | PHILLIPS 66 | 0.05 | SUNOCO | 0.04 | HESS | 0.03 |
| CITGO | 0.02 | MOBIL | 0.04 | EXXON | 0.03 | LUKOIL | 0.03 |
| EXXON | 0.02 | CIRCLE K | 0.02 | TEXACO | 0.03 | TEXACO | 0.03 |
| | | | | MARATHON ASHLAND | 0.02 | VALERO | 0.02 |

Note: Brands as reported by OPIS.

Table 10: Distribution of observed brands across states

| | Regular | Mid-grade | Premium | Diesel |
|---|---|---|---|---|
| $d_{ij}< 1mi.$ | | | | |
| Number of obs. | 20,131 | 3,808 | 2,779 | 2,280 |
| Avg. rank reversal | 0.146 | 0.157 | 0.121 | 0.131 |
| Avg. spread | 4.83 | 5.68 | 8.90 | 6.82 |
| $d_{ij}< 2mi.$ | | | | |
| Number of obs. | 60,686 | 11,349 | 8,133 | 6,126 |
| Avg. rank reversal | 0.156 | 0.165 | 0.130 | 0.143 |
| Avg. spread | 5.22 | 6.14 | 9.47 | 7.79 |

Note: Rank reversals calculated only when prices change with respect to the last observation

Table 11: Summary statistics, rank reversals conditioning for price changing

| | | 1 mile | | 2 miles | |
|---|---|---|---|---|---|
| fuel-type | Ha | D | p-value | D | p-value |
| RU | 1 | 0.1251 | 0.0000 | 0.1581 | 0.0000 |
| | 2 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| | KS | 0.1251 | 0.0000 | 0.1581 | 0.0000 |
| MU | 1 | 0.1164 | 0.0000 | 0.1275 | 0.0000 |
| | 2 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| | KS | 0.1164 | 0.0000 | 0.1275 | 0.0000 |
| PU | 1 | 0.1096 | 0.0004 | 0.1188 | 0.0000 |
| | 2 | -0.0028 | 0.9949 | 0.0000 | 1.0000 |
| | KS | 0.1096 | 0.0008 | 0.1188 | 0.0001 |
| DI | 1 | 0.1181 | 0.0001 | 0.1436 | 0.0000 |
| | 2 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| | KS | 0.1181 | 0.0002 | 0.1436 | 0.0000 |

Note: Rank Reversals calculated only when both stations change prices with respect to the last observation.

Table 12: Rank reversals test conditioning on price changing. Corners vs. Markets.

| Depvar=Range | All fuel-types |
|---|---|
| Cost | -0.33 |
| | (98.71)** |
| Average Price | 0.271 |
| | (82.53)** |
| Rival Firms | 0.255 |
| | (39.33)** |
| Premium gasoline | 2.145 |
| | (27.51)** |
| Regular gasoline | -3.854 |
| | (66.20)** |
| Constant | 15.687 |
| | (207.37)** |
| N | 196335 |

Note: T-Statistics in brackets. All regressions include state fixed effects. Each station appears in exactly one market. Prices are demeaned. Cost refers to the state-wide average residual.

Table 13: Price Dispersion regression for all fuel-types

| Depvar=Markup | Prices | | Demeaned prices | |
|---|---|---|---|---|
| | Diesel | Regular | Diesel | Regular |
| Spot Price | -0.288 | -0.144 | -0.316 | -0.18 |
| | [70.13]** | [91.01]** | [95.28]** | [122.82]** |
| Rival Firms | -0.005 | -0.148 | -0.083 | -0.028 |
| | [0.12] | [9.77]** | [2.61]** | [2.02]* |
| Constant | 161.208 | 116.574 | -143.603 | -177.186 |
| | [179.07]** | [316.48]** | [197.44]** | [516.60]** |
| N | 32532 | 144364 | 32532 | 144364 |

Note: T-Statistics in brackets. All regressions include state fixed effects. Markups in the first two columns are constructed using raw prices; those in the next two columns are constructed using demeaned prices.

Table 14: Regression of Markups, non-overlapping markets only