RAND Journal of Economics
Vol. 42, No. 4, Winter 2011
pp. 681–704

Consumer search and dynamic price dispersion: an application to gasoline markets

Ambarish Chandra*

and

Mariano Tappata**

This article studies the role of imperfect information in explaining price dispersion. We use a new panel data set on the U.S. retail gasoline industry and propose a new test of temporal price dispersion to establish the importance of consumer search. We show that price rankings vary significantly over time; however, they are more stable among stations at the same street intersection. We establish the equilibrium relationships between price dispersion and key variables from consumer search models. Price dispersion increases with the number of firms in the market, decreases with the production cost, and increases with search costs.

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 homogeneous good and have identical production costs.[1]

Establishing evidence that price dispersion results from costly consumer search is a challenging task. The comparative statics from search models do not usually identify the role of search in generating price dispersion. For example, the theoretical relationship between search intensity and price dispersion is nonmonotonic and so, in principle, any estimated relationship is consistent with the predictions of a search model. A better test of consumer search comes from

*University of Toronto; ambarish.chandra@rotman.utoronto.ca.
**University of British Columbia; mariano.tappata@sauder.ubc.ca.
We thank Dan Ackerberg, Jim Brander, Tim Hannan, Keith Head, Thomas Hellmann, Phil Leslie, Alan Sorensen, Hal White, and various seminar and conference participants for their comments and suggestions.

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

exploiting temporal price dispersion. Most search models predict firms playing mixed strategies each period, and thus we expect search to be related to changes in the ranking of firms' prices over time. The traditional approach in the literature has been to characterize the 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.[2] Moreover, temporal price dispersion can also be generated in a product differentiation environment with idiosyncratic demand or cost shocks. Therefore, it is important to find a control group or benchmark with which to compare the observed temporal price dispersion.[3]

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 gas stations for which imperfect information may possibly play a role to a control group of stations where imperfect information is absent. In particular, we construct for each pair of stations a rank-reversals statistic, which measures the proportion of times that the usually low-price station sets the high price. We control for imperfect information by using the distance separating the two stations. We use a new and extremely rich panel data set, which provides daily retail prices for U.S. 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, something that is rarely emphasized in the literature.[4]

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 because information depreciates at a higher rate. Sorensen (2000) follows a similar approach, although the identification comes from comparing 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 who are associated with lower search costs. Hortaçsu and Syverson (2004) use a structural model to estimate the search cost distribution in the S&P 500 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.[5] 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.[6]

The gasoline market is an appealing industry in which to study the role of search for two main reasons.[7] 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. We find price-ranking changes in more than 90% of the pairs of stations in a given market. Moreover, rank reversals are significantly lower for stations located at the same street intersection, despite these stations having lower price spreads, than for stations farther apart although still in the

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

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

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

[5] 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). Whereas 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.

[6] Other studies of online markets include Clay, Krishnan, and Wolff (2001), Smith and Brynjolfsson (2001), and Baye, Morgan, and Scholten (2004).

[7] 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.)

same local market. This is consistent with the predictions of search models as, for stations in the same intersection, differences in prices should be driven only by product characteristics and not by imperfect information. Second, because 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 nonmonotonic. 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 models: (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) firms have no capacity constraints.[8]

In addition to the contributions to the literature on search and price dispersion, our article also adds to the empirical literature on price dispersion in gasoline industries.[9] We employ new data which provide key advantages over prior data sets. First, we have daily data on gasoline prices; this is important because existing studies rely on weekly observations.[10] 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 that are associated with higher search costs exhibit higher equilibrium levels of price dispersion, suggesting relatively high levels of search intensity (low information rents). Markets with more firms are associated with greater price dispersion. Interestingly, we find that consumers should search less when pump prices increase as a result of shocks to the wholesale price. 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 article 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 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.[11] 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 intensity change and become nonmonotonic once we allow for endogenous search.[12] 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.

Assume a homogeneous-goods 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 consumers who

[8] See Borenstein, Cameron, and Gilbert (1997) and Noel (2007b).

[9] See, for example, Barron, Taylor, and Umbeck (2004), Hosken, McMillan, and Taylor (2008), Lewis (2008), and Lach and Moraga-Gonzalez (2009).

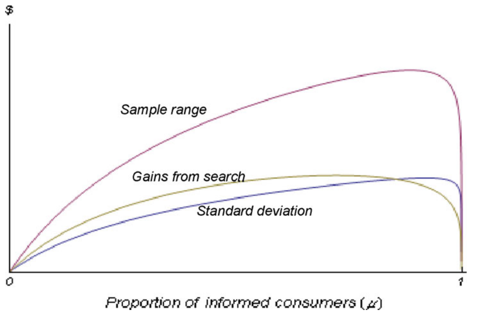
[10] Our data confirm the importance of high-frequency observations to analyze price dispersion; on average, a third of the gas stations change their prices with respect to the previous day.

[11] See Tappata (2009) for a formal analysis of the comparative statics discussed in this section.

[12] 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.

FIGURE 1

PRICE DISPERSION METRICS ($n = 5$)



have inelastic demands with valuations v.[13] A fraction $\lambda \in (0, 1)$ of them have zero search cost and are called "shoppers," whereas 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.[14] The proportion of consumers who decide to become informed, μ, defines the search intensity in the market. 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 (NE), and (ii) the search intensity in the market, μ^*, 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) shows that, given a proportion of informed consumers μ, there is a unique symmetric NE that involves firms playing mixed strategies. In each period, firms simultaneously draw prices from

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

where $p \in [p^* = \frac{cn\mu + (1-\mu)v}{1+(n-1)\mu}, v]$. For prices below $p^*(c, v, n)$, a firm always prefers to charge a monopoly price and sell to $(1 - \mu)/n$ consumers. The intensity of price competition is directly related to the amount of search. 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).

The gains from search (*GS*) for consumers associated with equation (1) are given by

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

It is easy to see that *GS* is a nonmonotonic 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. Figure 1 plots the traditional measures of price dispersion as a function of the search intensity. Both the sample range (SR) and the standard deviation (SD) resemble the shape of the GS.[15] From the same figure, we can see that dispersion

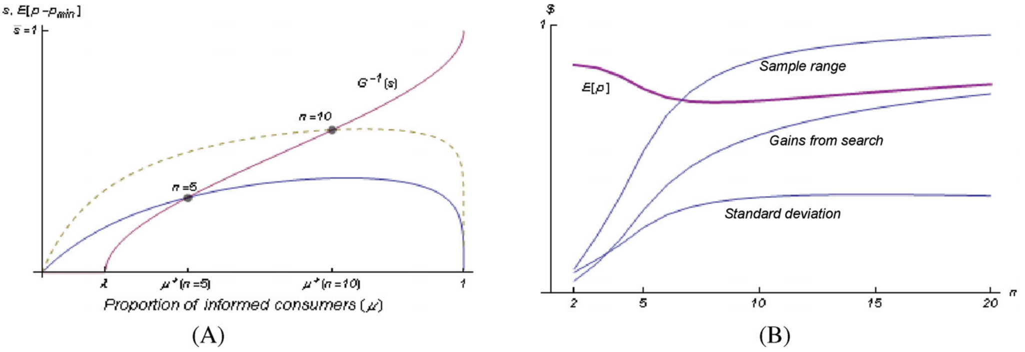
[13] This is a simplifying assumption, and the results in this section hold for a large set of downward-sloping demand functions.

[14] We focus on the case of nonsequential search here. The empirical predictions from sequential search models are very similar and we describe them in a supplementary note available from the authors' webpage.

[15] 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.

FIGURE 2

EQUILIBRIUM PRICE DISPERSION AND THE NUMBER OF FIRMS (n)



measures alone cannot be used to predict the search intensity or level of competition in a market. To see this, define $\widehat{\mu}$ as the search intensity that maximizes GS for a given production cost and number of firms.[16] The relationship between the search intensity and price dispersion is positive when $\mu^* < \widehat{\mu}$ and negative if $\mu^* > \widehat{\mu}$.

The second requirement for an equilibrium is that the amount of search in the market, μ^*, needs to be consistent with the firms' pricing strategies. That is, each consumer anticipates (1), taking the market search intensity as given, and compares the gains from search as in (2) with her search cost. Nonshoppers 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 2A shows the equilibrium when search costs are drawn from a beta distribution.[17] The proportion of informed consumers is measured on the horizontal axis, whereas the search costs and gains from search are on the vertical axis. The solid and dashed concave curves represent the gains from search to consumers in markets with 5 and 10 firms, respectively. 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.

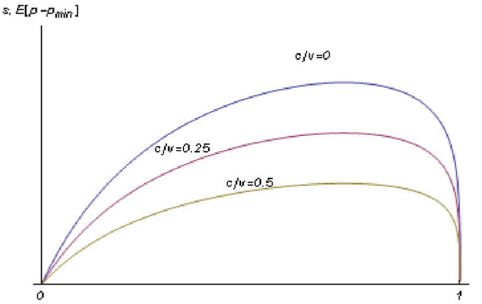
Comparative statics with respect to search cost can be thought of as changes in the number of shoppers or in the distribution of search costs for nonshoppers. In terms of the curves in Figure 2A, a decrease in search costs implies either a shift or a rotation of $G^{-1}(\cdot)$ to the right, arising 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, the price dispersion will increase or decrease depending on the initial equilibrium ($\mu^* \lesseqgtr \widehat{\mu}$). 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, as any relationship—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 ($\mu^* < \widehat{\mu}$)

[16] We show this visually in Figure 8 in Section 5.

[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.

FIGURE 3

PRODUCTION COST (c) AND PRICE DISPERSION



or decrease price dispersion ($\mu^* > \widehat{\mu}$). In Section 5, we use the price dispersion observed in the data for each fuel type to pin down this relationship. As 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.

We now analyze how changes in the number of firms and production costs 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 2A). Moreover, because 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.

However, the complete effect of the number of firms on price levels and dispersion needs to incorporate the reaction by rational consumers. For a given μ, price dispersion increases with n and so 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.[18] However, the total effect on the average posted price cannot be signed because more consumers searching pushes prices down. As shown in Figure 2B, the relationship between the average price and the number of firms is not monotonic.[19] Figure 2B 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 $c = p^* = \upsilon$). Given the search intensity, the cost pass-through is lower than 100% and increases with μ. Figure 3 shows—for any given level of search intensity—the negative effect of production cost on GS. 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

[18] Note that the number of firms in the market has no effect on consumers' search costs in a model of nonsequential search, although it may have an effect in a model of sequential search.

[19] Note that assuming a fixed consumer search intensity will imply a negative relationship between average prices and market concentration.

TABLE 1 Comparative Statics

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

Notes: $\overline{\mu}$ and μ refer to fixed and endogenous search intensity, respectively. n, number of firms. c, production costs. s, search costs. λ, consumers with zero search costs (shoppers). $PD = E[p - p_{\min}]$, price dispersion.

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%.

Table 1 summarizes the comparative statics predicted by our model.[20] 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 nonmonotonic 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).

As Varian (1980) points out, the main implication of imperfect information is that prices are dispersed across both sellers and time. The latter can be seen in a dynamic setting where the static game presented in this section becomes the stage game that is repeated in every period with the additional assumption that firms observe competitors' past prices before choosing their current prices. In this setting, a fixed price is a dominated strategy and there exists a Markov perfect equilibrium where firms play mixed strategies and draw new prices every period to avoid systematic undercutting by rival firms (Baye et al., 2004). The logic is similar to the one used to discard pure pricing strategies in the static game for any interior market search intensity. At the same time, "hit-and-run" pricing strategies keep consumers guessing as to which firms have sales in a particular period. It follows from the above description that the position of each firm in the price ranking varies over time.

The dynamic version of the model in this section predicts, all else constant, that firms draw *iid* prices over time. Note, however, that any change in μ^* affects the function from which firms draw prices. Additionally, we have assumed that consumers know firms' costs and so they observe any common cost shock experienced by the firms. The model becomes more involved when the current production cost is unknown and consumers update their priors from information acquired from previous purchases (Yang and Ye, 2008; Tappata, 2009). In these models, firms play mixed strategies in each period but the *iid* property is no longer valid, because the search intensity varies over time with consumers' expectations about firms' production cost.[21] In Section 4, we propose a simple and general test to assess the importance of search in the gasoline market based on changes in relative prices over time. We shall return, in Section 5, to analyzing the comparative statics predictions of the model, when we compare them to estimated results in the gasoline industry.

3. Data

■ In this section, we describe the data set 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

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

[21] The combination of product differentiation and costly consumer search is another reason why prices might not be *iid* over time. We discuss this in Section 4.

retail gasoline market. Our data set is unique in the sense that it covers more cross sectional observations and with 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 to May 2007).[22]

The data were originally collected by (Oil Price Information Service) OPIS 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 United States, 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 station-date-fuel type triple. We dropped some observations that 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 data set 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. As we do not observe daily data on rack or DTW prices, we use the spot price—which is observed daily—as a proxy for the shifts in the wholesale cost faced by stations. Figure A1 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 (the average spread is less than 1 cent and the correlation is above 0.99).[25]

Figure 4 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. 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 data set. 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, whereas 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, whereas premium is about 12–18 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.[26] This measure exhibits considerably less variation. As noted above, this value includes taxes and other state- and county-specific price

[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 data set, and 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 data set for 2006–2007.

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

[26] 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 4

WEEKLY RETAIL GASOLINE PRICE SERIES

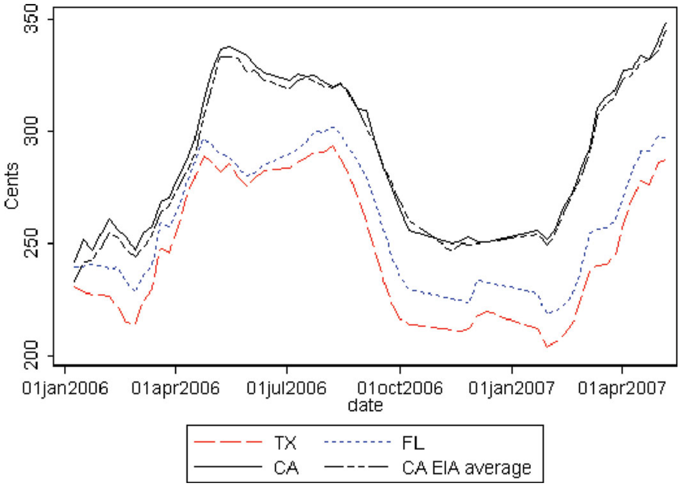


TABLE 2 **Station-Level Summary Statistics**

	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	

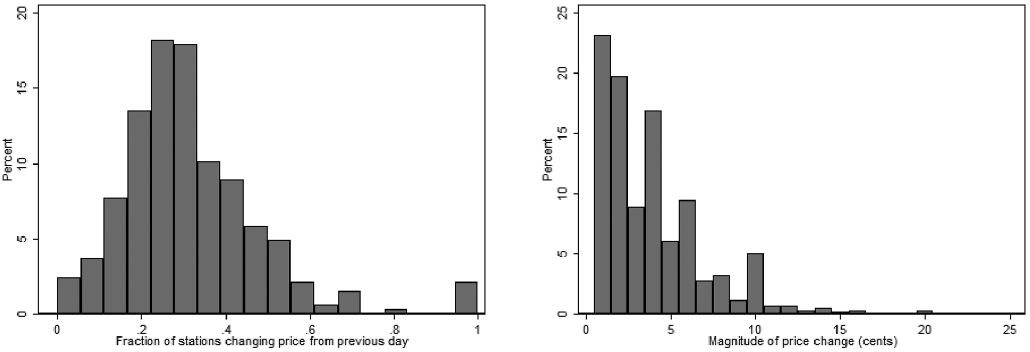
Notes: Standard deviations are in parentheses. Markups refer to retail price less spot price on that day.

differences. Nevertheless, it provides a better picture of the variation in prices according to fuel types and states.

Given the high volatility of wholesale costs, we expect firms to adjust gas prices frequently. Indeed, about a third of the gas stations in our data set change their prices with respect to the previous day. These price adjustments are not independent across stations, and some days represent a disproportionate share of stations changing prices. Figure 5A shows the distribution of the fraction of firms that change their price from the previous day. We find that the probability of such a price change is around 35% on weekdays but only around 20% on weekends. Additionally, the implied joint probability for two firms to both change their price from the previous day is around 12% on weekdays and 5% on weekends. We also examined the magnitude of price

FIGURE 5

DISTRIBUTION OF THE FREQUENCY AND MAGNITUDE OF DAILY PRICE CHANGES



(A) Frequency (B) Magnitude

changes. Figure 5B shows the distribution of the absolute value of price changes for regular gasoline. Conditional on a gas station having changed its price from the previous observation, the mean price change is 4.2 cents and the median is 3 cents.[27]

The data indicate that the timing of price changes is correlated with the distance between rival gas stations. In general, stations that are closer together are more likely to both change their prices from the previous day. Additionally, conditional on both stations adjusting their prices from the previously observed price, stations that are closer together are more likely to change their prices in the same direction.[28]

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 price levels, the station density in each market is similar across states. Table A2 shows the distribution of gasoline brands in the data set. 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% of the observed stations.[29]

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.[30,31]

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

[27] The unconditional mean and median are 1.4 and 0 cents, respectively.

[28] For example, gas stations that are within 262 feet of each other have a joint probability of both changing their prices within 1 day of 16.9% and an 85% probability of changing prices in the same direction, as compared with probabilities of 14.8% and 80% for a pair of stations that are more than 492 feet but less than 1 mile apart, respectively.

[29] This is consistent with other studies; Verlinda (2008), for example, reports around 9% of the station population in Orange County as being unbranded.

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

[31] 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 that 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 containing three or more stations.

TABLE 3 Market-Level Summary Statistics, 1 Mile Radius

	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)
Standard deviation 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	112,089		159,758		36,298		207,231	
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)
Standard deviation 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	15,341		25,799		593		9,871	
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)
Standard deviation 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	10,877		20,030		1,748		6,995	
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)
Standard deviation 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	9,786		13,626		4,092		25,202	

Notes: Standard deviations are in parentheses. Markets are restricted to having a minimum of three stations.

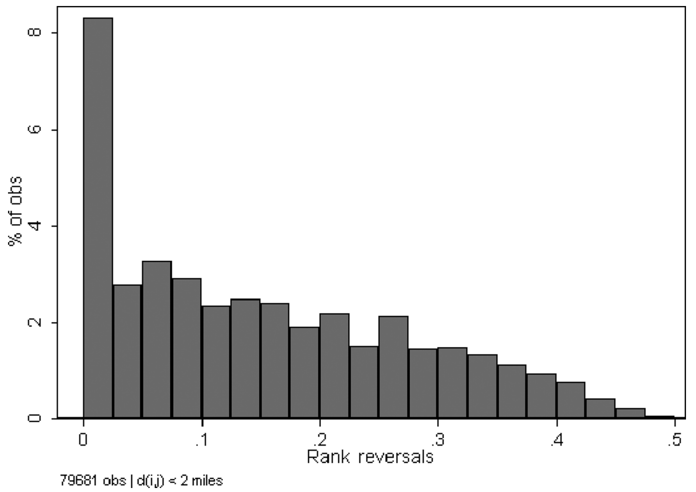
minimum price), and the gains from search in these markets. The last statistic is defined in equation (2). Table 3 contains summary statistics of price dispersion at the market level. The radius used to define markets in Table 3 was 1 mile; in Table A1, we present summary statistics on markets defined 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 three grades of gasoline, markets in California have the highest price dispersion among the four states, whereas 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, 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 data set 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 data set 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 four 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 data sets, weekly data from a single region, and small samples for a unique fuel type.[32]

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

FIGURE 6

TEMPORAL PRICE DISPERSION, REGULAR UNLEADED GASOLINE



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.[33] By contrast, models of consumer search, including the one presented in Section 2, predict temporal price dispersion (sales), because 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 6 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.[34] 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 octane rating, suggesting that the intensity of price competition is different across fuel types. We return to this point in the next

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

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

TABLE 4 Summary Statistics, Rank Reversals

	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

section. Figure A2 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 citywide price distributions. Therefore, the fact that we use station pairs that are within 1 or 2 miles of each other is an improvement over citywide samples. Additionally, the transition probabilities in prior studies are calculated for *residual prices*, that is, 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, because 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.[35] 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),[36] positive reversals could be reflecting the fact that we observe prices for stations at different moments.[37] 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 6 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.

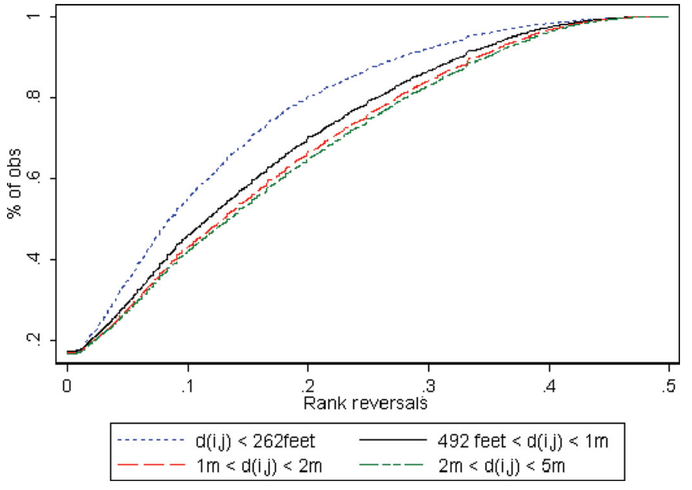
[35] These regressions are only valid if station fixed effects are additively separable from stations' costs. See Wildenbeest (this volume) for a specific example where this is the case. However, suppose the data generation process corresponds to a monopolistic competition model with firms setting deterministic 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.

[36] www.nytimes.com/2005/09/27/nyregion/27prices.html.

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

FIGURE 7

CDF OF RANK REVERSALS AND DISTANCE, REGULAR UNLEADED GASOLINE



The nature of the gasoline retailing market allows for such a test. Gas prices are prominently posted outside stations and visible to all drivers. Therefore, we expect that the price spread between gas stations that are located at the same street intersection reflects only product differences. Stations at 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 because every driver at that corner knows their prices.[38] However, these stations located at the same corner might choose their prices to compete for informed consumers with other distant stations and then distribute their captive customers based on product differences.[39] In other words, rank reversals are expected to happen less frequently for stations that are close to each other than for stations that are farther apart but still in the same market.

Assume that the rank reversals between stations at 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 nonparametric test that evaluates the alternative hypothesis of $F_1 \gtrless F_2$ against the null hypothesis H_0: $F_1 = F_2$. Figure 7 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.[40]

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

[39] 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.

[40] Note that the rank reversals for stations at the same corner are not zero (Figure 7). This may be related to measurement error because we define stations as being at 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 farther apart.

TABLE 5 Equality of Distributions Test for Rank Reversals: Corners versus Market

		1 Mile		2 Miles	
Fuel Type	Ha	D	p-Value	D	p-Value
Regular	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
Mid-grade	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
Premium	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
Diesel	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

Notes: H_0: $F_1(r_{ij}) = F_2(r_{ik})$, where $d_{ij} < 262$ ft and 492 ft $< d_{ik} < 1/2$ mi; $H_a(1)$: $F_1(r) > F_2(r)$; $H_a(2)$: $F_1(r) < F_2(r)$; $H_a(KS)$: $F_1(r) \neq F_2(r)$. D: Kolmogorov-Smirnov test statistic.

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 at 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, namely station pairs at 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 farther 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.

The rank-reversals 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 farther 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, because price spreads increase with the distance between stations.[41] That is, 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:

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

[41] 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 4).

TABLE 6 OLS and Quantile Regressions of Measures of Price Dispersion

Sample	Dependent Variable	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]**
$N = 25{,}345$	σ_{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]**
$N = 78{,}920$	σ_{ij}	−0.007	−0.009	−0.008	−0.007	−0.005
		[15.19]**	[29.45]**	[22.80]**	[11.34]**	[3.40]**

Notes: T statistics are in brackets. Values represent coefficients from a regression of the dependent variable on an indicator for whether the pair of stations is at the same corner. r and σ denote rank reversals and standard deviation of price spreads for station pairs, respectively.
*Significant at 5%.
**Significant at 1%.

where $s_{ijt} \in \mathbf{s}_{ij}$ and represents the price spread between stations (i, j) at day t, and $\bar{s}_{ij}$ is the average of T_{ij} days observed. Whereas 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 Figure A2B, where the price spread between the pair of gas stations is highly volatile but there are zero rank reversals.

Ordinary least squares 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.[42] The coefficient on the indicator, for the corner is negative and highly significant in all specifications.[43] The table presents results for regular unleaded; the results for mid-grade and diesel are similar, and not reported here. The results for premium using σ_{ij} are also similar. However, results for premium using r_{ij} are weaker, as was the case with the Kolmogorov-Smirnov 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). 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 farther apart. These parametric tests bear out the previous results using the nonparametric 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. Although this does not seem to be the case for the retail gasoline market, temporal cost shocks, if any, are expected to be correlated across stations carrying the same brand.[44] Given that stations of the same brand are unlikely to locate at the same corner

[42] Results using σ_{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.

[43] The results are robust to adding various controls, such as the distance between stations and state fixed 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.

[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

(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 7.

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 farther 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, rank reversals for pairs of stations located in the same market (1 mile apart) 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 stations fill their underground storage tanks at different moments. Despites doubts about the rationality of this behavior, we carry out the same test of rank reversals but restrict the observations to those for which both prices in the station pair change with respect to the last reported prices.[45] Table A4 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 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 A4 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 data set, that allow us to conduct this test.

5. Equilibrium price dispersion

■ The previous section used temporal price dispersion to establish the importance of imperfect information in gasoline markets. We now examine price dispersion across markets.[47] We have two goals in this section. First, we characterize the equilibrium relationships between price dispersion and key parameters: marginal costs and the number of firms in the market. We compare our estimates to the predictions of our stylized search model of Section 2; refer to Table 1 for

time), as stations are located within 1–2 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, which 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" (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; Noel, 2007a). The evidence in the United States is not as clear (Hosken et al., 2008), and research indicates that Edgeworth cycles in the United States are mainly concentrated in the Midwest, which does not apply to our data set (Lewis, 2009).

[47] The empirical exercise in this section uses both cross sectional and temporal variation in price dispersion. However, we compare our results to the stage game in Section 2.

TABLE 7 Price Dispersion Regressions

Sample	Prices		Residuals		Residuals[a]	
Dependent Variable	Range	Standard deviation	Range	Standard deviation	Range	Standard deviation
Cost	−0.013	−0.006	−0.053	−0.027	−0.061	−0.033
	[4.64]**	[4.94]**	[6.98]**	[6.94]**	[7.96]**	[7.97]**
Rival firms	0.317	0.089	0.382	0.109	0.252	0.11
	[16.98]**	[10.64]**	[17.51]**	[10.97]**	[8.22]**	[7.01]**
Premium	2.551	1.343	1.75	0.947	2.061	1.103
	[12.40]**	[12.90]**	[9.93]**	[10.36]**	[10.32]**	[10.33]**
Regular	−0.664	−0.587	−3.535	−2.082	−3.723	−2.000
	[4.94]**	[8.58]**	[16.13]**	[18.72]**	[16.59]**	[16.69]**
Constant	12.243	6.706	15.243	8.377	15.655	8.393
	[18.06]**	[21.39]**	[34.58]**	[35.34]**	[31.41]**	[30.65]**
N	606,630		606,630		196,335	
R^2	0.07	0.07	0.15	0.15	0.16	0.16

Notes: T statistics are in brackets. Standard errors are clustered in two ways: by central gas station and by date. Columns 1–2 use observed prices. Columns 3–6 use demeaned prices. All regressions include state fixed effects.
[a] Nonoverlapping markets.
**Significant at 1%.

these predictions.[48] Second, we attempt to pin down the equilibrium relationship between price dispersion and search intensity.

Our specification is

$$PRICEDISP_{jkt} = \beta_0 + \beta_1 MC_t + \beta_2 N_j + \gamma_k + \varepsilon_{jt}, \tag{4}$$

where $PRICEDISP_{jkt}$ denotes various measures of price dispersion in market j for fuel type k on date t, MC_t is a measure of the marginal cost faced by gas stations on date t, N_j is the number of stations in market j, and γ is a fuel type indicator. The results of estimating equation (4) are presented in columns 1 and 2 of Table 7. Results are presented separately for two dependent variables: the range and the standard deviation of prices. We have combined observations across all four states and the three gasoline fuel types, with fixed effects for fuel type and state included. Our measure of marginal cost is the same-day spot price of the wholesale market corresponding to the state. Standard errors are clustered in two ways: by market (i.e., central station) and by date, as all firms on the same date experience the same marginal cost shock.

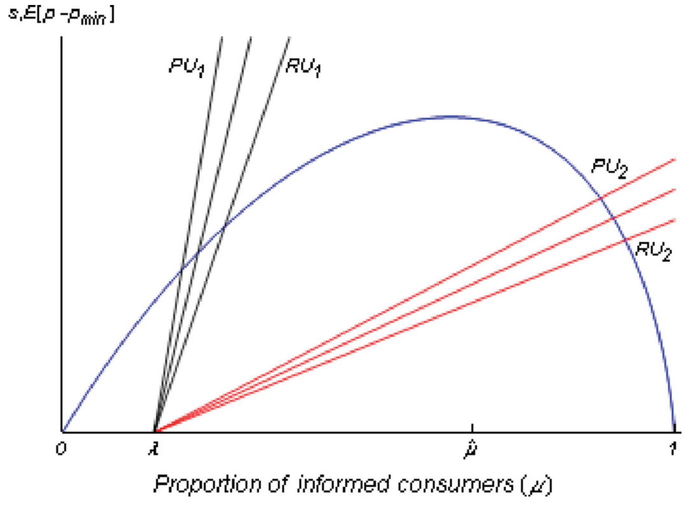
The coefficient on firms' marginal costs is negative, whereas that on the number of rival firms within 1 mile is positive, both of which are consistent with our search model of Section 2. Note that production costs and the number of firms in the market were taken to be exogenous in the model. Whereas it is reasonable to treat the spot price of gasoline as exogenous in our empirical exercise as well, the number of firms is likely to depend on market characteristics that may also affect price dispersion. Although there is precedent in the literature for treating the number of firms (and the existence of products) as exogenous, Table 7 must be interpreted with caution.

We now turn to robustness checks. In columns 3 and 4, we use residual or "cleaned" prices rather than actual prices to construct the dependent variable. This has been the traditional approach in the empirical literature, that is, the analysis has focused on the price dispersion that remains after product characteristics are controlled for. As noted in Section 4, this exercise makes certain assumptions about the interaction between product differentiation and imperfect information, namely that they are additively separable. The results using residuals are consistent with those that use raw prices.

[48] Note that the model assumed homogeneous products, whereas in reality retail gasoline is differentiated along a number of dimensions. We attempt to control for product differentiation below, but we acknowledge that we cannot directly test the model of Section 2.

FIGURE 8

CONSUMER SEARCH COST AND EQUILIBRIUM SEARCH INTENSITY



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. This implies 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. Doing so leaves us with approximately 30% of the original markets. Results using these nonoverlapping markets are presented in columns 5 and 6 and are consistent with the previous results.

The results relating market structure to price dispersion are counter to those of Barron et al. (2004), but similar to those of Lach and Moraga-Gonzalez (2009), and Lewis (2008) for local submarkets. 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.[49]

Using the results of column 5, the average price dispersion for premium gasoline is 13% larger than for mid-grade (which is the omitted category), which in turn is 30% larger than for regular.[50] We expect the demand for higher-octane fuel to be stronger for people driving expensive cars, which should be correlated with higher income and search costs.[51] If this assumption is correct, it implies that the price dispersion for higher-octane grades will be lower (higher) when the search intensity is low (high). This can be seen in Figure 8, 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 ($\mu^* < \widehat{\mu}$), whereas in scenario 2 it is high.

[49] Interestingly, Lewis and Marvel (2011) report that traffic to websites where gasoline prices are self-reported increases when oil prices are high.

[50] These results were obtained by controlling for other variables as in equation (4). However, even the unconditional differences in price dispersion have similar patterns, as can be seen in Table 3.

[51] 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.

As noted earlier, the nonmonotonic 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 8. Thus, our results suggest that the relationship between search intensity and price dispersion in the gasoline market is negative.[52] We also conclude that the market for regular gasoline is the most competitive, followed by mid-grade and then premium.

6. Conclusions

■ Studies that link price dispersion and costly consumer search usually address two questions: to what extent is observed price dispersion consistent with pricing strategies from costly consumer search models? And what is the equilibrium relationship between price dispersion and search costs, production costs, and the number of firms in the market? The focus in the literature has primarily been on the latter question, and in this article we shift the attention to the former. We argue that identifying the role of consumer search in explaining price dispersion requires a 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 street intersection. This is consistent with the theory of consumer search, as the dispersion in the latter group is driven only by product differentiation.

We use a unique high-frequency panel data set to examine equilibrium price dispersion in the U.S. retail gasoline industry. Our results imply that consumers could save as much as 5% by price shopping within a 1 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 and are less competitive. Premium gasoline has 13% more price dispersion than mid-grade, which itself has 30% more price dispersion than regular gasoline. 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 websites where users periodically list stations' gas prices may be one step in this direction.[53] Moreover, our results indicate that price dispersion decreases when the aggregate level of prices rises, 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 suboptimal. Increased information along these lines may help consumers to make better decisions regarding their search strategies.

[52] 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.

[53] 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.

Appendix

FIGURE A1

REGULAR UNLEADED RETAIL AND WHOLESALE PRICES IN CALIFORNIA

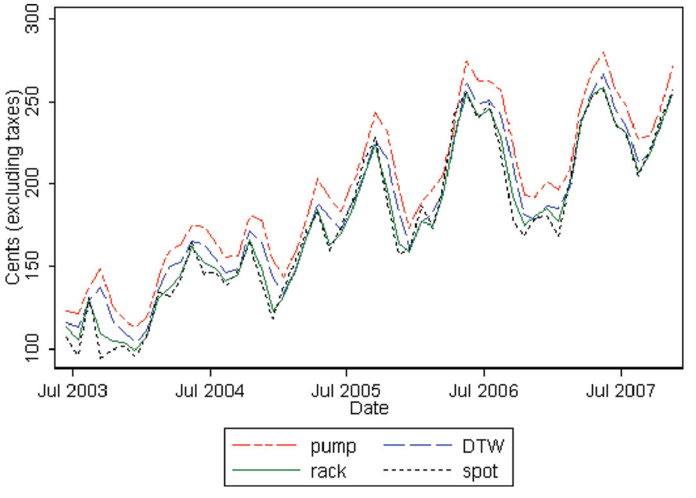
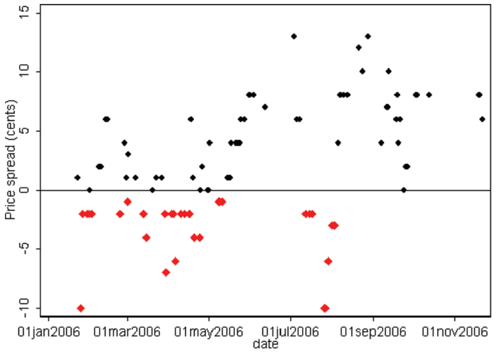
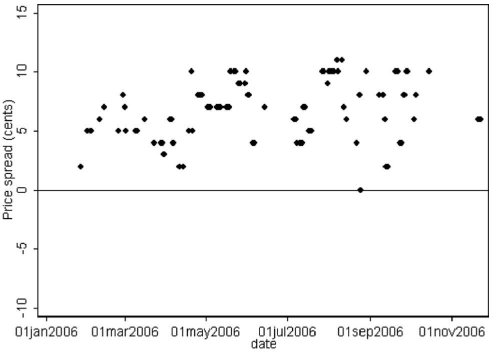


FIGURE A2

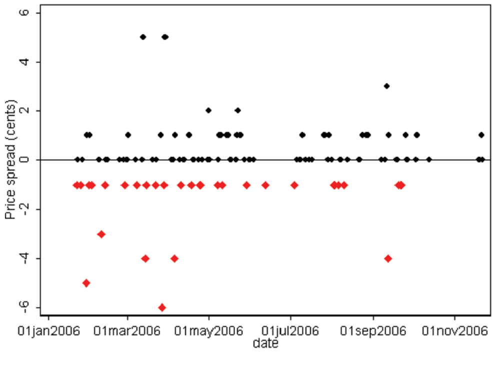
PRICE SPREAD PATTERNS BETWEEN PAIRS OF STATIONS



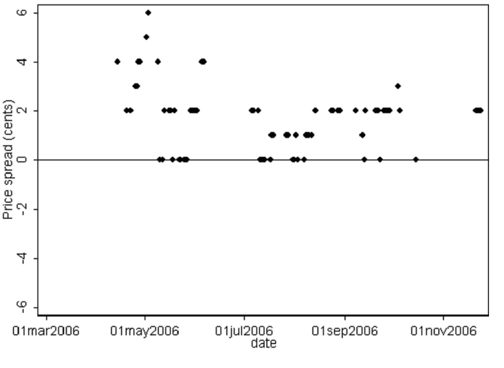
(A) High average spread, high $r_{i,j}$



(B) High average spread, zero $r_{i,j}$



(C) Low average spread, high $r_{i,j}$



(D) Low average spread, zero $r_{i,j}$

TABLE A1 Market-Level Summary Statistics, 2 Mile Radius

	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)
Standard deviation 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	318,314		386,031		96,275		496,152	
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)
Standard deviation 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	71,469		94,492		3,470		44,171	
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)
Standard deviation 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	49,753		73,772		8,194		30,085	
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)
Standard deviation 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	51,605		53,971		15,168		101,676	

Notes: Standard deviations are in parentheses. Markets are restricted to having a minimum of three stations.

TABLE A2 Distribution of Observed Brands across States

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 A3 Summary Statistics, Rank Reversals Conditioning for Price Changing

	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 A4 Rank-Reversals Test Conditioning on Price Changing: Corners versus Markets

		1 Mile		2 Miles	
Fuel Type	Ha	D	*P*-Value	D	*P*-Value
Regular	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
Mid-grade	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
Premium	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
Diesel	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

Notes: Rank reversals calculated only when both stations change prices with respect to the last observation. D: Kolmogorov-Smirnov test statistic.

References

BARRON, J.M., TAYLOR, B.A., AND UMBECK, J.R. "Number of Sellers, Average Prices, and Price Dispersion." *International Journal of Industrial Organization*, Vol. 22 (2004), pp. 1041–1066.

BAYE, M.R., MORGAN, J., AND SCHOLTEN, P. "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site." *Journal of Industrial Economics*, Vol. 52 (2004), pp. 463–496.

———, ———, AND ———. "Information, Search, and Price Dispersion." In T. Hendershott, ed., *Handbook in Economics and Information Systems*. Amsterdam: Elsevier Press, 2006.

BORENSTEIN, S., CAMERON, A.C., AND GILBERT, R. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *Quarterly Journal of Economics*, Vol. 112 (1997), pp. 305–339.

BROWN, J.R. AND GOOLSBEE, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy*, Vol. 110 (2002), pp. 481–507.

CARLSON, J.A. AND MCAFEE, R.P. "Discrete Equilibrium Price Dispersion." *Journal of Political Economy*, Vol. 91 (1983), pp. 480–493.

CLAY, K., KRISHNAN, R., AND WOLFF, E. "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry." *Journal of Industrial Economics*, Vol. 49 (2001), pp. 521–539.

DAHLBY, B. AND WEST, D.S. "Price Dispersion in an Automobile Insurance Market." *Journal of Political Economy*, Vol. 94 (1986), pp. 418–438.

ECKERT, A. "Retail Price Cycles and the Presence of Small Firms." *International Journal of Industrial Organization*, Vol. 21 (2003), pp. 151–170.

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

HORTAÇSU, A. AND SYVERSON, C. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." *Quarterly Journal of Economics*, Vol. 119 (2004), pp. 403–456.

HOSKEN, D.S., MCMILLAN, R.S., AND TAYLOR, C.T. "Retail Gasoline Pricing: What Do We Know?" *International Journal of Industrial Organization*, Vol. 26 (2008), pp. 1425–1436.

LACH, S. "Existence and Persistence of Price Dispersion: An Empirical Analysis." *Review of Economics and Statistics*, Vol. 84 (2002), pp. 433–444.

——— AND MORAGA-GONZALEZ, J. "Asymmetric Price Effects of Competition." Discussion Paper no. 7319, CEPR, 2009.

LEWIS, M. "Price Dispersion and Competition with Differentiated Sellers." *Journal of Industrial Economics*, Vol. 56 (2008), pp. 654–678.

——— "Temporary Wholesale Gasoline Price Spikes Have Long-Lasting Retail Effects: The Aftermath of Hurricane Rita." *Journal of Law and Economics*, Vol. 52 (2009), pp. 581–605.

——— AND MARVEL, H. "When Do Consumers Search?" *Journal of Industrial Economics*, Vol. 59 (2011), pp. 457–483.

MASKIN, E. AND TIROLE, J. "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles." *Econometrica*, Vol. 56 (1988), pp. 571–599.

NOEL, M. "Edgeworth Price Cycles: Evidence from the Toronto Retail Gasoline Market." *Journal of Industrial Economics*, Vol. 55 (2007a), pp. 69–92.

———. "Edgeworth Price Cycles, Cost-Based Pricing and Sticky Pricing in Retail Gasoline Retail Markets." *Review of Economics and Statistics*, Vol. 89 (2007b), pp. 324–334.

PNG, I.P.L. AND REITMAN, D. "Service Time Competition." *RAND Journal of Economics*, Vol. 25 (1994), pp. 619–634.

SETIAWAN, W. AND SPERLING, D. "Premium Gasoline Overbuying in the U.S.: Consumer-Based Choice Analysis." Discussion Paper Research Report UCD no. ITS-RR. 93-25, Institute of Transportation Studies, UC Davis, 1993.

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

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

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

STAHL, D.O. "Oligopolistic Pricing with Heterogeneous Consumer Search." *International Journal of Industrial Organization*, Vol. 14 (1996), pp. 243–268.

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

TAPPATA, M. "Rockets and Feathers: Understanding Asymmetric Pricing." *RAND Journal of Economics*, Vol. 40 (2009), pp. 673-687.

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

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

VERLINDA, J. "Do Rockets Rise Faster and Feathers Fall Slower in an Atmosphere of Local Market Power? Evidence from the Retail Gasoline Market." *Journal of Industrial Economics*, Vol. 56 (2008), pp. 581–612.

YANG, H. AND YE, L. "Search with Learning: Understanding Asymmetric Price Adjustments." *RAND Journal of Economics*, Vol. 39 (2008), pp. 547–564.