The Determinants and Consequences of Search Cost Heterogeneity: Evidence from Local Gasoline Markets*

Mitsukuni Nishida† Marc Remer‡

February 25, 2016

Abstract

Information frictions play a key role in an array of economic activities and are frequently incorporated into formal models as search costs. However, little is known about the underlying source of consumer search costs and how heterogeneous they are across consumers and markets. This paper documents the sources and magnitude of heterogeneity in consumer search costs and analyzes how the heterogeneity shapes firms' pricing and consumers' search behavior. By identifying hundreds of geographically isolated markets, we are the first to estimate the distribution of consumer search costs for many geographic markets. We directly recover the distribution of consumer search costs market by market using retail gasoline price data in the United States. We find that the distribution of consumer search costs varies significantly across geographic markets and that the distribution of household income is closely associated with the search cost distribution. We find that a policy that reduces both the standard deviation and mean of the search cost distribution has heterogeneous and potentially unintended consequences on prices across markets.

Keywords: Search Costs, Consumer Search; Price Dispersion; Pricing; Retail Gasoline.

*We have benefited from discussions with Tat Chan, Maqbool Dada, Babur De Los Santos, Emin M. Dinlersoz, Anthony Dukes, Amit Gandhi, Joe Harrington, Elisabeth Honka, Han Hong, Ali Hortaçsu, Sergei Koulayev, José Luis Moraga-González, Aviv Nevo, Zsolt Sándor, Philipp Schmidt-Dengler, Stephan Seiler, Matt Shum, Chad Syverson, Matthijs Wildenbeest, Xiaoyong Zheng, and seminar participants at Swarthmore College, Lehigh University, Wake Forest University, US Department of Justice, Kyoto University, Johns Hopkins University, North Carolina State University, Osaka ISER, University of Tokyo, Center for Economic Studies at US Census Bureau, 2013 IIOC, and the 5th workshop on consumer search and switching costs. We thank Reid Baughman, Autumn Chen, Sam Larson, Jianhui Li, and Karry Lu for research assistance. The retail gasoline price data for this project were generously provided by Mariano Tappata.

†The Johns Hopkins Carey Business School, 100 International Drive Baltimore, MD 21202. Email: nishida@jhu.edu.

‡Swarthmore College, 500 College Avenue, Swarthmore, PA 19081. Email: mremer1@swarthmore.edu.

1 Introduction

Information frictions play a key role in explaining many aspects of economic activity. Since Stigler's (1961) seminal article, a number of influential theoretical papers, such as Varian (1980), Burdett and Judd (1983), and Stahl (1989), demonstrate that consumer search costs can lead to competing firms that sell homogeneous products playing mixed strategies in price to capture sales from low and high search cost consumers, thereby resulting in price dispersion.[1] A key determinant of prices and welfare in such models is the distribution of consumers' search costs.

Consumer search costs are not just a theoretical curiosity; firms are often constrained by government regulations that seek to lower the cost of consumers' search. For instance, attempting to increase price transparency, some regulators have created online retail gasoline price aggregators that are accessible by smartphones. The purpose of such regulation is to decrease the cost of consumer search, but it also serves to make search costs more homogeneous across consumers. As we detail in this article, this regulation can have the unintended consequence of substantially increasing prices and hurting consumers. Similarly, governments often regulate how gas stations display gas prices for different payment methods (i.e., cash and credit card), which also affects both the level and variance of consumer search costs.

Despite the importance of search costs in theoretical models and its implications for firms and policy, we know surprisingly little about the nature of search costs and how they vary across markets. In particular, there is little evidence documenting the determinants and consequences of heterogeneous consumer search costs. Consequently, there still exists competing hypotheses as to the underlying source of search costs: opportunity costs of time or search efficiency (Goldman and Johansson 1978). We fill this gap in the literature by documenting the determinants of search cost heterogeneity within and across markets. In doing so, we provide evidence on the underlying source of consumer search behavior, and find that the opportunity cost of time is an important factor. To perform the analysis, we structurally estimate search cost distributions for hundreds of markets and then analyze the extent to which demographics explain the shape of the distributions. We then conduct counterfactual experiments in each market to demonstrate how policy affecting the mean and variance of the search cost distribution impacts equilibrium prices.

The analysis proceeds in three steps. First, we document the heterogeneity of search cost distributions both within and across retail gasoline markets. Because competition between retail gasoline stations is highly localized, we identify 367 geographically isolated local markets in the spirit of the firm entry literature initiated by Bresnahan and Reiss (1991). We then structurally estimate the search cost distribution separately for each market, and find that the distributions vary significantly across markets. This finding suggests that search cost estimates from a single market may be of limited value in making generalizable policy predictions.

Second, we explore the underlying source of consumer search costs. For this purpose, we analyze the extent to which observable consumer characteristics can explain variation in the distribution

[1] See Baye, Morgan, and Scholten (2006) for a broad review of the consumer search and price dispersion literature. See Ratchford (2009) for a review of consumer search and pricing in the marketing literature.

of consumer search costs across markets. We find that both the mean and variance of the search cost distribution in a market is closely related to the mean and variance of household income in a market; these results suggest opportunity costs partly drive consumers' search costs.

Finally, using counterfactual experiments, we analyze the consequences of heterogeneity in search costs on prices and consumer welfare. Interestingly, we find that a policy that decreases both the mean and standard deviation of search costs, as would most policy aimed at increasing price transparency, may have the unintended consequence of raising prices and decreasing consumer welfare. On one hand, a reduction in the mean of search costs incentivizes consumers to search more, which puts downward pressure on prices. On the other hand, a reduction in the standard deviation reduces the fraction of (relatively) low search cost consumers, which, in turn, results in firms increasing market prices. We demonstrate that the latter effect may dominate the former in some markets.

This article benefits from extensive information on daily retail gasoline prices from almost every station in the states of California, Florida, New Jersey, and Texas. Consumer search is an important feature of retail gasoline markets (Marvel 1976; Chandra and Tappata 2011), and therefore offers a natural setting in which to investigate the prevalence and heterogeneity of search costs. Other features of retail gasoline markets that are favorable to our analysis is that prices change frequently, product characteristics are largely fixed over the sample period, and consumer stockpiling is uncommon.

To perform the analysis, we leverage a model of fixed-sample search similar to Burdett and Judd (1983) and directly recover the consumer search cost distribution that rationalizes observed prices as an equilibrium outcome by firms and consumers (Hong and Shum 2006; Wildenbeest 2011). Only price data are required to estimate search costs with this methodology, which is useful because station-specific sales or indirect measures of search behavior, such as internet usage (Brown and Goolsbee 2002), are typically unavailable for local retail gasoline markets. To rationalize search costs from variation in retail prices, we first control for intertemporal variation in wholesale costs and time-invariant price differences across firms in a market that result from vertical product differentiation. In doing so, we extend the Wildenbeest (2011) model to directly incorporate cost shifters in a manner consistent with the theoretical model. Controlling for wholesale costs and vertical product differentiation removes 81% of observed price variation; we exploit the remaining variation to estimate the structural model. We therefore recover the search cost distribution for each market utilizing all intertemporal and cross-sectional variation in prices within a market, conditional on station fixed effects and wholesale costs.

We proceed as follows. The remainder of this section discusses how this article relates to the existing literature. Section 2 describes the data and details how we choose markets within which to perform the estimation. Section 3 presents a reduced-form analysis of price dispersion and mixed strategies. Section 4 details the model. Section 5 presents the estimation strategy. Section 6 describes the estimation results. Section 7 conducts the counterfactual experiments and discusses policy implications. Finally, Section 8 concludes.

1.1 Related Literature

This article relates to at least four strands of existing literatures. First, it is a continuation of a recent body of research in economics and marketing that structurally estimates consumer search costs from price data. Existing work focuses on recovering a single search cost distribution from one product or geographic market, such as the U.S. mutual funds (Hortacsu and Syverson 2004), online textbooks (Hong and Shum 2006), online memory chips (Moraga-González and Wildenbeest 2008), and a chemical product (Zhang, Chan, and Xie 2015). Because these studies estimate one search cost distribution, typically for a single market, how search costs vary across different geographies and populations remains unknown. We, by contrast, identify hundreds of geographically isolated markets and estimate substantial differences between the search cost distributions across markets. Appendix Table A1 compares this paper with recent work that structurally estimates search costs.

Second, our work relates to the broad literature on consumer search. Several studies, such as Ratchford and Srinivasan (1993), Ratchford, Lee, and Talukdar (2003), Fox and Hoch (2005), Gauri, Sudhir, and Talukdar (2008), and Brynjolfsson, Dick, and Smith (2010), focus on the consumer side and document the marginal gains of search. We consider a model in which both consumers and firms optimize their choices in an equilibrium. There is also a growing literature on search costs and competition (see, e.g., Bakos 1997; Brynjolfsson and Smith 2000; Brown and Goolsbee 2002; Brynjolfsson, Hu, and Rahman 2009; Anderson, Fong, Simester, and Tucker 2010; Brynjolfsson, Hu, and Simester 2011). These studies focus on the level of search costs in online marketplaces, whereas we analyze the role of search cost heterogeneity in traditional retail markets. We complement Kuksov (2004) and Branco, Sun, and Villas-Boas (2012), who theoretically study the impact of lowering search costs on pricing. Moraga Gonzalez, Sándor, and Wildenbeest (2016) show through numerical simulations how the effect of firm entry on prices depends on underlying search cost heterogeneity within a market.

Third, this article fits into the literature on the estimation of consumer search and consideration-set formation (Roberts and Lattin 1991; Mehta, Rajiv, and Srinivasan 2003; Kim, Albuquerque, and Bronnenberg 2010; De Los Santos, Hortacsu, and Wildenbeest 2012; Koulayev 2014; Honka 2014). Given that we only observe prices, our approach abstracts away from the decision of which firms to search, and instead models the size of the consideration set (i.e., how many price quotes to obtain). In contrast to these search models of consumer demand, we estimate a model that jointly characterizes search decisions by consumers and firm pricing decisions, which enables us to examine the equilibrium effect of search costs on pricing.

Finally, this article relates to the literature on price dispersion, consumer search, and pricing in retail gasoline markets. A number of studies, such as Marvel (1976), Barron, Taylor, and Umbeck (2004), Lewis (2008), Lewis and Marvel (2011), Chandra and Tappata (2011), and Pennerstorfer, Schmidt-Dengler, Schutz, Weiss, and Yontcheva (2015), have identified patterns of temporal and cross-sectional price dispersion in retail gasoline markets that are consistent with models of costly consumer search. Because search costs are not directly observed in these studies, the evidence has been limited to carefully executed reduced-form testing of comparative static relationships implied by a particular theoretical model. By contrast, we structurally back out the search cost distributions that rationalize the model and observed data. We contribute to the literature that examines more generally the pricing of retail gasoline. Firgo, Pennerstorfer, and Weiss (2015) estimate the pricing power afforded to retail gasoline stations that are centrally located in a network. Chan, Padmanabhan, and Seetharaman (2007) propose an empirical model of pricing and location decisions of gas retailers in Singapore. Iyer and Seetharaman (2003) examine price discrimination in the Greater St. Louis retail gasoline area. Using a model of product and pricing decisions, Iyer and Seetharaman (2008) investigate why some markets observe similar gas prices and other markets do not. We contribute to this literature by better understanding the role that search frictions play in retail gasoline pricing.

2 The Data

2.1 Price and Wholesale Cost Data

We utilize a large panel data set of daily, regular fuel gasoline prices for individual gas stations. The data originate from the Oil Price Information Service (OPIS), which obtains data either directly from gas stations or indirectly from credit card transactions. Academic studies of the retail gasoline industry have frequently relied upon OPIS's data (e.g., Lewis and Noel 2011; Taylor, Kreisle, and Zimmerman 2010; Chandra and Tappata 2011). Using a subset of the price data employed in Chandra and Tappata (2011), we estimate the model with data from gas stations in California, Florida, New Jersey, and Texas from February 27 through March 28, 2007. We refer interested readers to Chandra and Tappata (2011) for a more detailed description of the data.

The structural model requires only price data, and reconciles the observed price dispersion as a consequence of vertical product differentiation and heterogeneity in consumers' cost of search. In the full set of price data, which spans January 4th, 2006 through May 16th, 2007, prices may vary over time, in part, in response to changes in the wholesale cost of gasoline. To minimize the impact of input cost changes on observed price dispersion, we incorporate the spot price of wholesale gasoline directly into the empirical analysis and use only the 30 consecutive days of data that minimize variation in wholesale costs.

We utilize the spot price of wholesale gasoline traded on the New York Mercantile Exchange (NYMEX) as the measure of wholesale costs. These data are commonly employed as a measure of marginal cost in studies of the retail gasoline industry (e.g., Borenstein, Cameron, and Gilbert 1997; Velinda 2008; Lewis 2011). February 27th through March 28th, 2007 are the dates that minimize wholesale cost variation; the standard deviation of the daily New York Harbor spot price of wholesale unleaded fuel was 6.3 cents per gallon over that time. To address regional differences in wholesale prices, we use the daily spot price of reformulated gasoline delivered to the New York Harbor, Gulf Coast, and Los Angeles[2] for gas stations located in those respective regions. To

[2] These prices are available publicly through the U.S. Energy Information Administration (eia.gov). City-level rack wholesale prices are likely a more accurate measure of marginal cost, but such data are proprietary.

further account for marginal cost movement creating price dispersion in the data, we estimate the model using a firm's markup of price over wholesale cost ($p - c$) rather than its price. Section 5 develops how using markups rather than prices is consistent with the theoretical model. Estimation results are qualitatively similar whether the model is estimated using prices or markups, likely due to a priori selecting a time period with minimal wholesale cost variation.

2.2 Location Data and Selecting Isolated Markets

In the structural model, prices are generated by an equilibrium in which a defined set of gas stations compete for the same set of customers. Markets in the data must therefore be carefully defined to be consistent with this assumption. In retail gasoline markets, competition among stations is highly localized (Eckert 2013), and typically the literature defines a market as a specified radius around each firm in the data (see, e.g., Hastings 2004; Barron, Taylor, and Umbeck 2004; Lewis 2008; Remer 2015). A potential source of estimation bias inherent in this market definition is the existence of overlapping markets; two firms within a specified distance that compete against each other may belong to multiple markets, but each firm may not belong to the same total sets of markets.

To circumvent this problem, we focus the analysis on "isolated" markets, which we define in the spirit of Bresnahan and Reiss's (1991) geographic markets. We use two strict criteria. First, we define an isolated market as a set of firms all within 1.5 miles of each other, and no other competitor is within 1.5 miles of any firm in the market. Formally, let J be a set of firms and let $d(i, j)$ be the distance between any two firms $i, j \in J$. Then, J is an isolated market if (a) for all $i, j \in J$, $d(i, j) \leq X$, and (b) for all $k \notin J$ and $i \in J$, $d(i, k) > X$. We choose 1.5 miles as the maximum distance between firms in a market (X), which is consistent with Barron, Taylor, and Umbeck (2004), Hosken, McMillan, and Taylor (2008), and Lewis (2008). Restricting the analysis to isolated markets ensures that prices in a market are not influenced by competition with unobserved competitors, and therefore stands as an improvement over previous literature that allows firms to belong to multiple overlapping markets. Nonetheless, Moraga-González and Wildenbeest (2008) do find that the estimation methodology employed in this article may be robust to missing data entirely from some firms in the market. To implement the isolated market definition, we converted the street address of each gas station to longitude and latitude coordinates using ArcGIS, and then cross-referenced them with coordinates outputted from Yahoo maps, and finally, we calculated the geodesic distance between each pair of gas stations.

Second, to analyze the relationship between search costs and census tract-level characteristics, we further restrict isolated markets to only include markets where all gas stations are located within a single census tract.[3] Figure 1a depicts a map containing an isolated market with two firms within a single census tract ("Market 1"). Using only isolated markets, we estimate the model using prices from $1,145$ stations in 367 isolated (and single census tract) markets. Figure 1b shows

[3] Manuszak and Moul (2009) use a census tract as the unit of observation to analyze the relationship between consumer heterogeneity and travel costs in retail gasoline markets.

the distribution of these markets by the number of gas stations. Defining isolated markets does not solve all potential issues; for example, people may purchase gasoline not only where they live, but also along their commute to work. Still, the isolated market definition is more stringent than those used by most of the existing literature. A notable exception is Houde (2012), who directly incorporates information on commuting routes in the Quebec City gasoline market.

[Insert Figures 1a and 1b about here.]

Table 1 presents census tract population summary statistics for isolated markets and the remaining non-isolated markets in the data; the structural estimation of consumer search costs in isolated markets incorporates this data. The data are taken from the 2006-2010 American Community Survey (ACS), an ongoing survey conducted under the auspices of the US Census Bureau. Mean income and age are taken directly from the survey, while mean years of education are calculated by taking a weighted average of the people in a census tract who have reached a particular educational attainment. The ACS reports the number of households that fall within a particular income bracket; by assuming that average household income within a bracket is the mid-point of the bracket, we calculate the standard deviation of income.

[Insert Table 1 about here.]

Table 1 demonstrates that isolated and non-isolated markets are similar in terms of income, education, and age, suggesting the subsequent results are generalizable to other markets. The only substantive difference between isolated and non-isolated markets is that gas stations tend to be closer together in isolated markets. This is because we restrict isolated markets to have stations that fall within a single census tract, which tend to be smaller than a circle of 1.5 mile radius.

3 Evidence of Price Dispersion

3.1 Reduced-form Evidence

A large body of economic literature documents price dispersion in retail gasoline markets, and our data are no exception. The isolated markets in the data are characterized by substantial price dispersion, which is summarized in Table 2. The first column of Table 2 pools the prices across all isolated markets and 30 days of data from February 27 through March 28, 2007, and presents summary statistics for the distribution of all prices in the data. Column 2 presents the same for markups of price over wholesale costs.[4] The average price (markup) in the data we use to estimate search costs is $3.12 ($0.76) and the range (i.e., maximum minus minimum price) and standard deviation of all prices (markups) is $1.00 and $.099 ($1.03 and $.089), respectively. To facilitate comparison of prices across states, state fixed effects are removed from the prices, which

[4]When we estimate the structural model using markups rather than prices the estimates converge in 13 more markets (367 vs. 354). We present summary statistics only for markets where we obtain estimates, and therefore columns 2 and 4 have more observations.

largely controls for state taxes. It is of note that the price and markup distributions have a very similar range and standard deviation. If price dispersion were largely a consequence of wholesale price movement, retail prices would be much more dispersed than markups. In the extreme, if gas stations set price as a fixed markup over cost then we would observe price dispersion, but no variation in markups.

[Insert Table 2 about here.]

Columns 3 and 4 in Table 2 characterize the distribution of prices and markups, respectively, within isolated markets and provide evidence on how the distributions vary across markets. For each statistic, such as the mean retail price, we first calculate the statistic within a market, and then take the average of it across all isolated markets. For example, on average across markets, the mean price within a market for all 30 days of data is $3.12 and the standard deviation of the mean price across markets is $0.07. The average range and standard deviation of prices across markets are only slightly larger than those of markups, again suggesting that wholesale cost changes are not a major determinant of price dispersion in our data. This finding is likely due to selecting a time of minimal wholesale cost movement, as, in general, wholesale gasoline prices change over time and are an important determinant of retail gasoline price movement.

Figures 2a and 2b compare the distribution of all prices and markups to the distributions within three isolated markets, each with three gas stations. The first panel of Figure 2a depicts the distribution of all observed prices in isolated markets. The distribution of all prices has a much larger range and variance than the price distributions depicted for three individual markets. This pattern demonstrates that there are important differences in the distributions of prices across markets. It is also interesting that the markup distributions appear smoother than their price counterparts, perhaps suggesting that the markup specification is more appropriate for estimating search costs.

[Insert Figures 2a and 2b about here.]

Figures 3a and 3b more fully characterize the heterogeneity of price distributions across markets. Figure 3a depicts histograms of the mean and standard deviation of prices in each isolated market, and Figure 3b presents a scatter plot of the standard deviation of prices against the mean market price. As in Table 2, mean market prices are net of state fixed effects. The histogram of standard deviations illustrates that there is a mass of markets wherein the standard deviation of prices is about 5-6 cents per gallon, and the distribution is skewed right. The histogram of average prices also shows substantial variability in average market prices, which range from just under $3.00 to more than $3.30. In total, there appears to be a substantial degree of heterogeneity in both the mean and standard deviation in prices across markets.

[Insert Figures 3a and 3b about here.]

Interestingly, Figure 3b demonstrates, and simple regressions confirm, that in general there is no meaningful relationship between price levels and the standard deviation of prices. This result highlights the degree of heterogeneity in price distributions across the isolated markets. As the structural model maps the observed distribution of price-cost markups into the distribution of consumer search costs, these figures suggest that the estimation will yield meaningful differences in search cost distributions across markets.

In the structural model presented below, the mapping from the consumer search cost distribution to the equilibrium price distribution is non-linear. For example, as shown in Chandra and Tappata's (2011) fixed-sample search model, there is no price dispersion when the search costs are very low or high, because the former leads to competitive pricing and the latter leads to monopoly pricing. At intermediate search costs price dispersion exists. Taken as a whole, this implies a non-monotonic relationship between the moments of the search cost and price distributions. Appendix A1 demonstrates the same property in the model we estimate. It follows that studies that relate simple proxies for consumer search costs, such as the distance between stations or the number of commuters, to measures of price dispersion may be limited in testing how changes in the cost of search affect prices. Also, the complex relationship between price and search distributions implies that simple measures of price dispersion may poorly proxy for the search cost distribution in identifying the relationship between search costs and market characteristics.

[Insert Table 3 about here.]

To underscore this point, the results of regressing different measures of price dispersion on market and population characteristics are presented in Table 3, and we find that the significance of regressors is sensitive to how price dispersion is measured. For example, mean and log-mean income are positively and significantly correlated with the sample range of prices in the second, third, and fourth specifications, but are not significantly related to the sample standard deviation of prices. The significance of other variables, such as the average distance among stations and the log of education, is also sensitive to how price dispersion is measured. Furthermore, all specifications in the regressions fail to find a correlation between the standard deviation of income and price dispersion. This reduced-form result may not be surprising, because, as developed in Chandra and Tappata (2011), the equilibrium relationship between price dispersion and consumer search costs is theoretically non-monotonic. In the structural analysis that takes into account this non-monotonicity, we find a robust connection between the standard deviation of income and the variance of consumer search costs. It therefore appears that the distribution of prices is, at best, an imperfect proxy for the distribution of search costs.

Finally, we analyze the extent to which cross-sectional and temporal price variation is explained by differences across gas stations and wholesale cost changes. To do so, we run simple regressions of prices on wholesale costs and firm fixed effects. Using data from all stations and dates in the data set, column 1 in Appendix Table A2 shows station fixed-effects explain 30% of the variation in prices. Column 2 shows that adding wholesale prices increases the r-squared to .81, suggesting that much of the price dispersion in retail gasoline markets may be explained by marginal cost changes and station characteristics. Restricting the data to only stations in isolated markets and the 30-day time window for which we estimate search costs, we find that station fixed effects and wholesale costs explain 96% of the price variation. Thus, search costs likely explain at most between 4% and 19% of observed price dispersion.

When using all of the data and including firm fixed effects (column 2), we estimate a coefficient of 0.89 on wholesale costs. Previous studies of retail gasoline markets have estimated the coefficient to be one (Remer 2015; Bachmeier and Griffin 2003; Borenstein et al. 1997). Furthermore, industry reports find that oil and refining account for almost 80% of retail fuel costs and the accounting margins are typically slim.[5] Thus, it would be unprofitable in the long-run for gas stations to maintain a wholesale cost pass-through rate below 1. To construct the markup variable in the structural estimation, we effectively assume that the wholesale coefficient is equal to one, which is consistent with previous empirical findings and industry reporting.

3.2 Evidence of Mixed-Strategies

A large body of empirical literature analyzing retail gasoline markets documents price dispersion as a consequence of gas stations playing mixed strategies to extract informational rents from consumers with search costs (e.g., Lewis 2008; Hosken et al. 2008; Chandra and Tappata 2011; Lach and Moraga-Gonzalez 2012). Using much of the same data as in this article, Chandra and Tappata (2011) demonstrate that retail gasoline price dispersion can be explained by a model where firms use mixed strategy pricing. As evidence of price dispersion and mixed strategies, Chandra and Tappata (2011) calculate rank-reversal statistics by computing the percentage of days the typically lower-priced firm sets a higher price, and show that pairs of closely located gas stations often switch between which sets a higher price. Hosken et al. (2008) and Lewis (2008) similarly find that gas stations often switch places in the market price distribution.

We subset the data used in Chandra and Tappata (2011) into isolated markets and use a narrower time frame, and find similar evidence of mixed strategies. First, we replicate the rank-reversal analysis. Appendix Table A3 shows that there are a mass of firms which never switch price rankings, and on average the low-price firm sets a high price 8.2% of days. As we detail in the model below, due to vertical product differentiation, competing firms may play mixed strategies but, due to quality differences, draw prices from non-overlapping price distributions; such firms would never reverse price rankings. For the set of firms with positive rank-reversals, the rank-reversal statistic is distributed almost uniformly between zero and 50%. In theory rank-reversals could be generated from Edgeworth price cycles, which have been observed in gasoline markets (Noel 2007; Lewis and Noel 2011); however, we find no evidence of Edgeworth price-cycles in any of the markets included in the analysis.

[5] These numbers are reported by the Association for Convenience & Fuel Retailing (www.nacsonline.com/YourBusiness/FuelsReports/2015/Prices/Pages/The-Price-Per-Gallon.aspx). Last accessed December 2015.

In the structural model, firms compete by playing mixed strategies in utility, rather than in prices, and utility is empirically identified by removing a firm fixed effect and wholesale price. As such, we calculate utility rank-reversals for firms in isolated markets and find more support for mixed strategies. We find that the low-utility firm, on average, sets the higher utility on 25% of days. Furthermore, firms effectively never offer the same utility on a particular day; we find ties in utility on only 0.4% of days, which further evidences firms playing mixed strategies in utility space. We find that two firms in an isolated market offer the same retail price on 28% of days.

Finally, we analyze whether prices, and therefore utility, exhibit autocorrelation, which would be evidence against mixed strategies. Appendix Table A4 shows that when pooling prices from the 367 markets for which we estimate search costs, we fail to reject the null hypothesis of no autocorrelation with even 90% confidence. Taken as a whole, we believe the evidence supports gas stations in our data playing mixed strategies, which is consistent with many previous studies of retail gasoline markets.

4 A Model of Consumer Search with Vertical Product Differentiation

We introduce a fixed-sample search model based on the presentation in Hong and Shum (2006) and Wildenbeest (2011).[6] The empirical fixed-sample search model developed in Wildenbeest (2011) modifies Burdett and Judd (1983)'s theoretical model in two ways; the number of firms is discretized and vertical product differentiation is incorporated. We first introduce several assumptions on consumers' and firms' behavior, and then discuss optimality conditions for both.

We consider N gas stations ("firms") selling a vertically differentiated product to a continuum of consumers. Consumers have inelastic demand for one unit of gasoline, and are identical except for their search costs. Consumers draw an i.i.d. search cost, $c \geq 0$, from the cumulative distribution, F_c. Firms do not observe an individual consumer's search cost, but know the cumulative distribution of search costs.

Consumers have the same preferences over quality, which makes products vertically differentiated. The utility from purchasing from station j is given as follows:

$$u_j = v_j(w_j) - p_j. \tag{1}$$

Here, u_j is the utility of purchasing from firm j at a price of p_j. The value consumers obtain from a firm of quality w_j takes the following form, $v_j(w_j) = x + w_j$, where x can be thought of as the minimum quality across firms and additional quality, w_j, enters additively and separably. Consumers know a priori v_j, which is constant over time, but do not observe p_j, and hence the utility firm j will provide unless that station is searched for its price.

We make two assumptions about firms' production: quality inputs are purchased in per-

[6] Appendix A3 discusses fixed-sample versus sequential search.

fectly competitive markets, and the quality production function exhibits constant returns to scale. Perfectly competitive input markets imply that quality inputs are paid their marginal product. Constant returns to scale imply that firms choose w_j to maximize the valuation-cost markup, $v_j(w_j) - r(w_j)$, where $r(w_j)$ is the marginal cost of offering a product of quality w_j. From these two assumptions and Euler's theorem, it follows that the cost of obtaining a given level of quality is simply $r(w_j) = w_j$, and the value provided by a firm can be denoted as:

$$v_j = x + r_j. \tag{2}$$

Given equations (1) and (2), firm j's profit margin is $p_j - r_j = (v_j - u_j) - r_j = x - u_j$, which implies firm j chooses u_j to maximize its profit margin. As such, the relevant strategy space for firms can be formulated in terms of utility rather than prices, and we can focus on mixed strategies in utility. In equilibrium, firm j varies its utility over time by changing p_j such that any systematic difference in prices relative to other stations, $E[p_j - p_{-j}]$, exactly reflects the difference in its value, $v_j - v_{-j}$, and thereby $r_j - r_{-j}$. Because the marginal costs reflect differences in quality, all firms earn the same expected profits, even under vertical product differentiation. Firms are therefore symmetric in the utility strategy space, and we focus on symmetric equilibria.

As firms differ in quality, in order to provide the same expected utility, a higher quality firm will tend to set a higher price. The equilibrium generates a utility distribution, $F_u(u)$, that is i.i.d. across stations; $\underline{u}$ and $\bar{u}$ are the lower and upper bound, respectively, of the support of F_u.

In equilibrium, consumers know the distribution of utility, $F_u(u)$, but must engage in costly search to learn the price, and thereby utility, offered by a particular firm. Consumers receive one free price quote and then pay a cost, c, for each additional quote. Consumers do not observe any prices before choosing the number of searches that maximize expected utility, which consists of u_j minus the total cost of search. Consumers purchase from the firm in its sample that offers the highest utility.

A consumer's problem is to maximize overall expected utility by choosing the number of quotes, l:

$$l = \arg\max_{l \geq 1} -c \cdot (l-1) + \int_{\underline{u}}^{\bar{u}} l \cdot u \cdot F_u(u)^{l-1} f(u) du.$$

The first term, $-c \cdot (l-1)$, is the total cost of actively searching $l-1$ stations. The second term is the expected utility from consumption when a consumer has l quotes, and $F_u(u)^{l-1}$ is the probability that $l-1$ firms offer utility lower than u. By searching $l+1$ rather than l firms, a consumer obtains an expected marginal benefit, which is denoted as $\Delta_l \equiv Eu_{1:l+1} - Eu_{1:l}$, $l = 1, 2, ..., N-1$. Here, $Eu_{1:l}$ represents the maximum expected utility when a consumer takes l draws from F_u. Accordingly, a consumer with search cost c will sample l stores when $\Delta_{l-1} > c > \Delta_l$. A consumer with $c = \Delta_l$ is indifferent between searching l and $l+1$ firms. The proportion of consumers with l utility quotes, q_l, will be $q_1 \equiv 1 - F_c(\Delta_1)$ for $l = 1$, and $q_l \equiv F_c(\Delta_{l-1}) - F_c(\Delta_l)$ for $l \geq 2$. Therefore, q_l is the total market demand from consumers that actively search $l-1$ firms prior to purchasing.

Given consumers' optimal search behavior, each firm chooses a symmetric mixed-pricing strategy, F_u, for all $u \in [\underline{u}, \bar{u}]$ to maximize profits. In a symmetric equilibrium, all utility in the support of F_u generates the same expected profit, and therefore a firm is indifferent between offering the lowest utility $\underline{u}$ and any other utility $u \in [\underline{u}, \bar{u}]$. Firm j's total profit can then be denoted as $\Pi_j(u) = (x - u_j)[\sum_{l=1}^{N} q_l \cdot \frac{l}{N}(F_u(u_j)^{l-1}]$. The highest price firm j may set is $\bar{p}_j = v_j$, and therefore offers no utility, $\underline{u}_j = v_j - \bar{p}_j = 0$. In this case, firm j attracts only consumers with one price quote, and its profit will be $\Pi_j(\underline{u}) = \frac{x}{N} q_1$. Accordingly, we can represent the indifference equilibrium condition for firms as:

$$\Pi(u) = x \cdot \frac{q_1}{N} = (x - u)[\sum_{l=1}^{N} q_l \cdot \frac{l}{N} F_u(u)^{l-1}], \text{ for } u \in [\underline{u}, \bar{u}]. \tag{3}$$

The equal profit condition in equation (3) does not have a closed-form solution in F_u. As we detail in the next section, however, points on the distribution can be identified. Note that this condition does not depend on market size, as consumers have an inelastic demand for one unit and therefore profits are perfectly linear in market size. Thus, variation across markets in time-invariant demand-side characteristics, such as income and population, will only affect the mixed strategy equilibrium, and therefore price dispersion, indirectly through the distribution of search costs. Equation (3) can be solved implicitly for utility as:

$$u = x \cdot \frac{\sum_{l=2}^{N} l q_l F_u(u)^{l-1}}{\sum_{l=1}^{N} l q_l F_u(u)^{l-1}}. \tag{4}$$

Finally, given the equilibrium utility distribution, $F_u(u)$, and valuation at the firm level, v_j, we can recover the price distribution for firm j, $F_j(p)$, as

$$F_j(p) = \Pr[p_j \leq p] = \Pr[p \geq v_j - u_j] = \Pr[u_j \geq v_j - p] = 1 - F_u(v_j - p). \tag{5}$$

5 Estimation Strategy

5.1 First Stage: Identification and Nonparametric Estimation of Search Costs at the Market Level

The first-stage estimation strategy extends Moraga-González and Wildenbeest (2008) and Wildenbeest (2011) to control for intertemporal price dispersion that arises from changes in production cost. The Hong and Shum (2006) framework recovers the distribution of consumer search costs using only price data by rationalizing all observed price dispersion as a consequence of search costs. Also relying solely only price data, Wildenbeest (2011) converts prices to utility by controlling for vertical product differentiation with firm-specific fixed effects, and then maps the dispersion in utility onto the distribution of search costs. Controlling for vertical product differentiation is important in the retail gasoline industry; within a market firms may vary in the quality of service, location, and brand. Previous research, such as Lewis (2008), demonstrates that these characteristics are fixed over time and explain a substantial degree of price dispersion. We extend Wildenbeest (2011) to incorporate data on wholesale costs, which allows us to control for price variation that results from changes in marginal costs. This change stands as an improvement over previous structural estimates of consumer search costs, which map all price dispersion remaining after controlling for vertical product differentiation onto the distribution of search costs. We apply this methodology separately to each geographically isolated retail gasoline market. We treat price observations as independent across days and stations after controlling for station fixed effects and wholesale costs. We find in Section 3 that prices, and therefore utilities, are not autocorrelated, which supports this treatment.

To estimate the model, we first use the price and wholesale cost data to estimate the utility provided by a gas station on a particular day. To do so, we follow Wildenbeest (2011) and solve equation (1) for price and add a time subscript: $p_{jt} = v_{jt} - u_{jt}$. To control for changes in wholesale costs, we assume that the unit cost at the firm level takes the form $r_{jt} = r + c_t + \delta_j$, where r is a time-invariant component of marginal cost that is common across firms, c_t is a time-varying cost component shared by all firms, such as wholesale prices, and δ_j varies with firms' quality, but is fixed over time. Equation (2) gives $v_{jt} = x + r_{jt} = x + r + c_t + \delta_j$, and plugging v_{jt} into $p_{jt} = v_{jt} - u_{jt}$ yields:

$$p_{jt} = x + r + c_t + \delta_j - u_{jt}.$$

To identify utility, we estimate the following equation via fixed-effect regression:

$$\begin{aligned} p_{jt} - c_t &= x + r + \delta_j + \epsilon_{jt} \\ &= \alpha + \delta_j + \epsilon_{jt}, \end{aligned} \tag{6}$$

where $\alpha = x + r$ is a constant term, δ_j is a firm-specific fixed effect, and ϵ_{jt} is an idiosyncratic shock. We construct the markup of price over wholesale cost by subtracting the wholesale price data from a firm's retail price. We then recover the utility offered by station j at time t by running the fixed-effect regression and setting $\hat{u}_{jt} = -(p_{jt} - c_t) + \hat{\alpha} + \hat{\delta}_j = -\hat{\epsilon}_{jt}$. Therefore, the utility offered by a firm at a point in time is identified as the negative of the residual in equation (6). The systematic differences in firms' markups are attributed to differences in the firms' product quality.

With these estimates in hand, we identify the model parameters using the variation in utilities over time and across firms. Formally, we utilize the equilibrium condition specified in equation (3), and conduct nonparametric estimation of this optimality condition. The estimation methodology follows Moraga-González and Wildenbeest (2008), who extend Hong and Shum (2006) through the maximum likelihood estimation (MLE) approach to achieve more favorable convergence properties. Denoting the number of gas stations in a market by N and the number of price observations in that market by I, we employ the MLE to estimate the model parameters $\hat{\theta}_{MLE} = \{\hat{q}_l\}_{l=1}^{N-1}$ such that:

$$\hat{\theta}_{MLE} = \arg \max_{\{q_l\}_{l=1}^{N-1}} \sum_{i=2}^{I-1} \log f_u(u_i; q_1, .., q_N),$$

where f_u is the density function of F_u. Following Wildenbeest (2011), we apply the implicit function theorem to equation (3) to yield the density for equilibrium utility:

$$f_u(u_i) = \frac{\sum_{l=1}^{N} l q_l (F_u(u_i))^{l-1}}{(x - u_i)\sum_{l=1}^{N} l(l-1) q_l (F_u(u_i))^{l-2}}, \tag{7}$$

where $F_u(u_i)$ solves equation (3) for $i = 1,..,I$. Because all firms draw utility from the same equilibrium distribution, F_u, we pool all estimated utilities in a market to estimate the model parameters. The maximum likelihood routine yields parameter estimates that can be used to construct points on the non-parametric search cost CDF, $\{\hat{\Delta}_l, F(\hat{\Delta}_l)\}$. These estimates have the asymptotic properties of standard maximum likelihood estimation, and we compute the standard errors accordingly.

Identification of the first-stage model. To identify the model's primitives, we utilize the restrictions imposed by the theoretical model to map dispersion in utilities onto the distribution of consumer search costs. Intuitively, we proceed in two steps. First, we subtract wholesale costs from prices to control for intertemporal price dispersion due to marginal cost changes. Using the markups of price over wholesale cost, we then exploit the cross-sectional variation in markups across stations within a market to identify the firm-specific fixed effect δ_j, which controls for vertical product differentiation. By subtracting the estimated fixed effect from the markup, we recover the utility provided by station j at time t, which is then used in the MLE routine.[7] Second, by assuming firms play a mixed strategy in utility, the intertemporal variation in utility within a firm and cross-sectional variation in utility across firms identifies the remaining structural parameters of the first-stage model through the equilibrium restrictions in equation (3).

More formally, for a market with N stations and I price observations, there are $N-1$ unknown parameters in the model $\{q_1, q_2, .., q_{N-1}\}$, where $q_N = 1 - \sum_{l=1}^{N-1} q_l$. Using the station-level fixed-effect regressions in equation (6), we identify for each price observation the implied utility $\hat{u}$ as the residuals from the regression multiplied by -1, which we use to construct the empirical distribution of utility $\hat{F}_u(u)$ through equation (5). Then, using the empirical distribution of utility, we identify the marginal utility of search, $\{\hat{\Delta}_l\}_{l=1}^{N-1}$, such that $\hat{\Delta}_l \equiv E\hat{u}_{1:l+1} - E\hat{u}_{1:l}$. We then set $\underline{u}$' and $\bar{u}'$ to be the lowest and highest observed utility, and then order the I utility estimates: $\underline{u}$'$= \hat{u}_1 \leq \hat{u}_2 \leq ... \leq \hat{u}_I = \bar{u}'$. The empirical counterpart of the firms' indifference condition is:

$$\hat{x} \cdot \frac{\hat{q}_1}{N} = (\hat{x} - \hat{u}_i)[\sum_{l=1}^{N} \hat{q}_l \cdot \frac{l}{N} \hat{F}_u(u_i)^{l-1}],\ i = 1, 2, .., I. \tag{8}$$

Applying the implicit function theorem to this equation yields the likelihood function, equation

[7] Because of the station fixed effect in equation (6), the empirical model does not rely on systematic differences in prices across stations within a given market.

(7), and $I-1$ corresponding moment restrictions implied by the model equilibrium. Accordingly, for all markets such that $I-1 \geq N-1$, we identify the model parameters, $\{\hat{q}_1, \hat{q}_2, .., \hat{q}_{N-1}\}$, and the corresponding non-parametric cumulative distribution of search costs: $F_c(\hat{\Delta}_l) = F_c(\hat{\Delta}_{l-1}) - \hat{q}_l$. We then evaluate the empirical indifference condition in equation (8) at the highest utility $\bar{u}$ to identify the time-invariant and common component of the valuation-cost margin:

$$\hat{x} = \bar{u}\frac{\sum_{l=1}^{N} \hat{q}_l \cdot l}{\sum_{l=2}^{N} \hat{q}_l \cdot l}. \tag{9}$$

The lower bound of the utility distribution, $F(u)$, is normalized to zero; a station offering zero utility will only sell to consumers with one price quote, and the station maximizes profit by setting the maximum price $\bar{p}_j = v_j$, such that $\underline{u}= 0$. We then compute the maximum utility $\bar{u}$ by solving for $F_u(\bar{u}) = 1$ in equation (4).

5.2 Second Stage: Estimation of Parametric Search Cost Distribution That Allows for Variation Across Markets

The first stage estimation routine produces estimates of points on the distribution of search costs separately for individual isolated markets. We use these estimates to understand how the distribution of search costs relate to observable market and consumer characteristics. To achieve this goal, we pool the estimated points of the search cost CDFs across all markets, and use nonlinear least-squares (NLS) regression to fit a parametric distribution. Following previous work, including Hong and Shum (2006), Chen, Hong, and Shum (2007), and Wildenbeest (2011), we use the log-normal distribution as the main specification. As a robustness check, we also fit a normal distribution.

Given the first stage search cost estimates, $\{\hat{\Delta}_{l,m}, \hat{q}_{l,m}\}_{l=1}^{N-1}$ in each market, m, we employ NLS to estimate the log-normal parameters, (β, γ), using the following equations,

$$\begin{aligned}
\hat{y}_{l,m} &= \frac{1}{\hat{\Delta}_{l,m} \cdot \sigma_m \sqrt{2\pi}} e^{-\frac{(ln(\hat{\Delta}_{l,m})-\mu_m)^2}{2\sigma_m^2}} \text{ for } l = 1, .., N-1, \text{ where,} \\
\mu_m &= X_m^{\mu}\beta \\
\sigma_m &= X_m^{\sigma}\gamma \\
\hat{y}_{1,m} &= 1 - \hat{q}_{1,m}, \hat{y}_{2,m} = 1 - \hat{q}_{1,m} - \hat{q}_{2,m}, ..., \hat{y}_{N-1,m} = 1 - \sum_{l=1}^{N-1} \hat{q}_{l,m}
\end{aligned}$$

In the regression, the mean and variance parameters, β and γ, respectively, depend upon market-level characteristics, X_m^{μ} and X_m^{σ}. The cross-sectional variation in search cost distributions across markets allows us to identify the second stage parameters. These parameters describe how market characteristics influence the mean and variance of the consumer search cost distribution, and therefore are informative about the underlying cause of consumer search behavior.[8]

[8] We assume that the intertemporal and cross-sectional variation in prices, after removing wholesale cost variation and station fixed effects, is generated by firms' mixed strategy given the search cost distribution. In Appendix

6 Estimation Results

6.1 Search Cost Estimates

We estimate the proportion of consumers that search l stations, $\hat{q}_l$, and a combination of points, $\{\hat{\Delta}_l, \hat{F}_c(\hat{\Delta}_l)\}$, on the search cost CDF for each of 367 markets. Table 4 summarizes the estimates of $\hat{q}_l$ and $\hat{\Delta}_l$, and the per-gallon valuation-cost margin per gallon, $\hat{x}$, across isolated markets. We find that there is not much consumer search in retail gasoline markets; across markets, on average, 66.4% of consumers receive only one price quote. These consumers visit only one gas station, and therefore do no price comparison shopping before purchasing. The average marginal expected savings from visiting a second station is \$0.032 per gallon. While the average search intensity is low, there is substantial heterogeneity across markets; the standard deviation of $\hat{q}_1$ is 0.155. Moreover, its range is $0.936 - 0.004 = 0.932$; therefore in at least one market almost all consumers search and in another no one searches.

[Insert Table 4 about here.]

To illustrate how search cost distributions vary across markets, we randomly pick five markets in Texas that have three stations. Figure 4a plots $(\hat{\Delta}_1, F_c(\hat{\Delta}_1))$ and $(\hat{\Delta}_2, F_c(\hat{\Delta}_2))$ for each market, and the figure displays substantial differences between the distributions. For instance, the fraction of non-searchers ($q_1 = 1 - F_c(\hat{\Delta}_1)$) in market 1 is 0.09, and the gain from the first search is \$0.02 ($= \hat{\Delta}_1$) per gallon of regular gasoline. In market 4, on the other hand, the fraction of non-searchers is 0.89, and the gain from the first search is \$0.04 ($= \hat{\Delta}_1$) per gallon. Thus, for similar marginal savings, there is a wide disparity between these two markets in the propensity to engage in costly search, indicating that there is a large difference in consumer search costs.

[Insert Figures 4a and 4b about here.]

Finally, the valuation-cost margin parameter, x, and price distributions implied by the model are well estimated. The estimated maximum valuation-cost markup is on average \$0.261 per gallon across markets. Appendix Figure A1 displays the distribution of estimated valuation-cost margin for duopoly markets (Figure A1a) and markets with more than two firms (Figure A1b). We find that the valuation-cost margins are skewed to the right for both markets. Furthermore, the equilibrium price distribution from the estimated model approximates the empirical distribution of prices in most markets. For instance, Figure 4b presents the empirical and estimated price distribution from a market in Florida.

A2, we assess the possibility that time-invariant, market-level characteristics affect intertemporal variation through means other than search costs; we do not find evidence that these characteristics directly and meaningfully affect intertemporal pricing.

6.2 Estimation of Parametric Search Cost Distribution That Allows for Variation Across Markets

This subsection sheds light on the underlying source of search costs. Our aim is to determine the extent to which (i) the mean and variance of search cost distributions vary across markets and (ii) this variation can be explained by observable market and population characteristics.

Heterogeneity of search costs. Table 5 presents the results of pooling the first stage search cost estimates, $\{\hat{\Delta}_{l,m}, \hat{q}_{l,m}\}_{l=1}^{N-1}$, from all $m = 366$ markets,[9] and using nonlinear least-squares to fit a log-normal (columns 1 through 3) and normal (column 4) CDF. The log-normal specifications (columns 1 through 3) deliver lower mean squared residuals than the normal distribution specification (column 4). We regard column 1 as the baseline specification, and columns 2 and 3 examine whether the results are robust to the exclusion of imprecisely estimated parameters.

[Insert Table 5 about here.]

Using the estimates from column 1, we predict for each market the mean and variance of the search cost distribution, and present the distribution of results in Figures 5a and 5b. In Figure 5a, the histogram of mean search costs is skewed to the right and exhibits heterogeneity across markets around the median of \$0.127 per gallon. Interestingly, the histogram of mean market prices presented in Figure 3a exhibits very little skewness, again reinforcing that price distributions are not a perfect proxy for search cost distributions. Figure 5b similarly demonstrates substantial heterogeneity in the variance of search costs across markets.

[Insert Figures 5a and 5b about here.]

To quantify the degree of heterogeneity, we compare percentiles within the distributions presented in Figures 5a (distribution of estimated means) and 5b (distribution of estimated standard deviations). The 75-25 and 90-10 percentile ratios of the estimated means are 1.630 and 2.572, respectively. Similarly, the 75-25 and 90-10 percentile ratios of the estimated standard deviations are 2.109 and 4.695, respectively. These ratios confirm a great amount of variation in search cost distributions across markets, and suggests caution when generalizing search cost estimates from a single geographic market to other markets.

Potential sources of search cost heterogeneity. Having established the presence of heterogeneity in search cost distributions, we now investigate the source of consumer search costs. While the empirical model in the first stage is agnostic as to the underlying source of search costs, the second-stage estimation is designed to shed light on the potential causes. Following Goldman and Johansson (1978), we argue that search costs may derive from the (1) cost of information acquisition, such as opportunity cost of time and other expenditures (e.g., driving, phone calls, internet

[9] We exclude one market from the second stage estimation because it was missing demographic variables included as controls in the NLS regression.

access, etc.), which may vary with household income, and (2) information-processing or cognitive costs, which may vary with education or age. Note that these two sources of consumer search costs are not mutually exclusive. For example, online price search likely possesses both components; it requires time in front of a computer (cost of information acquisition) and knowledge of which websites to search and which gas stations are located near the desired commuting path (information processing costs). During the sample time period, 2007, retail gasoline search was a mixture of online and offline; websites such as gasbuddy.com, mapquest.gasprices.com, and autos.msn.com listed prices for individual gas stations searchable by zip code.

The results in Table 5 demonstrate which population and market characteristics influence the cost of search. Of first note in Table 5 is positive relationship between population mean income and search costs. Across all specifications, the mean income in a market positively affects the mean of the consumer search cost distribution, and is statistically highly significant. In terms of magnitude, a 10% ($6,410) increase in household income increases the expected search costs by $0.071 per gallon. Furthermore, the standard deviation of household income within a market positively affects the standard deviation of the search costs in a market. These findings tightly link the market income and search cost distributions. Appendix Table A5 presents results when wholesale data are not incorporated into the estimation, and therefore variation in retail prices rather than markups are used to identify the model. The income results are robust to this alternative specification, and income estimates are the only results that are robust to estimating the model with prices or markups. The results are also robust to different measures of household income, such as the median and median absolute deviation of income.

To get a better sense of the extent to which household income and other demographics explain variation in the mean and standard deviation of search cost distributions across markets, we regress the predicted market-level search cost mean and standard deviation on demographics. Columns 1 and 5 in Appendix Table A6 demonstrate that the household income mean and standard deviation explain about 63% and 57% of the variation in the mean and standard deviation, respectively.

The positive relationship between the distribution of search costs and income suggests that information acquisition costs are an important component of search costs. First, income is likely to be positively correlated with the opportunity cost of time, and therefore increases the time-investment cost of search. Furthermore, people with higher income may have a lower marginal utility of wealth and thereby gain less utility from saving money on gasoline. It follows that higher income consumers may have a lower incentive to spend time searching for a lower price, which supports the information cost interpretation.

Meanwhile, we do not find as much support for information processing as a major source of search costs. Mean years of education is not significant in any of the specifications, and the standard deviation of education only enters significantly at the 10% level when a normal distribution is fitted and wholesale costs are not included. It seems reasonable to expect consumers with a higher education to have lower information processing costs, yet the regressions do not produce this finding. There is modest support for older populations having a lower cost of search, as we estimate a negative relationship between mean age and mean search costs at the 5% or 10% level in some specifications (although not the baseline). It may be that older consumers, with more shopping experience and a better understanding of road patterns, all else equal, have lower information processing costs; however, this interpretation of age is somewhat speculative. In general, it may not be possible to strictly separate demographics and market characteristics into either information acquisition cost or information processing categories. Accordingly, we do not rule out the information processing interpretation of search cost.

Table 5 does contain results that may not fit with either story. The average distance between stations is found to negatively relate to the mean of the search cost distribution. This finding seems to contradict consumers engaging in costly search by driving between stations. However, if stations endogenously locate close together in markets where demand is mostly met by travelers with high search costs, perhaps near interstate highway exits, a negative relationship between mean distance and search costs may arise. This finding, however, is not robust to estimating the model using price data rather than markups. Similarly, we find a negative relationship between the standard deviation of age and the standard deviation of search costs when estimating the model with markups, but not prices. We see no direct relationship between this finding and either potential determinant of search costs.

Finally, the results in this section provide a rationale for time-invariant demand-side characteristics, such as income and age, affecting intertemporal price dispersion. While these variables are widely recognized to impact cross-market price dispersion, we show that they can also indirectly affect price changes over time through their impact on the distribution of search costs.[10]

Overall, in this section, we document the heterogeneity in search cost distribution across markets. In the following section, we analyze, through simulation, how exogenous changes in the search cost distribution affect equilibrium prices.

7 Policy Experiments: A Reduction in the Costs of Search

This section adopts the search cost distribution parameter estimates from column 1 of Table 5 to perform "what-if"experiments that quantify the effect of changes in search costs on equilibrium prices and consumer search behavior. We analyze a policy that simultaneously reduces the mean and variance of the search cost distribution, and the unique impact within each of the 366 markets included in the stage 2 regressions. Many policies aimed at reducing consumer search costs will reduce both the mean and variance of search costs in a market. For example, a regulation requiring gas stations to post prices on a website may decrease search costs for consumers at the high end of the distribution, but leave unchanged search costs for consumers at the low end that may already be using a similar website, such as gasbuddy.com. Similarly, a policy that requires gas stations to clearly post the price for customers paying with cash or credit may reduce both the mean and

[10] We thank the Associate Editor for this observation. Appendix A2 provides empirical evidence that demand-side characteristics do not directly influence intertemporal price dispersion outside the scope of our model.

variance of the search cost distribution, as it may lower the cost of search for credit card consumers and leave unchanged the cost for cash customers.

The policy decreasing both the mean and standard deviation of the search cost distribution can either increase or decrease the expected price. The intuition is as follows. Lowering the mean of the search cost distribution, all else equal, leads to more intense search and lower prices. Conversely, decreasing the variance of the search cost distribution, all else equal, leads to less search and higher prices. The reason for higher prices when search costs become less heterogeneous is that firms face a trade-off between setting a low price and selling to all consumers and setting a high price and selling to only consumers with low search intensity. When search costs become sufficiently homogeneous the trade-off always favors setting the highest price, as the sales from relatively lower search cost consumers becomes too small. The incentive to only set a high price may remain even if the average cost of search decreases. In Appendix A1, we closely examine the impact of lowering the mean and/or variance of the search cost distribution within a hypothetical market of average demographic characteristics.

For the policy counterfactuals, we consider a situation in which consumers' search costs become less heterogeneous (but not completely homogeneous) and the expected search costs decrease, which we define as $\int c \; dF_c$. We conduct the policy experiment by reducing the standard deviation by 10% and the expected value of search costs by 20% separately for each of the 366 isolated markets. A 20% decrease in the expected value of search costs for a market with average demographics is equivalent to a 26% (or $\$17,164$) reduction in the household income. To perform the experiments, we first recover the mean and variance of the search cost distribution for each market by using the second-stage estimates and pertinent demographic variables. We then solve for the equilibrium firm price distribution and consumers search behavior using the structural model. Because there may be multiple price dispersion equilibria, we try 10 unique starting values for a given market.

[Insert Figures 6a and 6b about here.]

Figure 6a summarizes the changes in the mean price in each market. The distribution is bimodal with the two modes falling on opposite sides of zero. Furthermore, the distribution exhibits heterogeneity in both the direction and magnitude of changes across markets. In left hump of the distribution, the new price dispersion equilibria are characterized by an average price decrease due to a lower expected cost of search and an increased number of searches. In the right hump, however, there are higher average prices; these are markets in which the price dispersion equilibrium vanishes, and only the monopoly equilibrium remains. The distribution is skewed to the right, illustrating that the magnitude of the positive price changes are, on average, larger than the negative changes. Indeed, the average price change across markets is positive and \$0.051. To confirm the finding, Figure 6b presents the changes in average prices when we reduce the standard deviation by 20%, while maintaining a 20% reduction in mean search costs. As expected, there are even more markets with higher prices, and the average price change across markets increases to \$0.088.

In summary, we show a reduction in the mean of the search cost distribution coupled with a lower variance may result in higher average prices for a non-negligible number of markets. Therefore, a well intentioned policy aimed at decreasing the cost of search may unintentionally lead to higher prices for consumers and profits for firms.

8 Conclusions

We document significant variation in the distribution of consumer search costs across geographical markets for retail gasoline, and explain the variation using observable demographics and market characteristics. Based on a fixed-sample search model and daily price data, we recover the distribution of search costs for a set of geographically isolated markets. The empirical approach extends the existing fixed-sample search models by (1) incorporating wholesale prices and demographics and (2) utilizing multiple geographic markets. We find that the distribution of income explains the shape of the search cost distribution, suggesting that opportunity costs are an important driver of consumers' search costs. Our results also provide a rationale for how time-invariant characteristics at the market level, such as consumer income, may contribute to intertemporal price variation.

We use the estimated search costs in each market to conduct counterfactual policy experiments that examine the impact of changing the mean and variance of the search cost distribution. The counterfactual simulations for 366 geographic markets confirm that subtle changes to the shape of the search cost distribution can have a large impact on firm prices and consumer welfare, and the effect of the policy can vary widely across markets. In some markets, policy that reduces the mean and variance of search costs may have the unintended consequence of reducing consumer welfare.

The empirical and counterfactual results have two managerial implications for a firm entering a new geographic market. First, perhaps counterintuitively, entering a market with uniformly low income might yield higher profit margins than entering a market with high and dispersed income. Second, an entrant may want to calculate the expected profits of entry, and understanding how demographics affect search cost distributions, and therefore pricing, allows firms to more precisely project revenues. The findings suggest a list of demographic variables, such as income and age, that may help predict the cost of consumer search when firms are formulating their pricing strategies.

References

Anderson, E. T., N. M. Fong, D. I. Simester, and C. E. Tucker (2010): "How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet," Journal of Marketing Research, 47(2), 229–239.

Bachmeier, L., and J. Griffin (2003): "New Evidence on Asymmetric Gasoline Price Responses," Review of Economics and Statistics, 85(3), 772–776.

Bakos, J. Y. (1997): "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," Management Science, 43(12), 1676–1692.

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), 1041–1066.

Baye, M. R., J. Morgan, and P. Scholten (2006): "Information, Search, and Price Dispersion," in Handbook on Economics and Information Systems, ed. by T. Hendershott, vol. 1, chap. 6. Elsevier, Amsterdam.

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

Branco, F., M. Sun, and J. M. Villas-Boas (2012): "Optimal Search for Product Information," Management Science, 58(11), 2037–2056.

Bresnahan, T. F., and P. C. Reiss (1991): "Entry and Competition in Concentrated Markets," Journal of Political Economy, 99(5), 977–1009.

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

Brynjolfsson, E., A. A. Dick, and M. D. Smith (2010): "A Nearly Perfect Market?," Quantitative Marketing and Economics, 8(1), 1–33.

Brynjolfsson, E., Y. Hu, and M. S. Rahman (2009): "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-channel Competition," Management Science, 55(11), 1755–1765.

Brynjolfsson, E., Y. Hu, and D. Simester (2011): "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," Management Science, 57(8), 1373–1386.

Burdett, K., and K. L. Judd (1983): "Equilibrium Price Dispersion," Econometrica, 51(4), 955–969.

Chan, T. Y., V. Padmanabhan, and P. Seetharaman (2007): "An Econometric Model of Location and Pricing in the Gasoline Market," Journal of Marketing Research, 44(4), 622–635.

Chandra, A., and M. Tappata (2011): "Consumer search and dynamic price dispersion: an application to gasoline markets," The RAND Journal of Economics, 42(4), 681–704.

Chen, X., H. Hong, and M. Shum (2007): "Nonparametric Likelihood Ratio Model Selection Tests between Parametric Likelihood and Moment Condition Models," Journal of Econometrics, 141(1), 109–140.

De los Santos, B., A. Hortaçsu, and M. Wildenbeest (2012): "Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior," American Economic Review, 102(6), 2955–2980.

Eckert, A. (2013): "Empirical Studies of Gasoline Retailing: A Guide to the Literature," Journal of Economic Surveys, 27(1), 140–166.

Firgo, M., D. Pennerstorfer, and C. R. Weiss (2015): "Centrality and Pricing in Spatially Differentiated Markets: The Case of Gasoline," International Journal of Industrial Organization, 40, 81–90.

Fox, E. J., and S. J. Hoch (2005): "Cherry-Picking," Journal of Marketing, 69(1), 46–62.

Gauri, D. K., K. Sudhir, and D. Talukdar (2008): "The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data," Journal of Marketing Research, 45(2), 226–240.

Goldman, A., and J. K. Johansson (1978): "Determinants for Search of Lower Prices: An Empirical Assessment of the Economics of Information Theory," Journal of Consumer Research, 5(3), 176–186.

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

Hong, H., and M. Shum (2006): "Using Price Distributions to Estimate Search Costs," The RAND Journal of Economics, 37(2), 257–275.

Honka, E. (2014): "Quantifying Search and Switching Costs in the US Auto Insurance Industry," The RAND Journal of Economics, 45(4), 847–844.

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," The 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. (2012): "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," The American Economic Review, 102(5), 2147–2182.

Iyer, G., and P. Seetharaman (2003): "To Price Discriminate or Not: Product Choice and the Selection Bias Problem," Quantitative Marketing and Economics, 1(2), 155–178.

——— (2008): "Too Close to be Similar: Product and Price Competition in Retail Gasoline Markets," Quantitative Marketing and Economics, 6(3), 205–234.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online Demand under Limited Consumer Search," Marketing Science, 29(6), 1001–1023.

Koulayev, S. (2014): "Search for Differentiated Products: Identification and Estimation," The RAND Journal of Economics, 45(3), 553–575.

KUKSOV, D. (2004): "Buyer Search Costs and Endogenous Product Design," Marketing Science, 23(4), 490–499.

LACH, S., AND J. L. MORAGA-GONZÁLEZ (2012): "Heterogeneous Price Information and the Effect of Competition," Discussion paper, mimeo.

LEWIS, M. S. (2008): "Price Dispersion and Competition with Differentiated Sellers," Journal of Industrial Economics, 56(3), 654–678.

——— (2011): "Asymmetric Price Adjustment and Consumer Search: An Examination of the Retail Gasoline Market," Journal of Economics & Management Strategy, 20(2), 409–449.

LEWIS, M. S., AND H. P. MARVEL (2011): "When Do Ccnsumers Search?," Journal of Industrial Economics, 59(3), 457–483.

LEWIS, M. S., AND M. NOEL (2011): "The Speed of Gasoline Price Response in Markets with and without Edgeworth Cycles," Review of Economics and Statistics, 93(2), 672–682.

MANUSZAK, M. D., AND C. C. MOUL (2009): "How Far For a Buck? Tax Differences and the Location of Retail Gasoline Activity in Southeast Chicagoland," The Review of Economics and Statistics, 91(4), 744–765.

MARVEL, H. P. (1976): "The Economics of Information and Retail Gasoline Price Behavior: An Empirical Analysis," Journal of Political Economy, 84(5), 1033–1060.

MEHTA, N., S. RAJIV, AND K. SRINIVASAN (2003): "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," Marketing Science, 22(1), 58–84.

MORAGA GONZALEZ, J. L., Z. SÁNDOR, AND M. R. WILDENBEEST (2016): "Nonsequential Search Equilibrium with Search Cost Heterogeneity," Discussion paper.

MORAGA-GONZÁLEZ, J. L., AND M. R. WILDENBEEST (2008): "Maximum Likelihood Estimation of Search Costs," European Economic Review, 52(5), 820–848.

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

PENNERSTORFER, D., P. SCHMIDT-DENGLER, N. SCHUTZ, C. WEISS, AND B. YONTCHEVA (2015): "Information and Price Dispersion: Theory and Evidence," Discussion paper.

RATCHFORD, B. T. (2009): "Consumer Search and Pricing," in Handbook of Pricing Research in Marketing, ed. by V. R. Rao, pp. 91–107. Edward Elgar Publishing Limited, Cheltenham, UK.

RATCHFORD, B. T., M.-S. LEE, AND D. TALUKDAR (2003): "The Impact of the Internet on Information Search for Automobiles," Journal of Marketing Research, 40(2), 193–209.

RATCHFORD, B. T., AND N. SRINIVASAN (1993): "An Empirical Investigation of Returns to Search," Marketing Science, 12(1), 73–87.

REMER, M. (2015): "An Empirical Investigation of the Determinants of Asymmetric Pricing," International Journal of Industrial Organization, 42, 46–56.

ROBERTS, J. H., AND J. M. LATTIN (1991): "Development and Testing of a Model of Consideration Set Composition," Journal of Marketing Research, 28(4), 429–440.

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

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

Taylor, C. T., N. M. Kreisle, and P. R. Zimmerman (2010): "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California: Comment," The American Economic Review, 100(3), 1269–1276.

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

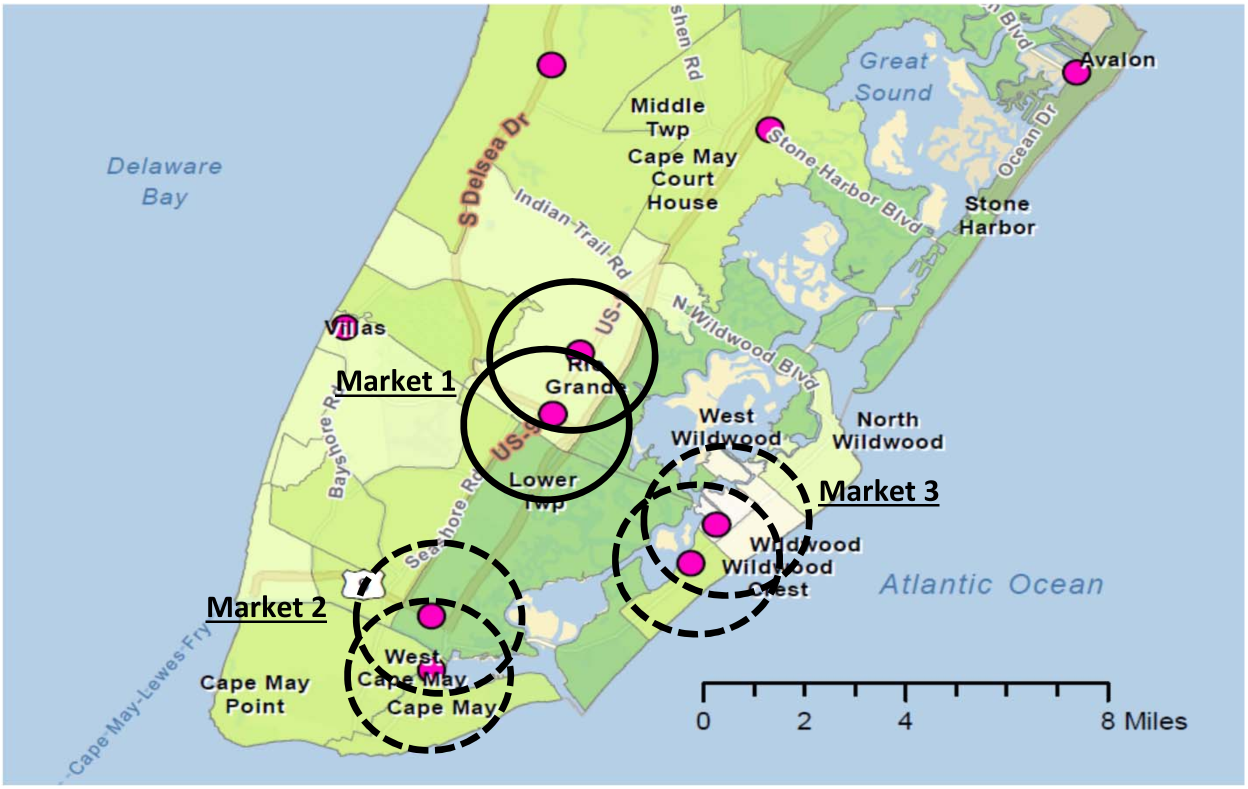
Verlinda, J. A. (2008): "Do Rockets Rise Faster and Feathers Fall Slower in an Atmosphere of Local Market Power? Evidence from the Retail Gasoline Market," Journal of Industrial Economics, 56(3), 581–612.

Wildenbeest, M. R. (2011): "An Empirical Model of Search with Vertically Differentiated Products," The RAND Journal of Economics, 42(4), 729–757.

Zhang, X., T. Y. Chan, and Y. Xie (2015): "Price Search and Periodic Price Discounts," Discussion paper.

Figure 1a

Example of 1.5-Mile Radius Isolated Markets at Capemay, New Jersey



Note: A small red circle represents a gas station location. A large black circle has a 1.5-mile radius. Different terrain colors represent different census tracts. Out of Markets 1, 2, and 3 with more than one station, only Market 1 qualifies as an "isolated" market in our sample, because (1) every station is located within 1.5 miles of all other stations and (2) every gas station in the market belongs to the same census tract (Rio Grande for Market 1).

Figure 1b

Distribution of Number of Stations by Isolated Markets (1.5 Miles Radius, 2007)

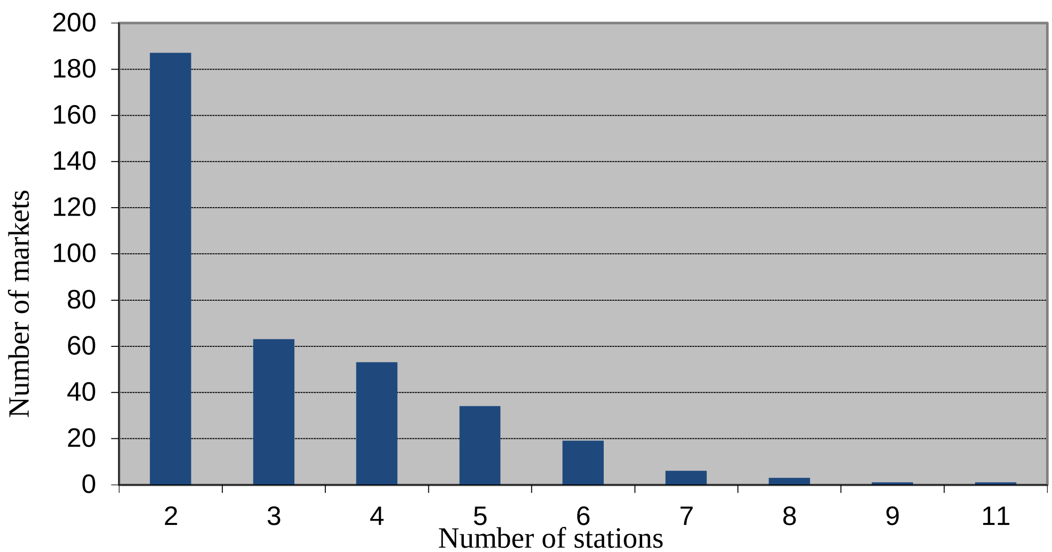
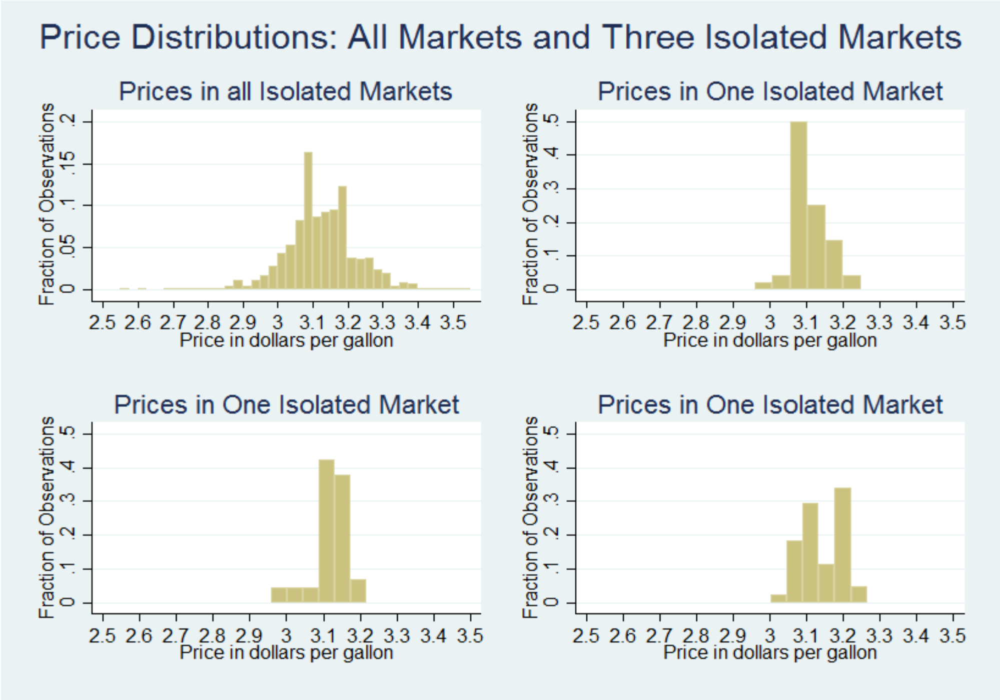
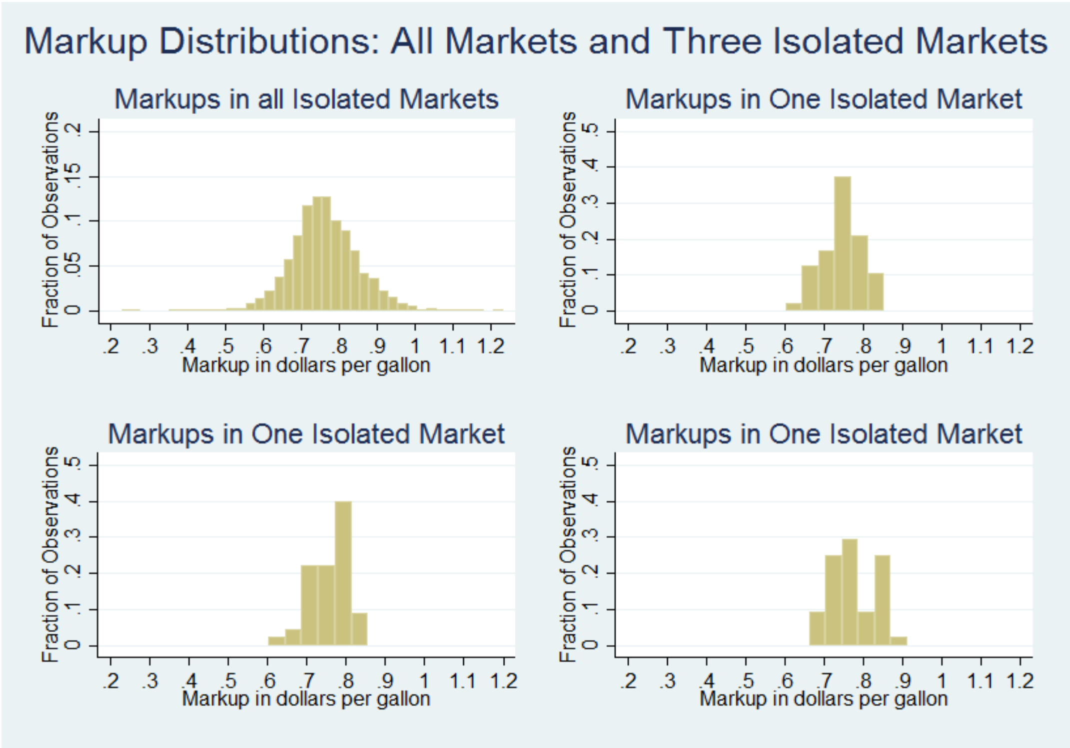


Figure 2a



Notes: The first panel is the distribution of prices across all isolated markets from February 27th to March 28th, 2007. The remaining panels are the distribution of prices over the same time in three sample isolated markets with three gas stations.

Figure 2b



Notes: The first panel is the distribution of markups across all markets from February 27th to March 28th, 2007. The remaining panels are the distribution of markups in three sample isolated markets with three gas station, and are the same markets as in Figure 2a.

Figure 3a

Histograms of Average Market Prices and Standard Deviation of Prices

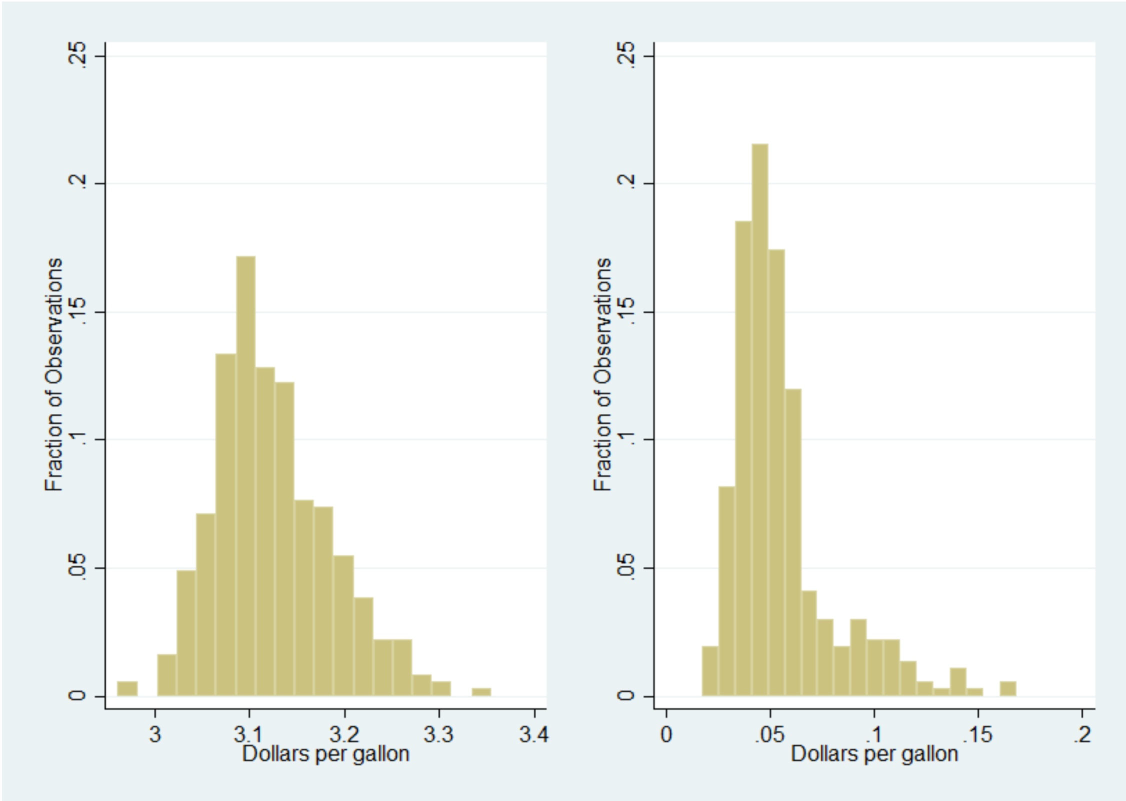
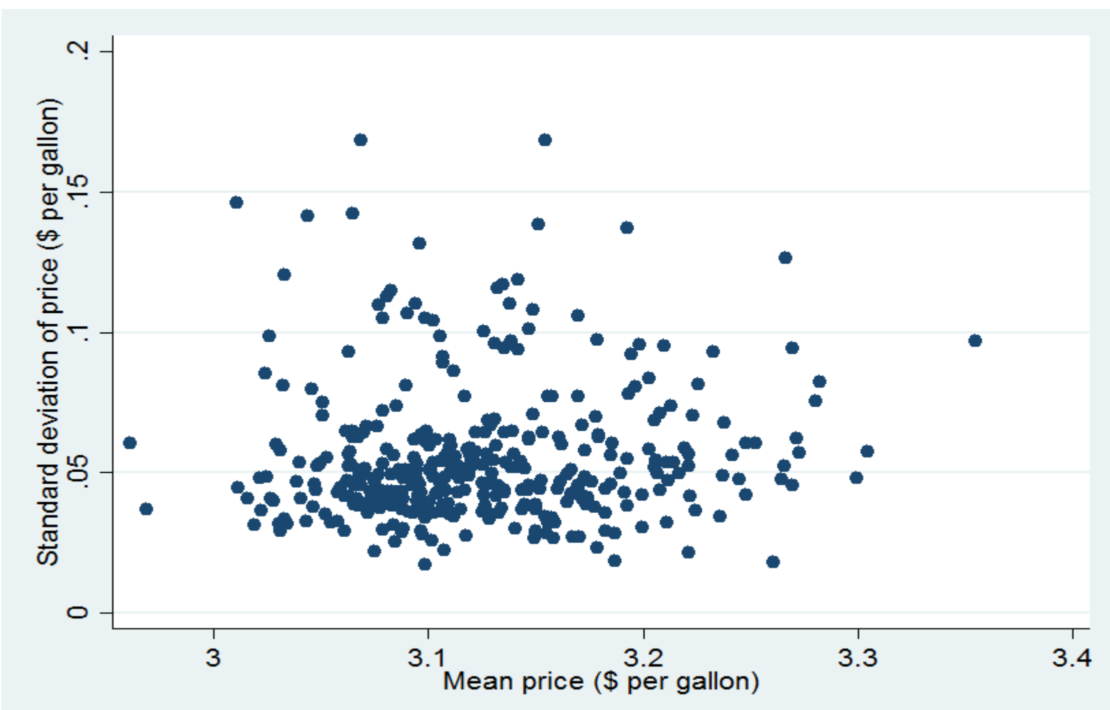


Figure 3b

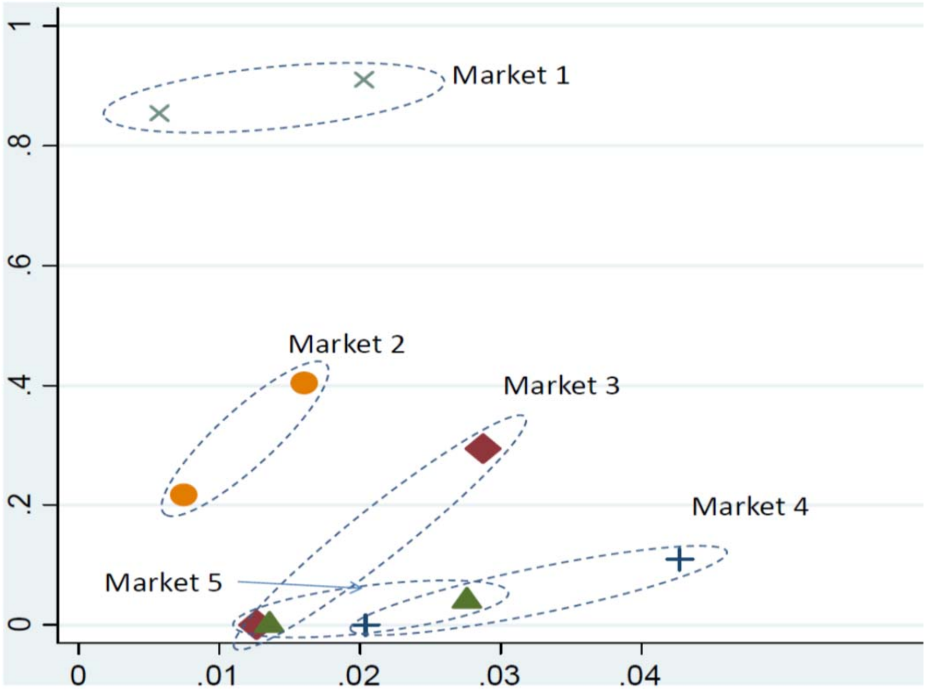
Scatterplot of Standard Deviation of Prices and Average Market Prices



Notes: The average market price has a state fixed effect removed. This largely controls for state taxes. The standard deviation is measured using all oberved station prices in an isolated market for the 30-day window. We use data from the 367 isolated markets used in the markup specification.

Figure 4a

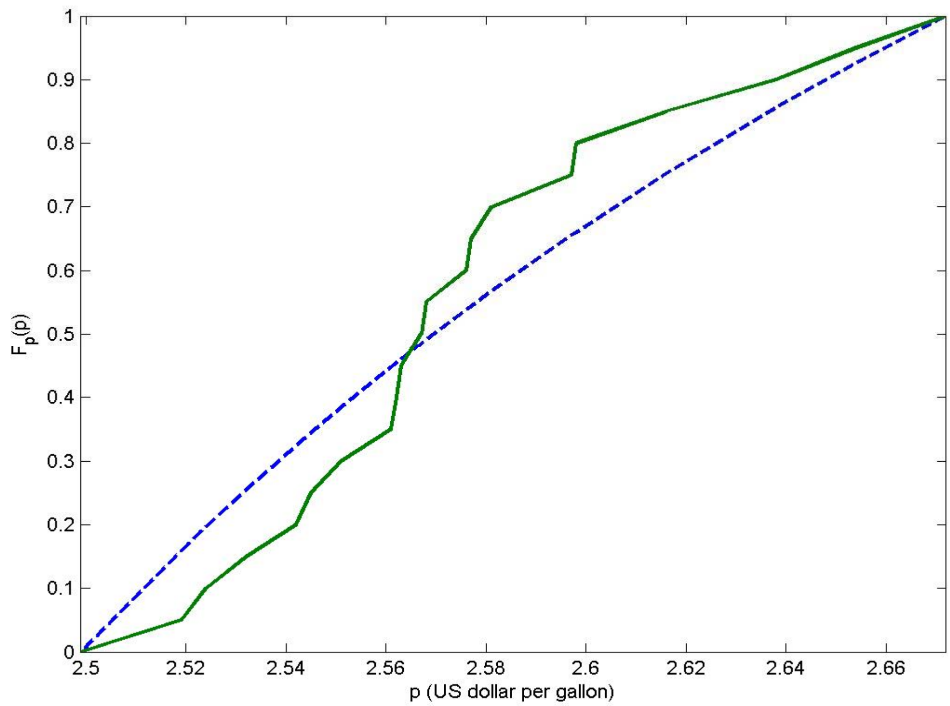
Estimated Search-Cost CDF from Five Randomly Picked Markets, February 27 through March 28, 2007



Note: The horizontal axis is the search cost in dollar per gallon. This figure plots the combination of estimated Δ_i and $Fc(\Delta_i)$ for each of five geographically isolated markets in Texas. All markets have three stations. For a given market, the right and left points correspond to $(\Delta_1, Fc(\Delta_1))$ and $(\Delta_2, Fc(\Delta_2))$, respectively.

Figure 4b

Empirical and Estimated (dash) Price Distribution, February 27th through March 28th, 2007



Note: A market in Florida. The horizontal axis is retail price per gallon.

Figure 5a

Histograms of Mean of Estimated Search Cost Distribution

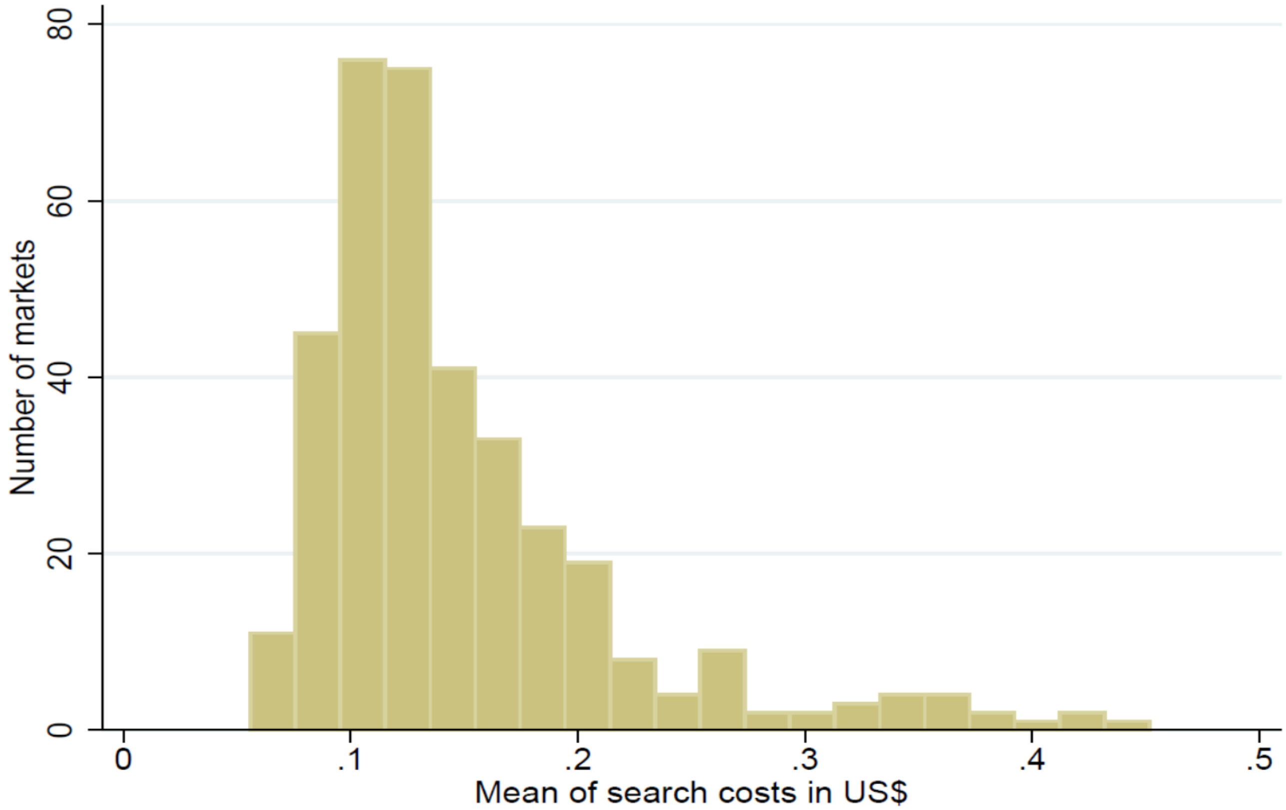


Figure 5b

Histograms of Standard Deviation of Estimated Search Cost Distribution

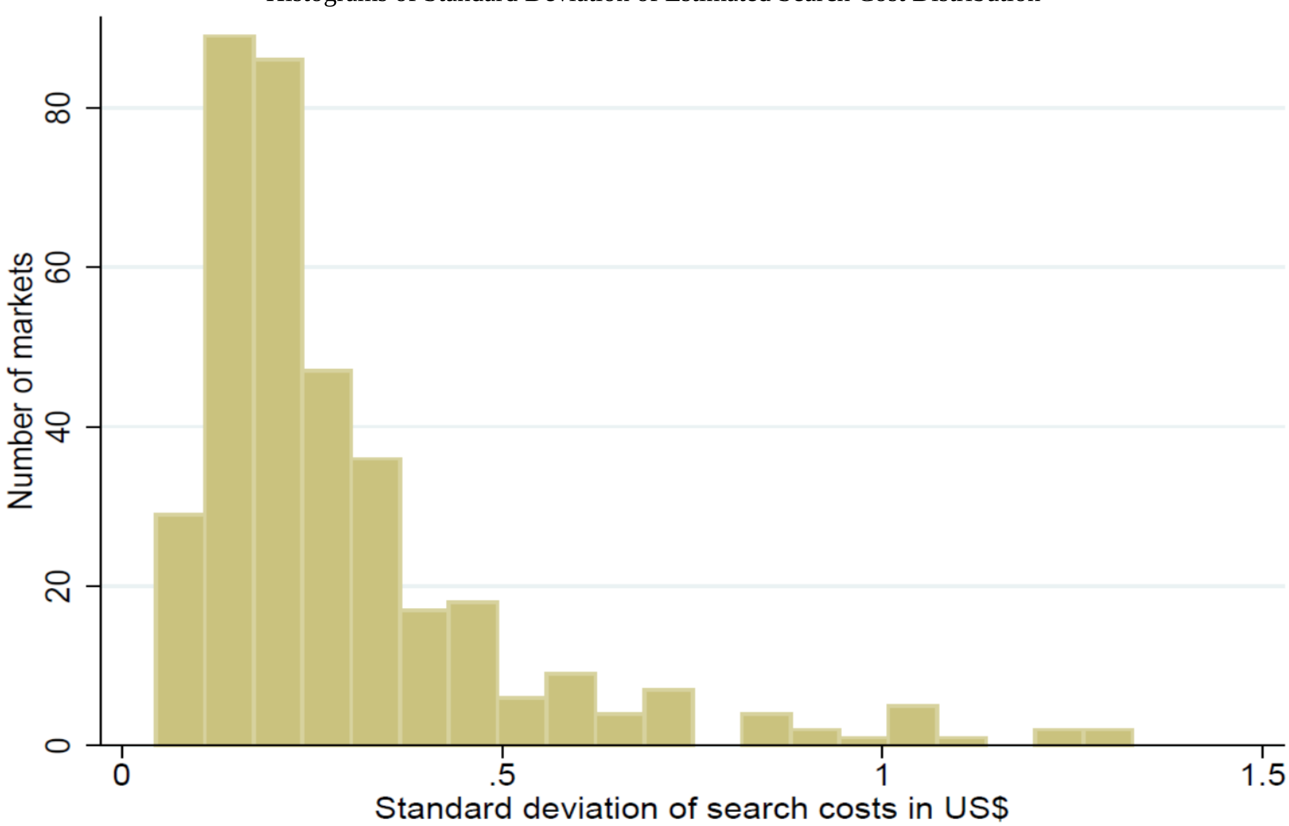


Figure 6a

Change in Average Price after a 20% Decrease in Expected Value and 10% Decrease in Standard Deviation of Search Costs

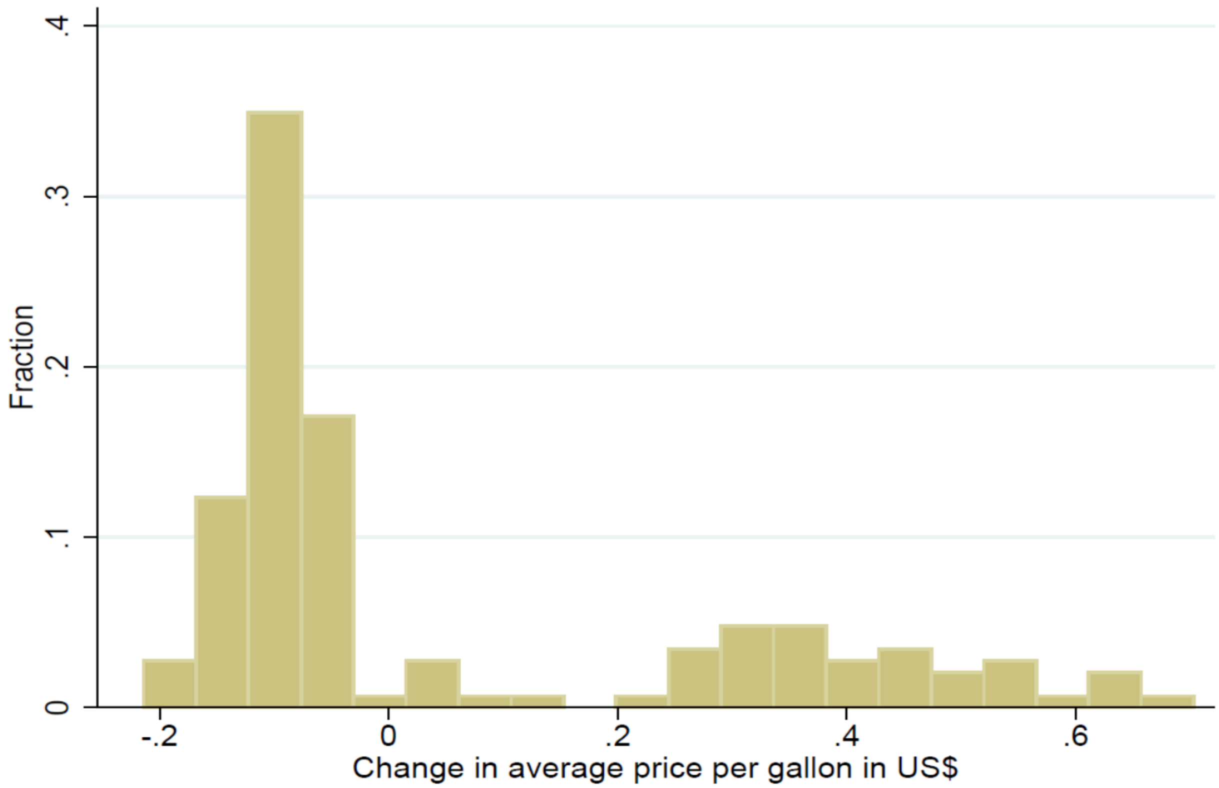


Figure 6b

Change in Average Price after a 20% Decrease in Expected Value and 15% Decrease in Standard Deviation of Search Costs

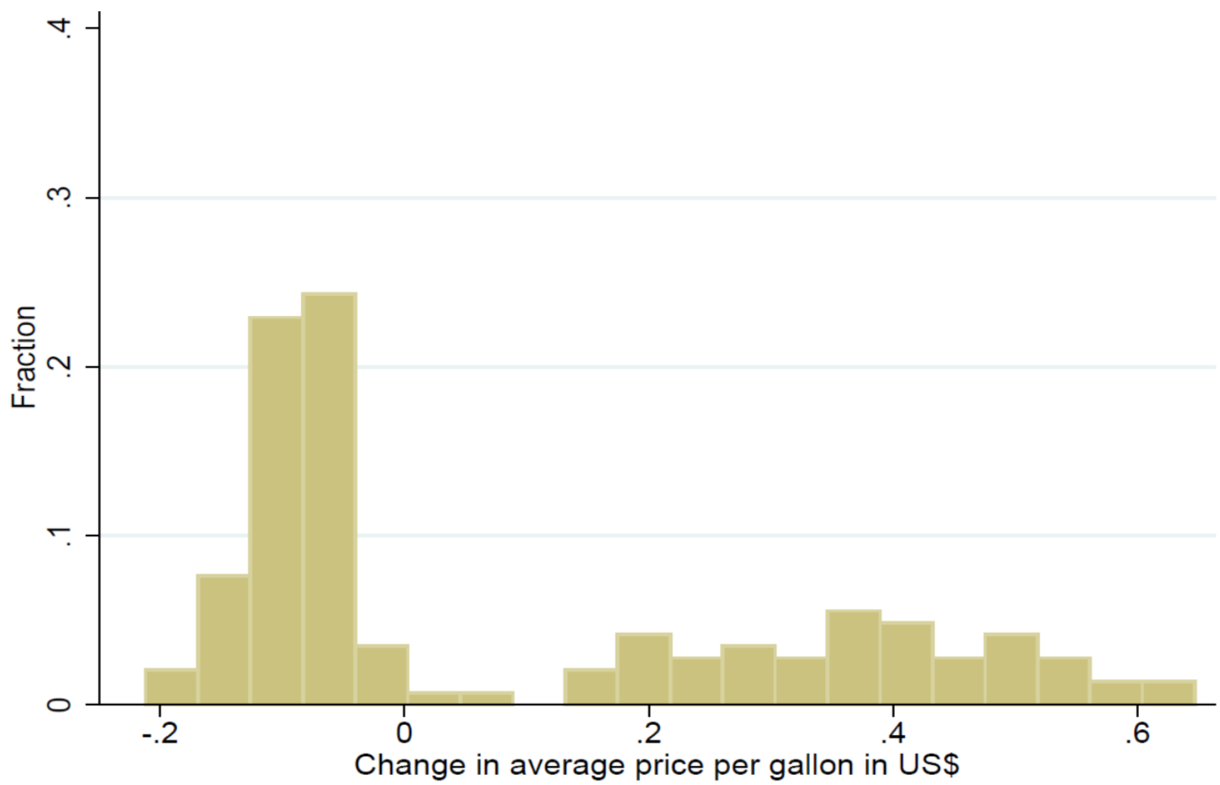


Table 1

Summary Statistics for Demographics across Isolated and Non-Isolated Geographical Markets

	Mean	Std dev	Minimum	Maximum
Isolated Markets				
Mean household income ($)	64,163	27,577	24,328	228,277
Years of education	12.625	1.158	8.643	15.761
Age	39.082	5.092	26.489	63.020
Distance between stations (miles)	0.433	0.339	0.002	1.476
Standard deviation of income	48,168	13,190	18,428	117,873
Non-Isolated Markets				
Mean household income ($)	70,608	32,907	11,306	396,574
Years of education	13.053	1.336	7.592	17.368
Age	37.177	5.525	17.602	77.193
Distance between stations (miles)	0.862	0.213	0.002	1.498
Standard deviation of income	50,085	16,137	6,822	263,025

Note: The data source is the 2006-2010 American Community Survey. The first set of summary statistics are for isolated markets, as defined in the second section; 367 isolated markets exist. The second set of statistics are for all markets that do not fit the isolated market critieria; 37,746 "non-isolated" markets exist. A firm can belong to more than one "non-isolated" market, because a market is defined as a 1.5-mile radius around each firm in the data; therefore, many overlapping markets exist. When the set of "non-isolated" markets is restricted such that each firm can belong to only one market, the summary statistics are largely unchanged; 5,138 of these "non-overlapping/non-isolated" markets exist.

Table 2

Price Dispersion Summary Statistics

	All Prices		Average Across Isolated Markets	
	Retail Price	Markup	Retail Price	Markup
Minimum	2.549	0.229	2.964	0.630
			(0.101)	(0.082)
Maximum	3.549	1.232	3.226	0.859
			(0.085)	(0.085)
Mean	3.127	0.764	3.124	0.761
			(0.070)	(0.061)
Range (= Max - Min)	1.000	1.003	0.262	0.229
			(0.105)	(0.026)
Standard deviation	0.099	0.089	0.067	0.057
			(0.033)	(0.102)
Number of markets			354	367

Note: The above statistics are for regular-grade gasoline prices from isolated markets in the states of CA, FL, NJ, and TX from February 27 through March 28, 2007. Markup is defined as the retail regular fuel price minus the wholesale cost. The first two columns present statistics when price observations are pooled across all isolated markets. There are a total of 12,250 price observations across all isolated markets. The last two columns give the average of the statistics across all isolated markets, and standard deviations are in parenthesis. For example, the minimum observed price in any isolated market is $2.549, and the average minimum price across all isolated markets is $2.964.

Table 3
Price Dispersion Regression Estimates

Dispersion measure	Sample range				Sample standard devation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean income	0.068	0.091**			0.011	0.015		
	(0.044)	(0.043)			(0.012)	(0.012)		
Mean years of education	-0.012	-0.010			-0.003	-0.003		
	(0.008)	(0.008)			(0.002)	(0.002)		
Mean age	-0.001	-0.000			0.000	0.000		
	(0.001)	(0.001)			(0.000)	(0.000)		
Mean distance among stations	0.007	0.000			-0.005	-0.006*		
	(0.013)	(0.013)			(0.003)	(0.003)		
Standard devation of income	-0.070	-0.116			-0.009	-0.016		
	(0.090)	(0.088)			(0.025)	(0.025)		
Standard devation of education	0.025	0.008			0.007	0.004		
	(0.017)	(0.017)			(0.005)	(0.005)		
Standard devation of age	0.000	-0.003			-0.001	-0.001		
	(0.003)	(0.003)			(0.001)	(0.001)		
Number of firms		0.018***		0.018***		0.003***		0.003**
		(0.003)		(0.003)		(0.001)		(0.001)
Log of mean income			0.067*	0.077**			0.015	0.016
			(0.038)	(0.037)			(0.010)	(0.010)
Log of mean years of education			-0.176*	-0.123			-0.046	-0.038
			(0.102)	(0.101)			(0.030)	(0.030)
Log of mean age			-0.041	-0.015			-0.001	0.003
			(0.044)	(0.043)			(0.011)	(0.011)
Log of mean distance among stations			0.007*	0.002			0.000	-0.000
			(0.004)	(0.004)			(0.001)	(0.001)
Log of standard devation of income			-0.048	-0.070			-0.010	-0.013
			(0.049)	(0.049)			(0.012)	(0.013)
Log of standard devation of education			0.059	0.024			0.017	0.012
			(0.048)	(0.047)			(0.013)	(0.014)
Log of standard deviation of age			0.027	-0.043			-0.008	-0.019
			(0.075)	(0.075)			(0.019)	(0.020)
Constant	0.323**	0.350***	0.457	0.568**	0.082**	0.086**	0.122	0.140
	(0.129)	(0.123)	(0.294)	(0.287)	(0.036)	(0.036)	(0.086)	(0.086)
Observations	367	367	367	367	367	367	367	367

Notes: The sample range is defined as the difference between the maximum and minimum observed price
from February 27th to March 28th, 2007 in an isolated market. The standard deviation is defined as the sample standard
deviation of all observed prices in a market over the same date range.
Robust standard errors are in parentheses. Income is standardized to $100,000
*** p<0.01, ** p<0.05, * p<0.1

Table 4
Estimated Search-Cost CDF for All Markets

	Mean	Std. Dev	Min	Max	# of Obs.
Proportion of people with l price quotes (q_l)					
q_1	0.664	0.155	0.004	0.936	367
q_2	0.301	0.130	0.031	0.755	367
q_3	0.063	0.112	0.000	0.781	367
q_4	0.073	0.095	0.000	0.396	180
q_5	0.107	0.139	0.000	0.682	117
q_6	0.105	0.089	0.000	0.273	64
q_7	0.100	0.152	0.000	0.512	30
q_8	0.064	0.060	0.000	0.124	11
q_9	0.098	0.138	0.000	0.195	5
q_{10}	0.000	.	0.000	0.000	2
q_{11}	0.166	.	0.166	0.166	1
Marginal expected savings from searching $l+1$ versus l stations (Δ_l)					
Δ_1	0.032	0.014	0.007	0.100	367
Δ_2	0.016	0.008	0.001	0.044	180
Δ_3	0.010	0.005	0.000	0.026	117
Δ_4	0.007	0.003	0.000	0.016	64
Δ_5	0.005	0.002	0.003	0.011	30
Δ_6	0.004	0.002	0.002	0.007	11
Δ_7	0.003	0.001	0.002	0.003	5
Δ_8	0.002	0.001	0.002	0.003	2
Δ_9	0.002	.	0.002	0.002	1
Δ_{10}	0.002	.	0.002	0.002	1
Valuation-cost margin ($ per gallon)	0.261	0.217	0.995	0.000	367

Note: The estimated valuation-cost margin measures the gap between the utility and the marginal cost of providing gasoline per gallon (i.e., $x = v_j - r_j$) for each market. This valuation-cost margin measures the maximum profit margin per gallon possible from that market.

Table 5
Nonlinear Least Square Estimation of Search-Cost Cumulative Distribution using Markup Price

Price	Markup price			
Distributional assumption	Lognormal			Normal
	(1)	(2)	(3)	(4)
Mean of distribution				
Constant	-6.926***	-5.838***	-5.647***	-0.019
	(1.980)	(1.905)	(1.733)	(0.088)
Mean income	0.716***	0.609***	0.408**	0.0196***
	(0.175)	(0.118)	(0.195)	(0.007)
Mean years of education	-0.361		-0.390	-0.002
	(0.962)		(0.690)	(0.034)
Mean age	-0.572	-0.797**		-0.030*
	(0.510)	(0.392)		(0.018)
Mean distance among stations	-0.594***	-0.596***	-0.5685***	-0.0155**
	(0.214)	(0.213)	(0.201)	(0.008)
Standard deviation of distribution				
Constant	-0.014	0.011	-2.656**	-0.948
	(2.590)	(2.631)	(1.183)	(2.585)
Standard deviation of income	0.502***	0.499***	0.351***	0.398**
	(0.123)	(0.124)	(0.110)	(0.163)
Standard deviation of years of education	-0.284		-0.420	-0.460
	(0.326)		(0.320)	(0.370)
Standard deviation of age	-1.392**	-1.492***		-1.768**
	(0.571)	(0.561)		(0.724)
Mean squared error	17.331	17.372	17.489	18.446
Number of observations	726	726	726	726

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The regressors are in logs.