Geographical Dispersion of Consumer Search Behavior

Hakan Yilmazkuday[*]

April 26, 2017

Abstract

This paper investigates whether consumer search behavior differs across zip codes within the U.S.. As an application, daily gasoline price data covering virtually all gas stations within the U.S. are employed to estimate the distribution of search costs in each zip code. The results show that there are significant differences across zip codes regarding the expected number of searches achieved before consumers purchase gasoline. In order to have a systematic explanation, such differences are further connected to geographic, demographic and economic conditions of the zip codes in a secondary analysis. The corresponding results imply several strategies for gas stations in order to maximize profits/markups; suggestions follow for policy makers and regulators to reduce redistributive effects of information barriers across locations.

JEL Classification: D12, D83, L81

Key Words: Consumer Search; Price Dispersion; Retail Gasoline.

[*] Department of Economics, Florida International University. E-mail: hyilmazk@fiu.edu. Phone: +1-305-348-2316. Fax: +1-305-348-1524.

1. Introduction

Prices for the very same (homogeneous) good can be different across retailers. This is most apparent in the gasoline market where gas stations post alternative prices even within the same zip code in the U.S.. For example, consider Figure 1 where the gasoline price spread has a median value of 14 cents with a range between 0 and 98 cents.[1] Since these retail prices are already controlled for gas-station and time fixed effects at the zip code level, they are independent of any gas station characteristics such as their location, brand, competition level, having a car wash or a convenient store as well as time-varying supply or demand shocks. One potential explanation to these price spreads is then the lack of information that consumers have, which has been connected to search costs in the literature following the seminal article of Stigler (1961) followed by other influential studies such as by Varian (1980), Burdett and Judd (1983), and Stahl (1989). In particular, if consumers do not search for lower prices, retailers may easily charge higher markups or get involved in collusive behavior. Accordingly, policy makers have considered this lack of information as a potential problem reducing consumer welfare due to information frictions.[2]

Within this picture, we investigate the search behavior across consumers in different five-digit zip codes within the U.S., where gasoline purchases account for approximately 5% of consumer spending.[3] First, we would like to know whether the search behavior of consumers differs across zip codes; we are particularly interested in the expected number of searches achieved by consumers before making a purchase. Accordingly, by using retail level gasoline price data obtained from virtually all gas stations with the U.S. as an application, we first estimate the expected number of searches and the corresponding search cost distributions at the zip code level. We achieve this by considering the implications of a non-sequential consumer search model with heterogeneous search costs. The model is the multi-region version of the model introduced by Moraga-Gonzalez and Wildenbeest (2008) which is an oligopolistic version of the model proposed in Hong and Shum (2006) who have generalized the non-sequential consumer search model of Burdett and Judd (1983) by adding search cost heterogeneity. The results show that the expected number of searches have a median of 1.66 across zip codes, which implies that consumers do not search much on average before

[1] Gasoline price spread is defined as the difference between the maximum and the minimum price in gas stations in a given zip code after controlling for gas-station and time fixed effects where the latter includes both day and hour fixed effects. See the data section below for further details.

[2] Consider the case of South Korea where gas stations are required to post their retail prices on Opinet. Policy makers in other countries such as Austria, parts of Australia, Luxembourg or parts of Canada have also adopted regulatory pricing rules for gas stations; see Haucap and Müller (2012).

[3] See Consumer Expenditure Survey, 2014, for example: http://www.bls.gov/cex/2014/combined/quintile.pdf

purchasing gasoline. However, the estimates for the expected number of searches range between 0.17 and 12.94 across zip codes; hence, there are significant differences in the search behavior of consumers across zip codes.

Understanding the reasons behind this heterogeneity in the search behavior of consumers across regions is the key to reduce the redistributive effects of information frictions. In particular, if information frictions are systematically higher in certain regions, policy makers can reduce them only by achieving region-specific policies. Instead, if a common multi-region policy is conducted, although it would reduce information frictions in all regions, it would not necessarily reduce redistributive effects of information frictions across regions.[4] Accordingly, in a secondary analysis, we investigate whether the heterogeneity in the estimated consumer search behavior can be explained systematically across zip codes. In particular, we attempt to connect the estimated expected number of searches to geographic, demographic and economic conditions of zip codes. It is found that geographical frictions due to factors such as the overall area of the zip code, population density or average distance between gas stations all contribute positively to the expected number of searches achieved by consumers. On the other hand, income and commuting time are shown to be negatively related to the expected number of searches across zip codes, potentially capturing the opportunity cost of time for making a search. Consumers in zip codes with individuals working in different industries are shown to have different search behavior as well, with industries such as retail trade and public administration contributing most to the expected number of searches. It is also shown that consumers in zip codes with higher percentage of Black or African American people search more compared to those with white or Asian people; consumers in zip codes with higher percentage of females are also shown to search more compared to other zip codes. Based on these results, several strategies are implied for gas stations in order to maximize profits/markups; suggestions follow for policy makers and regulators in order to reduce the redistributive effects of information frictions across locations.

This paper belongs to the literature estimating the search behavior of consumers by using only price data. Within this picture, Hortacsu and Syverson (2004) have estimated search cost distribution of the U.S. mutual funds market, Hong and Shum (2006) have estimated that of online textbooks, and Moraga-Gonzalez and Wildenbeest (2008) and Moraga-Gonzalez et al. (2013) have estimated that of computer memory chips. These papers have focused on the estimation of a single market, while this study deviates from them by considering the

[4] A recent example of such a local policy (to reduce information frictions) has been achieved by Palm Beach County, Florida (PBC) that has passed and put into effect an ordinance regarding the size of cash versus credit card prices at gas stations; however, this ordinance has been trumped by a common multi-location policy of the State of Florida due to its new law that binds all counties within Florida.

segmentation of the U.S. gasoline market based on the zip codes that the gas stations are located at. Such a strategy in this paper is essential to investigate the relationship between the consumer search behavior and geographic, demographic and economic conditions across geographical locations.

This paper also belongs to literature based on the consumer search behavior in the retail gasoline market. Studies such as by Marvel (1976), Lewis (2008), Chandra and Tappata (2011), and Pennerstorfer et al. (2014) have investigated how the price dispersion in the gasoline market can be connected to the models of costly consumer search. However, these studies have been limited due to the reduced-form testing of the comparative static relationships implied by their models. In contrast, this paper considers the information coming from the overall distribution of the consumer search behavior across zip codes within the U.S.. Within this literature, Nishida and Remer (2015; NR henceforth) is the closest study to this paper. By using the same estimation methodology, NR estimate the average and the standard deviation of search costs across markets using a similar data set on gasoline prices. Their results show 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. This paper deviates from NR in several dimensions. First and most importantly, we focus how the expected number of searches changes across markets, while NR focus on how the corresponding search costs change across markets. In other words, while we focus on the *quantity* of searches across markets, NR focus on the *price* of searches. Second, our data set, which is unique to this paper due to our efforts in collecting our own data, covers virtually all gas stations within the U.S., while their investigation focuses only on the states of California, Florida, New Jersey, and Texas by borrowing a subset of the data set that has been previously utilized by Chandra and Tappata (2011). Third, we investigate how the search behavior of consumers changes across zip codes, while NR focus on how the search behavior of consumers changes across geographic markets defined as great circles with a radius of 1.5 miles; having an investigation at the zip code level has the advantage of connecting zip code characteristics such as income, poverty, population density, commuting time, industries worked, area, race and sex to the search behavior of consumers as we achieve in this paper. Finally, while we control the retail level gasoline prices for retailer/station fixed effects and time fixed effects, where the latter include both day and hour (of data collection) fixed effects, NR controls only for retailer/station fixed effects. Regarding the testable implications of the estimated model as discussed by Moraga-Gonzalez and Wildenbeest (2008), missing to control for time fixed effects may lead to biased results, because, according to the model that is common between this paper and NR, (i) prices should be dispersed at any given moment in time, (ii) there should be variation in the position of a typical retailer in the price ranking, and (iii) supply or demand shocks should be absent during the sample period. These assumptions can only be satisfied by controlling the retail level gasoline prices for time fixed effects, together with retailer/station fixed effects, at the market level; accordingly, we control the retail level gasoline prices for both retailer/station and time fixed effects for each zip code individually. However, NR controls retail prices only for retailer/station fixed effects across all gas stations in their sample (rather than market by market); this may create an additional bias in their results due to not satisfying the assumptions mentioned above.

The next section introduces the consumer search model. Section 3 estimates the expected number of searches together with the distribution of search costs across zip codes using the daily gasoline price data. Section 4 connects the expected number of searches to the zip code characteristics depending on geographic, demographic and economic conditions. Section 5 concludes by providing suggestions for both gas stations (in order to maximize profit) and policy makers (for regulatory purposes).

2. Consumer Search Model

We employ a non-sequential consumer search model with heterogeneous search costs. The model is the multi-region version of the model introduced by Moraga-Gonzalez and Wildenbeest (2008) which is an oligopolistic version of the model proposed in Hong and Shum (2006) who have generalized the non-sequential consumer search model of Burdett and Judd (1983) by adding search cost heterogeneity. The economic environment consists of regions that are inhabited by retailers and consumers. Unlike Moraga-Gonzalez et al. (2013), who assume that different consumer markets have the same underlying search cost distribution, we focus on market segmentation where each region has its own search cost distribution; this is necessary to investigate whether consumer search behavior differs across zip codes within the U.S..

Region r is inhabited by N_r retailers who sell a homogenous good with a common unit cost of m_r, although the price charged by each retailer may be different. Consumers in each region know the distribution of retail prices, however they do not know which retailer charges which price; accordingly, they search for a subset of retailers to obtain information about prices. In order to obtain any price information beyond the first price observed, consumers in region r have to pay a randomly drawn search cost of c_r that differs across consumers in that region according to the distribution of search costs given by F_r^c. Total search cost $c_r i_r$ of a consumer in region r is simply determined by the multiplication of the search cost c_r and the number of retailers sampled i_r.

The symmetric mixed strategy equilibrium in region r is denoted by the distribution of prices F_r^p with density $f_r^p\left(p_r\right)$. Given the behavior of the retailers in region r, the consumer decides on the optimal number of retailers to search according to the following expression:

$$i_r\left(c_r\right) = \arg\min_{i_r>1} c_r\left(i_r-1\right) + \int_{\underline{p_r}}^{\overline{p_r}} i_r p_r \left(1-F_r^p\left(p_r\right)\right)^{i_r-1} f_r^p\left(p_r\right) dp_r \tag{2.1}$$

where $\underline{p_r}$ and $\overline{p_r}$ represent the lower and upper bound of the support of $F_r^p\left(p_r\right)$. Since $i_r\left(c_r\right)$ must be an integer, Equation 2.1 corresponds to the partition of consumers in region r into N_r subsets, each subset representing the fraction q_r^i of consumers searching for $i_r\left(=1,2,...,N_r\right)$ retailers; it is implied that $\sum_{i_r=1}^{N_r} q_r^i = 1$.

In order to calculate the fraction q_r^i in region r, consider the following search cost of a consumer who is indifferent between searching i_r retailers and i_r+1 retailers:

$$\Delta_r^i = Ep_r^{1:i_r} - Ep_r^{1:i_r+1} \tag{2.2}$$

where $Ep_r^{1:i_r}$ represents the expected minimum price in a sample of i_r prices drawn from the price distribution of $F_r^p\left(p_r\right)$. Since Δ_r^i is a decreasing function of i_r, the fractions of consumers sampling i_r prices in region r as implied as follows:

$$\begin{aligned} q_r^1 &= 1 - F_r^c\left(\Delta_r^1\right) \\ q_r^i &= F_r^c\left(\Delta_r^{i-1}\right) - F_r^c\left(\Delta_r^i\right), \quad i = 2,3,..,N_r-1 \\ q_r^N &= F_r^c\left(\Delta_r^{N-1}\right) \end{aligned} \tag{2.3}$$

where it is optimal for the retailers to mix in prices, given the search behavior of consumers.

The equilibrium price distribution in region r is obtained by considering the indifference condition that a retailer should obtain the same level of profits from charging any price in the support of $F_r^p\left(p_r\right)$:

$$\left(p_r - m_r\right)\left[\sum_{i_r=1}^{N_r} \frac{i_r q_r^{i_r}}{N_r}\left(1-F_r^p\left(p_r\right)\right)^{i_r-1}\right] = \frac{q_r^1\left(\overline{p_r}-m_r\right)}{N_r} \tag{2.4}$$

where q_r^1 represents the fraction of consumers who do not compare prices; thus, some of them end up with paying the upper bound $\overline{p_r}$ of the price distribution. It is implied that the minimum price charged in region r is given by:

$$\underline{p_r} = \frac{q_r^1\left(\overline{p_r}-m_r\right)}{\sum_{i_r=1}^{N_r} i_r q_r^{i_r}} + m_r \tag{2.5}$$

where the first term on the right hand side represents the additive markup on the minimum price. It is implied that the ratio of the maximum additive markup to the minimum additive markup within region r is given as follows:

$$\frac{\overline{p_r}-m_r}{\underline{p_r}-m_r} = \frac{S_r}{q_r^1}$$

where the numerator of the right hand side S_r $\left(= \sum_{i_r=1}^{N_r} i_r q_r^{i_r}\right)$ is the expected number of retailers searched in order to find lower prices, while the denominator is the fraction of consumers who do not compare prices. This ratio would be equal to one when the minimum price is equal to the maximum price (i.e., $\underline{p_r} = \overline{p_r}$), implying that none of the consumers would compare prices (i.e., $S_r = \sum_{i_r=1}^{N_r} i_r q_r^{i_r} = q_r^1 = 1$); this is due to having the same expected minimum price across different number of retailers sampled when prices are the same. As the maximum price gets higher compared to the minimum price (i.e., when the price dispersion increases across retailers), the fraction of consumers who do not compare prices q_r^1 would go down (i.e., some consumers would start searching for lower prices); this is due to positive potential gains out of making costly search. In an extreme case in which the price dispersion goes to infinity, q_r^1 would go to zero, implying that all consumers would make some search for lower prices.

We test the implications of this model on the dispersion of gasoline prices across gas stations next.

3. Estimation of Search Costs

3.1. Data and Estimation Methodology

Using gasoline price data obtained at the retail (i.e., gas station) level, the regions in the model are matched with five-digit zip codes within the U.S.. The gasoline prices have been downloaded at midnight of each day from MapQuest (http://gasprices.mapquest.com/) by using an automated procedure (written in Matlab) that scans the code of publicly available web pages, identifies relevant pieces of gasoline price information, and stores the data.[5] MapQuest receives gasoline prices from Oil Price Information Service (OPIS), a leading provider of petroleum data collecting gas price data based on fleet transaction data.[6] MapQuest gas prices are updated as qualifying transactions are processed by OPIS. We consider the daily gasoline price data for the whole month of July 2015. The data cover daily price observations from 112,515 gas stations within the U.S. for the whole month of July 2015 (i.e., for 31 days).

As shown by Hong and Shum (2006) and Moraga-Gonzalez and Wildenbeest (2008), Equations 2.1-2.5 provide enough information for the maximum likelihood estimation of the search cost distribution by using only retail price data; we refer the reader to these papers for

[5] This technique is commonly called "web scraping."

[6] Focusing on other topics and time periods, earlier studies such as by Abrantes-Metz et al. (2006), Doyle and Samphantharak (2008), and Chandra and Tappata (2011) have also used this data set.

technical details of the estimation.[7] Since we focus on the potential heterogeneity of search cost distributions across zip codes, we achieve the estimation for each zip code individually.

In order to match the gasoline price data with the consumer search model, it assumed that retailers in any zip code play a stationary repeated game of finite horizon so, in every period, the data should reflect the equilibrium of the static game analyzed in the model section. As shown by Moraga-Gonzalez and Wildenbeest (2008), this assumption has some testable implications at the zip code level such as (i) prices should be dispersed at any given moment in time, (ii) there should be variation in the position of a typical retailer in the price ranking, and (iii) supply or demand shocks should be absent during the sample period. On top of these assumptions, we also have an assumption coming from the consumer search model that the investigated good (i.e., gasoline) is a homogenous good. However, in a particular zip code, there are many factors that would violate these assumptions for retail level gasoline prices. The first two assumptions, together with the homogeneity assumption, may be violated due to some gas stations almost always setting higher prices due their brands and/or locations, while the second assumption may be violated due to daily changes in gasoline prices. Accordingly, we have to control for these factors before we can continue with the maximum likelihood estimation. By following the standard practice in many structural auction models (e.g. Haile et al., 2003; Bajari et al., 2006; An et al., 2010) and consumer search studies such as by Wildenbeest (2011), we achieve this by controlling the retail level gasoline prices for retailer/station fixed effects and time fixed effects, where the latter include both day and hour (of data collection) fixed effects.[8] In particular, for gas stations located in a particular zip code, we simply run a regression of gasoline prices on retailer/station fixed effects and time fixed effects in that zip code; we consider the residuals of this regression (plus the estimated constant that is specific to the zip code considered) as our measure of retail prices for the rest of this paper.[9]

Finally, it has been shown by Moraga-Gonzalez and Wildenbeest (2008) that the measurement error in the number of retailers may lead to biased estimates of the search cost distribution. Accordingly, we have to make sure that the number of gas stations in our sample in fact matches with the number of gas stations within the U.S.. We find that the number of gas stations (112,515) in our sample is very close to the number of gas stations in the 2013 County Business Patterns of the U.S. Census Bureau (i.e., the latest data available), which is 112,458. Therefore, we can safely claim that our daily gasoline price data cover virtually

[7] The Matlab codes for the estimation of search costs can be found at http://kelley.iu.edu/mwildenb/code.html.

[8] The approximate time of the gasoline price update is provided by MapQuest.

[9] Such a strategy is also important to control for gasoline markets that are inherently differentiated by the amenities offered and their locations (see e.g. Houde, 2012; Langer and McRae, 2014).

all gas stations within the U.S..

3.2. Estimation Results

The estimation is achieved for each zip code individually. The summary of the maximum likelihood estimations is given in Table 1 where the results across zip codes have been sorted with respect to the estimated expected number of searches S_r $\left(= \sum_{i_r=1}^{N_r} i_r q_r^{i_r}\right)$; the corresponding percentiles of zip codes are depicted. The corresponding estimates of critical search cost values across zip codes are given in Figure 2 (up to Δ^{10} to save space). Both Table 1 and Figure 2 show the importance of having an analysis at the zip code level, because the estimated values, which are all significant at the 5% level, are shown to be changing significantly across locations.

As is evident in Table 1, the median S_r across zip codes is 1.66 with range between 0.17 and 12.94. Therefore, on average, consumers do not search much for lower prices across gas stations. The median S_r estimate of 1.66 is consistent with other studies in the literature such as by Moraga-Gonzalez and Wildenbeest (2008) who have estimated S_r as 1.45, 1.60, 1.62 and 1.93 for different computer memory chips by using price data obtained from www.shopper.com. Compared to Hong and Shum (2006) who investigate the search costs for several economics and statistics textbook and estimate S_r as 1.06, 1.25, 1.26 and 1.47, however, the median S_r estimate of 1.66 in this paper is slightly higher.

In the zip code with the median S_r, the markups range between 7 cents and 22 cents. When we consider all other zip codes, although markups differ across stations, the median (across zip codes) difference between the minimum price and the unit cost is about 5.7 cents, while the median difference between the maximum price and the unit cost is about 24.13 cents. These markups, which have completely been obtained from the estimation of the proposed model, is consistent with the average markups discussed in the media or by organizations making research/surveys on gas stations; e.g., according to The Wall Street Journal, "The station owners, in turn, set their gas prices for consumers so that the average markup, or gross margin, on gas is typically around 15 cents or 16 cents a gallon."[10] Similarly, according to The National Association of Convenience Stores, "Over the past five years, the retail mark-up has averaged 17.1 cents per gallon."[11]

By going into more details in Table 1, we observe that about 58% of consumers do not search for lower prices in the zip code with the median S_r, although this percentage ranges between 10% and 80% in the zip codes revealed in this table. This value is consistent earlier

[10] http://www.wsj.com/articles/SB10001424052702303299604577323661725847318

[11] http://www.nacsonline.com/YourBusiness/FuelsReports/GasPrices_2014/ Documents/2014NACSFuelsReport_full.pdf.

studies in the literature such as by Hong and Shum (2006) who estimate q_r^1 ranging between 0.364 and 0.633 for different textbooks, while it is higher compared to studies by Moraga-Gonzalez and Wildenbeest (2008) or Moraga-Gonzalez et al. (2013) who estimate q_r^1 ranging between 0.22 and 0.34 for different computer memory chips.

Although these results are of interest by themselves, we particularly would like to focus on their distribution across zip codes. More specifically, we would like to understand whether the estimated expected number of searches S_r are systematically different across zip codes based on zip code characteristics; we achieve such an investigation next.

4. Number of Stations Searched across Zip Codes

Consumer search patterns may differ across zip codes due to several zip code characteristics. In this paper, we distinguish between such characteristics by focusing on geographic, demographic and economic conditions of zip codes.

The geographic indicators that we consider include the area of the zip code (measured in square miles) as well as the average distance between gas stations (measured in miles), although the latter may also be considered as an economic condition. The demographic indicators consist of population density (measured by workers over 16 years of age per square mile) as well as the distribution of race and sex in zip codes. The economic indicators consist of income and poverty level of individuals as well as their commuting time (measured in minutes) and the industries that they work.

The data for zip code area have been obtained from the U.S. Gazetteer ZIP Code file from the U.S. Census Bureau. The average distance between gas stations has been calculated using the gas station address information given in the OPIS data described above. The demographic and economic indicators have been obtained from U.S. Census Bureau 5-Year American Community Survey between 2009-2013.

4.1. Benchmark Case

We start with investigating the relationship between the dependent variable of log estimated expected number of searches S_r and the independent variables consisting of average distance between gas stations, area, population density, median income and average commuting time. The results of this regression is given in Table 2 where all variables enter the regression significantly. As is evident, consumers search more in zip codes where the average distance between gas stations is longer. In particular, as the average distance (in miles) between gas stations increases by 1%, consumers on average search for more stations by 0.063% across zip codes. This result suggests that consumers would double their expected number of stations searched when average distance goes up by about 15 times. Similarly, as the size of the zip code increases in by 1% in square miles, consumers search for more stations by 0.48%, suggesting that consumers would double their expected number of stations searched when the zip code area is tripled. According to the consumer search model, the last two results are mostly due to fact that gasoline price spreads (measured by the difference between the highest and the lowest prices) are higher in zip codes with spatially dispersed gas stations.[12] Hence, as consumers search more, there will be positive potential gains out of making search, which is in line with our discussion in the model section. Likewise, consumers would double their expected number of stations searched when zip code population density goes up by 2.5 times, mostly due to lower search costs when there are more gas stations per square mile (representing higher supply in such locations).

Median income is shown to be negatively related with the expected number of searches, where the coefficient is about -0.275; it is implied that consumers would halve their expected number of searches when their income is quadrupled. This is obviously due to the opportunity cost of searching for lower gasoline prices where higher income consumers do not find it profitable enough. The expected number of stations searched decrease with the commuting time across zip codes. Specifically, consumers halve their expected number of searches when commuting time is quadrupled. One possible reason may be the lack of time that consumers with longer commuting time have, while another reason may be methodological. Regarding the latter, we have so far employed average/median zip code characteristics in order to explain the expected number of searches across gas stations. Nevertheless, such an approach may suppress important information regarding the distribution of consumers having different characteristics within a given zip code. Accordingly, we investigate potential nonlinearities in some of our independent variables, below.

4.2. Income, Poverty and Industries Worked

We start with considering the effects of different income groups (in percentage terms) on the expected number of searches. We achieve this by keeping the benchmark case independent variables (except for the median income) in the regression.[13] The results are given in Table 3 where the benchmark case independent variables are still significant and very close to their

[12] In this paper, the correlation (across zip codes) between log average distance between stations and log price difference between the most and the least expensive stations is about 0.19. Moreover, such a positive correlation is not unique to this paper; e.g., studies such as by Chandra and Tappata (2011) have also shown similar evidence.

[13] Within the overall set of income groups, we also drop one group in the regression analysis in order to avoid any multicollinearity problem. We follow this strategy for the rest of tables in this paper.

estimated values in Table 2.

As is evident in Table 3, zip codes with higher percentage of groups with annual income between $10,000 and $34,999 search more, while other income groups do not contribute to the expected number of searches. Within the groups that have income between $10,000 and $34,999, the group with an income between $10,000 and $14,999 search most with a corresponding coefficient of 0.015, followed by groups with income levels ranging from $25,000 to $34,999 and from $15,000 to $24,999. One interesting observation belongs to the income group at the bottom of the income level with an annual income of at most $9,999. Potentially, people within this income group are the ones who cannot afford owning a car in the first place; therefore, it is not surprising that zip codes with higher percentage of these low income consumers do not search for lower gasoline prices compared to other income groups. Another result in Table 3 refers to the consumers in zip codes that do not search more than other zip codes due to having income levels higher than (or equal to) $35,000; this is due to the insignificant coefficients in front of such income groups. As in the benchmark case, this is again due to the opportunity cost of searching for lower gasoline prices where higher income consumers do not find it profitable enough.

The results based on the relationship between the log expected number of searches and poverty are given in Table 4. In terms of economic intuition, the results are similar to the ones that we have in Table 3. In particular, consumers in zip codes suffering from poverty search for more gas stations before purchasing gasoline, while consumers at or above 150 percent of the poverty level do not search more than other consumers.

Consumers working in different industries also have different search behavior, after controlling for benchmark case variables, according to Table 5. As is evident, consumers in zip codes that have higher percentage of individuals working in retail trade and public administration search most with a significant coefficient of 0.019, followed by transportation/warehousing/utilities and information/finance/insurance/real estate and rental. On the other hand, consumers in zip codes with higher percentage of individuals working in wholesale trade do not search more compared to other industries.

4.3. Commuting Time, Race and Sex

In this subsection, we further investigate the relationship between expected number of searches and zip code characteristics regarding the commuting time of individuals, this time by distinguishing among consumers having alternative commuting times within zip codes, together with focusing on other zip code characteristics such as race and sex.

The results for commuting time are given in Table 6 where we keep the independent variables in the benchmark case (except for the median commuting time). As is evident, the zip codes with higher percentage of consumers driving 10 to 14 minutes to work search most with a highly significant coefficient of 0.013, followed by those driving 45 to 59 minutes, 15 to 19 minutes, 30 to 34 minutes and less than 10 minutes. Hence, although it is hard to talk about a pattern across alternative commuting times, we can at least say that consumers in zip codes with commuting times between 45 to 59 minutes search for more gas stations before making a purchase compared to those with commuting times between 15 to 44 minutes or less than 10 minutes.

The results based on race are given in Table 7, while those based on sex are given in Table 8. We observe in Table 7 that consumers in zip codes with higher percentage of white, Asian and Black or African American people search for more stations compared to the other races, after controlling for benchmark case variables. Within these groups, zip codes with higher percentage of Black or African American consumers search most with a significant coefficient of 0.012, followed by white and Asian consumers. In Table 8, we observe that consumers in zip codes with higher percentage of female people search more compared to other zip codes with a significant coefficient of 0.007, again after controlling for benchmark case variables.

5. Concluding Remarks and Policy Implications

Retail prices differ significantly across retailers, even after controlling for retailer characteristics and time-varying shocks. This paper has considered the heterogeneity in the consumer search behavior as a potential explanation for the heterogeneity of retail price distributions across locations. Within this picture, we have focused on the determinants of the expected number of searches (that consumers achieve before making a purchase) across zip codes based on geographic, demographic and economic conditions. Based on the maximum likelihood estimation of a consumer search model, we recover the distribution of search costs for each zip code in the U.S. by considering the gasoline purchasing behavior of consumers as an application for which we have daily price data covering virtually all gas stations within the U.S..

The results have shown that geographical factors increasing the price dispersion across gas stations such as the average distance between them, overall area of the zip code or population density all contribute positively (across zip codes) to the expected number of searches achieved by consumers before making a purchase. On the other hand, income and commuting time have been shown to be negatively related to the expected number of searches across zip codes, potentially capturing the opportunity cost of time for making a search. Consumers in zip codes with individuals working in different industries have also been shown to having different search behavior, with industries such as retail trade and public administration contributing most to the expected number of searches. We have also shown that consumers in zip codes with higher percentage of Black or African American people search more compared to those with white or Asian people. Finally, consumers in zip codes with higher percentage of females have shown to search more compared to other zip codes.

Retailers can charge higher markups if consumers do not search for lower prices, which is one of the implications of the model used in this paper. Combining this information with the fact that gasoline is a relatively inelastic product (according to the U.S. Energy Information Administration[14]), it is implied by the results of this paper that gas stations can achieve higher profit margins if they would be located in zip codes in which gas stations are closer to each other; this partly explains why we observe gas stations located very close to each other in certain zip codes. Similarly, higher profit margins can be achieved in zip codes with smaller areas, lower population densities, higher income and/or higher commuting times; e.g., gas station profits would be maximized in zip codes with individuals having annual income levels above $35K. On the other hand, such profits would be lower in zip codes with higher percentage of Black or African American individuals, followed by those with higher percentage of white and Asian individuals. The profits would be lower also in zip codes with higher percentage of individuals working in industries such as retail trade and public administration. Finally, zip codes with a higher percentage of male population are also good locations to have a gas station in order to maximize profits.

Policy suggestions directly correspond to the duality of the results based on gas-station markups across zip codes. In particular, if the main objective is to reduce the redistributive effects of information frictions across locations, the corresponding suggestion is that the policy makers should consider the heterogeneity of consumer search behavior across markets (where the heterogeneity has been shown to depend on geographic, demographic and economic conditions) by conducting local policies rather than a common multi-location policy.

It is important to mention that all of these implications are robust to the consideration of gas station characteristics (e.g., its location, competition level, brand, having a car wash or a convenience store, etc.) as well as supply and demand shocks in the gasoline market, since we control for all of these factors in the investigation. However, the results are not without caveats. In particular, we are well aware of the situation that consumer search behavior may not be segmented at the zip code level, although such a strategy was necessary in order to understand whether the estimated expected number of searches change across zip codes and whether such estimates can further be connected to geographic, demographic and economic conditions. The attempts to address this issue in the literature in studies such as by Nishida

[14] See http://www.eia.gov/todayinenergy/detail.cfm?id=19191

and Remer (2015) are encouraging; however, they are subject to very similar criticisms, since they use other *ad hoc* market segmentation measures such as geographic markets defined as great circles with a radius of 1.5 miles. Therefore, unless the corresponding data for the market segmentation of consumers would be available (e.g., the geographical space covered by each consumer in order to make a search before making a purchase), together with data on geographic, demographic and economic characteristics of such consumers, the results in this paper are not subject to any further improvement.

References

[1] Abrantes-Metz, R., Froeb, L., Geweke, J., Taylor, C., (2006) "A variance screen for collusion," International Journal of Industrial Organization 24, 467–486.

[2] An Y, Hu Y, Shum M. (2010). Estimating first-price auction models with unknown number of bidders: a misclassification approach. Journal of Econometrics 157: 328–341.

[3] Bajari P, Houghton S, Tadelis S. (2006). Bidding for incomplete contracts: an empirical analysis. NBER Working Paper 12051.

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

[5] 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.

[6] Doyle, J.J. and Samphantharak, K. (2008), $2.00 Gas! Studying the effects of a gas tax moratorium. Journal of Public Economics 92: 869-884.

[7] Haile PA, Hong H, Shum M. (2003). Nonparametric tests for common values in first-price sealed-bid auctions. NBER Working Paper 10105.

[8] Haucap, Justus and Müller, Hans Christian (2012) : The Effects of Gasoline Price Regulations: Experimental Evidence, DICE Discussion Paper, No. 47, ISBN 978-3-86304-046-8.

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

[10] Hortacsu, 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.

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

[12] Marvel, H. P. (1976): “The Economics of Information and Retail Gasoline Price Behavior: An Empirical Analysis,”Journal of Political Economy, 84(5), 1033–1060.

[13] Moraga-González, J. L., and M. R. Wildenbeest (2008): “Maximum Likelihood Estimation of Search Costs,”European Economic Review, 52(5), 820–848.

[14] Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2013): “Semi-Nonparametric Estimation of Consumer Search Costs,”Journal of Applied Econometrics, 28(7), 1205–1223.

[15] Nishida, M. and Remer, M. (2015) "The Determinants and Consequences of Search Cost Heterogeneity: Evidence from Local Gasoline Markets," mimeo.

[16] Pennerstorfer, D., P. Schmidt-Dengler, N. Schutz, C. Weiss, and B. Yontcheva (2014): “Information and Price Dispersion: Evidence from Retail Gasoline,”Discussion paper.

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

[18] Stigler, G. J. (1961): “The Economics of Information,” Journal of Political Economy, 69(3), 213–225.

[19] Varian, H. R. (1980): “A Model of Sales,”The American Economic Review, 70(4), 651–659.

[20] Wildenbeest, M.R. (2011), “An empirical model of search with vertically differentiated products”, RAND Journal of Economics 42, 729-57.

Table 1 - Maximum Likelihood Estimation Results for Search Costs

	Minimum	10th Percentile	25th Percentile	**50th Percentile**	75th Percentile	90th Percentile	Maximum
Expected Number of Searches (S_r)	0.17	1.11	1.21	**1.66**	2.88	4.25	12.94
Minimum Price ($\underline{p}_r$)	2.54	2.98	2.91	**2.58**	3.16	2.38	2.18
Maximum Price ($\overline{p}_r$)	2.61	3.16	3.02	**2.73**	3.38	2.56	3.15
q_r^1	0.10	0.80	0.60	**0.58**	0.60	0.24	0.30
q_r^2	0.04	0.16	0.31	**0.34**	0.28	0.53	0.38
q_r^3	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00
q_r^4	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00
$q_r^{5...10}$	0.00	0.00	0.00	**0.08**	0.00	0.00	0.00
$q_r^{11...N_r}$	0.00	0.00	0.00	**0.00**	0.12	0.23	0.32
Unit Cost (m_r)	2.54	2.76	2.86	**2.51**	3.10	2.37	2.16
Number of Stations (N_r)	3	7	7	**5**	14	13	37
Sample Size	59	165	148	**81**	250	262	549
Log-likelihood	-88.70	-314.33	-345.44	**-161.00**	-473.67	-577.72	-508.69
Corresponding Zip Code	55046	10468	19064	**25315**	95351	76903	38305

Notes: The estimation has been achieved at the zip code level. The estimation results have been sorted across zip codes with respect to the estimated expected number of searches; the corresponding percentiles of zip codes are depicted in this table. All estimates are significant at the 5% level.

Table 2 - Determinants of the Expected Number of Consumer Searches

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between	0.063***
Gas Stations in the Zip Code	(0.009)
Log Zip Code Area	0.480***
(square miles)	(0.013)
Log Zip Code Population Density	0.398***
(workers per square mile)	(0.011)
Log Zip Code Median Income	-0.275***
(US$)	(0.027)
Log Average Commuting Time	-0.252***
in the Zip Code (minutes)	(0.032)
Sample Size	4332
Adjusted R-Squared	0.314

Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 3 - Number of Consumer Searches and Income

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between Gas Stations in the Zip Code	0.060*** (0.009)
Log Zip Code Area (square miles)	0.499*** (0.014)
Log Zip Code Population Density (workers per square mile)	0.415*** (0.011)
Log Average Commuting Time in the Zip Code (minutes)	-0.268*** (0.033)
$1 to $9,999 or less (percentage of workers)	0.001 (0.002)
$10,000 to $14,999 (percentage of workers)	0.015*** (0.004)
$15,000 to $24,999 (percentage of workers)	0.006*** (0.002)
$25,000 to $34,999 (percentage of workers)	0.009*** (0.003)
$35,000 to $49,999 (percentage of workers)	0.003 (0.003)
$50,000 to $64,999 (percentage of workers)	0.006 (0.004)
$65,000 to $74,999 (percentage of workers)	-0.007 (0.007)
Sample Size	4323
Adjusted R-Squared	0.323

Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 4 - Number of Consumer Searches and Poverty

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between	0.064***
Gas Stations in the Zip Code	(0.009)
Log Zip Code Area	0.483***
(square miles)	(0.014)
Log Zip Code Population Density	0.394***
(workers per square mile)	(0.011)
Log Average Commuting Time	-0.290***
in the Zip Code (minutes)	(0.032)
100 to 149 percent of the poverty	0.014***
level (percentage of workers)	(0.004)
At or above 150 percent of the poverty	-0.003
Level (percentage of workers)	(0.002)
Sample Size	4327
Adjusted R-Squared	0.315

Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 5 - Number of Consumer Searches and Industries Worked

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between Gas Stations in the Zip Code	0.053*** (0.009)
Log Zip Code Area (square miles)	0.522*** (0.015)
Log Zip Code Population Density (workers per square mile)	0.427*** (0.013)
Log Zip Code Median Income (US$)	-0.301*** (0.046)
Log Average Commuting Time in the Zip Code (minutes)	-0.311*** (0.039)
Agriculture, forestry, fishing and hunting, and mining	0.008* (0.005)
Construction	0.015*** (0.005)
Manufacturing	0.014*** (0.004)
Wholesale trade	0.008 (0.008)
Retail trade	0.019*** (0.005)
Transportation and warehousing, and utilities	0.017*** (0.005)
Information and finance and insurance, and real estate and rental	0.016*** (0.005)
Professional, scientific, management, and administrative and waste management services	0.014*** (0.005)
Educational services, and health care and social assistance	0.008* (0.004)
Arts, entertainment, and recreation, and accommodation and food services	0.012** (0.005)
Other services (except public administration)	0.019*** (0.007)
Public administration	0.019*** (0.006)
Sample Size	3896
Adjusted R-Squared	0.326

Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 6 - Number of Consumer Searches and Commuting Time

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between Gas Stations in the Zip Code	0.049*** (0.005)
Log Zip Code Area (square miles)	0.418*** (0.008)
Log Zip Code Population Density (workers per square mile)	0.348*** (0.007)
Log Zip Code Median Income (US$)	-0.204*** (0.019)
Less than 10 minutes (percentage of workers)	0.002* (0.001)
10 to 14 minutes (percentage of workers)	0.013*** (0.002)
15 to 19 minutes (percentage of workers)	0.005*** (0.002)
20 to 24 minutes (percentage of workers)	0.000 (0.002)
25 to 29 minutes (percentage of workers)	0.000 (0.002)
30 to 34 minutes (percentage of workers)	0.004** (0.002)
35 to 44 minutes (percentage of workers)	-0.002 (0.002)
45 to 59 minutes (percentage of workers)	0.005** (0.002)
Sample Size	9301
Adjusted R-Squared	0.343

Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 7 - Number of Consumer Searches and Race

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between	0.055**
Gas Stations in the Zip Code	(0.023)
Log Zip Code Area	0.545***
(square miles)	(0.040)
Log Zip Code Population Density	0.453***
(workers per square mile)	(0.036)
Log Zip Code Median Income	-0.345***
(US$)	(0.076)
Log Average Commuting Time	-0.409***
in the Zip Code (minutes)	(0.088)
White	0.008***
(percentage of workers)	(0.002)
Black or African American	0.012***
(percentage of workers)	(0.003)
Asian	0.007**
(percentage of workers)	(0.003)
American Indian and Alaska Native	0.007
(percentage of workers)	(0.009)
Native Hawaiian and Other Pacific Islander	-0.012
(percentage of workers)	(0.013)
Sample Size	957
Adjusted R-Squared	0.260

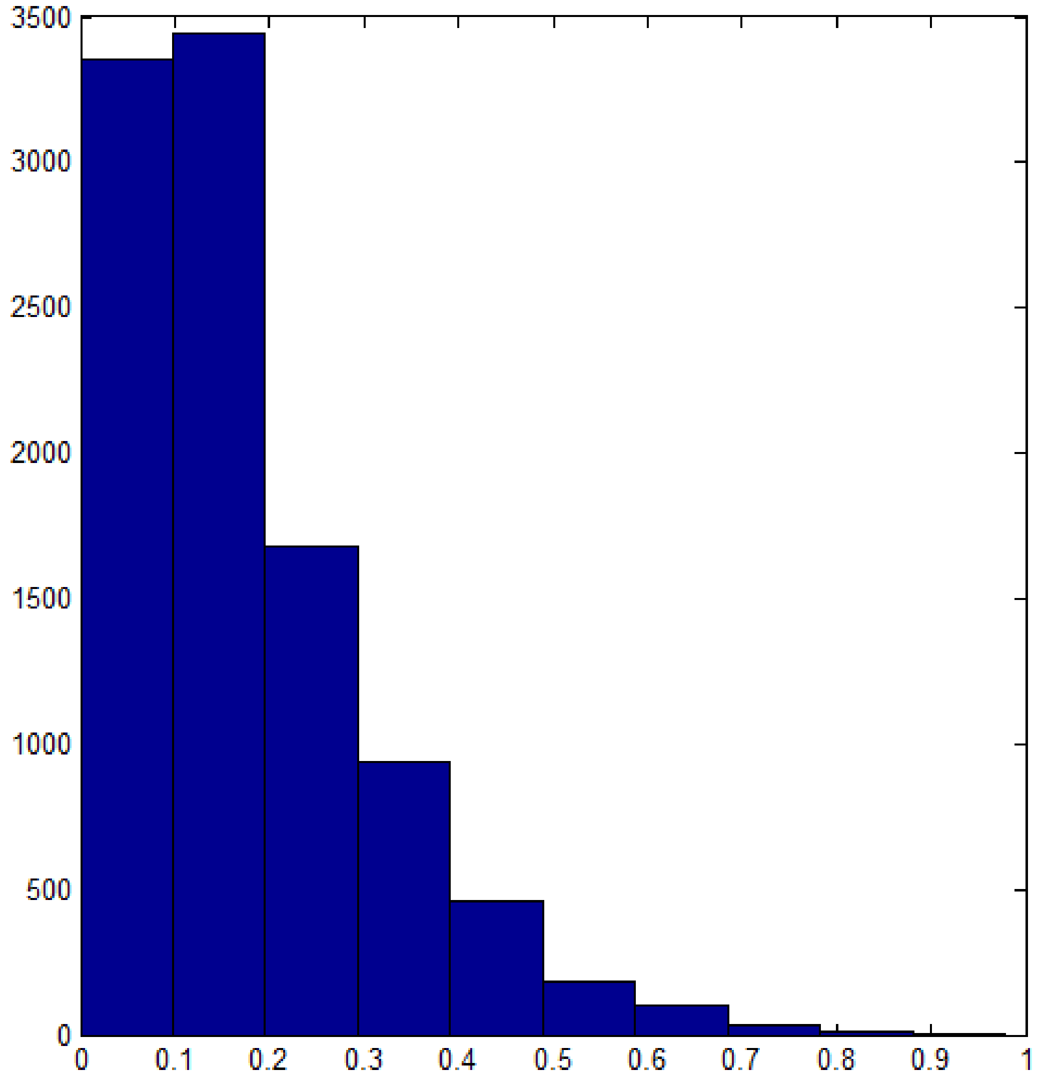
Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Table 8 - Number of Consumer Searches and Sex

	Dependent Variable: Log Expected Number of Searches in a Zip Code
Log Average Distance between	0.061***
Gas Stations in the Zip Code	(0.009)
Log Zip Code Area	0.485***
(square miles)	(0.014)
Log Zip Code Population Density	0.399***
(workers per square mile)	(0.011)
Log Zip Code Median Income	-0.264***
(US$)	(0.027)
Log Average Commuting Time	-0.248***
in the Zip Code (minutes)	(0.032)
Female	0.007***
(percentage of workers)	(0.002)
Sample Size	4332
Adjusted R-Squared	0.316

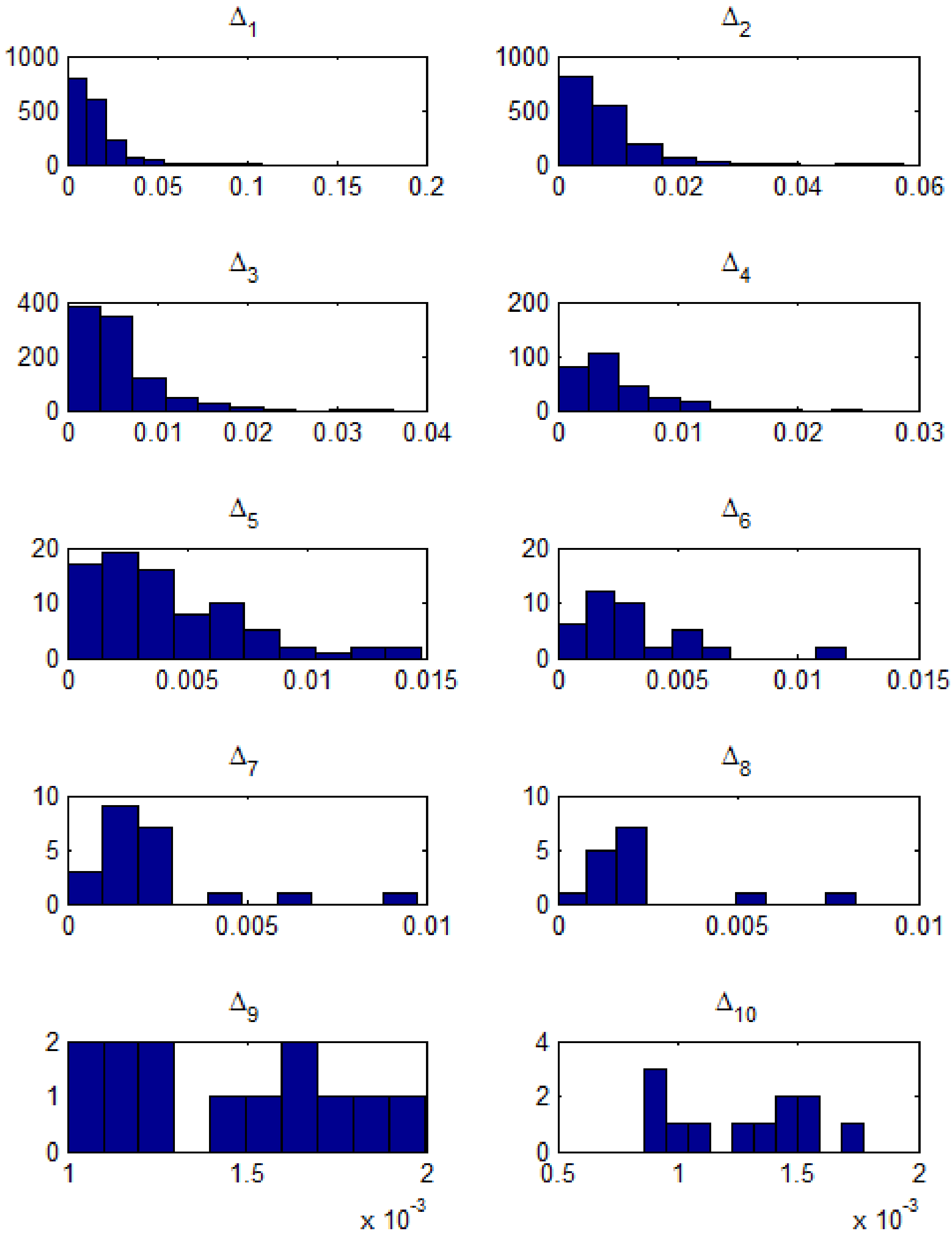
Notes: *, ** and *** stand for significance at the 10%, 5% and 1% levels. Standard errors are given in parenthesis. All regressions include constants that are not shown here. The regression is by OLS.

Figure 1 – Histogram of Gasoline Price Spreads in Zip Codes



Notes: The horizontal axis shows the gasoline price spread, while the vertical axis shows the number of zip codes. Gasoline price spread is defined as the difference between the maximum and the minimum price in gas stations in a given zip code after controlling for gas-station and time fixed effects where the latter includes both day and hour fixed effects.

Figure 2 – Histograms of Estimated Critical Search Cost Values



Notes: The horizontal axis shows the search costs in U.S. dollars, while the vertical axes show the number of zip codes.