YOUNGJUN JANG
jyjglory@mit.edu

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

OFFICE CONTACT INFORMATION
MIT Department of Economics
77 Massachusetts Avenue, E19-750
Cambridge, MA 02139
jyjglory@mit.edu
http://economics.mit.edu/grad/jyjglory

HOME CONTACT INFORMATION
70 Pacific Street, apt. 597A
Cambridge, MA 02139
Mobile: 650-248-5589

MIT PLACEMENT OFFICER
Professor Benjamin Olken bolken@mit.edu
617-253-6833

MIT PLACEMENT ADMINISTRATOR
Ms. Beata Shuster bshuster@mit.edu
617-324-5857

DOCTORAL STUDIES

Massachusetts Institute of Technology (MIT)
PhD, Economics, Expected completion June 2015
DISSERTATION: "Essays on Industrial Organization"

DISSERTATION COMMITTEE AND REFERENCES

Professor Glenn Ellison
MIT Department of Economics
77 Massachusetts Avenue, E18-269F
Cambridge, MA 02139
617-253-8702
gellison@mit.edu

Professor Sara Ellison
MIT Department of Economics
77 Massachusetts Avenue, E18-269B
Cambridge, MA 02139
617-253-3821
sellison@mit.edu

Professor Paulo Somaini
MIT Department of Economics
77 Massachusetts Avenue, E18-212
Cambridge, MA 02139
617-253-3362
psomaini@mit.edu

PRIOR EDUCATION

B.A.S. with Honors and Distinction — Stanford University — 2007
Economics and Mathematics

CITIZENSHIP Korea, Republic of

GENDER: Male

LANGUAGES English (fluent), Korean (native)

FIELDS Primary Field: Industrial Organization

Secondary Fields: Finance, Economics of Education

TEACHING EXPERIENCE

14.273 Advanced Topics in Industrial Organization — Spring 2014
Teaching Assistant to Professor Paulo Somaini (Graduate)
14.03 Microeconomic Theory and Public Policy — Spring 2014
Teaching Assistant to Professor Paulo Somaini (Undergraduate)
14.12 Game Theory — Fall 2012
Teaching Assistant to Professor Muhamet Yildiz (Undergraduate)
14.01 Principles of Microeconomics — Spring 2012
Teaching Assistant to Professor Jeffrey Harris (Undergraduate)
14.03 Microeconomic Theory and Public Policy — Spring 2012
Teaching Assistant to Professor Stephen Ryan (Undergraduate)
14.271 Industrial Organization I — Fall 2011
Teaching Assistant to Professor Glenn Ellison (Graduate)

RELEVANT POSITIONS

Research Assistant to Professor Glenn Ellison, MIT — 2010

OTHER EXPERIENCE

Military Service, Sergeant, Analysis and Assessment Group, Republic of Korea Army Headquarters — 2009-2011

FELLOWSHIPS, HONORS, AND AWARDS

Kwanjeong Educational Foundation Scholarship — 2007-2014
Commendation, Analysis and Assessment Group, Republic of Korea Army Headquarters — 2011
Special Combatant Award, Republic of Korea Army — 2010
MIT Presidential Fellowship — 2007-2009
Samsung Lee Kun Hee Scholarship — 2003-2007
Honorable Mention, William Lowell Putnam Competition, 2nd place at Stanford University — 2003
Silver Medal, 42nd International Mathematical Olympiad — 2001

PROFESSIONAL ACTIVITIES

Presentations:
Rising Stars Session, the 12th Annual International Industrial Organization Conference, Northwestern University — 2014
SSK International Conference on Competition and Information Economy, Yonsei University — 2014

RESEARCH PAPERS

"The Effects of the Internet and Mobile Search Technologies on Retail Markets: Evidence from the Korean Gasoline Market" (Job Market Paper)
Since 2008, a Korean government website has posted daily prices of all gasoline stations. Combined with the rapid increase of smartphone and mobile technologies, this price information service may have changed the consumer search environment significantly. This paper investigates the effects of these technological advances on the retail gasoline market, using daily price data, quantity data for select stations obtained from a credit card provider, and regional smartphone penetration rates. In daily price data for four regions from January 2010 to June 2012, price dispersion among gasoline stations and markup increase slightly when the smartphone penetration rate increases, even while additional descriptive evidence suggests that demand is becoming more price-sensitive. Structural estimation of a two-type consumer search model finds that the proportion of highly informed consumers increases as the smartphone penetration

rate increases. A counterfactual analysis confirms that observed price changes are consistent with theoretical models of pricing, given the structurally estimated parameters.

RESEARCH IN PROGRESS

"Mental Accounting and Consumer Choices in the Korean Retail Gasoline Market"
This paper presents a new approach to measure consumer choices, using a detailed transaction-level gasoline sales dataset for select stations in Korea. There are two unique market properties: first, almost all gasoline stations in Korea offer only full service; second, gas pumps can be set to fuel until pre-set integer dollar amounts. From January 2010 to December 2012, about 30% of regular gasoline consumers chose to simply fill up, while the remaining 70% of consumers spent pre-selected dollar amounts. Retail gasoline price changes had little effect on this tendency. I develop a discrete-choice utility model to explain the observed consumer behavior and discuss optimal pricing policies.

"Does Coeducation Reduce the Gender Gap? Evidence from Korean Secondary Schools"
In March 1997, the Korean government introduced a new policy to increase the proportion of co-ed schools. A dataset of Korean SAT scores for the entire pool of exam takers from 1999 to 2009 is used to examine the impact of the proportion of exam takers from co-ed high schools on the gender gaps for the whole score distribution, not only for mean or median scores. Focusing on two main subjects (Korean and mathematics), this research investigates the differences between the scores of male students and those of female students at 25%, 50%, 75%, 90% percentile levels. Preliminary results show that, as the number of exam takers attending co-ed high schools increases, so do the gender gaps at all percentile levels, and this tendency is significant, especially for mathematics.

"Ability Grouping and Student Achievement: Effects of the Equalization Policy in Korea"
This paper analyzes the effects of ability grouping on the academic performance of high school graduating students in Korea. About half of the regions in Korea adopted the equalization policy (EP), which means that high school students are randomly assigned. For the other non-EP regions, students are sorted among schools based on ability levels. I exploit that several regions adopted the EP during 2000-2008 as a result of the exogenous policy shifts, and utilize a difference-in-differences strategy. Studying a dataset of Korean SAT scores for the entire pool of exam takers from 1999 to 2009, I find that after the EP, students performed worse, especially those in high percentiles. In addition, there was an increasing trend in the achievement levels in the treatment regions, but after the introduction of the EP, this trend vanished.

Research Statement

Youngjun Jang

Dear Recruiting Coordinator,

I am a Ph.D. candidate in the Department of Economics at MIT writing to apply for the Digitization post-doctoral fellowship position listed in the September Job Openings for Economists.

My job market paper explores the effects of new technological advances on retail markets. Since 2008, a Korean government website has posted daily prices of all gasoline stations. While this price information service alone might have had limited effects on the market, the rapid increase of smartphone and mobile technologies significantly enhanced the impact of the service, since consumers may search for prices while driving. To investigate the effects of these two technological advances, I combine daily, individual station level prices and quantity data in four regions with smartphone penetration rates.

A natural or naïve intuition is that smartphones lower search costs and thus consumers search more, which would lead to intensify competition and finally result in lower price dispersion and markups. However, the data say the opposite: both price dispersion and markups actually increase slightly when the smartphone penetration rate increases, even while additional descriptive evidence suggests that demand is becoming more price-sensitive. To explain this apparent puzzle, I develop a two-type consumer search model. Structural estimation finds that the proportion of highly informed consumers increases as the smartphone penetration rate increases, and that gas stations modify their pricing strategies accordingly. Therefore, a new equilibrium, where both price dispersion and markups slightly increase, is reached.

I am also interested in consumer choices in retail markets, as a result of mental accounting. Using a detailed transaction-level gasoline sales dataset, I find that about 30% of regular gasoline consumers chose to simply fill up, while the remaining 70% of consumers spent pre-selected dollar amounts. Retail gasoline price changes had little effect on this tendency. I develop a discrete-choice utility model to explain the observed consumer behavior and discuss optimal pricing policies.

My other current work seeks to explain educational productivity changes due to an exogenous policy adoption. One project studies whether an increase in the number of co-ed schools affects the gender gap among high school graduating students. A dataset of Korean SAT scores for the entire pool of exam takers from 1999 to 2009 is used to examine the impact of the proportion of exam takers from co-ed high schools on the gender gaps for the whole score distribution, not only for mean or median scores. Preliminary results show that, as the number of exam takers attending co-ed high schools increases, so do the gender gaps at all percentile levels. I also analyze the effects of ability grouping on the academic performance. About half of the regions in Korea adopted the equalization policy (EP), which means that high school students are randomly assigned by location. For the other non-EP regions, students are sorted among schools based on ability levels. I exploit that several regions adopted the EP during 2000-2008 as a result of the exogenous policy shifts, and utilize a difference-in-differences strategy. I find that after the EP, students performed worse, especially those in high percentiles.

Looking forward, over the next three to five years, I will continue to focus on the impact of technology developments on market outcomes. Since technologies continue to progress, it would be interesting to study how consumers and firms adapt and how equilibrium changes as a result. In addition to studying how existing markets change, I will extend my research into new markets that are being created by innovations. While I mainly focus on empirical applications, I am also interested in developing theoretical models that could explain unprecedented market environment.

Among technological advances, I am particularly interested in search cost and delivery cost reductions. As the Internet and other mobile devices reduced search costs significantly, my future research may analyze other retail markets. Recently, new services such as Google Express and Amazon Prime Air (drone delivery system) are announced: these services will transform retail markets by cutting cut delivery time and cost significantly, and traditional brick and mortar firms will need to adjust their strategies.

The Effects of the Internet and Mobile Search Technologies on Retail Markets: Evidence from the Korean Gasoline Market*

Youngjun Jang

November 8, 2014

Abstract

Since 2008, a Korean government website has posted daily prices of all gasoline stations. Combined with the rapid increase of smartphone and mobile technologies, this price information service may have changed the consumer search environment significantly. This paper investigates the effects of these technological advances on the retail gasoline market, using daily price data, quantity data for select stations obtained from a credit card provider, and regional smartphone penetration rates. In daily price data for four regions from January 2010 to June 2012, price dispersion among gasoline stations and markup increase slightly when the smartphone penetration rate increases, even while additional descriptive evidence suggests that demand is becoming more price-sensitive. Structural estimation of a two-type consumer search model finds that the proportion of highly informed consumers increases as the smartphone penetration rate increases. A counterfactual analysis confirms that observed price changes are consistent with theoretical models of pricing, given the structurally estimated parameters.

JEL classification: L81, O33

*Department of Economics, Massachusetts Institute of Technology. Email: jyjglory@mit.edu. I am grateful to Glenn Ellison and Sara Ellison for their invaluable support and guidance throughout this project. I also thank Paulo Somaini, Nikhil Agarwal, Joonhwan Lee, and MIT Industrial Organization lunch seminar and workshop participants for helpful comments and suggestions. I gratefully acknowledge funding from Kwanjeong Scholarship Foundation, and thank anonymous company officials who provided quantity data.

1 Introduction

While the "law of one price" is an elegant and comprehensible economic theory, it is seldom true in practice. Since Stigler's seminal paper (1961) on search and price dispersion, many papers have noted that consumer search costs can lead to price dispersion. If there is a change in the search cost, consumer search behavior and the expected amount of price information consumers have will change. That will affect the supply side, and hence market outcomes such as price dispersion and markup will also change. Recent interest in the effects of consumer search costs has been fueled by the growth of e-commerce because the Internet can often supply settings with very low consumer search costs. However, there is no clear evidence that the Internet has made markets more efficient or decreased price dispersion. For example, Ellison and Ellison (2009) find that while consumers became extremely price-sensitive for some products, internet retailers developed obfuscation strategies to maintain their profit margins, and price dispersion has persisted. In fact, price dispersion seems to be increasing over time and price dispersion is an equilibrium phenomenon that depends on the market structure (Baye et al., 2004). A recent emergence of mobile technologies, in particular smartphones, reintroduces the topic of whether a new technology that seems to reduce consumer search costs may change the market structure and consumer behavior. Smartphone penetration rates have rapidly increased from 3% (2009) to 45% (2012) in the U.S., and similar growth rates are observed in other countries. In this paper, I study the effects of the increased usage of mobile technologies and the real time price information service on market outcomes.

The Korean gasoline retail market after 2008 is an excellent place to study the effects of search cost reduction on market outcomes for three reasons. First, gasoline is a fairly homogeneous good. Second, the Korean government has published free, real-time price information for all gasoline stations in the nation since 2008. As a consumer can access price information of all gas stations in the region of interest within 20 seconds via the Internet, his search cost is much lower than before. Furthermore, this price information service could be a paradigm shift in search instead of a mere search cost decrease, as most consumers previously had very limited price information and did not search at all before this price information service. Third, a rapid increase in mobile technologies helps to measure the effects of the price information service. While it is easy to check prices from the Internet, drivers usually choose which gas station to go to when they are on the road and running low on gas, instead of checking the prices before driving. Mobile devices, such as smartphones and in-car navigation systems that display gasoline prices, may significantly change the situation, as drivers can utilize the price information service whenever they want to. Moreover, there is enough variation in smartphone user population to estimate the effects. The smartphone penetration rate in Korea was below 5% before 2010 and reached 50% during 2012.

In addition, a combination of the unusually rich data environment created by the Korean government and private datasets that I could utilize provides an unique opportunity to examine market functioning. The Opinet information service offers daily retail prices of gas stations, station characteristics information, and weekly, national average distribution costs for all four major gasoline companies. Furthermore, I was able to obtain private daily sales data for select gasoline stations and one major telecommunication company. Since previous literature computes price dispersion and markup using only observed prices,[1] I compute more realistic measures that take into account quantity differences. Quantity-weighted price dispersion and markup measures ensure that included prices are real prices at which transactions occur, and prices with more frequent transactions are weighted more. Also, having a full set of station-level prices allows me to perfectly measure market-level price dispersion. Documenting stylized facts, I find an interesting phenomenon: price dispersion and markup levels did not decrease over time, even though consumer search costs decreased significantly. Graphs of all price dispersion measures and both unweighted and quantity-weighted markups show non-decreasing trends. To account for all factors which could drive price dispersion and markup measures, I run reduced-form regressions to examine these measures.

Regression results suggest that both price dispersion and markups slightly increase as smartphone penetration rates increase. This might be counter-intuitive, as it could be natural to assume that higher smartphone penetration rates would lead to more consumer search that would make stations compete more, and hence price dispersion and markup would decrease. The Opinet website and smartphone application usage trend confirms that more consumers utilized the price information service as smartphone penetration rates increased, which implies that the proportion of informed consumers increased. In addition, descriptive tests suggest that consumers became more price-sensitive, and the distribution of gas stations' markups moved toward a bimodal distribution. Intuitively, as the ratio of informed consumers grows, some stations set low prices to attract informed consumers, and some stations set high prices to serve less informed consumers. I find that search theory models could be consistent with these results. Many previous papers (see, e.g., Stigler, 1961; Diamond, 1971; Salop and Stiglitz, 1977; Varian, 1980; Morgan and Manning, 1985; Stahl, 1989 and 1996; Sorensen, 2000; Brown and Goolsbee, 2002; Baye et al., 2004) point out that the prices of homogeneous goods are quite dispersed and this price dispersion can be an equilibrium when there are some consumers who observe several prices while other consumers learn only one price. Search theory models with two types of consumers, where one type is more informed than the other, find that while there is no pure strategy Nash Equilibrium, there exists a symmetric mixed Nash Equilibrium where firms choose prices from an atomless distribution. In particular, it predicts a bimodal distribution of prices as firms either aim to be the lowest to attract informed consumers or to be the highest to get the maximum profit per customer. For instance, the

[1] Mainly due to the lack of quantity data.

Stahl model predicts that price dispersion increases when the ratio of informed consumers increases, as long as the ratio is below the critical value (Pennerstorfer et al., 2014).

Motivated by these descriptive results, I construct and estimate a discrete-choice demand model that includes both more and less informed consumers and a distance term that has different values for consumers at different locations. The structural model explains consumer choices and substitution patterns, and identifies a proportion of smartphone users who actually search for gasoline prices. These estimates are in line with the descriptive evidences: they indicate that the fraction of highly informed consumers has increased from 1.7% to 11.4% over the sample period, and that consumers have become more sensitive to prices as a result. A counterfactual analysis tells us what the new equilibrium prices would be if the informed consumer ratio changed. The structural model estimates that the increase in the fraction of informed consumers would be expected to result in an 0.57% increase in price dispersion, an 0.33% increase in the average markup, and an 0.09% increase in the quantity-weighted average markup. These findings are consistent with the descriptive analyses and the observed trends from the data. Lastly, the model interprets the effect of the distance between consumers and stations as how much more an average consumer is willing to pay to avoid driving an additional distance to visit a different gasoline station.

This paper attempts to measure the effects of new technological advances. While smartphones have become indispensable in our daily lives, no other research that I am aware of studies the impact of smartphones on the search environment. The Korean retail gasoline market is a quasi-experimental field where the effect of this new technology on search can be evaluated. Since smartphones combined with the real-time price information service directly change the consumer search situation and the subsequent decision, this paper helps us measure the effects of price information on retail market outcomes and how smartphones increase them. While there is no previous research that examines the effects of smartphone introduction on market outcomes, there are two empirical results about the introduction of mobile phones in developing countries. Jensen (2007) studies fish prices in several Indian towns and finds that the adoption of mobile phones by fishermen and wholesalers led to a significant reduction in price dispersion and an increase in social welfare. Aker (2008) also reports that the introduction of mobile phone service between 2001 and 2006 explains a 10 to 16 percent reduction in grain price dispersion in Niger. One difference, though, is that the effects of mobile phones and of smartphones are quite different. A mobile phone search, or making a phone call, provides one price quote at a time and it usually requires non-negligible time. On the other hand, a price search using a smartphone provides all the gasoline price information with a single search, and that single search only costs several seconds. Thus, the introduction of smartphones should not be interpreted as a mere decrease of search costs; rather, it changes the paradigm of searching and divides consumers into two groups: one with the full information, and the other with limited information.

I also contribute to the empirical literature on retail gasoline demand. Gasoline retail pricing has been studied in depth. Studies of price levels at individual stations have considered local market characteristics, region or time fixed effects, and individual station characteristics; see, for example, Barron et al. (2004), Eckert and West (2005), Hosken et al. (2008), and Lewis (2008) . In general, the findings have been mixed. In many cases, the impact of station characteristics on price is fairly small, and higher local station density implies lower price level and lower price dispersion. Also, there are some papers that study how search affects retail prices. Lewis and Marvel (2011) find that consumers search more as prices rise than they do when prices fall. As a result, when prices rise, margins are lower and there is less price dispersion. Similarly, Chandra and Tappata (2011) find that price dispersion increases with search costs. This paper provides an unusual case that station characteristics have significant effects on prices, and that price dispersion does not decrease with lower search costs.

Finally, this paper contributes to the empirical literature on search. While there are considerable theoretical results, relatively little empirical research has focused on measuring search costs and the effects of consumer search behavior changes in practice. As my empirical setting allows a two-type consumer search environment and two-type search costs, estimating search cost distribution is simplified to finding a proportion of informed consumers and this assists in understanding the change of search cost distribution easily. Recently, Hong and Shum (2006) presented structural methods to estimate search cost distributions using price data alone, and using similar methods, Wildenbeest (2011) focuses on the grocery markets in the United Kingdom and concludes that most of the observed price dispersion is explained by supermarket heterogeneity rather than search frictions. Yet I do not attempt to estimate the whole search cost distribution, nor to rationalize consumer search behavior. I focus on studying changes in market outcomes as a result of consumer search behavior changes.

The remainder of this article is organized as follows. In the next section I provide details about four data sets and discuss their merits and limitations. In section 3, I document stylized facts and present summary statistics. I construct regression models that estimate the effects of increasing smartphone penetration rates on price dispersion and gasoline station price levels. Section 4 explains consumer search behavior assumptions and the structural model setups. Structural estimation results and counterfactual analysis are reported in Section 5. Section 6 presents conclusions.

2 Data

In this paper, I combine four datasets: the Opinet price and station characteristics data; average wholesale prices; gasoline quantity data for select stations; and regional smartphone penetration rates. From the Opinet information service, I gathered daily retail prices and station characteristics of all gas stations in four regions. I utilized weekly national average distribution costs for all four major gasoline companies from the Opinet website to approximate the actual marginal costs. In addition, I obtained private daily sales data for select gasoline stations that are essential to perform daily level demand estimation and compute realistic price dispersion measures. Lastly, I use regional smartphone sales data to infer smartphone penetration rates.

2.1 Opinet Data

In 2008, the Korean government established an unusual resource for its citizens. Since April 15, 2008, every gas station in South Korea has been required to report its posted gasoline and diesel price at least once a day. Korea National Oil Corporation, a public institution, is in charge of collecting and publishing the real-time price information. I took advantage of this uncommon opportunity and scraped data from the Opinet website.[2] I obtained a complete set of historical daily prices for all gas stations in four regions from January 2010 to June 2012 (909 days) when mobile devices such as smartphones were diffusing rapidly. In particular, the smartphone penetration rate was nearly zero in January 2010, but went over 50% in June 2012.

Among the four regions, two regions are districts of Seoul, and the other two regions are small cities that are isolated by mountains.[3] The streets are laid out in a grid pattern for all regions, especially for the two districts of Seoul. On average, each region has 40 gasoline stations, and the four major gasoline companies have more than a 95% market share in total.

Each gas station owner can update the price information by calling the Opinet office, submitting the information on the Opinet website, or using an automated report system. According to an Opinet representative, most gas station owners use automated systems: for each credit or debit card transaction, price information for that transaction is electronically reported to the Opinet server and price information is automatically updated.[4] Consumers can access the price data via the website www.opinet.co.kr or other methods, such as car navigation systems or Opinet smartphone applications.[5] In addition to the price information, Opinet also provides station characteristics such as location, self-service or full-service, car wash, repair shop, and convenience store availability. I collected these data as well for all of the stations in my sample.

[2] I also contacted Opinet officials to supplement missing data.

[3] Except for two districts of Seoul, distances between any two regions are over 50 miles.

[4] While the Opinet website advertises that it offers real-time data, it does not update new price information every time transaction information comes in. Instead, it updates six times a day, based on price information during each time period. However, as gasoline stations rarely change prices more than once a day (change once in a week on average), updating six times a day basically provides the real-time information.

[5] As the name of this service is usually called "Opinet", following the name of its website, I will use the term "Opinet" to represent this real-time price service and the name of institution that provides the price data.

2.2 Average Wholesale Price

Since most, if not all, of Korean oil imports come from Middle Eastern countries, the raw price Korean oil companies pay when they buy crude oil closely follows Dubai oil price futures. After importing crude oil, oil companies refine it and distribute gasoline to retailers. According to the gasoline station owners, each gasoline company sets a base distribution price once a week, and most transactions between the company and stations are made at that price during the week. The Opinet website publishes these weekly national average distribution costs for all four major gasoline companies. From now on, I call the distribution price AWP, or average wholesale price. These average wholesale prices are used as an approximation of the actual marginal costs.

2.3 Gasoline Quantity Data

Quantitative data such as the number of smartphones sold for each region and the quantity of gasoline sold for each station are very difficult to gather.[6] For gasoline quantity data, I have daily credit card transaction numbers of select individual gas stations of one major gas company that has about a 30% market share: for about 20% of the total gas stations, I have the quantity information. According to the company official, credit cards are used for most of the transactions, and the proportion of credit card transactions has been stable and similar for all four regions during the period 2009-2012. Moreover, since the average transaction amount has been stable during the period, either using daily transaction numbers or using daily transaction amounts naturally delivers very similar results.[7] I present the results from using daily transaction numbers for the main analysis.[8]

Having quantity data provides two main advantages. First, quantity information is essential for any demand estimation. Second, more economically relevant statistics can be derived. In many markets, especially online markets, we cannot distinguish "real" prices at which transactions actually occur from unrealistic price listings that will never result in a sale. In addition, even after limiting the price listings to the ones at which trades are made, estimating price dispersion and average price without using the quantity information could be misleading.[9]

[6] I thank anonymous company officials who provided quantity data of one major company in each sector.

[7] Transactions with more than one million Korean Won (about 1000 USD) are treated as outliers and excluded.

[8] All gasoline stations had 200 or more transactions per day on average. When a station had 50 or fewer transaction numbers in a day, that station-day observation is excluded from the sample, under the assumption that either a card reader system did not work properly or the station opened for a very short time on that date.

[9] For example, Amazon.com's book listing for Jean Tirole's "The Theory of Industrial Organization" on October 20, 2014, contains 30 prices, ranging from \$69.54 (plus \$3.99 shipping) to \$263.64 (plus \$3.99 shipping). Among the thirty listings, eight listings have prices \$120 or higher. Since it is unlikely that anyone would pay \$120 or more, including these prices in the consumer choice set would overestimate price dispersion and average price. Moreover, even if we can identify the set of "real" prices, having quantity data is essential. For simplicity, suppose that these are two prices, \$70 and \$120, and 99% of consumers pay \$70 and 1% of consumers end up paying \$120. Quantity-weighted price dispersion gives a correct picture of actual price dispersion, which is close to zero (as the quantity-weighted price is close to \$70), while unweighted price dispersion does not properly reflect the market situation.

2.4 Smartphone Penetration Rates

I utilize data from a representative company in the telecommunication sector that has constant market share during the sample period to estimate total smartphone penetration rates. To compute daily smartphone penetration rates, I start from the quarterly number of regional smartphone users for one telecommunication company. This company is one of the three major telecommunication companies and its market share had been 30-35% during 2009-2012. Multiplying the number of users for this company by three, I estimate the total number of smartphone users of the region for each quarter. In regressions run at the daily level, I use a linear interpolation method to infer daily smartphone penetration rates.

Smartphones have become increasingly popular in the Korean market since late 2009. Before the introduction of iPhone 3G in November 2009, less than 1% of the population used a smartphone. Within 3 years, the smartphone penetration rate (the ratio of the number of smartphone users to the total population) went over 50%. Figure 1 shows smartphone penetration rates of four regions during the data period. Note that regions 1 and 2 have the same rates, as these two regions are neighboring districts of Seoul and the company treated them as a single region when they collected sales data. While all regions have increasing trends, there are some regional differences.

3 Descriptive Analyses on Price Dispersion and Markup

In Section 3, I present reduced-form evidence on price dispersion and markups. The most basic observation is that both price dispersion and markups do not decrease over time. Regression results suggest that higher smartphone penetration rates lead to slightly higher price dispersion and average markups. To investigate whether these relationships are causal or just because consumers do not use price information, I perform additional analyses: usage statistics; a demand regression to look for price sensitivity changes; and a test to check whether stations' markup distributions are bimodal.

3.1 Definitions and Summary Statistics

Table 1 defines the variables used in the analysis. *RetailP, AdjustedP*, and *Mkup* are retail gasoline prices, adjusted gasoline prices, and markups of gasoline stations. *AvRetailP* (*AvMkup*) is unweighted daily regional average retail prices (markups), and *QwRetailP* (*QwMkup*) is quantity-weighted average retail prices (markups). *AWP* stands for average wholesale prices (see section 2.2), and *SmartPen* denotes the ratio of smartphone users to the region's population, or the smartphone penetration rate, as described in section 2.4. *Self, Carwash, Repair,* and *Store* are station characteristic dummy variables. Since station characteristics remained constant for the sample period, these variables do not have t subscript.[10]

For the initial descriptive analysis, I aggregate my station-level data up to the regional level and compute several measures of price dispersion which vary at the region-day level. The last four variables of Table 1 *(Range, Std, IDR,* and *IQR*) are price dispersion measures that have two indexes: r stands for region and t for time. In addition to the standard measures of price dispersion, *Range* and *Std*, I also consider interdecile range (*IDR*) and interquartile range (*IQR*) to investigate the characteristics of price dispersion in detail. When calculating price dispersion, I use *AdjustedP*, prices adjusted for differences in station characteristics. I run a regression with day fixed effects, $RetailP_{jt} = \beta_0 + \Sigma\delta_t Day_t + \Sigma\gamma_j X_j + \epsilon_{jt}$, and define $AdjustedP_{jt} = RetailP_{jt} - \Sigma\hat{\gamma}_j X_j$, where X_j are station characteristics.

Table 2 presents summary statistics for all variables. I converted the data from Korean won per liter to U.S. dollars per gallon to aid interpretation.[11] While the mean and standard deviation of the quantity-weighted prices are very similar to the unweighted prices, the quantity-weighted markups are lower than the unweighted ones.

Summary statistics of the four price dispersion measures are shown in Table 3. The left side of Table 3 presents the unweighted statistics, and the right side presents descriptive statistics of the four dispersion measures that are computed with quantity-weighted data. For example, suppose that station 1 makes 2 sales at price p_1 and station 2 makes 6 sales at price p_2. For the quantity-weighted case, I consider that there are 2 entries of p_1 and 6 entries of p_2. The unweighted case assumes that two prices have equal proportions, and hence there are 4 entries of p_1 and p_2. For instance, if $p_1 = 2$ and $p_2 = 1$, then the unweighted standard deviation is 0.535 and the quantity-weighted standard deviation is 0.463. As expected, Table 3 confirms that the quantity-weighted dispersion measures have smaller values.

3.2 Average Markup and Price Dispersion Trends

I present average markup and price dispersion trend graphs and a brief interpretation of the graphs in this section.[12] Figure 2 shows trends of the four dispersion measures. None of the four measures decreases over time and they actually slightly increase in the second half of the sample period. Quantity-weighted dispersion measures also do not decrease over time; Figure 3 presents quantity-weighted and unweighted standard deviation measures. Note that the quantity-weighted dispersion is lower than the unweighted one,

[10] There are two cases that station characteristics changed for a short period of time. According to an Opinet official, it is likely that these are input errors.

[11] I used the following conversion rates: 1 gallon = 3.785 liter; 1 dollar = 1000 Korean won.

[12] Since there were nominal retail price differences between gasoline companies from 4/7/2011 to 7/7/2011, I treated that period separately (details in Appendix A). During this period, the Korean government asked the four major gasoline companies to cut their prices, and they agreed to reduce retail prices by 100 Korean won per liter. However, one gas company (SKE) chose to offer a rebate of 100 won per liter, instead of cutting the posted price directly. This policy caused artificial relative posted price differences between SKE stations and non-SKE stations.

as quantity-weighted dispersion reflects that stations with lower prices tend to make higher numbers of sales.

In addition, both quantity-weighted and unweighted markup levels remain steady (Figure 4). As stations with lower prices tend to have lower markups and sell more, it is not surprising that the quantity-weighted markups are lower than the unweighted markups. Also, the differences between two measures are fairly stable during the period.

There are two hikes: early 2012 and mid-2012, the latter being the end of the sample period. For both cases, while a rapid decline in the international oil prices caused a sharp drop of AWP, retail prices fell only slowly. For the second hike, I find that average retail prices decreased in the next month (right after the end of the sample period) and average markups went back to the usual levels. These cases are classic examples of asymmetric price adjustments, or "prices rise faster than they fall" (Peltzman, 2000).[13] For instance, AWP fell about 10% (about 64 cents per gallon) during the last four weeks of the sample period, but average retail prices fell only 3%.

3.3 Measures of Price Dispersion and Markup Changes

The graphs in the previous section were suggestive, but do not account for all factors which could drive price dispersion and markups. This section, therefore, will examine these quantities in a series of reduced-form regressions. These regressions are intended as descriptive in nature. I do, however, assume that smartphone penetration rates are exogenous with respect to gasoline price dispersion or markups of the stations. This assumption seems fairly innocuous – it is unlikely that people buy smartphones because gasoline price dispersion or markups are low or high, or are even aware of them for that matter.

As described in the section 3.1, I aggregate my station-level data up to the region level and compute several measures of price dispersion which vary at the region-day level. The base regression format is

$$PD_{rt} = \beta_0 + \beta_1 AWP_t + \beta_2 SmartPen_{rt} + \sum_{j=1}^{11} DM_j + \sum_{k=1}^{3} DR_k + \epsilon_{rt}$$

The dependent variable, PD_{rt}, is one of the four dispersion measures. $SmartPen_{rt}$ is a smartphone penetration rate at time (date) t for region r, and AWP_t is an average wholesale price of gasoline at date t.[14] DM_j are month of the year dummies and DR_k are region dummy variables. In Table 4, I present regression results with the dependent variable Std, standard deviation. Results from the unweighted data are presented in columns 1 and 2, and results from the quantity-weighted case are presented in columns 3 and 4. Both

[13] Many previous papers reported asymmetric price adjustments in the gasoline market. See, e.g., Borenstein et al. (1997), Noel (2007), Verlinda (2008), and Lewis (2011).

[14] Since it is possible that stations make purchases in advance and in fact pay AWP from a week or two weeks before, I also consider AWP a week before and AWP two weeks before as additional variables. Regression results show that there are no big differences, so I only report the regression with the AWP term.

month of the year dummies and region dummies are included in all regressions, and columns 2 and 4 have a time trend term included.

Positive *SmartPen* coefficients imply that dispersion is higher when there are more smartphone users. The regression results from the column 1 suggests that the standard deviation decreases by 0.252 cents per gallon when the smartphone penetration rate increases by 1%.[15] Including the time trend decreases the magnitude of *SmartPen* coefficients, but the significance levels do not change. Negative signs for the AWP coefficients mean that when marginal costs are higher (hence price levels are higher), price dispersion is smaller. For example, the results from column 1 suggest that the standard deviation decreases by 4.4 cents per gallon when the average wholesale cost goes up by one dollar per gallon.

As a robustness check, I present the regression results for all four price dispersion measures in Table 5. For the unweighted data, all coefficients are highly significant and have the same signs. For the quantity weighted case, coefficient estimates are less consistence across dispersion measures. This finding suggests that quantity-weighting might be important.

Since we are also interested in evidence on markups, not just price dispersion, so I also consider the following regression for completeness:

$$Mkup_{rjt} = \beta_0 + \beta_1 AWP_{jt} + \beta_2 SmartPen_{rt} + \sum_{i=1}^{4} DC_i + \sum_{j=1}^{11} DM_j + \sum_{k=1}^{3} DR_k + \sum_{k=1}^{4} DChar_{kj} + \epsilon_{rjt}$$

The dependent variable, $Mkup_{rjt}$, is markup of station j at time t for region r. $DChar_{kj}$ is a set of dummy variables representing characteristics k of station j. For instance, if station j offers full service, $DSelf_j = 0$. Similarly, *Carwash* denotes whether a station has a car wash, *Repair* denotes whether a station has a repair shop, and *Store* denotes whether a convenience store is located at the station.

Column 1 of the Table 6 shows the results of the above regression, and column 2 includes the time trend term t. Results of the two columns are almost identical except for the *SmartPen* coefficients. Since *SmartPen* is increasing over time during the data period, including the time trend term decreases the *SmartPen* coefficient. *SmartPen* has a positive coefficient and implies that a markup is 0.344 (0.266 for column 2) cents per gallon higher when a smartphone penetration rate is 1% higher.[16] *AWP, Self,* and *Store* coefficients are significant and have expected signs: higher markups are expected when the marginal cost is lower, when a station offers full service, or has a convenience store.[17]

[15]This result is contrary to the previous research on the effect of mobile phones (Jensen, 2000 and Aker, 2007). But as discussed before, the mechanism for potentially affecting price dispersion is much different here than in the previous cases.

[16]Since AWP is the same for all stations belonging to the same gasoline company, and the differences of AWP between gasoline companies are small, we would expect the markup regression results to be very similar to those for *RetailP*, and they are. The *RetailP* regression results also suggest that *RetailP* is increasing in *SmartPen*.

[17]While *Carwash* and *Repair* terms are not significant, their signs can also be interpreted. A station with car wash is more

Combined with the previous dispersion regression results, this descriptive analysis suggests that higher smartphone penetration rates are associated with higher markup levels and higher price dispersion. These results seem counter-intuitive: it is natural to assume that higher smartphone penetration rates would lead to more consumer search, and more price information would intensify competition, hence lower prices and price dispersion. There are two possible explanations. First, consumers do not, or seldom, use smartphones to search for the price information. If this were the case, the increase in the smartphone penetration rate does not necessarily change the search intensity and the proportion of informed consumers. Second, consumers search more, but gas stations change their pricing strategies since they face different demand. As a result, we reach a new equilibrium, and price dispersion and markups do not decrease. In the next section, I present three additional descriptive analyses that suggest changes in consumer search behavior and gas stations' pricing strategies.

3.4 Additional Descriptive Evidence

Before presenting the structural model, I offer three types of suggestive evidence that consumers indeed searched more during the sample period, and gas stations changed their pricing strategies as a result. I document stylized facts about the Opinet price information service usage to find that a stable fraction of consumers and smartphone users have utilized the Opinet service. I also discuss a simple regression model to confirm that consumers became more sensitive about the minimum prices. Lastly, I adduce that gas stations tend to charge either high prices or low prices over time, using a test for unimodality.

Opinet Usage Has Increased

In the data section, I presented a graph that shows a dramatic increase of smartphone users and discussed how Opinet real-time price information is provided. However, it is not certain what the combined effect of these two is. I start with the basic question that is motivated from the regression results: do people actually use smartphones to access the Opinet service? An even more fundamental question is: how many people use the Opinet service?

Figure 5 shows trends of four variables. The variable labeled *Opinet Web/Car* shows the number of Opinet website visitors for each week as a fraction of the total number of registered cars in the nation. It is slightly above 2% before 2011 and about 5-6% during 2011 and 2012. If we assume consumers search for gas prices only when they fill up and it happens once per week on average (Byrne and Roos, 2014), then on average, 5-6% of the gas purchasers use Opinet to become highly informed without using a smartphone.

desirable for consumers, and hence may charge higher price without losing consumers. For a station with a repair shop, it tries to attract consumers to the repair shop where profit margin is high. It is possible that this type of station uses gasoline prices as loss-leaders.

The *Total App/Car* variable, the ratio of the number of Opinet smartphone application (henceforth "app") downloads for each month (cumulative) over the number of cars, clearly shows an increasing trend. It has increased rapidly after 2010: it starts from less than 1% in December 2010, and reaches 12% at the end of the sample period. If we assume that all drivers who downloaded the Opinet app are informed and who did not download are not, this trend suggests that about 12% of drivers are highly informed at the end of the period.

iPhone App/iPhone represents the ratio of the number of Opinet iPhone app downloads to the number of iPhone users. The iPhone Opinet app was first introduced in May 2010, and the Android Opinet app was available from January 2011. After the initial release, about 5% of iPhone users downloaded the app, and this ratio remained fairly stable until January 2011. When Opinet also launched the Android version app and started advertising, it seems that more iPhone users became aware of the Opinet app: during 2011 and 2012, about 9.5% of iPhone users downloaded the app. The *Total App/Smartphone* variable, the proportion of the total Opinet app download numbers to the total number of smartphone users, shows a similar trend.[18]

In summary, these graphs imply that (i) the Opinet website received a steady stream of visitors, (ii) the proportion of smartphone users who downloaded the Opinet app was relatively stable, and (iii) with the rapid increase of smartphone penetration, the number of Opinet app downloads also increased.

While these results suggest that consumers actually utilized the Opinet service more frequently, and more consumers searched by using the service as the smartphone penetration rate increased, it should be noted that there are also other important channels through which the Opinet data have additional impacts. When the Opinet launched its Android version app in January 2011, it also started advertising the Opinet service and provided most of the Opinet information to other websites such as naver.com (Korean version of Google.com) and car navigation systems. While separate data for other methods that also provide the Opinet price information service are not available, Opinet officials suggested that the access rate trends for the other websites would be similar to that of the Opinet website. I construct estimates of the impact of Opinet as a function of smartphone penetration rate. Estimated effects should be thought of as reflecting the combined impact of the various channels, not just smartphones.

A Simple Demand Regression

In addition to the Opinet usage, I estimate a crude demand equation to suggest that consumers actually did search more in the latter periods when there were more smartphone users. If consumers search more,

[18] Since there was no Android or iPhone app available before May 2010, it is zero. From May 2010 to December 2010, there was only the iPhone app and the ratio remained stable at 1.5%. When the Android version was introduced, this ratio jumped to 8% and stayed 7-8% for the rest of the period. As the number of smartphone users was rapidly increasing, a stable *Total App/Smartphone* implies that the number of the app users was increasing also.

consumers will be more sensitive to the difference between its price and the minimum price. To examine this conjecture, I estimate a regression with station fixed effects on the 2010 quantity data and separately on the 2012 quantity data:

$$ln\left(Q_{jt}\right) = \alpha_1 p_{jt} + \alpha_2\left(p_{jt} - p_{min,t}\right) + f_j + \epsilon_{jt}$$

α_1 captures the effect of own price increases on the quantity sold, α_2 represents the effect of the difference between its price and the minimum price of the region at time t, and f_j are station fixed effects. Three instruments are used for the price and price difference terms: average wholesale costs of gasoline; the difference of the costs between station j and the costs of the lowest price station(s) in the market; and the sum of characteristics of other rival stations within 1 mile radius.

I emphasize that the main purpose of this regression is not to identify the exact effect of own price changes or relative price changes, but to show how these coefficients change over time. In particular, I compare estimates from the first year (2010) and ones from the last year (2012) at Table 7. While α_1 does not change much, the magnitude of α_2 is higher for the latter period (α_2 is negative, so it means that α_2 decreases) and this difference is statistically significant. These results suggest that consumers became more sensitive to the price difference between station j's price and the minimum price during 2012, compared to 2010. In other words, these results suggest that there are more informed consumers for the latter period.

The Bifurcation of Markups

In the previous sections, I find that while consumers become more price-sensitive and search more, price dispersion and average markups do not change. One possible explanation is a polarization of stations as suggested by the Stahl model (1989, 1996). The first group of gasoline stations focus on the consumers who do not know the prices (no Opinet price information) and happen to visit these stations. They try to maximize profit given these uninformed consumers and concede consumers with price information. On the other hand, the second group of gasoline stations adopt a low-price high-volume policy and try to attract consumers with price information.

If this were true, we might be able to observe bimodality in a price or markup distribution graph. Since I use the same marginal cost (daily average wholesale cost) for all stations belonging to one gasoline company for a given day, and the differences between AWP of gasoline companies are fairly small, shapes of price graphs and markups graphs are similar. The first graph in the Figure 6 shows a kernel density of quantity-weighted markups for gasoline stations in region 3 at March 1st, 2010, and the second graph shows the one at March 1st, 2012. The 2010 graph shows little evidence of bimodality while the 2012 graph is suggestive of a double peaked distribution.

The figures shown above are just from a single day, but these features are robustly present across dates. To show this formally, I use Hartigan's Dip Test (1985), which tests for multimodality in a sample by "the maximum difference, between the empirical distribution function, and the unimodal distribution function that minimizes that maximum difference". I compute proportions of days in a given period that a markup distribution of a given day is rejected to be unimodal, according to the test.[19] While the proportion of rejection rates of unimodality for unweighted markup distributions is low and does not vary significantly over time,[20] the proportion of rejection rates for quantity-weighted markup distributions shows an increasing trend. For example, average proportion of the rejection rates over the regions is 0.35 for 2010, 0.38 for 2011, and 0.56 for 2012: approximately 1/3 of the region-day observations reject the null hypothesis that a quantity-weighted markup distribution for a given region-day is unimodal at the beginning of the period, and more than half of the region-day observations reject the test at the end of the period. These results suggest that there were changes in the quantity-weighted markup distributions but not in the unweighted markup distributions, while average values for both markup measures remained stable during the period (the trend graphs in section 3.2).

4 Structural Model

While the descriptive analysis section suggests that consumers have become more price-sensitive and searched more, it also reports that both price dispersion and markup did not decrease. Structural model can help reconcile these descriptive results that seem counter to our naive intuition and provide more complete equilibrium pictures. Motivated by these results, I develop a structural model of consumer choice to estimate demand to answer the following questions: (i) Do mobile technologies and the price information service actually affect the proportion of informed consumers? If yes, then how much?, (ii) What are the effects of the factors related to the consumer choices in the retail gasoline market?, and (iii) How would gasoline stations adapt their pricing strategies in reaction to consumer search behavior changes?

In order to answer these questions, I study how the ratio of informed consumers changes, as smartphone penetration rates change, to measure the effect of the Opinet price information service on the market. As it is costly to gather prices by visiting gasoline stations, consumers who do not use the Opinet service typically have very limited price information. On the other hand, consumers who utilize the Opinet service may obtain all price information at once. Reflecting this price information distribution, I assume that there are two types of consumers (informed consumers and uninformed consumers). By estimating a modified version

[19] I calculate p-values for each region-day observation, under the null hypothesis that a given distribution is unimodal.
[20] These rejection rates are close to zero, as p-values for the test are higher than 0.2 for most of the time.

of a random-coefficients discrete-choice demand model (Berry et al., 1995, henceforth BLP) of the two types of consumers, I explain the effects of important factors such as prices and distances from consumers to stations. While I use the same taste parameters for both consumer types as there are no obvious differences between the two types except the amount of price information, consumers have different locations, and hence different distances to gasoline stations. Combining expected market shares from both types, I compare the total expected market shares with the observed market shares to find out the parameter values that minimize the objective function. In particular, I focus on the price sensitivity term, the distance sensitivity term, the proportion of informed consumers who search without smartphones, and the ratio of smartphone users who do search to get full price information.

4.1 Consumers on the Grid

To build a structural model of consumer search and purchase behavior, it is important to consider and incorporate factors that drive consumer choice in the retail gasoline market. There are three main factors: price, distance, and station amenities. Distance to a station is unique among these factors in horizontally differentiating gas stations: the value of a station's location depends on the consumers' location. Unlike other factors, such as the price at a station or whether that station offers a car wash, that are the same for all consumers, the distance to a certain station is different for consumers from different locations.

To consider how distances would affect consumer choices, I start with a simple model of location. Imagine a rectangle that covers the whole region. Divide it into $n-1$ by $m-1$ rectangles so that we have an n by m grid. I assume that consumers are located at each grid point according to a uniform distribution. For example, if $m = n = 10$ then each grid point has 1% of the consumers. Each of the four regions are approximate squares, so I employ square grids in each case.[21]

In contrast, Houde (2012) was able to use commuting patterns to assign consumer locations in a model of gasoline demand. He divides Quebec City into a grid and decomposes consumers into four components: workers, full-time students, unemployed, and outside commuters. Assuming non-congestion for travel paths, he computes the probability of commuting routes using road network and census data. However, the uniform distribution assumption without considering commuting patterns is realistic in my case, especially for two regions that are districts of Seoul. Figure 7 shows a map of one of the districts. As the streets are laid out in a grid pattern, the grid assumption reflects the actual road map well. Moreover, there are no clear commuting paths in this region as (i) residential and commercial areas are mixed, (ii) almost all roads are congested during rush-hours, and (iii) there are multiple ways to enter and exit the region.[22]

[21] Region 1 and 2 are $9\times$ 9 miles and Region 3 and 4 are $10\times$ 10 and $7\times$ 7, respectively.

[22] While it would be ideal to have traffic volume information to estimate demand distribution more precisely by putting different weights on different grid points, it is almost impossible to measure the volume of traffic separately for the roads where

stations are located.

4.2 Two Types of Consumers: Informed Consumers

There are two types of consumers, informed consumers and uninformed consumers.[23] While informed consumers know all station information by using the Opinet service, uninformed consumers only have limited information. For example, if someone starts to look for a gas station when he runs low on gas and chooses to go to one of the first two stations he finds, he is an uninformed consumer. Since uninformed consumers do not have information for most gas stations, or "products", their choice sets are limited. On the other hand, if someone uses her smartphone to check all the stations in the region and makes a decision to visit a certain station, she is an informed consumer. Since using the Opinet price service will provide all information (including gasoline prices) with a single search in several seconds, I assume that consumers who use the Opinet service gets all information without any cost.[24]

For informed consumers, following standard utility assumptions, the indirect utility of consumer at i for station (or product) j in market t is given by

$$u_{ijt} = d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where d_{ij} is a distance (in miles) from consumer at i (grid point i) to the station j.[25] X_j are observable station characteristics, β_c is a vector of consumer taste coefficients for station amenities, β_d is a distance coefficient, and α is a price coefficient (p_{jt} is price of station j in market t). Lastly, $\xi_{jt} = \xi_j + \Delta\xi_{jt}$ are unobserved station characteristics. As Nevo (2001) suggests, I include station-specific dummy variables as unobserved (by the econometrician) station fixed effects ξ_j. Market-specific unobserved components are included in $\Delta\xi_{jt}$ and are left as "error terms".

Market t is one day of a certain region. For example, Gangnam District, March 2nd, 2011 is one market. As the data from January 1, 2010 to June 27, 2012 are chosen, the number of markets is 909 times the number of regions. Note that this utility setup is a special case of the BLP model where consumers have different values (denoted by the subscript i) only for the $d_{ij}\beta$ term, except for the separable additive random shocks. As consumers are more likely to substitute toward stations that are close to each other, location of consumers with respect to stations is an important source of heterogeneity between consumers.

ϵ_{ijt} is an i.i.d random utility shock distributed according to a Type I extreme-value distribution. Then,

[23]The two-type consumer assumption is applied in other empirical settings. One example is an online computer memory chips market where Moraga-Gonzalez and Wildenbeest (2008) find that the consumer population can be split into two groups which either have high search cost or low search cost.

[24]This single search assumption is used in previous literature, including the retail prescription drugs study of Sorensen (2001). He uses transactions data of prescription drugs to estimate a discrete-choice demand model that embeds a simple search decision. In particular, he assumes that search is "all or nothing": consumers either search exhaustively to learn all pharmacies' prices, or not at all.

[25]$d = 1$ means that a distance between a consumer and a station is one mile.

the percentage of consumers at i who chooses station j in market t is[26]

$$\frac{\exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right)}{1+\sum_{j=1}^{J}\exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right)}$$

Note that in this setup, choosing the outside option in market t means that a consumer does not go to a gas station in market t. Let w_i denote the proportion of consumers at i in the population (for the base model, all $w_i = \frac{1}{n^2}$). The overall market share of station j is

$$s_{jt}^{informed} = \sum_i w_i \frac{\exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right)}{1+\sum_{j=1}^{J}\exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right)}$$

4.3 Two Types of Consumers: Uninformed Consumers

To explain how consumers search without the price information service, a consumer search behavior assumption plays an important role in the model setup. The theoretical literature typically models consumer search in two ways: the fixed sample size search model, where consumers sample a fixed number of stores and choose to buy the highest utility one, and the sequential search model, where consumers decide to search one more if the expected benefit from the next search is higher than the search cost. Both types certainly could be applied in this case. One could argue that when a driver needs gas, he would try to look for several stations nearby and decide where to go. One example is a driver who observes several prices on the way to work and chooses one of them on the way back home. On the other hand, it is possible that the driver sees the first gas station and checks the price and both observed and unobserved station characteristics, and decides whether to continue search or just stop by the station and fill his car up, depending on his expectation of prices of other stations and his own search cost.

Since both models are plausible, I chose a fixed sample search approach which is more straightforward to implement.[27] Let m be the number of stations whose information is known to an uninformed consumer. Choosing $m = 1$ would cause a station to set very high price in equilibrium so that it can make huge profit from uninformed consumers who only know its information and concede all other consumers.[28] Thus, I choose $m = 2$: each uninformed consumer learns $m = 2$ station prices among J stations in the market.[29]

[26] In this setup, choosing the outside option at market t means that a consumer does not go to gas station at date t. As usual, I normalize the utility from the outside good to zero.

[27] When there is a delay between the search decision and the search outcome, Morgan and Manning (1985) have shown that a fixed sample size search typically offers a better explanation of observed behavior than a sequential search. Santos, Hortacsu, and Wildenbeest (2012) argue that fixed sample size search models provide a better explanation of observed consumer search behavior in online book stores than sequential search models. While I do not access which model fits better in this paper, this result motivates me to apply the fixed sample size search model for the case when consumers do not conduct a smartphone price search and are not aware of the prices.

[28] Moreover, interviews with several drivers confirmed that consumers tend to avoid making choices when they are given only one price information, as they are not certain that whether the given choice is a complete rip-off.

[29] The estimation results are fairly robust to the choice of m: for example, $m = 3$ does not change the results significantly. Moreover, it is unlikely that a driver would observe only one station during his trip.

To determine which station information consumer at i gets, I assume that a probability of getting price information of the station j is proportional to the inverse of the distance, d_{ij}. In other words, Pr (consumer at i learns the price of the station j_1) / Pr (consumer at i learns the price of the station j_2) $= \left(\frac{d_{ij_2}}{d_{ij_1}}\right)$. Solving these equations, I get

$$r_{i,k} = \text{ Pr (consumer at } i \text{learns the price of the station } j_k\text{)} = \frac{\frac{1}{d_{ij_k}}}{\Sigma_s \frac{1}{d_{ij_s}}}$$

Using the previous result, I compute the probability of consumer at i learns the price of station j_{k_1} and j_{k_2}.

$$\begin{aligned} r_{i,k_1,k_2} &= \text{Pr (consumer at } i \text{learns the price of the station } j_{k_1} \text{and } j_{k_2}\text{)} \\ &= r_{i,k_1} r_{i,k_2} \left(\frac{1}{1 - r_{i,k_1}} + \frac{1}{1 - r_{i,k_2}} \right) \end{aligned}$$

I assume that after learning prices for two stations, an uninformed consumer would choose a station in this restricted set the same way an informed consumer would. Using the same distributional assumption, a market share for station j among uninformed consumers is

$$s_{jt}^{uninformed} = \sum_i w_i \left(\sum_{l \neq j} r_{i,j,l} \frac{\exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right)}{1 + \exp\left(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt}\right) + \exp\left(d_{il}\beta_d + X_j\beta_c - \alpha p_{lt} + \xi_{lt}\right)} \right)$$

4.4 Demand Side

As I only observe total quantity sold for each station, I need to combine the results from the informed consumers and uninformed consumers so that I can compute the total market share to match the observed values and the expected ones. I assume that the size of market t is the number of cars in the region at that time. Let IR_t, the informed ratio, denote the proportion of informed consumers in market t. Intuitively, IR_t should increase as time passes, since the proportion of informed consumers is bigger when there are more smartphone users. I allow for a linear growth in the informed ratio to change every day by assuming $IR_t = a_0 + a_1 SmartPen_t$. a_0 represents the fraction of drivers who are already informed without smartphones, and a_1 indicates the proportion of smartphone users who become informed consumers.

Then the total expected market share of station j is

$$\hat{s_{jt}} = IR_t \cdot s_{jt}^{informed} + (1 - IR_t) \cdot s_{jt}^{uninformed}$$

4.5 Supply Side

In addition to the demand side, I incorporate the profit-maximizing conditions for gas stations. As a profit for station j in market t is $(p_{jt} - mc_{jt})\, q_{jt} - F_{jt}$ where F_{jt} denotes fixed costs, the optimal behavior for gas stations is to follow the profit maximizing condition: $(p_{jt} - mc_{jt})\, \frac{\partial q_{jt}}{\partial p_{jt}} + q_{jt} = 0$. This first-order condition reduces to $p_j = mc_j - \frac{s_j}{\partial s_j / \partial p_j}$, and this requires the derivatives of the market share function with respect to price. Thanks to the analytical market share formula, I can compute $\frac{\partial s_{jt}}{\partial p_{jt}}$ explicitly (details in Appendix B).

For simplicity, I begin by assuming that the marginal cost is both independent of output levels and linear in cost characteristics: $mc_j = \gamma_{0,j} + \gamma_{1,j} AWP_j + \eta_j$. $\gamma_{0,j}$ is a fixed component that mainly consists of the transportation cost and the variable labor cost.[30] AWP_j is an average wholesale price for station j, and η_j is an unobserved cost shock. As AWP_j is what station j needs to pay today to replenish,[31] I set $\gamma_{1,j} = 1$. We have $\eta_j = p_j - (\gamma_{0,j} + AWP_j) + \frac{s_j}{\partial s_j / \partial p_j} = p_j - b_j(p, X, \xi; \theta) - W_j \gamma$ where $W_j = (1\ AWP_j)$ and $\gamma = (\gamma_{0,j}\ 1)'$. I assume that while the unobservable cost term η_j might be correlated with ξ_j, it is mean independent of the average wholesale cost (and the observable station characteristics).

4.6 Instruments

In this section, I explain which instruments are used and what their identifying assumptions are. Following previous literature, moment conditions are formed by interacting instruments with the unobservable error terms.

First, I exploit the panel structure of the data and use the average prices of the stations in different regions that belong to the same gasoline company as instruments.[32] The identifying assumption for these instruments is that the average prices of the stations that belong to the same gasoline company in two regions are correlated due to the common marginal cost changes, but they are independent of the region-specific unobserved valuation changes, $\Delta\xi_{jt}$. Since the regions in the data are far apart and no national demand shock is likely, the average prices of another region are independent of the unobserved valuation changes in the region.[33] Let these instruments be Z_1. Then, $E\left(\Delta\xi_{jt} | Z_1\right) = 0$ and we have the usual moment conditions $E\left(\Delta\xi_{jt} \cdot Z_1\right) = 0$.

Another set of instruments Z_2 are cost shifters, average wholesale prices for each gasoline company. As average wholesale prices directly change marginal costs but only affect demand through prices, I have

[30] In fact, this variable labor cost is fairly similar among gas stations, as most of the gas station workers are temporary part-time employees who are paid a national minimum hourly wage or slightly above the minimum.

[31] Also, this is the expected future average wholesale price given today's information.

[32] This is similar to the instruments used by Hausman (1996).

[33] Since region 1 and 2 are neighboring districts, I use region 3 and 4 prices as instruments for region 1 and 2. For region 3 and 4, all other region prices are valid instruments. As distances between regions are more than 50 miles except for the region 1 and 2, it is unlikely that consumers can switch to other regions.

$E(\Delta\xi_{jt}|Z_2) = 0$. In addition, I assume that this exogenous element of the marginal cost shifter is not correlated with the unobservable cost η. This assumption gives additional moment conditions $E(\eta \cdot Z_2) = 0$.

While Z_1 and Z_2 have excellent time-series variation, they have the same values for stations in the same gas company. The next set of instruments, Z_3, the sum of characteristics of other rival stations within one mile radius, provides the other type of necessary variation: they vary substantially by station.[34] The identifying assumption is that characteristics of other stations are correlated with prices as the markup levels are affected by these measures of isolation in the product space, and "location" of gas stations in the characteristics space is exogenous, or predetermined. Also, I assume that these observable characteristics are mean independent of the unobserved cost shock, as the effects of observables are reflected in the $b_j(p, X, \xi; \theta) + W_j\gamma$ part of the marginal cost equation. Then, I have $E(\Delta\xi_{jt} \cdot Z_3) = 0$ and $E(\eta \cdot Z_3) = 0$.

I conclude with an informal discussion on identification of the model. The main parameters of interest are the price sensitivity, the distance sensitivity, and the ratio of informed consumers (α, β_d, and IR_t). In this model, variation in the market shares is due to (i) variation in the prices and attributes of gasoline stations, (ii) variation in the smartphone penetration rates, and (iii) variation in distances among consumers. Note that the distance effects are assumed to be stable over time ($d_{ij}\beta_d$ does not have a t subscript). The distance parameter β_d is identified from the correlation of distances and market shares that are observed in the data. The price sensitivity term α is identified from time-series variation of market shares and the high-frequency variation in prices charged by a station and its competitors.

Moreover, I know that price changes occur once a week on average, but retail prices for stations move up or down in tandem by the same amount on many different days. Consequently, market share differences during periods with the same prices (or identical price movements) can be used to identify the informed ratio parameters (a_0 and a_1 from the variation of IR_t). Another way to explain the informed ratio identification is that the changes in IR_t capture the changes in the consumer price elasticity. As more consumers have the full price information, consumers become more price-sensitive. Since I assume that α is the same for all periods, movements in IR_t reflect this consumer price sensitivity changes by putting different weights on the informed consumers. For example, a high IR_t value assigns a large proportion of more elastic consumers (informed consumers) so that the consumer price elasticity for the market level is high.

4.7 Estimation

I estimate the parameters of the model by following the BLP algorithm, except that I include station-specific dummy variables and I utilize a different approach for the contraction mapping part. For the inner loop, I compute the mean utility level δ_j and run the IV regression $\delta_j = X_j\beta_c - \alpha p_j + \xi_j + \Delta\xi_{jt}$ with a set of

[34] However, they do not have good time-series variation.

instruments $Z_d = [z_1, \ldots, z_{M_1}]$. Then, the moment conditions are $E[Z_d \cdot \Delta\xi_{jt}(\theta)] = 0$, where $\theta = \{\alpha, \beta_c, \gamma, \sigma\}$ and $\sigma = \{\beta_d, a_0, a_1\}$. Another set of the moment conditions is from the mean independence assumption of η_j and a set of instruments $Z_s = [z_{M_1+1}, \ldots, z_{M_1+M_2}]$: $E[Z_s \cdot \eta(\theta)] = 0$. Let $h(\theta)$ be these moment conditions such that $E[h(\theta)] = 0$. For the outer loop, I calculate the empirical analogue of the moment conditions, $\hat{h}(\theta)$. Using the two-step generalized method of moments (GMM), I find the GMM estimate that minimizes the objective function $\hat{h}'(\theta)\Phi^{-1}\hat{h}(\theta)$, where Φ is a consistent estimate of $E[h(\theta)h'(\theta)]$.

The inclusion of the station fixed effects (in the inner loop IV regression) requires the minimum-distance procedure (Chamberlain, 1982) to estimate the taste parameters β_c. I follow Nevo (2001). Let f be the vector of station dummy coefficients (f_j captures both the quality of observed station amenities and the mean of the unobserved characteristics, $X_j\beta_c + \xi_j$). If we assume that $E[\xi|X] = 0$, the estimates of β_c and ξ are $\hat{\beta}_c = \left(X'V_f^{-1}X\right)^{-1} X'V_f^{-1}\hat{f}$ and $\hat{\xi} = \hat{f} - X\hat{\beta}_c$. $\hat{f}$ is the estimated coefficient vector in the regression $\delta_j = -\alpha p_j + f_j D_j + \Delta\xi_{jt}$, and V_f is the covariance matrix of these estimates.[35]

As α, β_c and γ enter the GMM objective function linearly, I only need to search for $\sigma = \{\beta_d, a_0, a_1\}$ in the outside loop. Formally, the error terms are calculated by $\Delta\xi_{jt} = \delta_j - X_j\beta_c + \alpha p_j - \xi_j$ and $\eta_j = p_j - b_j(p, X, \xi; \theta) - W_j\gamma$. Let $T = \begin{pmatrix} Y & 0 \\ 0 & W \end{pmatrix}$, where $Y = \{X_j, p_j\}$. Let Z denote instruments, and $P_Z = Z(Z'Z)^{-1}Z'$. There is an analytic solution for $\binom{\beta_c, \alpha}{\gamma}$ given $\delta(s, \sigma)$:

$$\binom{\beta_c, \alpha}{\gamma} = (T'P_ZT)^{-1} T'P_Z \begin{pmatrix} \delta(s, \sigma) \\ p - b(p, X, \xi; \sigma) \end{pmatrix}$$

As a result, the GMM objective function is a function of σ only. This reduces the number of parameters I need to search in the outside loop. The outside loop only updates 3 parameters: β_d, a_0, and a_1.

The estimation procedure closely resembles the nonlinear GMM approach developed by BLP, except that I made a modification in the mean utility step. When I calculate the mean utility level δ_j, I need to use a different approach, as I do not have the quantity data for many stations. For example, I know market shares for $J_0 \approx 10$ stations out of $J \approx 45$ stations in the region. Since I have only 10 observations for each date, I cannot estimate 45 δ's for the date with the traditional BLP contraction mapping. I utilize the unusual high frequency of the data to construct weekly mean utility levels.

I make an additional assumption that, for each week, the only variation of the mean utility level for station j comes from the variation in p_{jt}. In other words, $\delta_{j,Tue} = \delta_{j,Mon} + \alpha(p_{j,Mon} - p_{j,Tue}), \ldots, \delta_{j,Sun} = \delta_{j,Mon} + \alpha(p_{j,Mon} - p_{j,Sun})$. With this assumption, I can write $\delta_{j,Tue}, \ldots, \delta_{j,Sun}$ as a function of $\delta_{j,Mon}$. I

[35] D_j are station dummy variables.

have $J_0 \times 7\,(\approx 70)$ equations, $s_{jt} = \hat{s_{jt}}\left(\delta_{j,Mon}\right)$ for $J \approx 45$ parameters $\delta_{j,Mon}$ for each week:[36]

$$\begin{aligned} s_{jt} &= \hat{s_{jt}}\left(\delta_{j,Mon}\right) = IR_t \cdot s_{jt}^{informed} + (1-IR_t) \cdot s_{jt}^{uninformed} = IR_t \cdot \sum_i w_i \frac{\exp\left(d_{ij}\beta_d + \delta_{j,t}\right)}{1+\sum_{k=1}^{J}\exp\left(d_{ij}\beta_d + \delta_{j,t}\right)} \\ &+ (1-IR_t) \cdot \sum_i w_i \left(\sum_{l \neq j} r_{i,j,l} \frac{\exp\left(d_{ij}\beta_d + \delta_{j,t}\right)}{1+\exp\left(d_{ij}\beta_d + \delta_{j,t}\right) + \exp\left(d_{il}\beta_d + \delta_{l,t}\right)} \right) \end{aligned}$$

for $j = 1, \ldots, J_0$ and $t = Mon, Tue, \ldots, Sun$. I replace the BLP contraction mapping part by finding J parameters $(\delta_{j,Mon})$ that minimize squared distances between the actual market shares and the model expected ones. Note that the IV regression in the inner loop becomes $\delta_{j,Mon} = X_j\beta_c - \alpha p_{j,Mon} + \xi_j + \Delta\xi_{j,Mon}$, or $\delta_{j,Mon} = -\alpha p_{j,Mon} + f_j D_j + \Delta\xi_{j,Mon}$, where D_j are station dummy variables and f_j are station fixed effects. While this approach reduces the number of observations for the regression to the number of weeks in the dataset, thanks to the high frequency of the data set, I have enough observations for each j.[37]

Note that as long as at least one p_{jt} changes, the right-hand-side value changes as the denominator $1+\sum_{k=1}^{J}\exp\left(d_{ij}\beta_d + \delta_{j,t}\right)$ changes. In the case when all prices remain stable, say during t_1 to t_2, then this equation becomes $E\left(\hat{s_{jt}}\left(\delta_{j,Mon}\right)\right) = \frac{\Sigma s_{jt}}{t_2 - t_1}$. While this would decrease the number of equations by $t_2 - t_1 - 1$, it does not happen often and I have at least J equations in most cases. In rare cases when I do not have J equations, I assume that for two weeks (instead of one week), the only variation of the mean utility level for station j comes from the variation in p_{jt}.

[36] As a reminder, w_i is a weight on grid point i ($\frac{1}{n^2}$ for the uniform distribution n by n grid case). $r_{i,j,l}$ is a probability of that an uninformed consumer located at i learns station j and l prices.

[37] For the whole period, the number of observations for each j is 129 (909 days, so 129 weeks).

5 Structural Estimates

This section presents structural estimation results and provides interpretations of the estimated parameter values. Structural estimation results suggest that the fraction of informed consumers actually increases from 1.7% to 11.4% during the sample period, as the smartphone penetration rate increases from 1% to 53%. For the effect of distances, I find that an average consumer is indifferent between driving 0.25~0.4 miles more and saving ten cents per gallon. The counterfactual analysis section derives a new set of equilibrium prices under a different informed consumer ratio, and confirms that observed price dispersion and markup changes are consistent with theoretical models of pricing, given the structurally estimated parameters.

5.1 Results and Interpretation

I report parameter estimates from the structural model and connect them to the descriptive findings for a better interpretation. Estimation results are presented in Table 8. Column 1 shows estimates for January 2010 to December 2010 when no Android version of the Opinet app was available.[38] Column 2 provides estimates for 2011 to 2012, when both iPhone and Android Opinet apps were available and the Opinet service was advertised nationally. Column 3 presents estimates for all four regions for the whole sample period, and column 4 is for regions 1 and 2, two districts of Seoul, where average income level is higher, traffic congestion is severe, and smartphone penetration rates are higher.

First of all, a_1, the proportion of informed consumers to the number of smartphone users, is positive and significant. The estimated value of 0.093 in column 3 can be interpreted as indicating that 9.3% of smartphone users are highly informed consumers. The results in columns 1 and 2 indicate that this has changed over time: 3.5% in 2010 and 16.5% in 2011-2012. As it is likely that the ratio of informed consumers to smartphone users is proportional to the ratio of the Opinet app download numbers to the total number of smartphone users, this change is not surprising: in section 3.4 we show that *Total App/Smartphone* is about 1.1% in 2010 and 7.0% in 2011-2012. The model results suggest that even before the Android version Opinet app and national advertising of the Opinet service, 3.5% of smartphone users visited Opinet website using smartphones to get price information, and 16.5% of smartphone users were informed after 2010. The region 1 and 2 estimate of a_1 is slightly higher than the all-four-region estimate, as consumers in the two districts of Seoul are more information-sensitive than consumers in the two rural cities.

The baseline proportion of informed consumers, a_0, is estimated to be 1.8% in column 3 (full sample). Again, the estimates from the earlier and later subsamples are different: 1.4% in 2010 and 3.1% in 2011-2012. This pattern can be thought of as similar to what is captured by the *Opinet Web/Car* variable, the proportion

[38] From January 2010 to April 2010, no Opinet app was available, and only the iPhone app was available from May 2010 to December 2010.

of weekly Opinet website visitors to the number of cars. In fact, as performing price search without the Opinet service is too costly, these two variables should have a similar trend, and the results confirm this: *Opinet Web/Car* is 2.3% in 2010 and 5.1% in 2011-2012 (see section 3.4).

The increase of a_0 over time is likely due to the higher consumer awareness of the Opinet service, thanks to the Opinet advertising campaign and word-of-mouth information diffusion. Like in the a_1 case, the region 1 and 2 estimate of a_1 is slightly higher than the all four region estimate of a_1. Knowing a_0 and a_1, and the smartphone penetration rate, I can estimate the ratio of informed consumers (IR_t) in the market: the informed ratio started from 1.7% in January 2010 and reached 11.4% at the end of the sample period (June 2012).

Now let us look at two other key parameters, the price sensitivity term α and the distance sensitivity term β_d. For the full sample case (column 3), $\alpha = 26.131$ and $\beta_d = -6.715$. Both estimates are highly significant and have expected signs: consumer utility decreases as prices increase or traveling distances increase. Since α is a price sensitivity term and β_d is a distance sensitivity term, it is better to interpret magnitudes of these two terms together. As a distance between neighboring grid points is 1 mile, an average consumer is willing to travel $-\frac{0.1\alpha}{\beta_d}$ miles to save ten cents per gallon. For example, $\alpha = 10$ and $\beta_d = -2.5$ means that an average consumer is indifferent between traveling 0.4 miles more and saving ten cents per gallon. A distance an average consumer is willing to travel to save ten cents per gallon varies depending on the specifications, with the minimum of 0.254 miles (column 4) and the maximum of 0.407 miles (column 1). In terms of the willingness of a typical consumer to travel an additional mile, an average consumer asks for \$0.25~\$0.39 lower prices per gallon.[39] If we assume an average driving speed in these regions to be 20 miles/hour and consumers buy five gallons of gas per each purchase, then the implied opportunity cost of time is \$24.55~39.43 per hour, which is fairly reasonable.

As column 4 is for two district of Seoul where average income level is higher and traffic congestion is severe, it is not surprising that column 3 estimates of α and β_d are higher than column 4 estimates. Also, magnitudes of the α and β_d estimates of the first period (column 1) are smaller than those of the latter period (column 2); however, these differences are not statistically significant.

Finally, I discuss station characteristic coefficients, β_{Self}, $\beta_{Carwash}$, β_{Repair}, and β_{Store}. As consumers are likely to prefer stations with full-service, car wash, and convenience stores, negative β_{Self} and positive $\beta_{Carwash}$ and β_{Store} are expected results. Two coefficients, β_{Self} and $\beta_{Carwash}$, are significant at the 5%

[39]Compared to the results of Manuszak and Moul (2009), \$0.065~0.084 per gallon, these estimates have higher values. However, as average price level in my data is about six times higher (about \$7 per gallon) than the average price level of \$1.2 per gallon during their sample period (Chicago and northern Indiana, 2001), their estimates would become \$0.39~0.504 per gallon, considering these differences. Moreover, they focus on tax differences in the area and assume that taxes were stable and consumers were aware of the resulting price differences. In my case, since consumers are not aware of the amount of potential savings unless they know the prices by searching, it is likely that the willingness to pay amount is different.

level for all specifications. For the full sample case (column 3), estimates imply that an average consumer would travel 0.58 miles more for full-service and 0.52 miles more for a car wash. On the other hand, β_{Store} is less significant under some specifications, and β_{Repair} does not even have a constant sign.

5.2 Counterfactual

When I presented the reduced-form results, I noted that they raised a puzzle: why have price dispersion and markups not decreased as more consumers have become informed? In this section, I provide a resolution of this puzzle by presenting counterfactual simulations examining how equilibrium prices would be expected to change as more consumers became informed, given the degree of consumer substitution and the magnitudes of the proportion of informed consumers that seems to be present. Changing a_0 or a_1 leads to different informed ratios and hence results in different equilibrium prices. For example, I study what the equilibrium prices would have been if the informed ratio were 0.05 or 0.2, instead of the original value 0.1.

To do this, I first estimate implied marginal costs that are consistent with the stations' pricing decisions. Note that without daily, individual station shocks, it is impossible to establish that stations follow the profit-maximizing condition every day. For the previous section, I used the sum of average wholesale prices and a constant as approximations of marginal costs. If I denote daily differences between the real, unobserved marginal cost and the average wholesale cost as e_{jt}, I can find e_{jt} from the first order condition:

$$(p_{jt} - [AWP_{jt} + e_{jt}]) \frac{\partial q_{jt}}{\partial p_{jt}} + q_{jt} = 0$$

$$(p_{jt} - [AWP_{jt} + e_{jt}]) \frac{\partial s_{jt}}{\partial p_{jt}} + \frac{q_{jt}}{Q_t} = 0$$

For the $\frac{q_{jt}}{Q_t}$ term, I use the actual q_{jt} for the stations that I have quantity data, and model computed $s_{jt} = \frac{q_{jt}}{Q_t}$ for the stations without quantity data. Then I get

$$e_{jt} = p_{jt} - AWP_{jt} + \frac{q_{jt}}{Q_t} / \frac{\partial s_{jt}}{\partial p_{jt}}$$

For simplicity, I define $ex_t(i,j) = \exp(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt})$. Based on the analytical market share formula, I can compute $\frac{\partial s_{jt}}{\partial p_{jt}}$ (details in Appendix B).

$$\frac{\partial s_{jt}}{\partial p_{jt}} = \frac{1}{n^2} \left\{ IR_t \sum_i \left(\frac{-\alpha \cdot ex_t(i,j)(1 + \Sigma_{k \neq j} ex_t(i,k))}{\left(1 + \sum_{k=1}^{J} ex_t(i,k)\right)^2} \right) + (1 - IR_t) \sum_i \left(\sum_{l \neq j} r_{i,j,l} \frac{-\alpha \cdot ex_t(i,j)(1 + ex_t(i,l))}{(1 + ex_t(i,j) + ex_t(i,l))^2} \right) \right\}$$

Empirical distribution of e_{jt} suggests that $e_{jt} = f_j + \epsilon_{jt}$, where f_j is an unobserved station fixed cost and $\epsilon_{jt} \sim N\left(0, \sigma_e^2\right)$. This fixed component refers to characteristics of the station j that are invariant during the period. These include, for instance, rent for the space, and labor cost for managers and core workers. The transitory component ϵ_{jt} reflects daily temporary changes. Empirical distributions of ϵ_{jt} are similar for both stations that I observe quantities and stations that I do not, and the estimate of σ_e is 0.057. As an average retail price during the period is about 7 dollars per gallon, a daily cost shock is about 0.8% of the average retail price.

Using the real marginal cost $c_{jt} = AWP_{jt} + e_{jt}$, I derive a new equilibrium prices when informed ratio IR_t is a different value. For each market t, I have n parameters $(p_1, p_2, \ldots, p_n)$and n profit maximizing conditions:

$$(p_j - c_j)\frac{\partial q_j}{\partial p_j} + q_j\left(p_j, p_{-j}\right) = 0$$

Note that q_j is a function of $p_1, p_2, \ldots, p_n$ and for each equation, I calculate a best response: fix p_{-j} and find optimal p_j. For each step, we get new $p_1', p_2', \ldots, p_n'$ as best responses from $p_1, p_2, \ldots, p_n$. Using the $\frac{\partial s_{jt}}{\partial p_{jt}}$ formula (Appendix B), the profit-maximizing condition becomes

$$p_{jt}' = c_{jt} + \frac{1}{\alpha}\frac{IR_t s_{jt}^{informed} + (1 - IR_t)\, s_{jt}^{uninformed}}{IR_t s_{1t} + (1 - IR_t)\, s_{2t}}$$

where $s_{1t} = \sum_i \left(\frac{ex_t(i,j)(1+\Sigma_{k\neq j} ex_t(i,k))}{\left(1+\sum_{k=1}^{J} ex_t(i,k)\right)^2}\right)$ and $s_{2t} = \sum_i \left(\sum_{l\neq j} r_{i,j,l}\frac{ex_t(i,j)(1+ex_t(i,l))}{(1+ex_t(i,j)+ex_t(i,l))^2}\right)$. Repeating these steps, I can find a fixed point (converging point) as new equilibrium prices.[40] Examining price dispersion (standard deviation) and markup levels at the new equilibrium prices, I find that both price dispersion and markup levels are slightly increasing as the informed ratio goes up, which is consistent with the results form the descriptive analysis section.

Table 9 shows how price dispersion (as measured by standard deviation) and markup levels change when the fraction of informed consumers, IR, changes from the baseline case, 10%. As the structural model estimates show that IR was close to 1% at the beginning of the sample period and about 10% at the end of the period, I start with comparing statistics for the 1% and 10% case. The counterfactual results suggest that there are moderate changes in price dispersion and average markups. The change in equilibrium prices, due to changes in demand, would increase the standard deviation of prices by 0.57%, the average markups by 0.33%, and the average quantity-weighted markups by 0.09%. Hence, this model could explain the "puzzle", the slight increase in both measures that we observe in the actual data. One way to interpret these moderate changes, in terms of magnitude of changes as the informed ratio increases, is that consumers have strong

[40] I use $\Sigma\left(p_j - p_j'\right)^2 < 10^{-12}$ as a convergence criterion: in this empirical setting, convergence took less than 20 steps, or one minute.

location preferences. As consumers prefer certain gasoline stations or stations close to them, their choices would not change much even if they learned the prices of stations that are far away.

Since mobile technologies continue to be developed and could reduce consumer search costs even further, the fraction of informed consumers is likely to grow. Thus, it would be interesting to consider what would happen if IR goes beyond 10%. Table 9 shows the results up to 35%. While price dispersion increases for all IR levels, both unweighted and quantity-weighted markups start to decrease at some point. Unweighted markups stop increasing around 30%, and quantity-weighted markups show a reversal when the informed ratio reaches 20%. As the proportion of informed consumers goes up, setting low prices to draw informed consumers becomes more attractive to stations. Since stations with lower prices make more sales, these reversals in markup trends happen earlier for the quantity-weighted case.

6 Conclusion

In this paper, I investigate how real-time gasoline price information and the spread of mobile technologies affect market outcomes, e.g., consumer search behavior and price dispersion among gasoline stations. I combine daily, individual gasoline station price and quantity data with regional smartphone penetration data for the analysis. The universe of daily station-level prices for each region allows me to perfectly measure daily market-level price dispersion. In addition, I utilize daily station-level quantity information to compute more realistic measures of price dispersion and markup levels. Having both price and quantity data enables me to estimate daily demand in the gasoline market, which was impossible for the previous literature due to the lack of data.

This paper is motivated by two technological advances: a free, real-time gasoline price information service provided by the Korean government, and the introduction and rapid growth of mobile technologies, in particular, smartphones. The price information service reduces consumer search costs, and the existence of smartphones facilitates the use of this price information service by allowing consumers to search for prices while driving. In particular, from the standpoint of the gasoline retail market, the introduction of smartphones is an exogenous technology shock.

To measure the impact of these changes, I analyze the data and find interesting stylized facts: both price dispersion and markup did not decrease, even though smartphone penetration rates increased significantly. Similar trends are found in quantity-weighted measures. This observation is contrary to many previous research results that a search cost reduction leads to higher competition, and hence lower price dispersion and markup. There are two potential reasons: first, it is possible that consumers may not use smartphones to search; second, consumers indeed search more, but this change in consumer search behavior affects gas stations' pricing strategies and changes market outcomes.

Additional descriptive evidence suggests that consumers searched more and became more price-sensitive. I analyze the Opinet website visitor numbers and the Opinet smartphone application download numbers to find that 2-5% of consumers visited the website throughout the period and a constant fraction[41] of smartphone users downloaded the Opinet app. Since the smartphone penetration rate has increased rapidly during the sample period, the Opinet service usage has increased, and it is likely that the ratio of informed consumers has increased. Moreover, a simple regression test suggests that consumers became more sensitive to price differences, in particular, the difference between a price of a station and the minimum price of the region.

I find that search theory models can explain these surprising trends. As indicated by the counterfactual analysis, a model with differentiated products and two consumer types predicts that under a two-type consumer environment (one type more informed than the other) where the ratio of informed consumers is low, price dispersion can increase until the ratio reaches a critical point. The fact that quantity-weighted markup distributions moved toward a bimodal distribution coincides with theoretical search model results: intuitively, it is due to a bifurcation of firm strategies with some setting low prices to attract informed consumers and some setting high prices to serve less informed consumers.

I develop a discrete choice model of consumer demand for spatially differentiated gasoline stations to estimate how consumer choices and the informed consumer ratio change as smartphone penetration rates increase. The structural model results suggest that consumers became more price sensitive as the informed ratio increased from 1.7% to 11.4% during 2010-2012. The counterfactual analysis studies what the new equilibrium prices would be if the proportion of informed consumers were different. In particular, when the informed ratio moves from 1% to 10%, price dispersion, unweighted markups, and quantity weighted markup levels are expected to increase 0.57%, 0.33%, and 0.09%, respectively. These estimates are consistent with the reduced-form section results and the actual data: both price dispersion and markup levels increase slightly as the informed ratio increases, and the magnitude of the increases fit the observed patterns.

[41] About 1.5% until Dec 2010, and 7% after 2010.

References

[1] Jenny Aker. Does digital divide or provide? the impact of cell phones on grain markets in niger. *Center for Global Development Working Paper*, (154), 2008.

[2] John M Barron, Beck A Taylor, and John R Umbeck. Number of sellers, average prices, and price dispersion. *International Journal of Industrial Organization*, 22(8):1041–1066, 2004.

[3] Michael R Baye, John Morgan, and Patrick Scholten. Price dispersion in the small and in the large: Evidence from an internet price comparison site. *The Journal of Industrial Economics*, 52(4):463–496, 2004.

[4] Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

[5] Severin Borenstein, A Colin Cameron, and Richard Gilbert. Do gasoline prices respond asymmetrically to crude oil price changes? *The Quarterly Journal of Economics*, 112(1):305–339, 1997.

[6] Jeffrey Robert Brown and Austan Goolsbee. Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of Political Economy*, 110(3):481–507, 2002.

[7] David P Byrne and Nicolas De Roos. Search and stockpiling in retail gasoline markets. *Available at SSRN 2427556*, 2014.

[8] Ambarish Chandra and Mariano Tappata. Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics*, 42(4):681–704, 2011.

[9] Babur De los Santos, Ali Hortacsu, and Matthijs R Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review*, 102(6):2955–2980, 2012.

[10] Peter Diamond. A model of price adjustment. *Journal of Economic Theory*, 3:pp. 156–68, 1971.

[11] Andrew Eckert and Douglas S West. Price uniformity and competition in a retail gasoline market. *Journal of Economic Behavior & Organization*, 56(2):219–237, 2005.

[12] Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452, 2009.

[13] John A Hartigan and PM Hartigan. The dip test of unimodality. *The Annals of Statistics*, pages 70–84, 1985.

[14] Jerry A Hausman. Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*, pages 207–248. University of Chicago Press, 1996.

[15] Han Hong and Matthew Shum. Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275, 2006.

[16] Daniel S Hosken, Robert S McMillan, and Christopher T Taylor. Retail gasoline pricing: What do we know? *International Journal of Industrial Organization*, 26(6):1425–1436, 2008.

[17] Jean-François Houde. Spatial differentiation and vertical mergers in retail markets for gasoline. *The American Economic Review*, 102(5):2147–2182, 2012.

[18] Robert Jensen. The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The Quarterly Journal of Economics*, 122(3):879–924, 2007.

[19] Matthew Lewis. Price dispersion and competition with differentiated sellers. *The Journal of Industrial Economics*, 56(3):654–678, 2008.

[20] Matthew S. Lewis. Asymmetric price adjustment and consumer search: An examination of the retail gasoline market. *Journal of Economics & Management Strategy*, 20(2):409–449, 2011.

[21] Matthew S. Lewis and Howard P. Marvel. When do consumers search? *The Journal of Industrial Economics*, 59(3):457–483, 2011.

[22] Mark D Manuszak and Charles C Moul. 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, 2009.

[23] José Luis Moraga-González and Matthijs R Wildenbeest. Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848, 2008.

[24] Peter Morgan and Richard Manning. Optimal search. *Econometrica: Journal of the Econometric Society*, pages 923–944, 1985.

[25] Aviv Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342, 2001.

[26] Michael D. Noel. Edgeworth price cycles: Evidence from the toronto retail gasoline market. *The Journal of Industrial Economics*, 55(1):69–92, 2007.

[27] Sam Peltzman. Prices rise faster than they fall. *Journal of Political Economy*, 108(3):466–502, 2000.

[28] Dieter Pennerstorfer, Philipp Schmidt-Dengler, Nicolas Schutz, Christoph Weiss, and Biliana Yontcheva. Information and price dispersion: Evidence from retail gasoline. Technical report, 2014.

[29] Steven Salop and Joseph Stiglitz. Bargains and ripoffs: A model of monopolistically competitive price dispersion. *The Review of Economic Studies*, 44(3):pp. 493–510, 1977.

[30] Alan T. Sorensen. Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):pp. 833–850, 2000.

[31] Alan T Sorensen. An empirical model of heterogeneous consumer search for retail prescription drugs. Technical report, National Bureau of Economic Research, 2001.

[32] Dale O. Stahl. Oligopolistic pricing with sequential consumer search. *The American Economic Review*, 79(4):pp. 700–712, 1989.

[33] Dale O. Stahl. Oligopolistic pricing with heterogeneous consumer search. *International Journal of Industrial Organization*, 14(2):243–268, 1996.

[34] George J. Stigler. The economics of information. *Journal of Political Economy*, 69(3):pp. 213–225, 1961.

[35] Hal R Varian. A model of sales. *The American Economic Review*, pages 651–659, 1980.

[36] Jeremy A. Verlinda. Do rockets rise faster and feathers fall slower in an atmosphere of local market power? evidence from the retail gasoline market. *The Journal of Industrial Economics*, 56(3):581–612, 2008.

[37] Matthijs R Wildenbeest. An empirical model of search with vertically differentiated products. *The RAND Journal of Economics*, 42(4):729–757, 2011.

Appendix A. Government Intervention Period

There were two international issues that caused a big jump in Dubai oil prices in March, 2011. The series of protests and demonstrations across Middle East and North Africa caused social unrest. Also, trade sanctions against Iran directly affected oil supply. Following the international price spike, the Korean domestic gasoline prices went up by more than 150 Korean won per liter (about 57 cents per gallon) within a month. Since the price of gasoline plays a important role in a retail price index that people are interested in, the Korean government chose to intervene in the retail gasoline market to stabilize prices, and asked four major gasoline companies to cut gasoline distribution prices (average wholesale prices, or AWP. SKE, a leading gasoline company, announced a price cut of 100 Korean won per liter from 4/7/11 to 7/7/11 and other companies followed. According to an Opinet representative, discounts by three companies (GSC, HDO, and SOL) were reflected in the price data, as they cut the distribution price directly and posted prices went down. However, SKE offered refund bonus points that were equivalent to 100 Korean won per liter discount to customers after their purchases. Thus, posted prices for SKE stations did not reflect the discount. Note that I have not used dollar per gallon metric in this section to emphasize the impact of 100 Korean won per liter differences. Figure 8 presents that there are big gaps between national average prices among gas companies from April 2011 to July 2011.

To test this information, I compared mean prices for SKE gas stations and those of three other major companies. If the different discount methods were the only reason of the spike, we should be able to observe that compared to other periods, SKE average prices are around 100 won (per liter) higher than average prices of other companies from April 2011 to July 2011.[42] For the main analysis, I used to modified prices for SKE stations (retail prices were subtracted by 100 Korean won) from 4/7/11 to 7/7/11. As a robustness check, I also tried using the data without this period, and using the data without SKE stations from 4/7/11 to 7/7/11. The results were similar in all cases.

[42] It is difficult to understand what exactly happened during this period, as many issues such as supply chain networks and political considerations are involved.

Appendix B. $\frac{\partial s_{jt}}{\partial p_{jt}}$ term

I derive an analytical form of $\frac{\partial s_{jt}}{\partial p_{jt}}$. For simplicity, I define $ex_t(i,j) = \exp(d_{ij}\beta_d + X_j\beta_c - \alpha p_{jt} + \xi_{jt})$. Then $\frac{d(ex_t(i,j))}{dp_{jt}} = -\alpha \cdot ex_t(i,j)$. Then

$$s_{jt}^{informed} = \sum_i w_i \frac{ex_t(i,j)}{1+\sum_{k=1}^{J} ex_t(i,k)}, \; s_{jt}^{uninformed} = \sum_i w_i \left(\sum_{l \neq j} r_{i,j,l} \frac{ex_t(i,j)}{1+ex_t(i,j)+ex_t(i,l)} \right)$$

$$\begin{aligned}
\frac{\partial s_{jt}^{informed}}{\partial p_{jt}} &= \sum_i w_i \left(\frac{-\alpha \cdot ex_t(i,j)\left(1+\Sigma_{k\neq j} ex_t(i,k)\right)}{\left(1+\sum_{k=1}^{J} ex_t(i,k)\right)^2} \right) \\
\frac{\partial s_{jt}^{uninformed}}{\partial p_{jt}} &= \sum_i w_i \left(\sum_{l\neq j} r_{i,j,l} \frac{-\alpha \cdot ex_t(i,j)\left(1+ex_t(i,l)\right)}{\left(1+ex_t(i,j)+ex_t(i,l)\right)^2} \right) \\
\frac{\partial s_{jt}}{\partial p_{jt}} &= IR_t \frac{\partial s_{jt}^{informed}}{\partial p_{jt}} + (1-IR_t)\frac{\partial s_{jt}^{uninformed}}{\partial p_{jt}}
\end{aligned}$$

Using $w_i = \frac{1}{n^2}$ (the uniform distribution over the n by n grid), we can simplify further:

$$\frac{\partial s_{jt}}{\partial p_{jt}} = \frac{1}{n^2}\left\{ IR_t \sum_i \left(\frac{-\alpha \cdot ex_t(i,j)\left(1+\Sigma_{k\neq j} ex_t(i,k)\right)}{\left(1+\sum_{k=1}^{J} ex_t(i,k)\right)^2} \right) + (1-IR_t)\sum_i \left(\sum_{l\neq j} r_{i,j,l} \frac{-\alpha \cdot ex_t(i,j)\left(1+ex_t(i,l)\right)}{\left(1+ex_t(i,j)+ex_t(i,l)\right)^2} \right) \right\}$$

$$s_{jt} = \frac{1}{n^2}\left\{ IR_t \sum_i \frac{ex_t(i,j)}{1+\sum_{k=1}^{J} ex_t(i,k)} + (1-IR_t)\sum_i \left(\sum_{l\neq j} r_{i,j,l} \frac{-\alpha \cdot ex_t(i,j)\left(1+ex_t(i,l)\right)}{\left(1+ex_t(i,j)+ex_t(i,l)\right)^2} \right) \right\}$$

Then the profit-maximizing condition becomes:

$$\alpha\left(p_{jt} - mc_{jt}\right)\left\{ IR \sum_i \left(\frac{ex_t(i,j)\left(1+\Sigma_{k\neq j} ex_t(i,k)\right)}{\left(1+\sum_{k=1}^{J} ex_t(i,k)\right)^2} \right) + (1-IR_t)\sum_i \left(\sum_{l\neq j} r_{i,j,l} \frac{ex_t(i,j)\left(1+ex_t(i,l)\right)}{\left(1+ex_t(i,j)+ex_t(i,l)\right)^2} \right) \right\}$$

$$= \left\{ IR_t \sum_i \frac{ex_t(i,j)}{1+\sum_{k=1}^{J} ex_t(i,k)} + (1-IR_t)\sum_i \left(\sum_{l\neq j} r_{i,j,l} \frac{ex_t(i,j)}{1+ex_t(i,j)+ex_t(i,l)} \right) \right\}$$

where d_{ij} are given and $r_{i,j,l}$ are probabilities from section 4.3. α, β_c, β_d, and IR_t (or a_o and a_1) are the model estimates.

Appendix C. Figures and Tables

Figure 1: Smartphone Penetration Rates

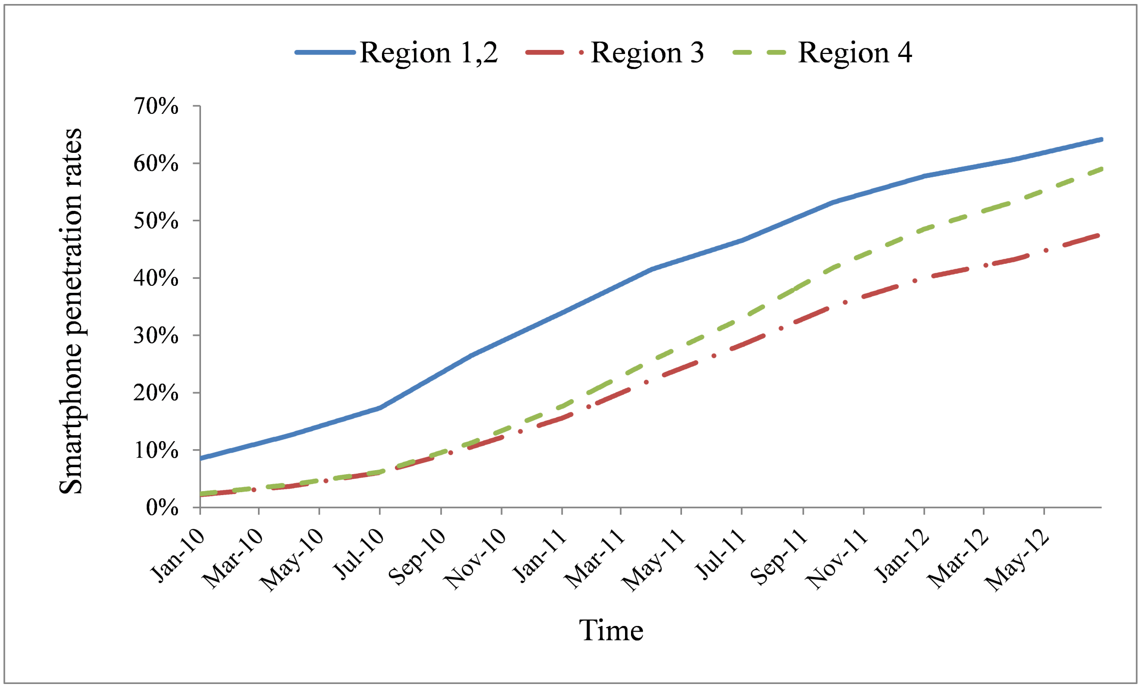


Figure 2: Four Price Dispersion Measures

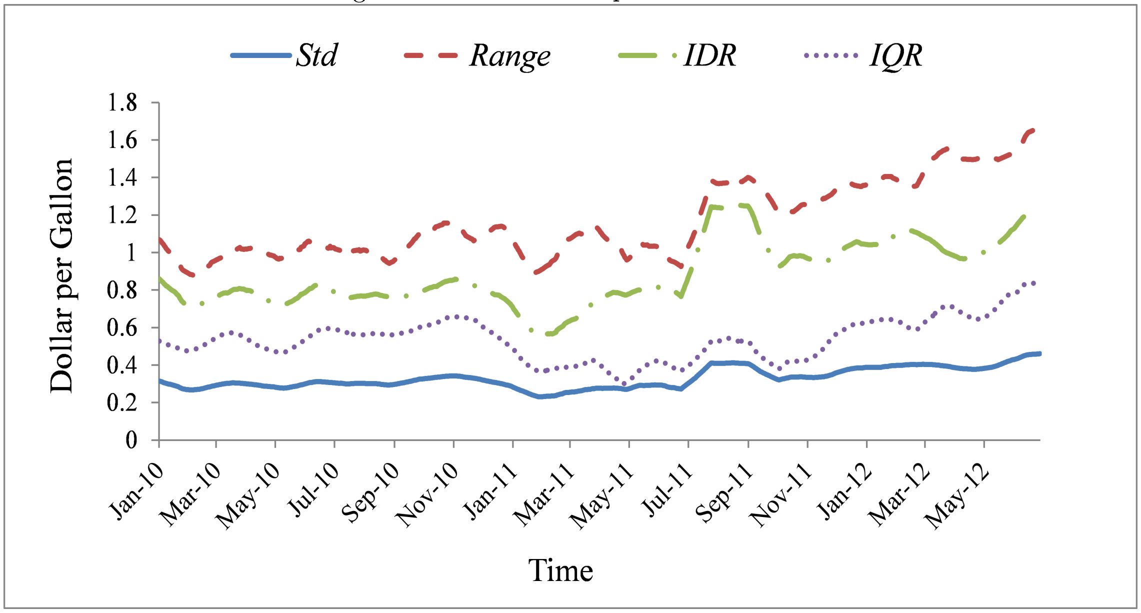


Figure 3: Quantity-weighted and Unweighted *Std*

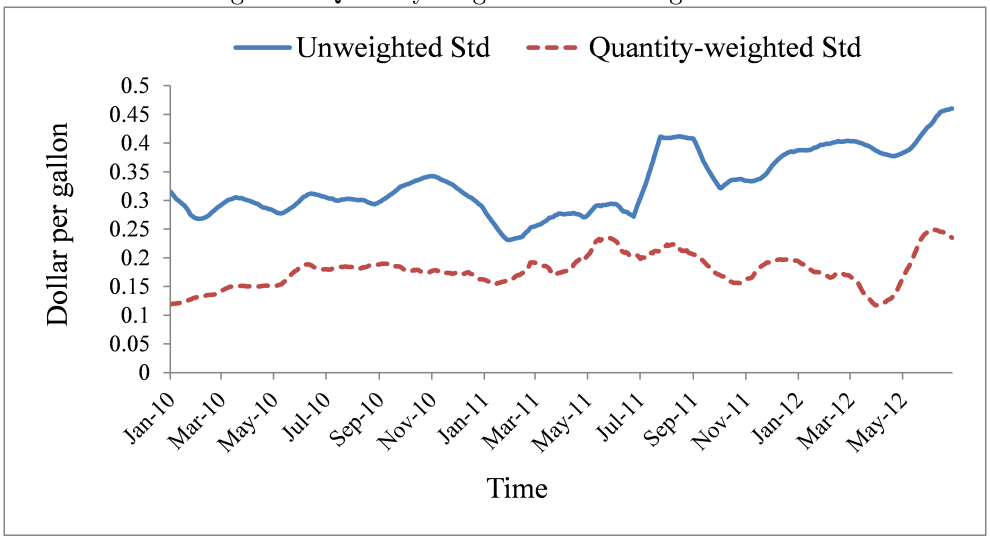


Figure 4: Quantity-weighted and Unweighted Markup Trend

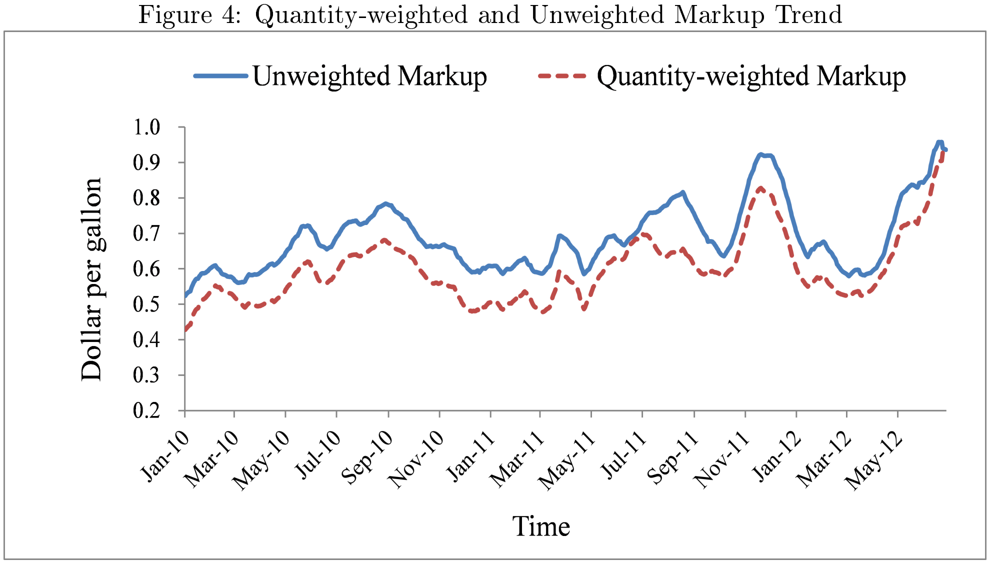


Figure 5: Opinet Website and Application Usage Trends

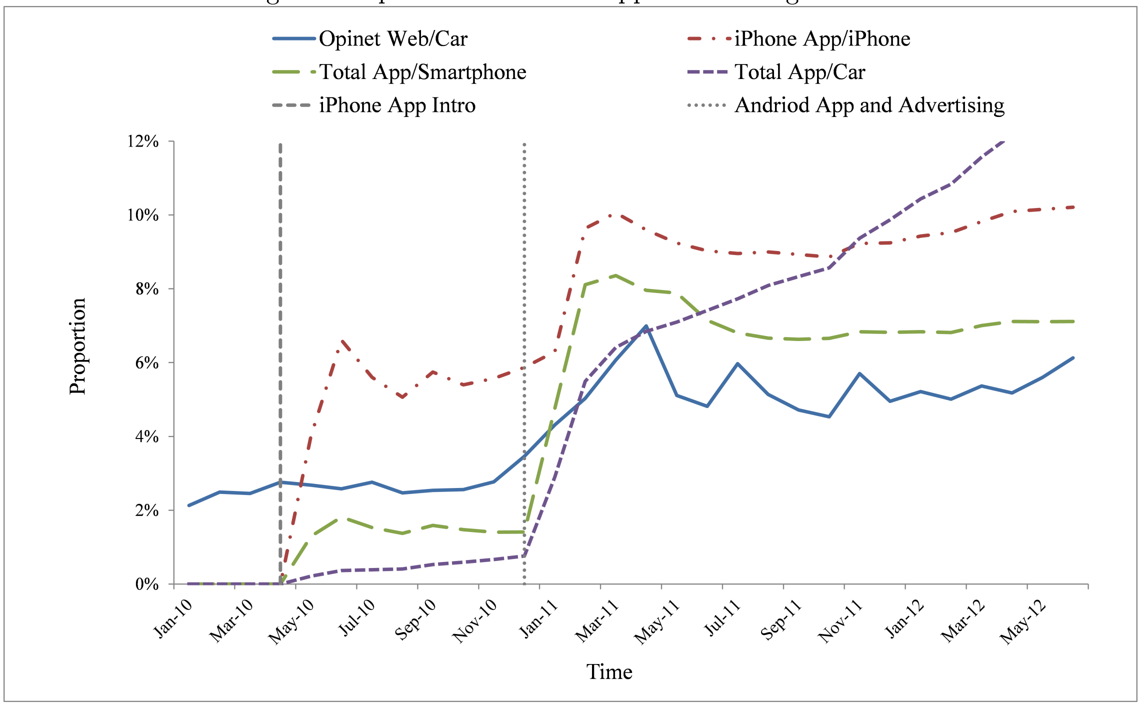


Figure 6: Quantity-weighted Markup Kernel Density (2010 March 1st, 2012 March 1st)

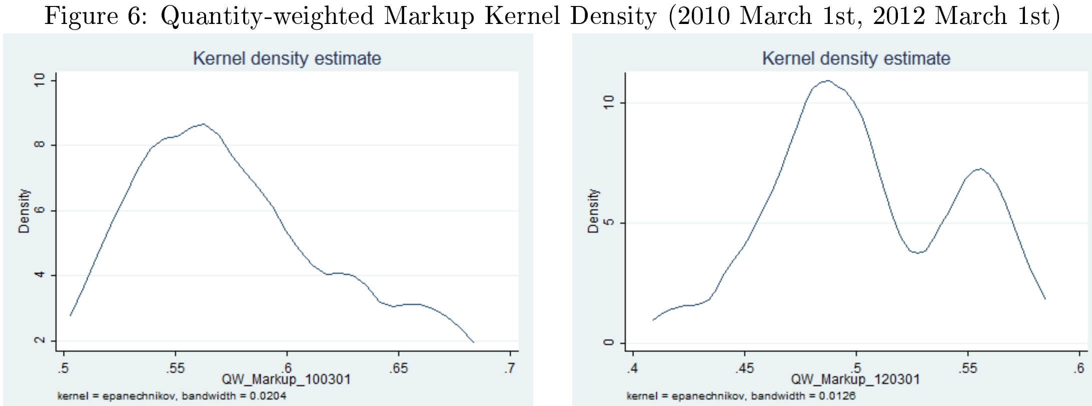


Figure 7: Region 1 (a district of Seoul) Map

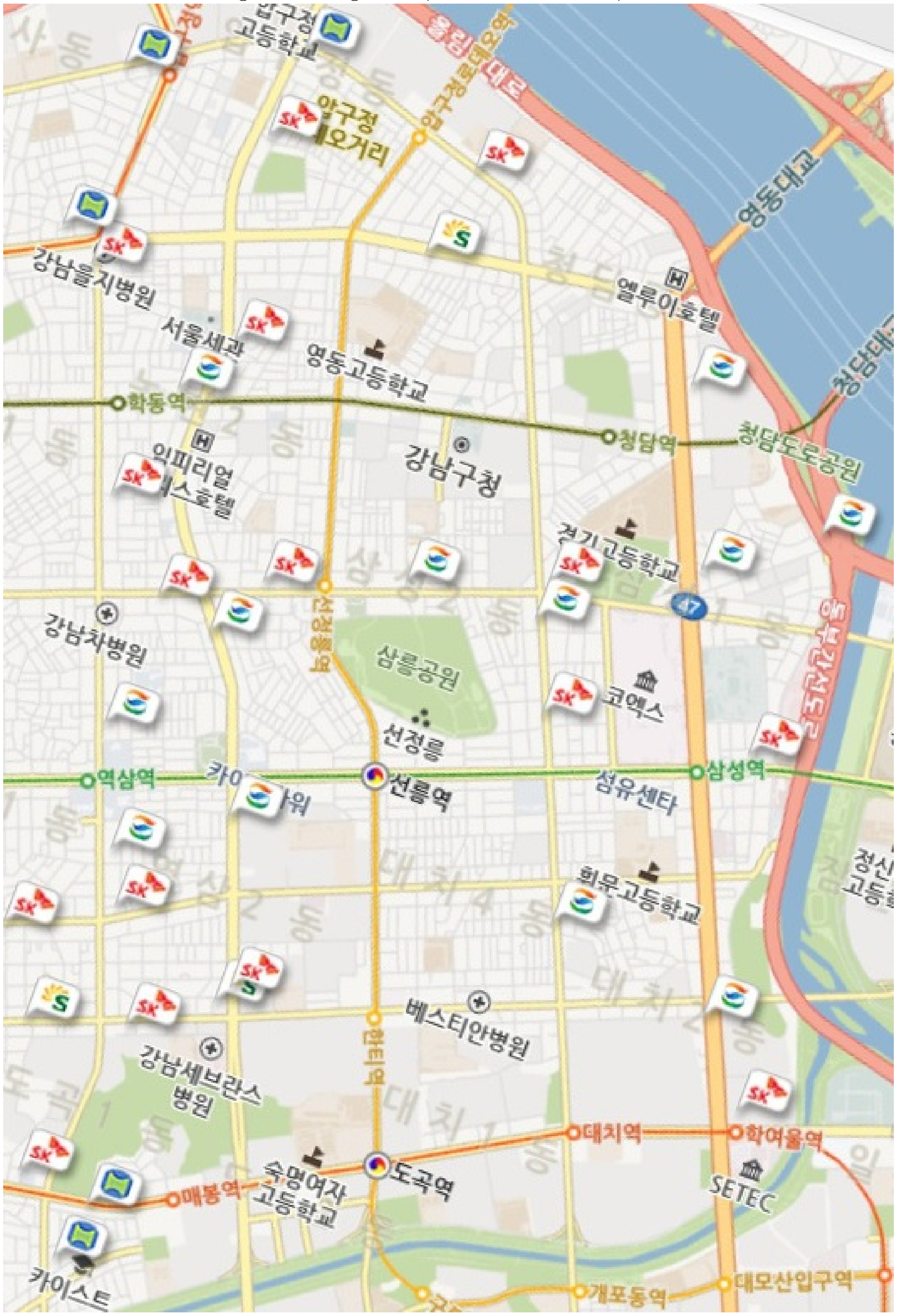
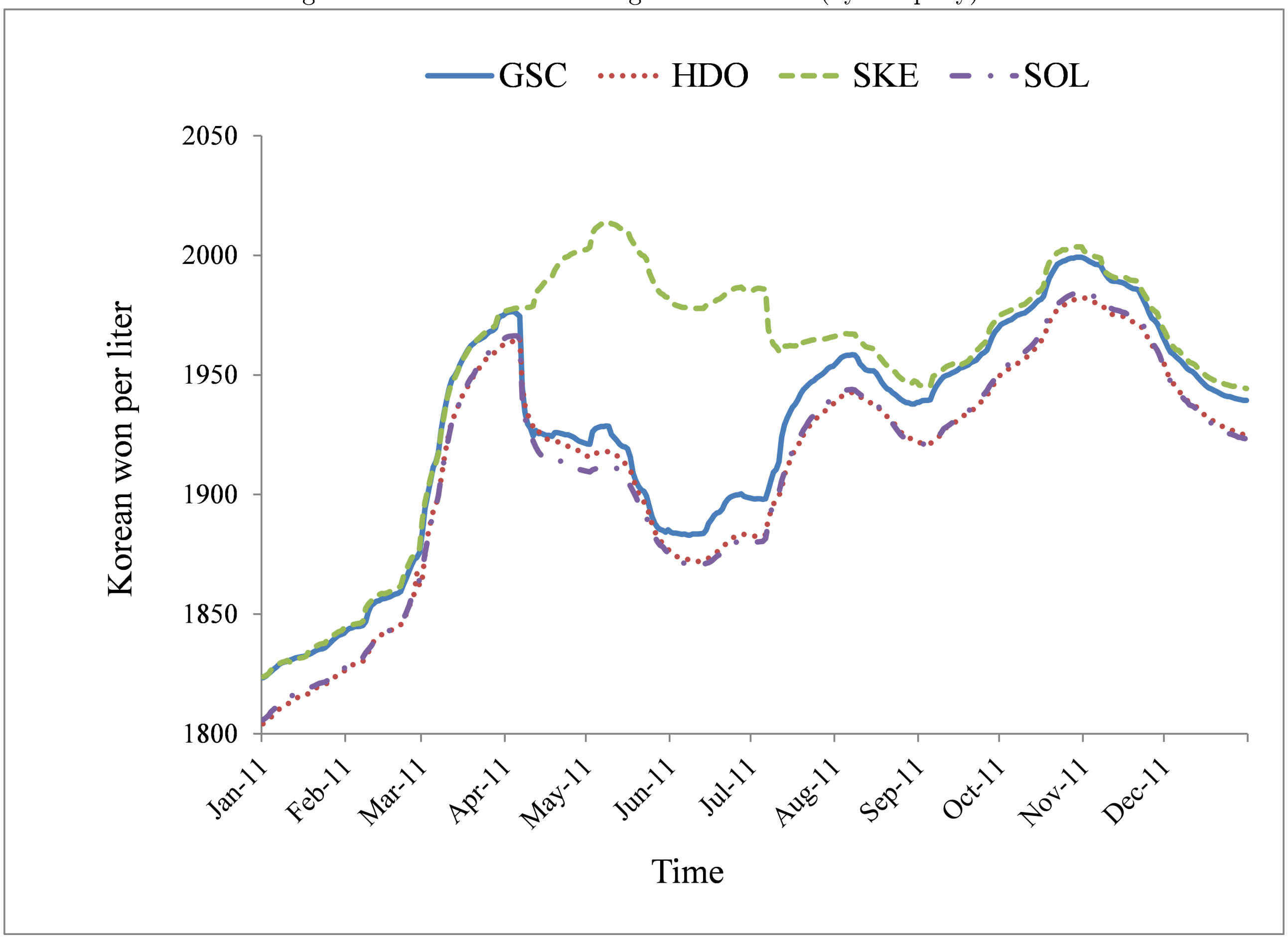


Figure 8: 2011 National Average Retail Prices (by company)



Notes: Korean won per liter metric is used to present the effects of the 100 Korean won price cut.

Table 1: Definition of Variables

Variable	Indexes vary over	Definition
RetailP	j, t	Retail gasoline price (posted price)
AvRetailP	r, t	Daily average retail gasoline price of each region
QwRetailP	r, t	Quantity-weighted daily retail prices of each region
AdjustedP	j, t	Price net of characteristics fixed effects
Mkup	j, t	Markup of station j (*RetailP* - *AWP*)
AvMkup	r, t	Daily average markups of each region
QwMkup	r, t	Quantity-weighted daily markups of each region
AWP	j, t	Average wholesale price
SmartPen	r, t	The ratio of smartphone users to the total population
Self	j	1 if a station offers self-service, 0 otherwise
Carwash	j	1 if a station has a car wash, 0 otherwise
Repair	j	1 if a station has a repair shop, 0 otherwise
Store	j	1 if a station has a convenience store, 0 otherwise
Range	r, t	Maximum price - minimum price
Std	r, t	Standard Deviation
IDR	r, t	Interdecile Range (90% percentile price - 10% percentile price)
IQR	r, t	Interquartile Range (75% percentile price - 25% percentile price)

Notes: r: region, t: market, j: station.

Table 2: Summary Statistics (Variables)

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
RetailP	7.279	0.626	6.014	9.076	150634
AvRetailP	7.279	0.565	6.213	8.473	3636
QwRetailP	7.223	0.580	6.176	8.587	3636
AdjustedP	7.342	0.571	5.963	8.666	150634
Mkup	0.667	0.394	-0.086	2.510	150634
AvMkup	0.667	0.281	0.155	1.574	3636
QwMkup	0.587	0.306	0.120	1.508	3636
AWP	5.943	0.435	5.226	6.780	150634
SmartPen	0.386	0.177	0.086	0.642	3636
Self	0.101	0.302	0	1	150634
Carwash	0.602	0.489	0	1	150634
Repair	0.205	0.404	0	1	150634
Store	0.133	0.339	0	1	150634

Notes: Dollar per gallon metric is used. Region-level variables have 3636 (909 times 4, or the number of days times the number of regions) observations, and station-level variables have 150634 observations.

Table 3: Summary Statistics (Price Dispersion Measures)

Unweighted				
Variable	Mean	Std. Dev.	Min.	Max.
Range	1.170	0.341	0.837	1.623
Std	0.329	0.057	0.226	0.496
IDR	0.882	0.182	0.506	1.283
IQR	0.536	0.123	0.240	0.935

Quantity-weighted				
Variable	Mean	Std. Dev.	Min.	Max.
Range	1.170	0.341	0.837	1.623
Std	0.178	0.034	0.098	0.383
IDR	0.447	0.106	0.193	0.992
IQR	0.245	0.096	0.053	0.799

Table 4: Price Dispersion Regression Result (*Std*)

	Unweighted		Quantity-weighted	
	(1)	(2)	(3)	(4)
Dependent var.	*Std*	*Std*	*Std*	*Std*
AWP	-0.044**	-0.043**	-0.005	-0.005
	(-8.78)	(-8.76)	(-1.09)	(-1.21)
SmartPen	0.252**	0.213**	0.195*	0.167*
	(19.47)	(7.93)	(1.76)	(1.54)
time trend	No	Yes	No	Yes
R^2	0.917	0.921	0.701	0.832

Notes: the number of observations is 3636. t-statistics are in parentheses. Month of the Year dummies and region dummies are included. $*$: Significant at the 10% level. $**$: Significant at the 5% level.

Table 5: Four Price Dispersion Measures Regression Result

	Unweighted				Quantity-weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent var.	*Std*	*Range*	*IDR*	*IQR*	*Std*	*Range*	*IDR*	*IQR*
AWP	-0.044**	-0.122**	-0.135**	-0.161**	-0.005	0.085**	0.043**	-0.065**
	(-8.78)	(-5.96)	(-7.60)	(-11.10)	(-1.09)	(2.99)	(3.15)	(-5.35)
SmartPen	0.252**	1.006**	0.689**	0.454**	0.195*	0.017	-0.167**	0.156**
	(19.47)	(19.00)	(15.30)	(12.00)	(1.76)	(0.23)	(-4.74)	(4.79)
R^2	0.917	0.905	0.829	0.816	0.701	0.798	0.491	0.387

Notes: the number of observations is 3636. t-statistics are in parentheses. Month of the Year dummies and region dummies are included. $*$: Significant at the 10% level. $**$: Significant at the 5% level.

Table 6: Markup Regression Result

	(1)	(2)
Dependent var.	*Mkup*	*Mkup*
AWP	-0.392**	-0.397**
	(-29.54)	(-35.34)
$SmartPen$	0.344**	0.266**
	(22.21)	(6.00)
Self	-0.038**	-0.038**
	(-3.14)	(-3.13)
$Carwash$	0.005	0.005
	(0.41)	(0.41)
$Repair$	-0.023	-0.023
	(-1.44)	(-1.44)
$Store$	0.036**	0.036**
	(3.02)	(3.02)
time trend	No	Yes
R^2	0.561	0.647

Notes: the number of observations is 150634. t-statistics are in parentheses. Month of the Year dummies and region dummies are included. $**$: Significant at the 5% level.

Table 7: A Simple Demand Regression Result

	2010		2012	
	(1)	(2)	(3)	(4)
Dependent var.	$ln\,(Q)$	$ln\,(Q)$	$ln\,(Q)$	$ln\,(Q)$
p_{jt}	-0.179**	-0.173*	-0.206**	-0.159**
	(-2.482)	(-1.734)	(-3.301)	(-2.615)
$p_{jt} - p_{min,t}$		-0.277		-1.246**
		(-0.875)		(-8.468)
R^2	0.542	0.681	0.562	0.762

Notes: the number of observations is 13476 for columns 1 and 2, and 7464 for columns 3 and 4. t-statistics are in parentheses. Month of the Year dummies and region dummies are included. $*$: Significant at the 10% level. $**$: Significant at the 5% level.

Table 8: Structural Model Estimates

	(1) No Android App Jan 2010 - Dec 2010	(2) Both Apps Jan 2011- Jun 2012	(3) All Regions 2010-2012	(4) Region 1 and 2 only 2010-2012
α	25.251**	29.238**	26.131**	22.709**
	(3.922)	(2.053)	(2.298)	(2.474)
β_d	-6.199**	-8.481**	-6.715**	-8.955**
	(1.252)	(0.528)	(0.534)	(1.201)
$-\frac{0.1\alpha}{\beta_d}$	0.407	0.345	0.389	0.254
a_0	0.014*	0.031**	0.018**	0.023**
	(0.008)	(0.003)	(0.005)	(0.004)
a_1	0.035**	0.165**	0.093**	0.121**
	(0.011)	(0.014)	(0.006)	(0.013)
β_{Self}	-4.196**	-4.297**	-3.900**	-4.696**
	(0.365)	(0.464)	(0.397)	(0.303)
$\beta_{Carwash}$	3.695**	3.394**	3.503**	4.101**
	(1.182)	(0.401)	(0.615)	(0.825)
β_{Repair}	-0.123	0.438*	0.516	-0.468
	(0.180)	(0.261)	(0.655)	(0.474)
β_{Store}	3.704**	2.300*	2.504**	1.905
	(0.758)	(1.197)	(0.923)	(1.187)

Notes: Station characteristics coefficients are estimates from a minimum-distance procedure. Other parameters are GMM estimates. Standard errors are given in parentheses. $*$: Significant at the 10% level. $**$: Significant at the 5% level.

Table 9: Counterfactual Results

IR	1%	5%	10%	15%	20%	25%	30%	35%
Std	-0.57%	-0.23%	0	0.13%	0.25%	0.34%	0.41%	0.43%
Mkup	-0.33%	-0.15%	0	0.10%	0.17%	0.21%	0.22%	0.20%
QwMkup	-0.09%	-0.02%	0	0.04%	0.06%	0.05%	0.03%	0.01%