Search Cost, Smartphones, and Consumer Choice: Evidence from Korean Gasoline Market

Youngjun Jang

This Version October 7, 2013

The Korean government has been publishing daily prices of all gasoline stations on a publicly available website since 2008. Combined with the rapid increase of the smartphone user population after 2009, this price information service has changed the consumer search environment significantly. I study the effects of these technology advances on price and markup dispersion among gasoline stations. Using daily price data of three regions from January 2010 to June 2012, I find that price and markup dispersion among gasoline stations tend to increase when the smartphone penetration rate increases. I utilize structural estimations of the two-type consumer search model to find that the proportion of searchers among consumers increases as the smartphone penetration rate increases. These findings suggest that the increase of smartphone users not only changed how consumers search, but also affected how gasoline stations compete and set prices. Empirical markup distributions provide some evidence that is consistent with the bimodal distribution result of the Stahl model.
1 Introduction

While “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 economists have discussed this topic. In particular, previous literature notes that consumer search cost is one of the main reasons that the law of one price does not hold and a certain degree of price dispersion exists in a market. Theoretically, if there is a change in the search cost, consumer search behavior and the expected amount of price information consumers have will change and that will affect the supply side and hence 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 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 still existed. 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 smartphones reintroduces the topic that 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 smartphone usages on price dispersion, markups, and consumer search behavior.

The Korean gasoline retail market after 2008 is a great place to study the effects of search cost reduction for three reasons. First, gasoline is a fairly homogenous good that can be regarded as a single good if one only considers regular gasoline (not premium). Second, the Korean government has published free real-time gasoline price 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 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 had very limited price information and did not search at all before this price information service. It was nearly impossible before to learn prices unless a consumer checks one by one with phone calls or the actual visits and such checking costs would be much bigger than a potential gain from finding
Lastly, a rapid increase in smartphone penetration allows me 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 when they are on the road and are running out on gas, instead of checking the prices before driving. Smartphones significantly change the situation, as drivers can utilize the price information service whenever they want to. Moreover, there are enough variations in smartphone user population to estimate the effects. Smartphone penetration rate was below 5% before 2010 and reached 50% during 2012 in Korea.

Many previous papers (see e.g. Stigler, 1961; Stahl, 1989; Sorensen, 2000; Brown and Goolsbee, 2002; Baye et al., 2004) point out that the prices of homogenous goods are quite dispersed and this price dispersion can be an equilibrium when there are some consumers who observe several prices (searchers, or shoppers in the Stahl model) while other consumers (non-searchers) learn only one price. Among search theory models, the Stahl model is the most relevant model for this paper. The model assumes two types of consumers and finds that while there is no pure strategy Nash Equilibrium, there exists a symmetric mixed Nash Equilibrium where firms choose prices from an atom-less distribution. In particular, it predicts a bimodal distribution of prices as firms either aim to be the lowest to attract searchers or to be the highest to get maximum profit per customer.

While there are considerable theoretical results, relatively little empirical research has focused on measuring search costs and search cost distribution in practice. 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 observed price dispersion is explained by supermarket heterogeneity rather than search frictions. Another empirical paper that discusses price dispersion in the real world is about S&P 500 index funds sector. Hortacsu and Syverson (2004) utilize a structural estimation of search-over-differentiated-products model and find that while investor search costs can explain the considerable price (fee) dispersion in the index funds sector, product differentiation is necessary to support the observed level of price dispersion. Compared to these papers, this article has two advantages. First, with the help of the quantity data in addition to the price information, I can estimate search cost distribution more precisely. Second, unlike supermarkets or index funds, gasoline is a fairly homogeneous good.

To explain how consumers search without the price information service, a consumer search
behavior assumption plays an important role in the model setup. When there is a delay between the search decision and the search outcome, Morgan and Manning (1985) have shown that fixed sample size search is typically better than the sequential search. Santos, Hortacsu, and Wildenbeest (2013) argue that fixed sample size search models provides better explanation of observed consumer search behavior of online book stores environment than sequential search models. While I do not compare 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.

Another required assumption is about the consumer distribution. I assume that there are two consumer groups, searchers and non-searchers. While searchers know all the necessary information including prices to make decisions, non-searchers have very limited information. This two-type consumer distribution assumption reflects this particular situation where a consumer can get all gasoline price information with a single search. 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’ price, or not at all. Moreover, this two-type assumption is already applied in other empirical settings. One example is an online computer memory chips market where Moraga-Gonzalez and Wildenbeest (2007) find that the consumer population can be split into two groups which either have high search cost or low search cost.

While there is no previous research that examines the effects of smartphone introduction on price dispersion, there are two empirical results about the introduction of mobile phones in developing countries. Jensen (2000) studies fish prices in several Indian towns and find that the adoption of mobile phones by fishermen and wholesalers led to a significant reduction in price dispersion and increase in social welfare. Aker (2010) 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 impact of mobile phones and that of smartphones are quite different. A mobile phone search, or making a phone call, provides one price information 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.
The contribution of this paper are twofold. First, 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. As smartphones combined with the real-time price information service directly changes consumer search situation and the following decision, this paper helps us measure the effects of price information and how smartphones enhance them. As my empirical setting allows a two-type consumer search environment and two-type search costs, estimating search cost distribution is simplified to find a proportion of searchers and this assists to understand the change of search cost distribution easily. Second, I construct and estimate a consumer choice model, which incorporates a distance term that has different values for consumers at different locations to understand consumer behavior in retail gasoline market. The existence of the distance term allows that searchers do not necessarily go to the station with the minimum price, which helps the model to reflect the observed consumer decisions better. Moreover, the model estimates how much an average consumer is willing to pay more to avoid driving additional distance for gas.

The structure of this article is as follows. In the next section I provide details about several data sets and discuss their merits and limitations. In section 3, I present summary statistics for the four price dispersion measures and 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 possible counterfactual analysis are reported in Section 5. Section 6 concludes.

2 Data

I have gathered data from three main sources for this paper. I describe those sources below as well as the structure of the various data sets.

2.1 Opinet Price Data

Opinet price data offer daily, individual station level gasoline price information and station characteristics. 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. I was able to obtain a complete set of historic prices for three regions from January 2010 to June 2012 (909 days). Two
regions are two districts of Seoul and the other region is a city that is isolated by mountains. The streets are laid out in a grid pattern for all three regions, especially for the two districts of Seoul. The number of stations in those regions are 42, 44, and 63.

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 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. 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, updating six times a day basically provides real-time information. It helps me to ensure that consumers receive the correct price information when they use the Opinet price service. In other words, I can eliminate the possibility of “cheats” that stations may report false prices to lure customers.

Consumers can access the price data via its website www.opinet.co.kr or other methods, such as car navigation systems or Opinet smartphone applications. 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. 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.

2.2 Average Wholesale Price

In this subsection, I explain how to construct approximated values of the average wholesale prices that individual gasoline stations pay. Since most, if not all, of Korean oil imports come from Arab 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. This price, an average distribution price from oil companies to retailers, is known and this is the best approximate average wholesale price that retailers pay in the market. One possible limitation is that this average distribution price cannot capture which oil company charges more or less compared to others. Unfortunately, there is no available data to estimate such differences since oil companies do not reveal distribution prices which are classified. The only reliable source
of these prices is the Opinet website: it offers monthly, national average distribution cost data collected from oil companies.

The main advantage of my data set is that it has daily gasoline retail price for individual gas stations. To fully utilize this advantage, daily cost data for individual gas stations is desired. Since such data is not available, using an interpolation method, I construct an approximate daily cost that retail gas stations pay by combining monthly national average distribution cost data with daily Dubai future prices.

Suppose we know average distribution prices at time $t_1$ and $t_2$. Since we know daily Dubai price $D_t$ and all taxes, we can compute daily after-tax Dubai price $b_t$ and daily oil company distribution price $d_t$ as

$$d_t = b_t + C + \alpha = \{D_t \times (1 + \tau_i) + \tau_t \times 1.41 + \tau_{sales}\} \times (1 + \tau_{sur}) + C + \alpha$$

where $\tau_i =$ import tax = 3%, $\tau_t =$ traffic tax = 529 (won), $\tau_{sales} =$sales tax = 36 (won), $\tau_{sur} =$ surtax = 10%, $\alpha =$oil company markup, and $C =$ a sum of other costs such as transportation costs and administrative costs. While the exact amount of other costs such as transportation costs are unknown, these costs are regarded as fixed lump-sum cost in general. This additional cost term will be estimated in the structural model.

In addition, according to an industry insider, markups are relatively stable and are usually constants. To confirm this, I check price differences between daily tax-adjusted Dubai future prices $b_t$ and average retailer prices. As expected, I find that they follow similar trends and price differences are relatively stable (Figure 1).

In summary, we can interpolate daily average oil company distribution prices as

$$\hat{p}_t = p_{t_1} + (p_{t_2} - p_{t_1}) \frac{b_t - b_1}{b_2 - b_1}$$

From now on, I call this $\hat{p}_t$ as AWP, or average wholesale price. These average wholesale prices are used as marginal costs in the rest of the paper. Admittedly, applying the same marginal cost for all stations in all regions for a given day is far from ideal. However, there are good reasons to approximate marginal costs with the simple AWP term. Details will be discussed in the structural estimation section.
2.3 Smartphone Penetration Rates and Quantity Data

Quantity data such as the number of smartphone sold for each region and the quantity of gasoline sold for each station are very difficult to gather as companies try to keep the data secret. I was fortunate to contact benevolent information providers who agreed to share quantity data of one major company in each sector.

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 rate, I start from the number of quarterly, regional smartphone users for one telecommunication company. This company is one of the three major telecommunication companies and its market share has been 30-35% during 2009-2012. Multiplying the number of users for this company by three, I estimate total number of smartphone users of the region for each quarter. Finally, using linear interpolation method as in the previous section, I calculate daily smartphone penetration rates. Smartphones have become increasingly popular in the Korean market since late 2009. Before the introduction of IPhone 3G at 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 2 shows smartphone penetration rates of three regions, from 2010 1st quarter to 2012 3rd quarter. Note that region 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.

For quantity data, I have daily credit card transaction numbers and total credit card transaction amounts of individual gas stations of one major gas company that has about 25% market share. 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, both using daily transaction numbers and using daily transaction amounts naturally deliver very similar results. I present the results from using daily transaction numbers for the main analysis.

3 Descriptive Evidence on Price Dispersion and Markup

In Section 3, I provide definitions, summary statistics, and trend graphs of the main variables. In addition, I discuss how I treated an unusual three months period in 2011 when the Korean
government intervened in gasoline prices for political reasons. Finally, I present reduced from
evidence on price dispersion.

3.1 Definitions and Summary Statistics

Table 1 defines the variable used in the analysis. For my initial descriptive analysis, 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. First four variables (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 study interdecile range and interquartile range
to investigate the characteristics of price dispersion in depth. The definition of AWP is given in
the section 2.2. Note that it only has one index, $t$, as I do not have company or region specific
average wholesale prices. Lastly, SmartPen denotes the ratio of smartphone users to the country’s
population, or smartphone penetration rate, as described in section 2.3. self, Carwash, Repair, and
Store are station characteristic dummy variables.

Table 2, 3, and 4 present summary statistics for all variables used in this section for the three
regions. There are two possible ways to define daily price dispersion in a given region. Given
that there are multiple retail outlets associated with each particular company, sometimes in close
proximity, one could imagine calculating dispersion in a geographical area within company or overall.
Each could be interesting in its own right, but here I present overall dispersion by computing daily
price dispersion among gas stations in one region, ignoring gas company information. In this case,
the number of observation is 909, as there are 909 days in the data period. I changed the metric
from Korean won per liter to US dollar per gallon using the following conversion rates: 1 gallon =
3.785 liter, 1 dollar = 1000 won for the approximation.

The left side of the Table 2 presents unweighted case and the right side presents descriptive
statistics of the four dispersion measure that are computed with quantity weighted data. For
example, suppose that station 1 makes 10 sales at price $p_1$ and station 2 makes 20 sales at price $p_2$.
The previous case considers that there are two entries (ignore the number of sales), $p_1$ and $p_2$. For
quantity weighted case, I consider that there are 10 entries of $p_1$ and 20 entries of $p_2$. Note that
dispersion measures have smaller values with the quantity weighted case.
3.2 Government Intervention Period

There were two international issues that caused a big jump in Dubai oil prices at March, 2011. The series of protests and demonstrations across Middle East and North Africa caused social unrest. Also, trade sanctions against Iran directly impacted oil supply. Following the international price spike, the Korean domestic gasoline prices went up more than 150 won per liter (about 57 cents per gallon) within a month. Since the price of gasoline plays a important role in the retail price index that people are interested in, the Korean government chose to intervene in the gasoline market to stabilize prices and asked four major gasoline companies to reduce the gasoline distribution price. SKE, a leading gasoline company, announced a price cut of 100 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, SOL) are reflected in the price data as they cut the distribution price directly. However, SKE offered refund bonus points that are equivalent to 100 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 understand the impact of 100 won per liter differences. Figure 3 presents that there are big gaps between 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 for other companies for April 2011 to July 2011. However, Figure 3 shows something different. While it is true that SKE prices are relatively higher than average prices of other companies, the gap sizes are much smaller than 100 won. Also, SOL prices are much lower during the April 2011 to July 2011 period.

It is almost impossible to understand what exactly happened during this period as many issues such as supply chain networks and political considerations are involved. For the purpose of this paper, it would be enough to eliminate outliers and compute price dispersion measures. From the graph, I decide to exclude SKE and SOL stations from 4/7/11 to 7/7/11. I also ran the same regressions using the data without this period and got similar results.
3.3 Markup and Price Dispersion Trends

I present markup and price dispersion trend graphs and a brief interpretation of the graphs in this section. Figure 4 shows that markups have been fairly stable until mid 2011 and tend to increase after mid 2011. While there are constant gaps between regions, the general shape of the graph is the same for all regions. Quantity weighted markups in Figure 5 show the same trend, except that a gap between region 2 and 3 disappears.

Figure 6 presents a graph of standard deviation of prices. The dispersion is fairly persistent and tends to slightly increase at the end of the period. Quantity weighted version in Figure 7 delivers the same message. We can observe that the average level of standard deviation is lower in the quantity weighted case. Figure 8 and 9 confirm that other dispersion measures present similar trends for both unweighted and quantity weighted Std. Figure 8 shows the four price dispersion measures of region 1 and Figure 9 depicts the quantity weighted version. Note that the average level of dispersion is lower in the quantity weighted cases again. This fact implies that some of the extreme prices have very small quantity sold and hence have smaller portions for the quantity weighted cases. After examining the data, I confirm that some high price stations tend to make fewer number of sales.

3.4 Regression Results

Since this paper examines impacts of reduced consumer search costs on gasoline price dispersion and markup, regressions should address how markup and various measures of price dispersion change as smartphone usage changes. 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

\[ Y_{rt} = \beta_0 + \beta_1 AWP_t + \beta_2 SmartPen_{rt} + \sum DM_j + \sum DR_k + \epsilon_{rct} \]

A dependent variable, \( Y \), is one of the four dispersion measures. \( SmartPen_{rt} \) is smartphone penetration at time (date) \( t \) for region \( r \). \( DM_j \) are month of the year dummies and \( DR_k \) are region dummy variables. Results from the unweighted data are presented at the column (1)-(4) of Table 5 and results from the quantity weighted case are presented at the column (5)-(8) of Table 5. Both
monthly dummies and region dummies are included in all regressions.

For the unweighted data, all coefficients are highly significant. Negative signs for the AWP coefficients mean that when marginal costs are higher (hence price levels are higher) price dispersion is smaller. This finding is consistent with an empirical observation (Lewis and Marvel, 2011) that consumers tend to be more cautious about the prices and search more when the prices are higher. For example, the results from regression (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).

Positive SmartPen coefficients imply that dispersion is higher when there are more smartphone users. The results from regression (1) suggests that the standard deviation decreases by 0.252 cents per gallon when the smartphone penetration rate increases by 1%. 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 with smartphones than the cases with mobile phones.

One possible explanation is a polarization of stations that is suggested by the Stahl model: 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 out of these consumers and ignore 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 graph. Since I use the same marginal cost (daily average wholesale cost) for all stations for a given day, shapes of price graphs and markups graphs are basically the same. The first graph in the Figure 10 shows a kernel density of markups for gasoline stations in region 1 during the March, 2010 and the second graph shows the one during the March, 2012. While this trend of moving toward “double-peak shape” is observed in the most parts of the data, it is difficult to quantify a degree of “double-peak” shape. Hartigans Dip test (1985) confirms that the markups are not unimodal, but it does not test explicitly for bimodality but for unimodality.

For the quantity weighted case, several coefficients are not statistically significant and signs are different for different dispersion measures. It suggests that the effects of the explanatory variables (AWP and SmartPen) need to be examined more carefully and motivates structural estimation approaches.
Table 6 shows the regression results for IQR, the interquartile range dispersion measure with and without dummy variables and confirms that monthly or regional dummies do not affect the results of main parameters. The coefficients of the main parameters have the same sign for all cases and t-statistics are actually higher when dummy variables are excluded from the regression. The table presents the IQR case and the same tendency is observed in other dispersion measure cases.

My assumption is that smartphone penetration rates are exogenous with respect to gasoline price dispersion or markups of the stations. It is unlikely that people buy smartphones because gasoline price dispersion is low or high. The same argument can be used for the markup case. In fact, most consumers and smartphone buyers would not even be aware of gasoline price dispersion or gasoline station markups when they make gasoline purchases.

Since the regression results of dispersion measures encourage further study on firms’ pricing decisions, I start with the following regression:

\[
Price_{rtj} = \beta_0 + \beta_1 AWP_t + \beta_2 SmartPen_{rt} + \sum_{i=1}^{3} DC_i + \sum_{j=1}^{11} DM_j + \sum_{k=1}^{3} DR_k + \sum_{k=1}^{K} DChar_{kj} + \epsilon_{rtj}
\]

\(DChar_{kj}\) are dummy variables that represent characteristics of station \(j\). For example, if a 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 7 shows the results of the main regression and column (2)-(4) show the result of the regression without one type of dummy variables. Since all columns present very similar results, it is enough to focus on the column (1). All coefficients are highly significant as the number of observation is huge. AWP and all characteristics variables except \(Repair\) show the expected signs: higher prices are expected when the marginal cost is higher, when a stations offers full service, and when a station has amenities like car wash, repair shop, and convenience store. As price minus AWP is markup and the AWP coefficient is less than one, higher AWP implies lower markup, according to the regression result.

The \(SmartPen\) variable has a positive coefficient and implies that a price is 1.66 cents (per gallon) higher when a smartphone penetration rate is 1% higher. Combined with the previous dispersion regression results, the reduced form analysis suggests that higher smartphone penetration rates lead to higher price level and higher price dispersion. These results could be counter-intuitive as it might
be natural to assume that higher smartphone penetration rates would lead to more consumer search and more price information would make stations compete more and hence price dispersion would decrease. There are three potential explanations. First one is a transformation of the competitive landscape which lead to a bimodal distribution of prices among stations which is consistent with the Stahl model. Another point is that the regression model does not consider the distance between consumers and gas stations that certainly affect consumer choices and pricing decisions. Lastly, it is uncertain that higher smartphone penetration rates lead to higher search activities. In the next section, I will incorporate the distance term in the model setup and estimate the proportion of searchers with structural models.

4 Structural Model

Since the main result of the regression that price dispersion is actually increasing in smartphone penetration rates motivates further research, I develop a structural model of consumer choice to estimate how the proportion of searchers changes over time and how the distance term affects consumer decisions. I assume that there are two-type of consumers (searchers and non-searchers) and define the searcher ratio as the proportion of searchers in the market. Also, I need to specify how searchers and non-searchers make decisions and where consumers are located. I start this section with the consumer location assumption (grid assumption) and explain two types of consumers, searchers and non-searchers. Combining expected market shares from both types, I compare expected quantity sold with the actual quantity sold to find out the parameter values that minimize the objective function.

4.1 Consumers on the Grid

To build a structural model of consumer search and purchase behavior, it is important to consider and incorporate the likely factors that drive consumer choice in this market. There are three main factors: price, distance, and station amenities, or characteristics. Among these three factors, a distance to a station has an unique nature that it is different for everyone. Unlike other factors such as price or car wash availability that are the same for all consumers, a distance to a certain station is different for consumers from different locations.

To consider how distances would affect consumer choices, I start with the consumer distribution
assumption. Imagine a rectangle that covers the whole region. Divide it to \( n - 1 \) by \( n - 1 \) rectangles so that we have an \( n \) by \( n \) grid. I assume that consumers are located at each grid point according to a uniform distribution. For example, if \( n = 10 \) then each grid point has 1% of the consumers. For the three regions of interest, I can assume that a square with a side length of 10 miles represent each region.

Since all three regions, in particular two districts of Seoul, have checkerboard-like road designs, the grid assumption reflects the actual road map well. While it would be nice 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 for the roads where stations are located.

4.2 Two Types of Consumers: Searchers

There are two types of consumers, searchers and non-searchers. While searchers know all station information by using Opinet service, non-searchers only have limited information. For example, if one starts to look for a gas station when one runs out of gas and chooses to go one of the first two stations he finds, he is a non-searcher. On the other hand, if one uses her smartphone to check all the stations in the region and makes a decision to visit a certain station, she is a searcher. Since non-searchers do not have information for most gas stations, or “products”, standard BLP setup or 2-type discrete model such as business and leisure travelers for airplanes (Berry and Jia, 2008) cannot be applied.

For searchers, following standard utility assumptions, the utility of consumer \( i \) for station (or product) \( j \) in market \( t \) is given by

\[
 u_{ijt} = d_{ij} \beta + X_j \gamma_i - \alpha p_{jt} + \xi_j + \epsilon_{ijt}
\]

\( d_{ij} \) is a distance from consumer \( i \) (grid point \( i \)) to the station \( j \). \( X_j \) are station characteristics and \( \gamma_i \) represents preferences of consumer \( i \). \( p_{jt} \) is price of station \( j \) at date (market) \( t \). Lastly, \( \xi_j \) are unobserved station characteristics.

Market \( t \) is one day (date \( t \)) of a certain region. For example, Gangnam District, March 2nd, 2011 is one market. If data from January 1, 2010 to June 27, 2012 are chosen, the number of markets is 909. Note that this utility setup is basically the same as the BLP model, except the \( d_{ij} \beta \)
term.

Since the main parameters of interest are $\alpha$ and $\beta$, I use simplified model below:

$$u_{ijt} = d_{ij} \beta - \alpha p_j + \xi_j + \epsilon_{ijt}$$

Now $\xi_j$ denote station fixed effects. Assuming the logit model, the percentage of consumers at $i$ who chooses station $j$ in market $t$ is

$$\frac{\exp(c + d_{ij} \beta - \alpha p_j + \xi_j)}{D_{it}} = \frac{\exp(c + d_{ij} \beta - \alpha p_j + \xi_j)}{\sum_{k=1}^{J} \exp(c + d_{ik} \beta - \alpha p_k + \xi_k)}$$

where $c$ is a constant term to consider the outside option. In this setup, choosing the outside option at market $t$ means that a consumer does not go to gas station at date $t$.

Then the share of consumers at $i$ who make the purchase is $s_{i,t}(d_t, p_t, \xi_t, \theta_d) = \frac{D_{it}}{1 + D_{it}}$. 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

$$s^*_j(d_t, p_t, \xi_t, \theta_d) = \sum_i w_i \frac{\exp(c + d_{ij} \beta - \alpha p_j + \xi_j)}{D_{it}} s_{i,t}(d_t, p_t, \xi_t, \theta_d)$$

4.3 Two Types of Consumers: Non-Searchers

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 easier to im-
plement. Let \( m \) be the number of stations whose information is known to a non-searcher. Choosing \( m = 1 \) would cause a station to set very high price in equilibrium so that it can make huge profit from non-searchers who only know its information and ignores all other consumers. Thus, I choose \( m = 2 \): each non-searcher consumer learns \( m = 2 \) station prices among \( J \) stations in the market. To determine which station information consumer \( i \) gets, I assume that probability of getting price information of the station \( j \) is proportional to the inverse of the distance, \( d_{ij} \).

In other words, \( \Pr (\text{consumer at } i \text{ learns the price of the station } j_1) / \Pr (\text{consumer at } i \text{ learns the price of the station } j_2) = \left( \frac{d_{ij_1}}{d_{ij_2}} \right) \). Solving these equations, I get

\[
p_{i,k} = \Pr (\text{consumer at } i \text{ learns the price of the station } j_k) = \frac{\frac{d_{ij_k}}{\sum s_{ij_s}}}{\sum_{j} \frac{d_{ij_s}}{\sum_s s_{ij_s}}}
\]

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

\[
p_{i,k_1,k_2} = \Pr (\text{consumer at } i \text{ learns the price of the station } j_{k_1} \text{ and } j_{k_2}) = p_{i,k_1}p_{i,k_2} \left( \frac{1}{1 - p_{i,k_1}} + \frac{1}{1 - p_{i,k_2}} \right)
\]

I assume that after learning prices for two stations, a non-searcher would choose stations as a searcher would do. Using the same logit model assumption, market share for station \( j \) among non-searchers is

\[
s_{jt}^{ns} = \sum_i \left( \sum_{i \neq j} \frac{\exp (c + d_{ij} \beta - \alpha p_{jt} + \xi_j)}{1 + \exp (c + d_{ij} \beta - \alpha p_{jt} + \xi_j) + \exp (c + d_{il} \beta - \alpha p_{lt} + \xi_l)} \right)
\]

### 4.4 Demand Side

As I only observe total quantity sold, I need to combine the results from the searchers and non-searchers so that I can compute the total market share and quantity sold to match the observed values and the model-expected ones. Let \( R_t \) denote the proportion of the searchers at market \( t \). To start, I divide the data set into three time periods: 2010, 2011, and 2012. I assume constant searcher ratios within each period and denote the searcher ratios as \( SR \). Intuitively, \( SR \) should
increase as time passes, as the proportion of searchers is bigger when there are more smartphone users. An alternative specification is to allow the searcher ratio to change every day by assuming a linear relationship: \( R_t = a_0 + a_1 \text{SmartPen}_t \). \( a_0 \) represents the fraction of people who are already searchers and \( a_1 \) indicates the proportion of smartphone users who become searchers.

Then the total expected market share from the model is

\[
 s_{jt} = R_t \cdot s_{jt}^s + (1 - R_t) \cdot s_{jt}^{ns}
\]

Since I only have partial quantity data, I do not observe the total quantity in the market and have to estimate it within the model. Moreover, as the searcher ratio should be between zero and one, I put a restriction to change \( R_t \) to zero if \( R_t < 0 \) and to one if \( R_t > 1 \). At the actual estimates, these restrictions should not bind.

As both population size and average income for any region were stable during the data period, I assume that the total quantity for a given region stays the same for the period and let \( Q_0 \) be the total quantity in the market.

Given the expected market shares and the total quantity parameter, I can compute the expected quantity for station \( j \) at market \( t \) as \( \hat{q}_{jt} \). Suppose that quantity data for station 1, 2,..., \( J_0 \) are available. Then the objective function to minimize is

\[
 \text{Obj} = \sum_{j,t} J_0 \sum_{j=1}^{J_0} (q_{jt} - \hat{q}_{jt})^2
\]

### 4.5 Supply Side

In this section, I incorporate the profit maximizing conditions for gas stations. As profit for station \( j \) at market \( t \) is \((p_{jt} - mc_{jt})q_{jt} - F_{jt}\) where \( F \) denotes fixed cost, the optimal behavior for gas stations is to follow the profit maximizing condition:

\[
(p_{jt} - mc_{jt}) \frac{dq_{jt}}{dp_{jt}} + q_{jt} = 0
\]

I seek parameters that would minimize

\[
 \sum_{j,t} \left((p_{jt} - mc_{jt}) \frac{dq_{jt}}{dp_{jt}} + q_{jt}\right)^2
\]
To compute this term, I numerically approximate \( \frac{dq_{jt}}{dp_{jt}} \) term as
\[
\frac{dq_{jt}}{dp_{jt}} \approx q_{jt} - q'_{jt} \epsilon_0
\]
where \( q'_{jt} \) is the expected quantity when \( p_{jt} \) is changed to \( p'_{jt} = p_{jt} + \epsilon_0 \) keeping other prices unchanged (\( \epsilon_0 \) is a very small value.) For \( mc_{jt} \), this term is in fact \( mc_t \), since no station specific cost information is available and the best approximation of the marginal cost term is \( AWP \).

While it would be ideal to have individual station marginal cost information, I believe that using the same marginal cost for all the stations in the region, as marginal costs between stations in the same region are indeed very similar. Marginal cost mainly consists of the transportation cost and the variable labor cost. The transportation cost is based on the shipping distance and the most of the shipping distance happens when gasoline is transported from the city where oil refinery is to the city where the gas stations are. Also, within the same region, labor cost is fairly similar as most gas stations offer the same conditions. Most of the gas station workers are temporary part-time employees and they are paid minimum hourly wage or slightly above the minimum.

Combine the demand side and the supply side, I minimize the following objective function:

\[
\sum_t \sum_{j=1}^{J_0} (q_{jt} - \hat{q}_{jt})^2 + \sum_{j,t} \left( (p_{jt} - mc_{jt}) \frac{dq_{jt}}{dp_{jt}} + q_{jt} \right)^2
\]

The main parameters of interest are \( \alpha, \beta, \) and \( R_t \) (or \( a_0 \) and \( a_1 \)).

### 4.6 Station Fixed Effects

For the computation issues, I used pre-set station fixed effects in the main analysis. To minimize the objective function of the previous section, I need to consider the following parameters: \( \alpha, \beta, R_t \) (or \( a_0 \) and \( a_1 \)), \( c, q_0 \), and station fixed effects \( \xi_j \). While it is ideal to include all those parameters in the optimization routine, including too many variables of interest tend to results in unstable outcomes that are local optima. In this case, since there are more than 40 stations per region, if I include all station fixed effects, more than 40 variables will be added to 6 main variables that I am primarily interested in.

As market shares of the stations are fairly stable during the sample period, using “historical” station fixed effects is a viable option in this case. For the stations with quantity data, I define the historical share of station \( j \), \( \bar{\xi}_j \) as a ratio of total quantity of station \( j \) throughout the period to the
total quantity observed throughout the period:

$$\xi_j = \log \left( \frac{\sum_{t=1}^{T} q_{jt}}{\sum_{j=1}^{J_0} \left( \sum_{t=1}^{T} q_{jt} \right)} \right)$$

For the stations without quantity data, I assign an average value, \( \frac{1}{J_0} \sum_{j=1}^{J_0} \xi_j \), as an approximate value.
5 Structural Estimates

This section presents structural estimation results and provide interpretation of the estimated parameter values. I also briefly discuss basic counterfactual analysis and future research topics. As the main motivation for estimating this structural model was to estimate the effects of the distance term and the proportion of the searchers, the estimates of $\alpha$, $\beta$, $SR$ ($a_0$ and $a_1$ for alternative specification) will be of particular interest.

5.1 Results and Interpretation

Estimation results for region 1 are presented at Table 8. For the estimates that converge to different points depending on the starting points, I choose ones with the minimal objective function value. I plan to apply bootstrapping to derive estimates of standard errors.

Column (1), (3), (5), and (7) assume varying searcher ratios ($R_t = a_0 + a_1SmartPen_t$) and column (2), (4), (6), and (8) assume constant search ratios ($SR$) during the sample period. First of all, note that $a_1$ is positive, which means that the proportion of searchers increases as the number of smartphone users increase. Estimates of $a_0$ imply that there are only 1-2% of searchers without the smartphones. For yearly specifications (column 3, 5, and 7), $a_1$ coefficients are greater than one, which means that there are more than one new searchers per one new smartphone user. This result suggests that a proportion of consumers who receive price information from other resources is not negligible and desires further attention. For example, you have a friend who bought a smartphone recently and saw that she did a price search with her smartphone. Now you learn about the existence of the price information service, and search over the internet before you need gas. Another possibility is that the growth of searching is lagging smartphone penetration. It takes a time for smartphone users to learn how to access the Opinet service or learn the existence of the price information service. The increasing trend of $a_1$ is consistent with this explanation.

Combining $SR$, $a_0$, and $a_1$ estimates, I compute the model estimates of the average smartphone penetration rates for 2010, 2011, 2012, and all period (January 1, 2010 to June 27, 2012) to be 11.75%, 18.97%, 38.47%, and 15.52% respectively. These numbers are slightly lower than the actual smartphone penetration rates, but they are fairly close.

Since $\alpha$ is a price sensitivity term and $\beta$ is a distance sensitivity term, it is better to interpret these two terms together. As the size of grid is assumed to be 10 miles, $\alpha = 1$ and $\beta = 20$ means that an average consumer is indifferent between traveling 0.5 miles more and saving one dollar per
gallon. A distance an average consumer is willing to travel to save one dollar per gallon varies depending on the specifications, with the minimum of 0.28 miles (column 4) and the maximum of 0.79 miles (column 6).

### 5.2 Counterfactual

I conduct counterfactual exercises by setting a restriction on the searcher ratio term and study how it affects other parameter estimates. Instead of estimating it during the optimization routine, I set the searcher ratio as a constant. I consider the two extreme cases: the first case is the no searchers case \( R_t = 0 \) and the second case is the all searchers case \( R_t = 1 \).

Table 9 shows the counterfactual analysis results for region 1. Note that \( \frac{2}{3} \times 10 \) term denotes that a distance (in miles) an average consumer is willing to travel to save one dollar per gallon. For all time periods \( R_t = 1 \) cases have higher \( \frac{2}{3} \times 10 \) values. One possible explanation is that knowing the prices, consumers know in advance that they can save money and hence they are willing to travel more.

The next step is to estimate what price levels and price dispersion would have been if \( R_t = 0 \) or \( R_t = 1 \). In theory, I could solve for an equilibrium of the pricing game between the stations. With this analysis, I could estimate the impact of price information technology combined with the rapid smartphone penetration and calculate consumer and producer surplus changes. This is a very interesting topic for the future research.

### 6 Conclusion

In this paper, I have investigated how real-time gasoline price information and smartphones affect consumer search behavior and price dispersion among gasoline stations. I combined daily, individual gasoline station prices with quantity data and regional smartphone penetration data for the analysis.

Both markups and price dispersion measures seem to be stable for the first half of the period and slightly increasing for the second half of the period. I find that higher marginal costs lead to higher price level and lower price dispersion. Also, higher smartphone penetration rates imply higher price dispersion and average price level. Since price, all four price dispersion measures, and smartphone penetration rates increase for the second half of the sample period, these positive relationships are not very surprising.
There are two possible explanations. First, the increase of smartphone users does not lead to the increase of the searcher population. Second, it does increase the proportion of searchers, but it has affected the competition between gasoline stations at the same time and this leads to the increased price dispersion. Since structural estimation models report that the proportion of searchers does increase as smartphone penetration rate increases, the second case is plausible. While the empirical markup distribution shows some evidence of a bimodal distribution which is expected from the Stahl model result, further research is required to confirm this hypothesis.

The structural estimation and counterfactual analysis also provide an interesting relationship between the potential saving from going to a station with a lower price and driving additional distance. An average consumer is indifferent between saving one dollar per gallon and traveling 0.28-0.79 miles more. This additional distance tends to be lower when a consumer knows the prices and hence know in advance that she can save by going farther.
References


[19, 7, 16, 12, 17, 1, 2, 4, 3, 5, 6, 8, 9, 10, 11, 13, 14, 15, 18, 20]
Tables and Figures

Figure 1: Dubai price and Average prices for 3 regions
Table 1: Definition of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Indexes Varies Over</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>r, t</td>
<td>Maximum price - minimum price</td>
</tr>
<tr>
<td>Std</td>
<td>r, t</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>IDR</td>
<td>r, t</td>
<td>Interdecile Range 90% percentile price - 10% percentile price</td>
</tr>
<tr>
<td>IQR</td>
<td>r, t</td>
<td>Interquartile Range 75% percentile price - 25% percentile price</td>
</tr>
<tr>
<td>AWP</td>
<td>t</td>
<td>Average wholesale price</td>
</tr>
<tr>
<td>SmartPen</td>
<td>r, t</td>
<td>The ratio of smartphone users to the total population</td>
</tr>
<tr>
<td>Self</td>
<td>j</td>
<td>1 if a station offers self-service, 0 otherwise</td>
</tr>
<tr>
<td>Carwash</td>
<td>j</td>
<td>1 if a station has a car wash, 0 otherwise</td>
</tr>
<tr>
<td>Repair</td>
<td>j</td>
<td>1 if a station has a repair shop, 0 otherwise</td>
</tr>
<tr>
<td>Store</td>
<td>j</td>
<td>1 if a station has a convenience store, 0 otherwise</td>
</tr>
</tbody>
</table>

r: region, t: market (date), j: station

Table 2: Summary Statistics (Region 1)

<table>
<thead>
<tr>
<th>Unweighted</th>
<th>Quantity Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>1.170</td>
</tr>
<tr>
<td>Std</td>
<td>0.329</td>
</tr>
<tr>
<td>IDR</td>
<td>0.882</td>
</tr>
<tr>
<td>IQR</td>
<td>0.536</td>
</tr>
<tr>
<td>AWP</td>
<td>5.943</td>
</tr>
<tr>
<td>SmartPen</td>
<td>0.386</td>
</tr>
</tbody>
</table>
Table 3: Summary Statistics (Region 2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unweighted</th>
<th></th>
<th></th>
<th></th>
<th>Quantity Weighted</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>1.272</td>
<td>0.206</td>
<td>0.833</td>
<td>1.915</td>
<td>1.139</td>
<td>0.279</td>
<td>0.375</td>
<td>1.745</td>
</tr>
<tr>
<td>Std</td>
<td>0.346</td>
<td>0.046</td>
<td>0.232</td>
<td>0.483</td>
<td>0.168</td>
<td>0.027</td>
<td>0.093</td>
<td>0.296</td>
</tr>
<tr>
<td>IDR</td>
<td>0.837</td>
<td>0.182</td>
<td>0.286</td>
<td>1.446</td>
<td>0.354</td>
<td>0.115</td>
<td>0.083</td>
<td>0.719</td>
</tr>
<tr>
<td>IQR</td>
<td>0.493</td>
<td>0.076</td>
<td>0.117</td>
<td>0.696</td>
<td>0.119</td>
<td>0.066</td>
<td>0</td>
<td>0.344</td>
</tr>
<tr>
<td>AWP</td>
<td>5.943</td>
<td>0.435</td>
<td>5.226</td>
<td>6.780</td>
<td>5.943</td>
<td>0.435</td>
<td>5.226</td>
<td>6.780</td>
</tr>
<tr>
<td>SmartPen</td>
<td>0.386</td>
<td>0.177</td>
<td>0.086</td>
<td>0.642</td>
<td>0.386</td>
<td>0.177</td>
<td>0.086</td>
<td>0.642</td>
</tr>
</tbody>
</table>
Figure 5: Quantity Weighted Markup Trend (3 regions)

Figure 6: Standard Deviation for 3 Regions

Table 4: Summary Statistics (Region 3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0.433</td>
<td>0.099</td>
<td>0.269</td>
<td>0.742</td>
<td>Range</td>
<td>0.314</td>
<td>0.072</td>
<td>0.182</td>
<td>0.670</td>
</tr>
<tr>
<td>Std</td>
<td>0.112</td>
<td>0.021</td>
<td>0.073</td>
<td>0.188</td>
<td>Std</td>
<td>0.089</td>
<td>0.021</td>
<td>0.021</td>
<td>0.044</td>
</tr>
<tr>
<td>IDR</td>
<td>0.292</td>
<td>0.054</td>
<td>0.179</td>
<td>0.479</td>
<td>IDR</td>
<td>0.238</td>
<td>0.063</td>
<td>0.083</td>
<td>0.647</td>
</tr>
<tr>
<td>IQR</td>
<td>0.164</td>
<td>0.036</td>
<td>0.077</td>
<td>0.314</td>
<td>IQR</td>
<td>0.140</td>
<td>0.054</td>
<td>0.011</td>
<td>0.443</td>
</tr>
<tr>
<td>AWP</td>
<td>5.943</td>
<td>0.435</td>
<td>5.226</td>
<td>6.780</td>
<td>AWP</td>
<td>5.943</td>
<td>0.435</td>
<td>5.226</td>
<td>6.780</td>
</tr>
<tr>
<td>SmartPen</td>
<td>0.229</td>
<td>0.149</td>
<td>0.022</td>
<td>0.476</td>
<td>SmartPen</td>
<td>0.229</td>
<td>0.149</td>
<td>0.022</td>
<td>0.476</td>
</tr>
</tbody>
</table>
Table 5: Price Dispersion Regression Result

<table>
<thead>
<tr>
<th></th>
<th>Unweighted Data</th>
<th>Quantity Weighted Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std</td>
<td>Range</td>
</tr>
<tr>
<td><strong>AWP</strong></td>
<td>-0.044**</td>
<td>-0.122***</td>
</tr>
<tr>
<td></td>
<td>(-8.78)</td>
<td>(-5.96)</td>
</tr>
<tr>
<td><strong>SmartPen</strong></td>
<td>0.252**</td>
<td>1.006***</td>
</tr>
<tr>
<td></td>
<td>(19.47)</td>
<td>(19.00)</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.917</td>
<td>0.905</td>
</tr>
</tbody>
</table>

Note: the number of observations is 2727. t-statistics are in parentheses. Month of the Year dummies and region dummies are included.

*: p < 0.1, **: p < 0.05, ***: p < 0.01
Table 6: Price Dispersion Regression Result (IQR)

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Data</th>
<th>Quantity Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>M and R</td>
<td>R</td>
<td>M</td>
<td>M and R</td>
</tr>
<tr>
<td>AWP</td>
<td>-0.161</td>
<td>-0.101</td>
<td>-0.588</td>
</tr>
<tr>
<td></td>
<td>(-11.10)</td>
<td>(-9.36)</td>
<td>(-56.89)</td>
</tr>
<tr>
<td>SmartPen</td>
<td>0.454</td>
<td>0.315</td>
<td>1.584</td>
</tr>
<tr>
<td></td>
<td>(12.00)</td>
<td>(11.01)</td>
<td>(61.79)</td>
</tr>
<tr>
<td>Monthly</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.816</td>
<td>0.805</td>
<td>0.625</td>
</tr>
</tbody>
</table>

Note: the number of observations is 2727. t-statistics are in parentheses.

*: p < 0.1, **: p < 0.05, ***: p < 0.01
### Table 7: Price Regression Result

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWP</td>
<td>0.536***</td>
<td>0.320***</td>
<td>0.553***</td>
<td>0.500***</td>
</tr>
<tr>
<td></td>
<td>(88.46)</td>
<td>(76.84)</td>
<td>(111.8)</td>
<td>(79.88)</td>
</tr>
<tr>
<td>SmartPen</td>
<td>1.660***</td>
<td>2.219***</td>
<td>1.593***</td>
<td>1.736***</td>
</tr>
<tr>
<td></td>
<td>(106.54)</td>
<td>(219.0)</td>
<td>(123.9)</td>
<td>(107.3)</td>
</tr>
<tr>
<td>Self</td>
<td>-0.140***</td>
<td>-0.165***</td>
<td>-0.139***</td>
<td>-0.147***</td>
</tr>
<tr>
<td></td>
<td>(-57.15)</td>
<td>(-61.08)</td>
<td>(-55.79)</td>
<td>(-65.83)</td>
</tr>
<tr>
<td>Carwash</td>
<td>0.014***</td>
<td>0.064***</td>
<td>0.014***</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(7.05)</td>
<td>(29.19)</td>
<td>(6.98)</td>
<td>(21.34)</td>
</tr>
<tr>
<td>Repair</td>
<td>-0.089***</td>
<td>0.030***</td>
<td>-0.089***</td>
<td>-0.055***</td>
</tr>
<tr>
<td></td>
<td>(-33.51)</td>
<td>(10.67)</td>
<td>(-32.78)</td>
<td>(-20.44)</td>
</tr>
<tr>
<td>Store</td>
<td>0.137***</td>
<td>0.089***</td>
<td>0.137***</td>
<td>0.130***</td>
</tr>
<tr>
<td></td>
<td>(62.10)</td>
<td>(33.68)</td>
<td>(59.80)</td>
<td>(56.27)</td>
</tr>
<tr>
<td>monthly</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>region</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.783</td>
<td>0.739</td>
<td>0.775</td>
<td>0.771</td>
</tr>
</tbody>
</table>

Note: the number of observations is 120627. t-statistics are in parentheses.

*: p < 0.1, **: p < 0.05, ***: p < 0.01

### Table 8: Structural Estimates for Region 1

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.017</td>
<td>0.512</td>
<td>1.001</td>
<td>0.425</td>
</tr>
<tr>
<td>$\beta$</td>
<td>26.12</td>
<td>14.18</td>
<td>20.78</td>
<td>16.71</td>
</tr>
<tr>
<td>$SR$</td>
<td>0.144</td>
<td>0.160</td>
<td>0.560</td>
<td>0.560</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.013</td>
<td>0.025</td>
<td>0.005</td>
<td>0.013</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.844</td>
<td>1.149</td>
<td>1.392</td>
<td>1.422</td>
</tr>
</tbody>
</table>

### Table 9: Counterfactual Estimates for Region 1

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.366</td>
<td>0.579</td>
<td>0.317</td>
<td>0.541</td>
</tr>
<tr>
<td>$\beta$</td>
<td>8.32</td>
<td>11.34</td>
<td>21.30</td>
<td>16.29</td>
</tr>
<tr>
<td>$\gamma \times 10$</td>
<td>0.440</td>
<td>0.511</td>
<td>0.148</td>
<td>0.332</td>
</tr>
<tr>
<td>$R_t$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

32