

# Asymmetric Price Effects of Competition<sup>1</sup>

Saul Lach  
The Hebrew University and CEPR

José Luis Moraga-González  
University of Groningen

May 28, 2009

<sup>1</sup>We are indebted to Pim Heijnen and Marco Haan for providing us with the gasoline price data and the list of gas stations operating in the Netherlands. Thanks also go to Corine Hoeben for giving us the data on municipal taxes in the Netherlands. Guillermo Caruana, Riemer Faber, David Genesove, Gerard Llobet, Jan de Loecker, Marielle Non, Dirk Stelder, Matthijs Wildenbeest and Shlomo Yitzhaki gave us useful comments. We also thank seminar participants at Ben Gurion University, CEMFI, 10<sup>th</sup> CEPR Conference on Applied Industrial Organization, CERGE-EI, The Hebrew University, Tel-Aviv University and The World Bank. Yaron Aronshtan, Danny Bahar, Eli Berglas and Sarit Weisburd provided excellent research assistance. Moraga-González, also affiliated with CESifo and the SPSP Center at IESE Business School, gratefully acknowledges financial support from the Marie Curie Excellence Grant MEXT-CT-2006-042471. E-mails: <saul.lach@huji.ac.il> and <j.l.moraga.gonzalez@rug.nl>.

## **Abstract**

This paper examines how the distribution of prices changes with the number of competitors in the market. Using gasoline price data from the Netherlands we find that as competition increases, the distribution of prices spreads out: the low prices go down while the high prices go up, on average. As a result, competition has an asymmetric effect on prices. These findings, which are consistent with a theoretical model where consumers differ in the information they have about prices, imply that consumers' gains from competition depend on their shopping behavior. In our data, all consumers, irrespective of the number of prices they observe, benefit from an increase in the number of gas stations. The magnitude of the welfare gain, however, is greater for those consumers that are aware of more prices. We conclude that an increase in the number of gas stations has a positive but unequal effect on the welfare of consumers in the Netherlands.

# 1 Introduction

Economists have dedicated a significant amount of effort to analyze the relationship between the number of firms and prices. Standard oligopoly models assume consumers are perfectly informed about all prices in the market and predict that an increase in the number of firms will lower the equilibrium price. Alternative, and more realistic, models depart from the assumption that all consumers have the same information and describe equilibria characterized by non-degenerate price distributions.<sup>1</sup> In such markets where price dispersion is prevalent the question of what happens to “the” price when the number of firms changes is not even well defined. An increase in the number of firms usually affects the sellers’ pricing strategies and this alters the whole distribution of equilibrium prices.

Empirical research of markets with price dispersion has usually proceeded by estimating the impact of competition on the mean and variance of prices.<sup>2</sup> In this paper we take a broader view and study how the distribution of gasoline prices changes with the number of competitors in the market. Specifically, we examine how the various percentiles (or quantiles) of the price distribution vary with the extent of competition as measured by the number of firms operating in a market.

We think this broader approach is important for at least two reasons. First, some theoretical models based on imperfect consumer information and search costs carry the implicit prediction that the effect of competition on “high” prices differs from its effect on “low” prices.<sup>3</sup> Indeed, we construct a simple model based on similar premises where it is possible for the frequency of quoting relatively low *and* high prices to increase with the number of firms operating in the market. If this prediction is verified in the data then it has important welfare implications. Analyzing these implications constitutes the second motivation for this paper.

---

<sup>1</sup>See Baye et al. (2006) for a recent survey of models that rationalize price dispersion.

<sup>2</sup>See, for example, Borenstein and Rose (1994), Barron, Taylor and Umbeck (2004), Baye et al. (2004), Lewis (2008), and Gerardi and Shapiro (2009).

<sup>3</sup>See, for example, Varian (1980), Stahl (1989) and Janssen and Moraga-González (2004). In these models some consumers know all the prices in the market while others know only one or two prices.

In markets exhibiting a single price in equilibrium, an increase in the number of firms reduces price and this unambiguously increases welfare for all consumers. When price dispersion is prevalent, different consumers may experience distinct welfare effects depending on how different parts of the price distribution respond to changes in the number of competitors. If, as mentioned above, the frequency of low *and* high prices increases with competition then whether consumers are successful in paying the lower prices depends on their shopping behavior. Increased competition is likely to favor more those consumers exposed to several prices because they may end up paying one of the lower prices. By the same token, increased competition may even hurt consumers that observe very few prices (e.g., only one price) because they may end up paying one of the higher prices. Theoretically, price changes originating from an increase in the number of firms can result in welfare gains for some consumers and, at the same time, in welfare losses for others. Analyzing the effects of entry-promoting policies just on the mean and dispersion of prices cannot capture these distinct welfare effects.

We use gasoline (Euro 95) prices posted on a daily basis by about 3,100 gas stations in the Netherlands during May 2006. For a given gas station, the relevant market is defined as the municipality where the gas station is located. For each of such 423 markets, we compute a number of percentiles of the price distribution, including the minimum and maximum prices. We then regress these statistics on the number of gas stations in the market as well as on municipality characteristics to control for common determinants of prices and the number of stations. We also use population size and local taxes as instruments for the (endogenous) number of stations.

The empirical findings suggest that as competition increases the distribution of prices spreads out; therefore competition has *asymmetric effects on prices*. Specifically, as the number of gas stations in a market increases, the low prices decrease while the high prices increase, on average. For example, adding 4 additional gas stations to a single-station market would, on average, lower the minimum price of a liter of Euro 95 by 0.93 cents and increase the maximum price by 0.83 cents. Likewise, the 90 percentile would go up by 0.16 cents and the 10 percentile would go down by 0.16 cents. These are small changes relative to the mean price of 142 cents, but these changes are quantitatively

significant relative to the dispersion in (residual) prices which is about 1 cent. This characterization of the effect of competition on prices is in line with the theoretical predictions of our model where consumers have imperfect information about prices and observe different numbers of prices.

In addition, we estimate the gains from increased competition to consumers observing different numbers of prices. In our data, all types of consumers benefit from an increase in the number of stations. The magnitude of the welfare improvement due to price changes depends, however, on their shopping behavior and is larger for those consumers that observe more prices. The decline in the expected price paid by consumers that observe 4 or 5 prices is about twice as large as that for consumers that observe only 2 prices.

We believe the message of this paper goes beyond the present application to the gasoline market in the Netherlands. Since imperfect price information is prevalent in many markets (telecommunications, health, gas, electricity, etc.), the price effects of competition-enhancing policies (industry deregulation, trade liberalization, etc.) might not be as straightforward as those implied by standard models. Moreover, since increased competition can potentially have unequal effects among consumers, distributional issues become a central part of the welfare assessment of these policies. This advocates the importance of taking a broader view where the interaction between competition and consumer policy is taken into consideration (Armstrong, 2008; Waterson, 2003; Wu and Perloff, 2007).

In Sections 2 and 4, we present a model, inspired by Varian's (1980) model of sales, of the distribution of prices in an oligopolistic market where consumers differ in the amount of prices they are exposed to. In Sections 3 and 5 we describe the gasoline price data for the Netherlands and explain how markets are defined for the empirical analysis. In Section 3 we also present evidence that gas stations in our data appear to be using mixed pricing strategies as implied by Varian-style models. Our empirical strategy is outlined in Section 6, while the empirical results are presented in Section 7. An empirical assessment of the welfare implications of increased competition is presented in Section 8. Conclusions close the paper.

## 2 A model of the distribution of prices

The market for gasoline is a good example of a homogenous good market where price dispersion is observed.<sup>4</sup> Many consumers are informed about a few prices only, and this gives some monopoly power to the gas stations. In many instances, consumers run out of fuel and have no option but to fill their gas tanks at the first gas station they encounter and this gives additional market power to the gas stations. Prices change quite frequently and it is not trivial to tell which gas station is the cheapest in a given market. The model we consider below, inspired by Varian's (1980) model of sales, has these characteristics.

Suppose we have a market with  $N \geq 2$  identical firms that compete in prices to sell a homogeneous good to a large number  $L$  of consumers. We assume firms' unit selling costs,  $c$ , are the same across all firms.<sup>5</sup> Each consumer wishes to purchase at most a single unit of the good (e.g., a full tank). The maximum willingness to pay for the good is common across consumers and is denoted by  $v > 0$ .

The entire population of consumers  $L$  can be divided into various types, each type consisting of all the consumers with similar exposure to price information. In particular, we assume that a fraction  $\mu_s \geq 0$  of the consumers is informed about  $s$  prices in the market, with  $s = 1, 2, \dots, N$ ; by construction  $\sum_{s=1}^N \mu_s = 1$ . The rationale behind this assumption is that the typical consumer is exposed to a number of prices that depends on the number of gas stations she observes while driving to her work. We view search in this market as "passive" in the sense that consumers do not deliberately drive to various gas stations to observe their prices. We will nevertheless refer to the different consumer types as consumers exhibiting different "shopping behavior", but we emphasize that this term is not meant to convey that consumers are actively engaged in searching for the

---

<sup>4</sup>Price dispersion in gasoline markets has been widely documented. Recent papers on this topic are, for example, Barron, Taylor and Umbeck (2004), Chandra and Tappata (2008), Hosken et al. (2008), and Lewis (2008).

<sup>5</sup>We therefore abstract from cost differences across firms as an explanation for price dispersion in gasoline markets. For a recent paper where price dispersion reflects differences in marginal costs see Goldmanis et. al. (forthcoming). In the empirical part we control for station-specific unobserved effects as well as for aggregate effects which capture fluctuations in the wholesale price of gasoline.

lowest price of gasoline.<sup>6</sup> The vector  $\boldsymbol{\mu}(N) = (\mu_1(N), \mu_2(N), \dots, \mu_N(N))$  represents the distribution of price information in the market. The argument  $N$  indicates that  $\boldsymbol{\mu}(N)$  has up to  $N$  coordinates and that the value taken by each coordinate may change with the number of stations in the market (the dependency of  $\mu$  on  $N$  will sometimes be left implicit). For example, if a new station opens on the way to work of some consumers then these consumers will observe an additional price.<sup>7</sup>

Firms play a simultaneous-moves Bertrand game. An individual firm  $i$  chooses a price  $p_i$  taking the prices of the rival firms as given. To rule out pure-strategy equilibria, we shall assume  $1 > \mu_1 > 0$  (as in Varian, 1980).<sup>8</sup> The intuition is as follows. Consider the position of a firm  $i$  and suppose all its rivals were charging a price  $\tilde{p}$ , with  $c \leq \tilde{p} \leq v$ . There are two forces that affect the price-setting decision of firm  $i$ . First, there is a desire to *steal business* from competitors and this pushes this firm to undercut the rivals' price. This desire arises because there exist consumers who are exposed to various prices and choose the cheapest gas station to tank (i.e.,  $\mu_s > 0$  for at least one  $s \in \{2, 3, \dots, N\}$ ). Second, the possibility of *extracting surplus* from consumers who do not compare prices prompts firm  $i$  to set higher prices than the rivals' price. This desire arises because there exist consumers (in particular a fraction  $\mu_1/N > 0$ ) who have no other option but to tank at firm  $i$ . It is easy to see that either of these deviations destabilizes any such price  $\tilde{p}$ . Therefore a single price cannot accommodate these two incentives.

Denote the mixed pricing strategy of a firm  $i$  by a distribution of prices  $F_i$ . We shall only study symmetric equilibria, i.e., equilibria where  $F_i = F$  for all  $i = 1, 2, \dots, N$ .<sup>9</sup>

---

<sup>6</sup>We do not model the consumer's decision of how many prices to observe. For models of this kind where there are two types of consumers see, for example, Stahl (1989) and Janssen and Moraga-González (2004). See also Hortaçsu and Syverson (2004) for an empirical model of price dispersion where product differentiation is present.

<sup>7</sup>It is likely that in small markets – markets with up to, say, 6-7 gas stations – each coordinate of  $\boldsymbol{\mu}(N)$  will be strictly positive. However, in markets where there is a very large number of gas stations (e.g., Rotterdam has 80 gas stations), one would expect  $\mu_s$  to be zero for large  $s$ .

<sup>8</sup>If  $\mu_1 = 1$ , then  $p_i = v$  for all  $i$  is a pure-strategy equilibrium. If  $\mu_1 = 0$ , then  $p_i = c$  for all  $i$  is a pure-strategy equilibrium.

<sup>9</sup>It is easy to see that the support of  $F$  must be a convex set and that  $F$  cannot have atoms. Details are available from the authors upon request.

To calculate the expected profit to firm  $i$  from charging a price  $p$  when its rivals choose a random pricing strategy according to the cumulative distribution  $F$ , we first consider the chance that firm  $i$  sells to a consumer of type  $s$ , i.e., to a consumer that observes  $s$  prices in the market. The chance that such a consumer observes the price of firm  $i$  is  $s/N$  and, conditional on this, the probability that firm  $i$  sells to this consumer at price  $p$  is  $(1 - F(p))^{s-1}$ . Therefore, the profits to firm  $i$  from all types of consumers is

$$\Pi_i(p; F) = L(p - c) \left[ \sum_{s=1}^N \frac{s\mu_s}{N} (1 - F(p))^{s-1} \right] \quad (1)$$

In equilibrium, a firm must be indifferent between charging any price in the support of  $F$  and charging the upper bound  $\bar{p}$ . Therefore, any price in the support of  $F$  must satisfy  $\Pi_i(p; F) = \Pi_i(\bar{p}; F)$ . Since  $\Pi_i(\bar{p}; F)$  is monotonically increasing in  $\bar{p}$ , it must be the case that  $\bar{p} = v$ . As a result, equilibrium pricing requires

$$(p - c) \left[ \sum_{s=1}^N s\mu_s (1 - F(p))^{s-1} \right] = (v - c)\mu_1. \quad (2)$$

This equation cannot be solved explicitly for  $F$ , except in special cases. However, the lower bound can easily be found by setting  $F = 0$  in (2) and solving for  $p$  which gives,

$$\underline{p} = c + \frac{(v - c)\mu_1}{\sum_{s=1}^N s\mu_s}$$

Existence and uniqueness of an equilibrium price distribution  $F$  can easily be proven.<sup>10</sup>

A closer look at equation (2) serves to make an important point to which we will come back later in Section 4 when we study the effects of entry on the price distribution. What truly matters for determining the equilibrium price distribution is not  $N$  – the number of firms – but the distribution of information among consumers, that is  $\boldsymbol{\mu}(N)$ . Changes in  $\boldsymbol{\mu}(N)$  – holding  $N$  constant – will alter the equilibrium price distribution while changes (increases) in  $N$  – holding  $\boldsymbol{\mu}(N)$  constant – will not. It is therefore changes in the distribution of information that cause more or less “competitive pressure” in the

---

<sup>10</sup>To prove that  $F$  exists and is unique, rewrite equation (2) as  $\sum_{s=1}^N s\mu_s (1 - F(p))^{s-1} = \frac{\mu_1(v-c)}{p-c}$ . The LHS of this equation is positive and decreases in  $F \in [0, 1]$ . At  $F = 1$  it takes on value  $\mu_1$ , while at  $F = 0$ , the LHS equals  $\sum_{s=1}^N s\mu_s > \mu_1$ ; by contrast, the RHS is a positive constant. As a result, for every  $p \in (\underline{p}, v)$ , there is a unique solution to equation (2) satisfying  $F \in [0, 1]$  and  $F$  increases in  $p$ .

market. In Section 4 we shall discuss plausible ways in which a change in the number of gas stations  $N$  may affect the information structure  $\boldsymbol{\mu}(N)$  and the pricing implications of such a change.

In Figure 1 we plot an example of an equilibrium price density function.<sup>11</sup> A desirable feature of the model is that it allows for bell-shaped density functions. This is desirable because bell-shaped densities are a typical feature of real-world price data. In particular, Figure 2 shows that the density of gas prices in the Netherlands is bell-shaped. Other studies have also found bell-shaped price density functions for various products (e.g., Roberts and Supina, 2000; Lach, 2002, Hosken et al., 2008). Varian's (1980) model corresponds to the case where there are only two types of consumers: fully-informed consumers  $\mu_N$  who observe all  $N$  prices in the market, and uninformed consumers  $\mu_1 = 1 - \mu_N$  who observe just one price.<sup>12</sup> In this case, the price density is roughly U-shaped (see Figure 2 in Varian (1980)). Thus, because empirical price densities are usually bell-shaped, the simple Varian model is inconsistent with the data. Assuming additional consumer heterogeneity in the form of additional consumer types, besides being a more realistic assumption, allows for plausible price density distributions to arise in equilibrium.

The main goal of this paper is to examine the relationship between the number of stations and the distribution of prices. Before we do this, however, it is important to examine the data for evidence that firms are using mixed pricing strategies. Lacking such evidence the model presented in this Section would not be appropriate. In the next Section, we first describe the gasoline price data and then analyze what the data say on the use of mixed strategies.

---

<sup>11</sup>For a market with  $N = 5$  firms,  $\boldsymbol{\mu}(N) = (0.73, 0.2, 0.01, 0.01, 0.05)$ ,  $v = 1$  and  $c = 0$ .

<sup>12</sup>Under these assumptions we can solve explicitly for the distribution of prices:

$$F(p) = 1 - \left( \frac{(1 - \mu_N)(v - p)}{N\mu_N(p - c)} \right)^{\frac{1}{N-1}}$$

with support

$$c + \frac{(1 - \mu_N)(v - c)}{(1 - \mu_N) + N\mu_N} \leq p \leq v.$$

### 3 Data on gasoline prices and mixed strategies

We use daily prices for Euro 95 gasoline from a large sample of gas stations in the Netherlands.<sup>13</sup> The price data were obtained from Athlon Car Lease Nederland B.V., the largest private car leasing company in the Netherlands with over 129,000 cars as of the end of 2008 ([www.athloncarlease.com](http://www.athloncarlease.com)). The typical contract between Athlon and its lessees stipulates that Athlon pays for the gasoline consumed (up to a limit) as well as for car maintenance, insurance, etc. In order to do this, Athlon gets the lessees' gas receipts and it is from these receipts that the fuel prices are retrieved. Athlon's lessees do not get special discounts so the prices reported by Athlon are actual prices paid by drivers at the pump.

Price distributions are defined for a given market. In our application, we define markets as the geographical area comprised by a municipality. There are 440 municipalities in the Netherlands for which we have gasoline price data. The majority of the municipalities are quite small in terms of population: 55 percent have less than 25,000 inhabitants, and the population in 91 percent of the 440 markets is under 75,000. This definition of the market ignores stations that may be geographically close (or in the way to work) but located in different municipalities. Because of this, we will examine the robustness of our findings to the inclusion of measures of competitors in neighboring municipalities.<sup>14</sup>

Another important reason for choosing to work at the municipality level is that we have economic, geographic and demographic data for almost each municipality. This is very convenient for our purposes since we will be able to control for common determinants of the number of stations and prices.

We view our sample of prices in a given market as random draws from the distribution  $F$  corresponding to that market. We are able to assume this because of several

---

<sup>13</sup>We also use prices for Diesel to examine the robustness of our results (see Section 7.1).

<sup>14</sup>It is not necessary for the gas stations to be located in a given municipality, provided every gas station in a given municipality factors into its pricing strategy the same distribution of information  $\mu(N)$ . This would approximately be true as long as neighboring municipalities do not differ much.

reasons. First, Athlon's lessees do not pay themselves for the gas (it is part of the contract) and therefore it is reasonable to assume that they have no incentives to search for gas stations offering the lowest prices. This is important because otherwise our sample could be seen as a sample of the cheapest gas stations. Second, all gas stations in the Netherlands are self-service and therefore there is a single price for gas in each station. Finally, we believe that the extent to which various prices in a given market are set by a single firm (because of joint-ownership) and/or reflect collusive agreements is minor, implying that prices can be viewed as independent draws.

On this last point we first note that although we do not have information on the gas stations' owners, according to the Dutch Competition Authority about 62 percent of the gas stations are owned and operated by independent dealers (NMa, 2006). The remaining stations belong to the main oil producers: BP, Esso, Shell, Texaco and Total. But even among these branded stations, most are dealer-operated. For example, Shell serves fewer than 15 percent of the gas stations and about 2/3 of the Shell-branded gas stations are operated by dealers who are free to set their own prices. This suggests that joint ownership of gas stations is not such a prevalent phenomenon as one may be led to believe from casual observation (although we have no data on joint ownership of gas stations by independent owners). An exception to this is the highway market, where most gas stations, 63 percent, are owned and operated by the large oil producers (NMa, 2006). Second, although there is no direct information on the extent of collusion (if at all) in the Dutch gasoline market, the Dutch Competition Authority has only identified the highway market as potentially collusive. The main reason for this assessment is again that most of the gas stations located on highways are owned and operated by the major oil producers. In fact, since 2002, the Dutch government has forced divestitures of highway stations but not in other markets. Because of these two last reasons, potential collusion and joint decision-making, we decided to remove the stations that were located in highways from our sample. We think that excluding these stations leaves us with a large sample of gas stations whose prices can be considered as independent.

Prices were obtained from 3,300 gas stations for the period May 5–26 2006, except for May 10 and May 17, for a total of 20 days. After removal of the 217 stations located

in highways we are left with 3,083 gas stations. Because the price information arrives directly from the lessees, not all stations are sampled every day, which results in an unbalanced panel data of gas stations.<sup>15</sup> There are 32,348 station-day observations on Euro 95 prices.

As an illustrative device, the left panel in Figure 2 displays a kernel estimate of the density function of prices. The average price of Euro 95 gas in our sample is 142.04 cents and the average of the within-market standard deviations is 2.52 cents. The lowest price is 119 cents while the highest price in the sample is 167.<sup>16</sup> Not surprisingly, there is dispersion in gasoline prices but, as evidenced by the coefficient of variation, it is not very large. However, the daily variation in the total cost of filling-up a 50 liter tank – the difference in cost between buying at the highest-priced and lowest-priced station in a given day – is between 8.5 and 24 euros which is not a trivial amount.<sup>17</sup>

Because price differentials among stations are likely to be driven by time-invariant factors (e.g., brand, ownership structure, location, availability of a convenience store, additional services, etc.), it is problematic to compare prices of different gas stations, even within the same market. The same is true when comparing gas prices in different days. We therefore remove day and station-specific effects from actual (raw) prices to obtain a residual price which is more comparable across days and stations.<sup>18</sup> These residual prices are obtained by regressing prices on station-specific dummies and on a

---

<sup>15</sup>We have one price of Euro 95 per station per day. The number of days or, equivalently, the number of price quotations per gas station in the sample ranges from 1 to 17 days with an average of 10.5 days and a median of 12 days.

<sup>16</sup>The 119 price is an outlier; the second lowest price is 129 cents. However, we do not think this is a typo since the very same gas station is also charging a very low price for Diesel on a *different* day (78 cents when the average is 108 cents).

<sup>17</sup>The stations exhibiting the extreme prices can be anywhere in the Netherlands so this calculation is purely illustrative. On May 11, the maximum and minimum price were 1.49 and 1.32 euros per liter. This 17 cent difference translates into a 8.5 euros saving in filling up a 50 liter tank. On May 14, the maximum and minimum price were 1.19 and 1.67 euros per liter, respectively, implying a 24 euro saving in filling up a 50 liter tank.

<sup>18</sup>As done, for example, by Lach (2002), Hosken et al. (2008), and Lewis (2008). A similar approach is taken when estimating auction data in order to generate “homogenized bids” or “normalized bids” which are comparable across auctions (see, for example, Haile, Hong, and Shum (2003)).

cubic trend, separately for each municipality.<sup>19</sup> Residual prices are therefore detrended prices net of station-level effects. The mean residual price for each station (and for each municipality) is then zero. The implicit assumption here is that station and day effects affect *only* the mean price charged by a gas station. By removing these effects, we “homogenize” stations *within* markets so that we can treat residual prices in a market as coming from the same distribution of prices  $F$ . Of course, the distribution of residual prices varies *across* markets due to differences in the number of stations and in the information structure (“shopping behavior”) as well as in  $v$  and  $c$ .<sup>20</sup>

We will use the residual prices in our empirical analysis; their distribution is plotted in the right panel of Figure 2. As expected, the average of the within-market standard deviations in residual prices, 1.05, is lower than that in the raw data. Nevertheless, as can be seen in the graph, residual prices still exhibit considerable variation.<sup>21</sup>

### 3.1 Evidence on mixed strategies

We now present empirical evidence on the use of mixed strategies by gas stations. This issue has been examined by Lach (2002) for three food products and an appliance in Israel, by Wildenbeest (2008) for groceries in the UK and by Hosken et. al. (2008), for gasoline prices in the suburbs of Washington DC. All these studies found evidence supporting the use of mixed pricing strategies.

Specifically, we check whether gas stations vary their relative position in the cross-sectional distribution of prices over time, as implied by the use of mixed strategies. Simply put, the use of mixed strategies implies that we should not observe gas stations always selling at high prices or always selling at low prices.

We observe the residual price posted by gas station  $i$  on day  $t$  and we locate this

---

<sup>19</sup>We do not use day dummies because in 5.4 percent of the observations there is only one station per day per municipality. Moreover, in one municipality (Reiderland) we only have one observation and therefore residual prices cannot be computed, leaving us with residual prices in 439 municipalities.

<sup>20</sup>Note, however, that the theoretical model in Section 2 implies that proportional changes in  $v$  and  $c$  affect only the location of  $F$ . In any case, removing station-specific effects also removes the effect of market-level factors affecting the location of  $F$ .

<sup>21</sup>The longer left tail of the distribution is due to the outlier price mentioned in footnote 16. Removing this station makes the density much more symmetric and lowers the standard deviation to 1.06.

residual price within the price distribution observed in  $i$ 's market on day  $t$ . We can then track the relative position of the station's residual price over time.<sup>22</sup> There are a number of ways of doing this.

We start by computing the number of days that a gas station was in the  $q^{th}$  quartile of the cross-sectional price distribution. We denote this statistic by  $T_q$ ,  $q = 1, 2, 3, 4$ .  $T_q$  is expressed as a percentage of the total number of days a station appears in the sample.<sup>23</sup> For example, if the station was never in the first quartile of the distribution then  $T_1 = 0$ , whereas if the station was always in the first quartile then  $T_1 = 1$ . Clearly, for each gas station,  $T_1 + T_2 + T_3 + T_4 = 1$ . Figure 3 plots the histograms of  $T_1 - T_4$ .

If many stations always remain in the same quartile of the (residual) price distribution we should observe a large number of firms with  $T_q = 1$ . Figure 3 indicates that this is not the case. 2.15 percent of the stations were always in the first quartile and, for higher quartiles, this percentage is even lower. The low number of stations always selling in the same quartile of the price distribution is consistent with the use of mixed strategies.<sup>24</sup>

We also observe in the top-left graph of Figure 3 that 14 percent of the stations were never in the first quartile of the price distribution ( $T_1 = 0$ ). This means that about 84 percent ( $100 - 14 - 2.15$ ) of the stations were part (but not all) of their time in the first quartile of the distribution and the remaining time in other quartiles. Similarly, 77, 81 and 75 percent of the stations were part (but not all) of their time in the second, third and fourth quartile of the distribution, respectively, and the remaining time in other quartiles.<sup>25</sup> Although this is evidence that a sizable number of stations moves around

---

<sup>22</sup>Recall that the time horizon is the 20 days in May 2006 but no station appears in the sample for more than 17 days.

<sup>23</sup>Note that the cross-sectional distribution in day  $t$  is defined for the stations which quoted prices in day  $t$ . Therefore the number of stations, and their identity, may change from day to day. The statistics were computed in all markets and days where the number of stations was at least 4.

<sup>24</sup>When using the actual prices, 20 percent of the stations always charge prices in the first quartile of the distribution ( $T_1 = 1$ ), 1.7 percent always in the second quartile ( $T_2 = 1$ ), 3.4 percent always in the third quartile ( $T_3 = 1$ ), and 7 percent always in the fourth quartile ( $T_4 = 1$ ). These figures are higher because actual prices reflect store-specific factors (e.g., location) that are fixed over time.

<sup>25</sup>The corresponding percentages for the actual price data are 42, 54, 47 and 33 percent.

the cross-sectional price distribution, the histograms of  $T_1 - T_4$  do not reveal how *long* a particular gas station stayed in each quartile of the price distribution. We examine this in Figure 4.

Figure 4 graphs, for 50 randomly selected gas stations, the percentage of days each of these stations was in the first, second, third and fourth quarter of the cross-sectional distribution of prices.<sup>26</sup> The changing bar colors indicate that only two stations remained in the same quartile during all the days they appear in the sample (stations number 2 and 9).<sup>27</sup>

These figures still do not reveal how gas stations “travel” across the quartiles of the price distribution over time, i.e., the extent of intra-distribution dynamics. The transition process from one cross-sectional distribution to another can be modelled by assuming that this transition is done in a Markovian fashion through a  $4 \times 4$  transition matrix whose  $(i, j)^{th}$  entry gives the probability that a gas station in the  $i^{th}$  quartile in day  $t$  moves to the  $j^{th}$  quartile in day  $t' > t$ . Consistent estimates of these probabilities are the sample proportions of stations moving from one quartile to another in each market. Assuming a time- and market-invariant transition matrix, the estimated transition matrices for each day and market are averaged to produce a single (estimated) transition matrix.

Table 1 presents estimates of 1-week ( $t' = t + 7$ ) transition probabilities.<sup>28</sup> Examination of the transition matrix gives a good idea on the extent of intra-distribution mobility. If stations keep their positions over time – lack of mobility – then the matrix should have “large” diagonal entries. If there is a lot of mobility across the quartiles of the distribution this would be reflected in “large” off-diagonal probabilities. The probability of remaining in the first quartile is 33 percent, which means that the probability

---

<sup>26</sup>We plot only 50 stations that were randomly sampled from the 2472 gas stations appearing in markets and days where the number of stations was at least 4. Plotting all the stations generates graphs that are too cluttered to be readable.

<sup>27</sup>Using the actual price data, 62 percent of the stations (31 stations) move between 2 or more quartiles of the price distribution.

<sup>28</sup>The entries are weighted averages of the estimated transition probabilities for each day and market with weights equal to the proportion of observations in each cell.

that a low-price station will be selling at a higher price a week ahead is 67 percent.<sup>29</sup> Overall, the diagonal entries do not appear to be large relative to the off-diagonal terms. This is indicative of significant intra-distribution dynamics, as found in the studies cited at the beginning of this subsection.

## 4 The distribution of prices and the number of firms

We are interested in how the pricing strategy of the firms, given by the distribution function that solves (2), changes with the number of competitors  $N$ . As explained before, changes in  $N$  have no effect on prices *per se*. It is via changes in consumer information  $\boldsymbol{\mu}(N)$  that an increase in the number of competitors results in more or less “competitive pressure” in the market.

Let an additional gas station enter a market operated by  $N$  stations. The effect of such entry on the exposure to prices of a single consumer will depend on a number of factors including the location of the new gas station, the location of the consumer’s home and that of her job. In our model, the change in overall consumer information is represented by the change from  $\boldsymbol{\mu}(N)$  to  $\boldsymbol{\mu}(N + 1)$ .

Let  $F(p|\boldsymbol{\mu}(N), v, c)$  be the equilibrium price distribution that solves equation (2). Even though we do not have an explicit expression for the equilibrium price distribution, we can perform comparative statics by examining the inverse of the price distribution. From (2), we obtain

$$p(\boldsymbol{\mu}(N), v, c) = c + \frac{(v - c)\mu_1(N)}{\sum_{s=1}^N s\mu_s(N)(1 - F)^{s-1}} \quad (3)$$

where  $F$  takes values on  $[0, 1]$ .

If we let  $F$  take on values  $\{0, 0.1, 0.2, \dots, 0.9, 1\}$  in equation (3) we obtain the appropriate percentile of the equilibrium price distribution. As shown below, we use the inverse expression to examine how different percentiles of the price distribution are affected by changes in the number of competitors. Let  $F(p|\boldsymbol{\mu}(N + 1), v, c)$  denote the

---

<sup>29</sup>Using the actual price data we obtain probabilities of remaining in the same quartile a week ahead equal to .76, .47, .51 and .69 for the first, second, third and fourth quartile, respectively.

price distribution after entry of one additional firm, and  $p(\boldsymbol{\mu}(N+1), v, c)$  its inverse. We can then compute the difference between the (inverse) price distributions with  $N$  and  $N+1$  firms:

$$\begin{aligned}
& \frac{p(\boldsymbol{\mu}(N), v, c) - p(\boldsymbol{\mu}(N+1), v, c)}{v - c} \\
&= \frac{1}{\sum_{s=1}^N s \frac{\mu_s(N)}{\mu_1(N)} (1-F)^{s-1}} - \frac{1}{\sum_{s=1}^{N+1} s \frac{\mu_s(N+1)}{\mu_1(N+1)} (1-F)^{s-1}} \\
&= \frac{\sum_{s=1}^N s \left( \frac{\mu_s(N+1)}{\mu_1(N+1)} - \frac{\mu_s(N)}{\mu_1(N)} \right) (1-F)^{s-1} + (N+1) \frac{\mu_{N+1}(N+1)}{\mu_1(N+1)} (1-F)^N}{\left( \sum_{s=1}^N s \frac{\mu_s(N)}{\mu_1(N)} (1-F)^{s-1} \right) \left( \sum_{s=1}^{N+1} s \frac{\mu_s(N+1)}{\mu_1(N+1)} (1-F)^{s-1} \right)} \quad (4)
\end{aligned}$$

If we let  $F$  take on values  $\{0, 0.1, 0.2, \dots, 0.9, 1\}$  in equation (4), we obtain the effect of an increase in  $N$  on the different percentiles of the equilibrium price distribution. For example, setting  $F = 0.8$ , (4) gives the change in the 80<sup>th</sup> percentile of the price distribution when the number of firms increases from  $N$  to  $N+1$ .

The main point we want to make here is that the way firm entry affects the different percentiles of the price distribution is intimately linked to the way  $\boldsymbol{\mu}(N)$  changes into  $\boldsymbol{\mu}(N+1)$ . The coordinates of  $\boldsymbol{\mu}(N)$  and  $\boldsymbol{\mu}(N+1)$  may differ in multiple ways but what we can reasonably expect is that entry causes the price information set of *every* consumer to become *weakly* larger, and that of *some* consumers to become *strictly* larger. Therefore, we can reasonably expect that entry implies,

$$(1) \mu_1(N+1) \leq \mu_1(N) \quad (5)$$

$$(2) \mu_s(N+1) > \mu_s(N) \text{ for at least one } s = 2, 3, \dots, N+1$$

where it is understood that  $\mu_{N+1}(N) = 0$ .

Unfortunately little can be said about how the other coordinates of  $\boldsymbol{\mu}(N)$  will change with  $N$ . To see this, consider for example the effect of an increase in  $N$  on the share of consumers exposed to  $s$  prices. After entry, some consumers who used to observe  $s-1$  prices will observe  $s$  prices, while some consumers who used to be exposed to  $s$  prices will be exposed to  $s+1$  prices. These two effects operate in opposite direction on

$\mu_s$  and therefore the net effect on  $\mu_s$  is in principle undetermined. This indeterminacy implies that, at this degree of generality, the different percentiles of the price distribution may increase or decrease with the number of firms, depending on parameters.

We provide two examples that illustrate this point.<sup>30</sup>

**Example 1** *In this example,  $N = 2$  and  $\boldsymbol{\mu}(2) = (0.5, 0.5)$ . Half of the consumers pass one gas station while driving to work, while the other half pass two stations. Both types of consumers fill their tank at the station offering the lowest price. Suppose that after entry of a new gas station, we have  $\boldsymbol{\mu}(3) = (0.45, 0.52, 0.03)$ . That is, the share of consumers exposed to just one price goes down to 45 percent, but the share of consumers observing two and three prices increases. The resulting equilibrium price distributions are depicted in Figure 5.*

In Example 1, the price distribution  $F_N$  dominates the price distribution  $F_{N+1}$  in a first-order stochastic sense. As a result, the effect of an increase in the number of competitors is to decrease all the quantiles of the price distribution. Example 1 accords with the usual intuition that markets with more firms have lower prices. The effect of increasing  $N$ , however, is stronger at the lower quantiles than at the higher quantiles of the price distribution. In other words, the high prices fall less than the low prices when we move from a market with  $N$  firms to a market with  $N + 1$  firms.

We can actually make this result more general. If the coordinates of  $\boldsymbol{\mu}(N)$  and  $\boldsymbol{\mu}(N + 1)$  satisfy the inequality

$$\frac{\mu_s(N + 1)}{\mu_1(N + 1)} \geq \frac{\mu_s(N)}{\mu_1(N)}, \text{ for all } s = 1, 2, \dots, N. \quad (6)$$

then the price distribution  $F_N$  first-order stochastically dominates the price distribution  $F_{N+1}$ . Condition (6) says that the share of consumers comparing  $s$  prices relative to the non-comparing (price-insensitive) consumers increases in  $N$ . This condition is sufficient for stochastic dominance, as can readily be seen upon inspection of equation (4).

---

<sup>30</sup>In all numerical examples (Figures 5-7) we set  $v = 1$  and  $c = 0$ .

**Example 2** *The initial situation is as in Example 1 where  $N = 2$  and  $\mu(2) = (0.5, 0.5)$  but we assume that after entry of a new gas station we have  $\mu(3) = (0.45, 0.1, 0.45)$ . Again the share of consumers exposed to just one price goes down to 45 percent, but, in contrast to Example 1, the share of consumers observing two prices decreases in favor of the share of consumers observing three prices, which goes up from 0 to 45 percent. The resulting equilibrium price distributions are depicted in Figure 6.*

What is striking in Figure 6 is that the price distributions cannot be ranked according to first-order stochastic dominance. We see, for example, that the 80<sup>th</sup> percentile of the price distribution increases in  $N$  while the 20<sup>th</sup> percentile decreases. That is, the effect of competition on “high” prices differs from its effect on “low” prices.

In order to understand the intuition behind these results, recall that the distribution of prices of a firm is chosen to maximize expected profits accruing from the various groups of consumers, given other firms’ strategies. These profits, as shown in equation (2), are constant at all prices chosen with positive probability. Note also that the elasticity of the expected demand of the consumers observing  $s$  prices is equal to  $p(s-1)f_N(p)/(1-F_N(p))$ .<sup>31</sup> Thus, keeping rivals’ strategies fixed, this elasticity is independent of  $N$ . By contrast, the elasticity of the expected demand of consumers observing  $N$  prices is  $p(N-1)f_N(p)/(1-F_N(p))$ , which increases in  $N$ . Therefore, if the number of firms increases, and keeping the rivals’ strategies fixed, only the elasticity of the expected demand of the fully informed consumers changes.

Consider now a firm contemplating how to change its strategy as a response to entry. This firm knows that the expected demand from the fully informed consumers becomes more elastic as  $N$  increases. This firm has therefore an incentive to offer even lower prices when  $N$  increases. This is what we see in the two examples above; the lower bound of the price distribution falls as we move from a market with  $N$  firms to a market with  $N + 1$  firms.

On the other hand, in order to compensate for offering lower prices and getting

---

<sup>31</sup>Every consumer demands at most a single unit so individual demands are inelastic. However, given the strategies of the rival firms, the expected number of units sold by a firm increases as this firm reduces its price.

lower profits, firms increase the frequency of their high prices in order to extract more profits from the less-informed consumers (e.g., those observing a single price). The latter’s lack of full information on prices allows firms to increase the frequency at which they charge high prices. The strength of this “compensating” effect depends on the mass of fully-informed consumers. This effect is therefore much stronger in Example 2 than in Example 1 because in the former case 90 percent of the consumers who used to observe two prices get to observe three prices when  $N$  increases from 2 to 3, as opposed to 6 percent in Example 1. As a result, the lower bound declines more in Example 2 than in Example 1, but also the frequency of high prices increases in Example 2 but not in Example 1. This effect may be strong enough so as to drive mean prices up. In fact, in Example 2 the mean price increases from 0.549 to 0.554.<sup>32</sup> Figure 7 shows that when we move from two to three firms, the very low and the high prices become more frequent in detriment of intermediate prices.

In sum, the competitive effect of an increase in  $N$  is to prompt firms to offer “lowest ever” prices by shifting some probability mass towards even lower prices (prices that had zero density before the increase in  $N$ ), and, at the same time, to increase the frequency of high prices to keep the firm’s incentives balanced. In Example 1 the first effect dominates and the price distributions do not cross each other. In Example 2 the second effect is sufficiently strong so that high prices become more frequent after entry. Thus, it is the presence of a sufficiently large mass of fully-informed consumers that prompt firms to more extreme pricing.

As these examples indicate, the overall implication of the changes in the equilibrium strategies of the firms on the prices the various consumers pay is subtle since we need to account for changes in the support of the price distribution and for changes in the price frequencies.

---

<sup>32</sup>Other models show a tendency for average prices to rise with  $N$ , e.g., Stahl (1989). In Stahl’s model, there are two consumer types (shoppers, who observe all the prices in the market, and non-shoppers, who just observe one), search is endogenous and the price distribution converges to a distribution degenerated at the monopoly price as the number of firms operating in the market goes to infinity. Our model, however, presents a more realistic information structure and, as seen in Example 1, mean prices need not rise with  $N$ .

Our goal is to estimate the causal effect of a change in  $N$  on the distribution of prices. Our theoretical model makes clear that the effect of a change in  $N$  works through changes in the distribution of price information among consumers. That is, we view the causal effect of  $N$  on prices as that effect arising solely from an induced change in the (unobserved) information-related variables  $\boldsymbol{\mu}$ . In other words, holding  $\boldsymbol{\mu}$  constant, changes in  $N$  should have no causal effect on prices. The theoretical model, however, is silent about how  $N$  affects the distribution of information  $\boldsymbol{\mu}(N)$  and therefore it does not offer testable implications (besides the use of mixed pricing strategies). Nevertheless, the model offers a framework for thinking about the mechanisms through which  $N$  affects prices, while the direction and magnitude of such effect remains an empirical matter.

In order to link with the empirical part of this paper we now probe into the relationship between  $\boldsymbol{\mu}$  and  $N$ . Let  $\mathbf{x}_1$  denote other determinants (e.g., economic and geographical variables) of the distribution of price information in the market. Thus, the relationship between  $\boldsymbol{\mu}$  and  $N$  is,

$$\boldsymbol{\mu} = \boldsymbol{\mu}(N, \mathbf{x}_1)$$

while the distribution of prices in equilibrium is given by  $F(p|\boldsymbol{\mu}(N, \mathbf{x}_1), v, c)$ .

If we were to observe  $\boldsymbol{\mu}$  directly we could first estimate the effect of  $\boldsymbol{\mu}$  on prices and then estimate the effect of  $N$  on  $\boldsymbol{\mu}$ . Because  $\boldsymbol{\mu}$  is not observed we cannot proceed in this way. Instead, we will use the variation in  $N$  across markets to estimate the effect of  $N$  on prices. This poses two immediate problems: we need to control for the possible correlation between  $\mathbf{x}_1$  and  $N$ , and we need to account for factors other than  $\mathbf{x}_1$  that affect both the number of firms and prices, e.g.,  $v$ . If we want to estimate the causal effect of a change in  $N$  on the distribution of prices, we need to ensure that changes in  $N$  are not accompanied by changes in other determinants of the distribution of prices  $F(p|\boldsymbol{\mu}(N, \mathbf{x}_1), v, c)$  except for direct changes in  $\boldsymbol{\mu}$ . The existence of such exogenous variation in  $N$  is crucial for interpreting the effect of a change in the number of firms on prices.

Economic theory is very useful in pinning down the determinants of  $N$  and thereby in suggesting sources of exogenous variation (Bresnahan and Reiss, 1991). In a long-run

market equilibrium with free entry, the number of stations is determined by a zero-profit condition, after covering the entry costs ( $E$ ),

$$L\pi(N, \mathbf{x}_1, v, c) - E = 0 \quad (7)$$

where  $\pi(N, \mathbf{x}_1, v, c) \equiv \pi(\boldsymbol{\mu}(N, \mathbf{x}_1), v, c)$  are profits per consumer (equation (1) divided by  $L$  and evaluated at the equilibrium price distribution).

This condition implies that the long-run market equilibrium value of  $N$  is determined by  $(\mathbf{x}_1, v, c, L, E)$ ,

$$N^* = N(\mathbf{x}_1, v, c, L, E) \quad (8)$$

It is therefore only when  $N$  changes because of changes in  $L$  and/or in  $E$  that the variation in  $N$  is exogenous to prices in the sense that nothing else affecting the price distribution  $F(p|N, \mathbf{x}_1, v, c)$  changes when such a change in  $N$  occurs. Thus,  $L$  and  $E$  are natural instruments for  $N$  since they help in determining the number of firms but, given  $c, \mathbf{x}_1$  and  $v$ , do not affect prices. Changes in  $L$  and/or in  $E$ , given  $c, \mathbf{x}_1$  and  $v$ , generate exogenous variation in  $N$  which will allow us to estimate its causal effect on prices.

## 5 Data on the number of gas stations

In order to estimate the effect of  $N$  on prices we need data on the number of all gas stations in each municipality for which we have prices. We obtained a list of *all* the gas stations and their addresses operating in the Netherlands in August 2007. These stations were assigned to municipalities according to their addresses. This allows us to know the number  $N_m$  of gas stations operating in each market  $m$ . We do not have price data on all  $N_m$  because we observe gas prices only from Athlon's lessees who do not patronize *all* the gas stations in a market. On average across markets, however, the number of stations in the sample represents 87 percent of all the gas stations in the Netherlands which is a reasonable coverage.

The mean number of stations by market is 8.2, respectively, and there is a lot of variation across markets – the standard deviation is almost as large as the mean, 7.6 stations. This variation is better seen in Table 2 where the distribution of the number of

stations per market (municipality) is tabulated.  $N_m$  ranges between 1 to 80 (Amsterdam has 59 stations and Rotterdam has 80). Sixty percent of the markets have 7 or less stations.

## 5.1 Preliminary evidence

In Table 3 we tabulate the average, over all markets having the same number of stations  $N$ , of various percentiles in the residual price distribution, as well as the averages of the minimum and maximum residual prices. The 10<sup>th</sup> and 25<sup>th</sup> percentile as well as the median price appear to decline with the number of stations, while the 90<sup>th</sup> percentile seems to be increasing in  $N$ . These effects are even stronger for the extreme prices. The magnitudes of these changes may not seem large – usually fractions of one cent – but they should be viewed relative to the small price dispersion of gasoline (residual) prices, which amounts to 1.05 cent.

Of course, simple differences in means cannot be used to infer the effect of changes in  $N$  on the distribution of prices. There are many factors that affect both  $N$  and prices and, if these are not controlled for, part of their effect is attributed to changes in  $N$ . In order to control for such confounding factors we proceed to a multivariate regression analysis of the data.<sup>33</sup>

## 6 Empirical strategy

We use residual prices in our empirical analysis;  $p_{it}$  now refers to the residual price of station  $i$  in day  $t$ . In each market  $m$ , we have  $K_m$  observations (by station and day) on residual prices  $p_{it}$ .<sup>34</sup> As explained in Section 4,  $(N_m, \mathbf{x}_{1m}, v_m, c_m)$  completely determines the equilibrium price distribution in market  $m$ . We assume that the sample

---

<sup>33</sup>Moreover, the minimum and maximum prices, being order statistics, tend to decrease and increase, respectively, with the sample size. This implies that there is a built-in tendency for the extreme prices to vary systematically with  $N$  (which is highly correlated with, but not equal to, the sample size in each market).

<sup>34</sup>The sample size  $K_m$  depends on the number of stations sampled in the market (which is always less than or equal to  $N_m$ ) and on the number of days each station appears in the sample (10.5 days on average). Thus,  $K_m$  is usually much larger than  $N_m$ . The mean value of  $K_m$  is 73.5 station-day observations.

of residual prices in market  $m$  is randomly drawn from the same distribution  $F_{N_m}(p) \equiv F(p|N_m, \mathbf{x}_{1m}, v_m, c_m)$ . This accords with the way in which prices were collected. Let  $q_m(\tau)$  be the  $\tau^{th}$  quantile of the residual price distribution in market  $m$ . The quantile  $q_m(\tau)$  is a function of  $(N_m, \mathbf{x}_{1m}, v_m, c_m)$ , the determinants of the price distribution. By analyzing how  $N$  affects  $q_m(\tau)$  at different values of  $\tau$ , we learn about the effect of changing  $N$  on the distribution of prices.

In order to do this we first estimate  $q_m(\tau)$  by the  $[n_m\tau]^{th}$  smallest value among all  $n_m$  observations in market  $m$ . We denote this estimator by  $\hat{q}_m(\tau)$ . This estimator has an asymptotic normal distribution with expected value  $q_m(\tau)$  and variance  $\frac{\tau(1-\tau)}{K_m f_{N_m}(q_m(\tau))^2}$ ,  $0 < \tau < 1$  where  $f_{N_m}$  is the conditional density of  $p_{it}$  given  $(N_m, \mathbf{x}_{1m}, v_m, c_m)$ . In a second step, we regress  $\hat{q}_m(\tau)$  on  $N_m, \mathbf{x}_{1m}, v_m, c_m$  using the municipality-level data. This provides us with an estimate of the effect of the number of stations on prices. We run a separate regression for each chosen value of  $\tau$ .<sup>35</sup>

An alternative estimator of the effect of the number of stations on prices could be obtained by estimating the quantile functions directly with the station-level price data  $p_{it}$  using a standard “quantile-regression” procedure. One should note, however, that although the two estimators are not numerically identical in finite samples, they are first-order asymptotically equivalent.<sup>36</sup> We do not adopt the quantile regression approach for a number of reasons. First, the regressors  $(N, \mathbf{x}_1, v, c)$  vary only at the level of the municipality, and this procedure might underestimate standard errors, even if the standard errors were clustered at the municipality level. Because market-level regressions are based on a much smaller number of observations (about 440 observations) than station-level regressions (about 31,000 observations), the former procedure generates

---

<sup>35</sup>See Chamberlain (1994) and Bassett, Tam and Knight (2002) for examples of this 2-step approach.

<sup>36</sup>If the function being estimated is the mean then a weighted regression of market level data produces the same estimates as those obtained from station level data. But this is not the case for quantiles. In our data, the store-level and weighted market level estimates are quite close to each other (i.e., well within 1 standard deviation of each other), when the weights are the number of observations in each market. This was also observed by Bassett, Tam and Knight (2002) in their study of ACT scores by school. See also Knight (2002) for the asymptotic equivalence results and for a comparison of both estimators using simulated data.

more conservative standard errors.<sup>37</sup> Second, the procedure based on station-level data gives more weight to the largest municipalities because the number of sample observations ( $K_m$ ) increases with the number of stations ( $N_m$ ). If the effect of the number of stations on prices is nonlinear, and this is not fully accounted for in the functional specification of the regression function, estimates based on station-level data will overly represent the effect of competition on prices in the largest markets. We want to give equal weights to all markets so as to be able to interpret the estimated coefficient as the effect of changing  $N$  on the price distribution of a market chosen at random.<sup>38</sup> Finally, it is known that the normal approximation is not adequate for extreme quantiles and values and this can affect the asymptotic distribution of the estimators based on station-level data. With market-level data, however, we are estimating an average of the extreme quantiles and values across markets so that a normal approximation applies when the number of markets is large. We therefore use unweighted market-level data to estimate the mean quantiles regressions  $E[\hat{q}_m(\tau)|N_m, \mathbf{x}_{1m}, v_m, c_m]$ .

Regarding functional form, we make a separability assumption on the conditional expectation function between  $N$  and  $(x_1, v, c)$  and specify the  $N$  part in natural logarithm, i.e.,

$$E[\hat{q}_m(\tau)|N_m, \mathbf{x}_{1m}, v_m, c_m] = \beta_0 + \beta_N \ln N_m + h(\mathbf{x}_{1m}, v_m, c_m) \quad (9)$$

The logarithmic specification implies that the marginal effect of  $N$  on a price quantile is decreasing in  $N$ . The justification for this is that a given increase in  $N$  is likely to generate a larger change in the information structure  $\boldsymbol{\mu}(N, \mathbf{x}_1)$  when  $N$  is small than when it is large. For example, adding a station to a 2-station market will likely change the distribution of information in a more drastic manner than when  $N = 10$ . The logarithmic specification is a parsimonious way of achieving this. In any case, we will check the robustness of our conclusions to alternative functional forms.

---

<sup>37</sup>An argument made by Guryan and Charles (2008).

<sup>38</sup>The station-level estimates can be interpreted as the effect of changing  $N$  on the price distribution faced by a consumer chosen at random. Although this is of interest and certainly important for welfare analysis, it is not the focus of our paper.

The most important econometric problem is that  $(\mathbf{x}_1, v, c)$  is unobserved and correlated with  $N$  (see (8)). Thus, ignoring the term  $h(\cdot)$  and treating it as error will bias our estimates of  $\beta_N$ . We approach this problem in two ways. First, we use an array of covariates  $\mathbf{x}$  to proxy for  $(\mathbf{x}_1, v, c)$  and, secondly, we use instruments for  $N$  to deal with the remaining correlation.

Specifically, we take a linear projection of  $h(\mathbf{x}_1, v, c)$  on  $\mathbf{x}$ ,

$$h(\mathbf{x}_1, v, c) = \mathbf{x}\pi + r \text{ with } Cov(r, \mathbf{x}) = 0$$

which results in

$$E[\widehat{q}_m(\tau)|N_m, \mathbf{x}_m, r_m] = \beta_0 + \beta_N \ln N_m + \mathbf{x}_m\pi + r_m \quad (10)$$

The problem with using proxies for  $h(\mathbf{x}_1, v, c)$  is that there is no guarantee that  $\ln(N)$  will be uncorrelated with the unobserved  $r$  in equation (10). The correlation between  $\ln(N)$  and the residual heterogeneity  $r$ , however, need not be strong if  $\mathbf{x}$  includes the main determinants of  $(\mathbf{x}_1, v, c)$ . That is, controlling for sufficient municipality-level characteristics can potentially ameliorate this endogeneity problem. The availability of economic, geographic and demographic data at the municipality level is one of the main reasons for defining markets as municipalities.

Nevertheless, because the proxies are not perfect, the omitted variable bias is never completely eliminated. As shown in Section 4,  $L$  and  $E$  are correlated with  $N$  but not with prices given  $(\mathbf{x}_1, v, c)$ . In order to use  $L$  and  $E$  as instruments we therefore need to assume that our proxy variables  $\mathbf{x}$  are capturing all the correlation between  $(\mathbf{x}_1, v, c)$  and  $(L, E)$  so that the instruments are uncorrelated with the residual heterogeneity  $r$ . In slightly stronger terms, we require that

$$E(r|E, L, \mathbf{x}) = 0 \quad (11)$$

This identifying assumption says that among markets with the same observed characteristics  $\mathbf{x}$ , variations in population size and entry costs are not associated with  $(v, \mathbf{x}_1, c)$ , i.e., with the willingness to pay for gasoline and with shopping behavior. If, for example, more affluent municipalities have higher willingness to pay, higher entry

costs and lower population then this assumption would be violated if we do not include measures of income or wealth among the controls  $\mathbf{x}$ . As usual, the strength of this assumption depends on what is included in the vector of proxies  $\mathbf{x}$  (we describe  $\mathbf{x}$  in detail in Section 7).

For the number of consumers  $L$  to be a valid instrument, the unobserved (part of) marginal cost must be independent of market size. In connection with this, we note that variable costs in gasoline retailing are mostly driven by the cost of gasoline. The typical brand in the Netherlands buys its gasoline from the Amsterdam-Rotterdam-Antwerp (ARA) spot market (this is true even for Shell which sells much more gasoline than it produces). The ARA market is a centralized marketplace where price discrimination mechanisms such as quantity discounts are unfeasible due to the anonymity of the traders. Therefore, it is reasonably safe to assume that most gas stations in the Netherlands face similar wholesale gasoline prices irrespective of the population level in the areas where the stations are located.

The standard errors of the estimators need to account for the heteroskedasticity induced by the sampling error in estimating the quantiles,  $\frac{\tau(1-\tau)}{K_m f_m(q_m(\tau))^2}$ . Instead of estimating the density function we use (White) standard errors that are robust to arbitrary heterogeneity.<sup>39</sup>

## 7 Empirical Results

Panels A-D in Table 4 present estimates of several variants of equation (10), as well as similar regressions for the minimum and maximum price in each market. The regressions for each price statistic are run separately since there are no efficiency gains to joint estimation when the regressors are the same across equations. Panel A presents OLS estimates of regressing a price statistic on  $\ln(N)$  only. These regressions are based on 439 observations (municipalities) because residual prices could not be computed for one

---

<sup>39</sup>As a check we also bootstrapped the standard errors of the 2SLS regressions by resampling from the  $(\hat{q}_m(\tau), N_m, \mathbf{x}_m)$  data. These bootstrapped standard errors, based on 1000 replications, are between 8 and 20 percent (15 percent on average) higher than the White standard errors. We chose to use the latter because they are easier to compute and do not alter any of our conclusions regarding statistical significance.

municipality (Reiderland) where only one station has prices in only one day. The effect of  $N$  is negative for the lower price percentiles and positive for the higher ones, as predicted by Example 2 of the theoretical model. The more extreme the price statistics, the more significant are the effects of  $N$ . The median price also decreases with  $N$  but this effect is marginally significant. As  $N$  increases, say from 1 to 2 stations, the minimum price is estimated to decrease by 0.35 cents ( $-0.508 \times \log 2$ ), while the maximum price is estimated to increase by 0.26 cents. These are not small changes relative to the standard deviation in residual prices (1.05 cent).

In Panel B we add 39 provincial dummies to control for unobserved time-invariant effects at the regional level.<sup>40</sup> These regional effects are always jointly significant at the 1 percent level (also in panels C and D). The estimated coefficients of  $\ln(N)$  increase somewhat, particularly for prices in the middle of the distribution.

In Panel C we add proxies for consumers' reservation values ( $v$ ) and for shopping behavior ( $\mathbf{x}_1$ ). We do not directly proxy for  $c$  because, as mentioned above, variable production costs are quite similar across markets. The regressions in panel C and D are estimated on 423 markets because of missing data on some of the covariates. The reasons for missing covariate data are unrelated to the price of gasoline and therefore there is no risk of sample selection bias. Indeed, Panel B was reestimated for the sample of 423 observations used in Panel C and D and the estimated coefficients are very similar to those reported in the table.

Perhaps among the main determinants of the willingness to pay and of shopping behavior for gasoline is income. We therefore include average household income as proxy for both  $v$  and  $\mathbf{x}_1$ . Because of income and substitution effects we expect this variable to be positively correlated with the willingness to pay and with the share of non-price sensitive consumers ( $\mu_1$ ). Thus, income should positively affect prices. We also include the share of cars registered to business (out of total cars in the municipality), which should be positively correlated with the willingness to pay for gas and therefore also affect prices positively.

---

<sup>40</sup>There are 40 regional areas (known as COROP areas) in the Netherlands. Each regional area comprises several municipalities.

An additional set of controls is related to the geographic or spatial characteristics of markets. The distribution of price information may vary with the geography of the market. Consumers' shopping behavior may be different in a geographically small, interconnected municipality than in a large, spatially-spread municipality. We therefore add controls for the total area of the municipality (in  $km^2$ ), the area that is land (also in  $km^2$ ), the share of land that is built (urbanized) and the share that is agrarian (the remainder is land for recreation and forests), and the kilometers of roads within the municipality borders.<sup>41</sup>

We also add the sample size  $K_m$  to the list of regressors. We do this because the sample minimum and maximum are monotonic functions of the sample size, while  $K_m$  is correlated with the number of stations in the market (the simple correlation between  $K_m$  and  $\ln(N_m)$  is 0.72), as well as to control for any sample size effects in the estimation of the quantiles. In this way, the estimates of  $\beta_N$  are less likely to reflect the built-in correlation between the price statistics and sample size.

The effect of adding these additional regressors is to lower the estimates of  $\beta_N$ , particularly for the lower prices. In addition, the precision of the estimates decreases because the additional estimated parameters do not contribute much to the regressions' explanatory power (we expand on this below) but reduce the left-over variation in  $\ln N$ . As a result, most of the estimated  $\beta_N$ 's, except in the minimum and maximum price regressions, are not significantly different from zero.

Our final set of regressions in panel D uses 2SLS to eliminate potential biases from unobserved common determinants of both prices and  $N$ . An additional reason for using 2SLS is that the number of stations is likely to be measured with error because our data for  $N$  correspond to stations operating during August 2007, while our price data were collected in May 2006.

As explained in Section 4, free entry and a zero profit condition predict a positive relationship between the number of stations in the market and population size  $L$ , and a negative relationship between  $N$  and entry costs. We do not have data on market-specific

---

<sup>41</sup>We would also like to have a measure of the distance between stations in a municipality but unfortunately we do not have these data.

entry costs but we have data on the level of municipality taxes imposed on business real estate and use these as a measure of  $E$ .<sup>42</sup> First-stage regression results appear in Table 5. In columns (1) we regress the number of stations (in logs) on population and the tax rate (both in logs), while in column (2) we add 39 provincial dummies.<sup>43</sup> Both instruments have coefficients with the predicted sign and are significantly correlated with  $\ln(N)$ . As controls are added to the regression, in columns (3) and (4), the effect of taxes is halved and loses its statistical significance. Nevertheless, in all regressions, the F-test for joint significance of population and taxes is very high indicating that these are strong instruments. Column (4) corresponds to the first-stage in the 2SLS procedure used in Table 4.<sup>44</sup> The overidentification tests in panel D in Table 4 do not reject the assumption that the instruments are excluded from the price equations.<sup>45</sup>

Panel D in Table 4 reports the 2SLS estimates of  $\beta_N$  using the same specification as in panel C. The 2SLS results are in line with the previous estimates but they are larger in absolute value than the OLS estimates in panel C; these differences are quantitatively important. On the one hand, because unobserved determinants of prices are likely to be positively correlated with the number of stations, using 2SLS should increase the absolute value of the negative estimates of  $\beta$  in the minimum and low percentiles' regressions and decrease the positive estimates of  $\beta$  in the maximum and high percentiles' regressions. On the other hand, in the presence of measurement errors in  $N$ , OLS estimators are biased towards zero in all the regressions. Since 2SLS also removes the correlation with the measurement error, 2SLS estimates of the coefficient in the lower price regressions

---

<sup>42</sup>Tax rates vary between 1.5 and 18 percent across municipalities with an average of 7.1 percent. We also experimented with multiplying the tax rate by the average value of land in the municipality to generate another proxy for  $E$ ; the final 2SLS estimates were almost identical to those obtained using the tax rate only.

<sup>43</sup>Using the tax rate in levels instead of logs works equally well. We treat  $L$  and  $E$  symmetrically as it would be suggested by a logarithmic approximation to the zero profit condition (7).

<sup>44</sup>There are 424 municipalities with data on all covariates but one of them (Reiderland) does not have residual price data. Thus, the total number of observations used in the price regressions in Table 4 is 423.

<sup>45</sup>Using  $\ln L$  as the only instrument for  $\ln N$  gives essentially the same estimates as in Table 4. But using only the tax rate as the sole instrument gives much larger (in absolute value) 2SLS estimates of  $\beta_N$  but also with much larger standard errors.

should be, on both accounts, more negative than the OLS estimates. This is indeed what we observe in panel D. For the higher price regressions, the biases in OLS due to omitted variables and to measurement error work in opposite directions and it is therefore not possible to predict in which direction the estimates should change with 2SLS.

The marginal effects of an increase in the number of stations from  $N$  to  $N + 1$  is

$$\beta_N \ln \left( \frac{N + 1}{N} \right).$$

These marginal effects are plotted in Figure 8 for  $N = 1, \dots, 20$ , along with a 2 standard deviation band. We observe that for small values of  $N$ , the marginal effects are indeed positive for the higher prices and negative for the smaller prices. These effects are also significantly different from zero. The logarithmic specification implies that marginal effects converge to zero as  $N$  increases.

Using these estimates we find, for example, that adding 1 additional gas station to a two-station market would lower the minimum price of a liter of Euro 95 by 0.24 cents ( $-0.58 \times \log 1.5$ ) and increase the maximum price by 0.81 cents. Recall that the standard deviation in the residual price distribution is 1.05 cents and therefore these estimated effects are quantitatively significant relative to the dispersion in prices.

If we were estimating the mean residual price, the effect of the market-level covariates would be identically zero because they are orthogonal to the residual price by construction. However, because we are estimating quantiles of the residual price distribution these effects need not be zero. Nevertheless, the estimated coefficients of the control variables (not reported) are usually not significantly different for zero, both individually and jointly. This is driven in part by the inclusion of regional dummies in the regression which correlate with the municipality-level characteristics. Indeed, if the provincial dummies are removed from the regressions in panel D, the controls are jointly significant in four of the seven regressions (while the estimates of  $\beta_N$  are virtually unaffected). Moreover, as we will see later, when the actual instead of the residual prices are used to compute the extreme prices and the quantiles, the controls are significant in all regressions (see bottom panel in Table 7).

Finally, we also estimate  $\beta_N$  at percentiles closer to the bounds of the price distri-

bution, namely, the 1<sup>st</sup> to 5<sup>th</sup> and 95<sup>th</sup> to 99<sup>th</sup> percentiles. Table 6 presents these results. The estimated effect of  $N$  on prices follows the same pattern as that in Table 4: negative for the low prices and positive for the high prices.

In Tables 4 and 6 we also note that the estimated  $\beta_N$  increase gradually in (absolute) size as we move from the 25<sup>th</sup>(75<sup>th</sup>) percentile regression to the minimum (maximum) price regression, while they have about the same (absolute) size between the 25<sup>th</sup> and 75<sup>th</sup> percentile regressions. Thus, increased competition is felt most strongly at the extremes of the price distribution.

An alternative interpretation of these results is that the estimated effects reflect our inability to fully control for the effect of sample size on the extreme prices and quantiles. Note, however, that our market sample sizes are quite large (50 percent of the markets have values of  $K$  above 53 observations; the minimum value of  $K$  is 7 while the maximum is 942) and therefore variations in  $K_m$  across markets should have minimal effects, if at all, on the estimated extreme prices and quantiles. Although estimates of the minimum and of maximum prices, and of extreme quantiles as well, depend on the sample size, these estimates will not change much if the sample size is already large. That is,  $K$  has a diminishing marginal effect. We therefore do not expect that variations in extreme prices and quantiles across markets reflect variations in sample size. In any case, we do control for sample size in the regressions but, as expected, the coefficient of  $K_m$  is never significantly different from zero, and excluding the sample size from the regressions does not affect any of the estimates of  $\beta_N$ . Moreover, when we add a quadratic of sample size ( $K_m^2$ ) to the regressions the estimates are not much affected and, if anything, they become slightly stronger (results not reported). Thus, we do not believe that variation in sample size across markets is driving our results.

To provide further support for this claim we estimated the specification in Panel D of Table 4 on extreme percentiles and prices computed from a sample consisting of randomly selected half the original number of observations in each market. If sample sizes are important for the estimation of  $\beta_N$ , then using half the sample should certainly affect the estimated parameters. We repeated this estimation 1,000 times – each time drawing 50 percent of the observations in each market – and averaged the 1,000 estimates

of  $\beta_N$ . For the minimum price regressions, this gives a mean estimate of  $\beta_N$  equal to  $-0.55$  with a standard deviation of  $0.11$ . For the maximum price regressions, the mean estimate of  $\beta_N$  is  $0.59$  with a standard deviation of  $0.08$ . The corresponding mean estimates for the  $10^{th}$  and  $90^{th}$  percentiles are  $-0.06$  (standard deviation  $0.046$ ) and  $0.13$  (standard deviation  $0.062$ ), respectively. Thus, the estimates of  $\beta_N$  based on samples half the sample size do not differ substantially from the original estimates in panel D of Table 4 supporting our conclusion that sample sizes in these data are large enough for them not to matter.

In terms of the theoretical model presented in Sections 2 and 4 we observe that Figure 6 (Example 2) predicts that the high percentiles of the price distribution increase with  $N$  while the low percentiles decrease with  $N$ . This is what we found in the data. Figure 6 also shows that the effect of  $N$  on the low percentiles increases as we move towards the lower bound; this “second-order” effect is also borne by the data. However, the effect of  $N$  on the very high percentiles, although positive, decreases with  $N$  and this is not what we see in the data. The reason for this difference in the theoretical results is that the lower bound of the price distribution decreases with  $N$  while the upper bound remains unchanged at  $v$ . Clearly, the model ought to be modified if we want to accommodate this “second-order” effect at the very high quantiles. It seems to us that the model should allow for the upper bound of the price distribution to increase with  $N$ , which could be obtained if consumers’ reservation value were made endogenous. We believe, however, that for the purposes of organizing and interpreting the empirical work the present modelling framework suffices.

In sum, the empirical findings suggest that as the number of stations in the market increases, the low prices tend to decrease while the high prices tend to increase. Although we cannot check what happens to the mean residual price as  $N$  increases (but see next Section for the effect of  $N$  on the mean raw price), we found that the median price is lower in markets with more stations. This characterization of the effect of competition on prices accords with models where some consumers have imperfect information about prices and observe different number of prices but is not consistent with the predictions of standard oligopoly models. To the best of our knowledge, this asymmetric effect

of a change in the number of firms has not been analyzed empirically. The welfare implications of such asymmetry will be studied in Section 8. We first perform a set of robustness checks.

## 7.1 Robustness checks

We now examine some of the assumptions underlying our baseline specification in panel D of Table 4 and verify the robustness of our conclusions to departures from these assumptions.

We first address functional form issues. Although entering the number of stations in logarithmic form is parsimonious as well as theoretically appealing – because  $N$  is likely to have a smaller impact on  $\mu(N, \mathbf{x}_1)$  when  $N$  is large – it may be practically restrictive. We therefore allowed the coefficient of  $\ln N$  to change for  $N \geq N_0$ , for various levels of  $N_0$  ( $N_0 = 2, \dots, 17$ ), but the interaction term was never large and usually not significantly different from zero.<sup>46</sup>

We also added the square of  $\ln(N)$  to the regression to allow for more flexibility in the marginal effect but this term was never significantly different from zero (its  $p$ -value ranged between 0.34 and 0.99) except in the maximum price regression ( $p$ -value 0.06).<sup>47</sup> Adding  $(\ln(N))^2$  made the coefficient of  $\ln(N)$  also insignificant. This is not surprising because  $\ln(N)$  and  $(\ln(N))^2$  are highly correlated; their simple correlation coefficient is 0.95. Although the individual parameters cannot be precisely estimated, the marginal effects track very closely the marginal effects estimated from the regression in panel D of Table 4 except, perhaps, for those of the 75<sup>th</sup> percentile price (see Figure 9).

We could avoid making strong functional form assumptions if we allow for the effect of  $N$  to vary non-parametrically with  $N$ . This can be achieved by using dummy variables for each value of  $N$ . The problem here is that  $N$  takes on 36 distinct values and the corresponding dummies would still be endogenous. Even if we had the large number of

---

<sup>46</sup>The main exception was the interaction coefficient in the median price regression which was significant in six out of the 16 cases. Interactions between the instruments (population and the tax rate) and the threshold dummies were added to the list of instruments.

<sup>47</sup>We added the squares of log population and log tax to the list of instruments.

instruments required (or use a control function approach), this approach is not practical given our sample size. We therefore group the number of stations into 4 size groups and add dummies corresponding to these groups. The four groups are defined as markets with 1 and 2 stations – the baseline group –, markets with 3-6 stations, markets with 7-10 stations and markets with more than 11 stations. In order to address the endogeneity of these group dummies we follow the procedure suggested by Wooldridge (2002, p. 623) and first estimate a probit equation for the probability that the number of stations in a market is in a given size group. We run a separate regression for each size group and compute the predicted probability of belonging to a size group. In this regression we include the same regressors as in the first stage of the 2SLS estimator in panel D of Table 4. We then run the regressions as in panel D using the predicted probabilities as instruments for the endogenous group dummies.<sup>48</sup>

Results of this two-stage 2SLS estimation are presented in the top panel of Table 7, where the coefficients represent the change in price in a given group size relative to the preceding group size. We see that markets with 3-6 stations have lower low prices but higher high prices than markets with 1-2 stations. Markets with 7-10 stations exhibit the same pattern, relative to markets with 3-6 stations, but the effects are of lower magnitude and less significant. Finally, the estimates indicate that prices in markets with 11 or more stations are not significantly different from prices in markets with 7-10 stations. These findings accord, at least in a qualitative sense, with the marginal effects depicted in Figure 8. In sum, changing the simple functional form used in Table 4 would not change our conclusions regarding the asymmetric effect of competition.

Recall that in (9) we made a separability assumption. We test for this assumption as follows. We added the interaction between  $\ln N$  and each of the controls (except the provincial dummies) one at a time and estimated the model by 2SLS adding the interaction between the control and  $\ln L$  to the list of instruments. In no case were the interaction terms significant. This test, albeit partial, suggests that the separability

---

<sup>48</sup>Wooldridge (2002, p. 623) shows that the first-stage probit regressions need not be correctly specified, and that inference based on the standard errors of the 2SLS procedure is correct even though the instruments are generated in a previous step.

assumption cannot be rejected in these data.

Next, we examine what happens to our estimates of  $\beta_N$  when the dependent variables are based on the actual (raw) prices and not on the prices net of station and day effects. We do not pool observations over time because the wholesale price may be changing over the sample period and therefore compute the percentiles and extremes of the price distribution for each market and for each day. We can now also compute a mean price for each market-day. But we cannot control for unobserved station-level effects because doing so wipes out all market-level regressors, including  $\ln(N)$ . We now have 7091 market-day observations. We estimate the same model as in panel D of Table 4.<sup>49</sup> The results using raw prices, which now include a regression for the mean price, are in the bottom panel of Table 7. The mean and median price do not appear to be significantly affected by the number of stations in the market but the low and high prices are. The estimated parameters follow the same pattern as in Table 4 but are much stronger than the estimates based on residual prices. This is to be expected simply because the distribution of raw prices in a market is centered around higher prices and is more dispersed than the distribution based on residual prices. Using raw prices in our analysis would then pick up the impact of competition on the mean level of prices which, although of interest and important for welfare, is not the main focus of this paper.

In contrast to the regressions in Table 4 where the dependent variables were based on residual prices, the covariates are now jointly significantly different from zero in all regressions. Average household income and the share of cars registered to business always have positive coefficients but only the latter are significant. The geographic controls are significant in four of the seven regressions. Interestingly, the sample size variable is negative and significant in the minimum price and the 10<sup>th</sup> quantile regressions while it is positive and significant in the 90<sup>th</sup> quantile and the maximum price regressions; it is not significantly different from zero in the other regressions (results not reported). This is as

---

<sup>49</sup>The only differences with the specification in Table 4 are that the dependent variable and the sample size regressor change over days, and that we added day dummies to control for the effect of the day in the month. The other regressors are constant over time. Standard errors were clustered at the market level to allow for arbitrary serial correlation and heteroskedasticity.

expected because now the sample size ( $K_{mt}$ ) in these regressions is quite small – it varies between 1 and 67 with a median of 3 stations ( $K_{mt}$  is now the number of stations in the sample by market and day). Thus, it is not surprising that sample size is significant in the regressions for the quantiles and extreme prices based on the raw price data. In fact, this observation reinforces our previous observation that sample size is not important when the sample size is large (as it is in Tables 4 and 6).

The number of stations in the market is defined as the number of stations in the municipality. It may well be that the “relevant” number of stations affecting prices in a market includes the stations in neighboring municipalities. In order to examine this possibility we computed, for each market  $m$ , the number of stations in all the municipalities sharing a border with market  $m$  and added the logarithm of this variable to the basic model. The results appear in the top panel of Table 8. We drop 3 municipalities that are islands and therefore have no neighbors. Essentially, the number of gas stations in neighboring markets has a much smaller effect on prices than the own number of neighbors and, in all cases, this effect is not significantly different from zero. Importantly, the estimated  $\beta_N$ ’s are almost unaffected by the inclusion of the number of neighboring stations in the regression.<sup>50</sup>

We also re-run the regression excluding the four largest markets in the Netherlands (Amsterdam, Rotterdam, ’s-Gravenhage, and Utrecht). Since these cities represent only 4 observations we do not expect to obtain very different results. And, indeed, the estimated coefficients based on restricted sample of smaller cities are very similar to those in panel D of Table 4.<sup>51</sup> In this vein, recall that the theoretical model presented in Section 2 does not examine markets with a single station. There are 16 municipalities where  $N = 1$ . We included these markets in the sample because it is quite likely that consumers in these markets do in fact observe more than 1 price (i.e., they observe prices in other municipalities as they travel to work). In any case, removing these observations

---

<sup>50</sup>Because provincial dummies pick up regional effects, we treat the number of neighboring stations as exogenous in the price regressions. The overidentification tests support this assumption.

<sup>51</sup>These new estimates, in the order in which they appear in Table 4, are: -0.669 (0.231), -0.131 (0.0595), -0.0435 (0.0395), -0.0568 (0.0337), 0.0276 (0.0569), 0.105 (0.0721) 0.523 (0.145).

from the regressions makes each of the estimated coefficients even stronger.<sup>52</sup>

Finally, our findings are not restricted to a particular gas product (Euro 95). The other popular product in gas stations is, of course, Diesel. The bottom panel in Table 8 replicates the regression in panel D of Table 4 for residual Diesel prices. The estimated coefficients are remarkably similar to those from the Euro 95 regressions.

## 8 Welfare implications

The evidence presented in Section 6 points to significant differences in the way increased competition – increased number of gas stations – affects different parts of the price distribution. Whether consumers are successful in paying the lower prices depends on their shopping behavior, i.e., on the number of prices they observe. Increased competition is likely to favor more those consumers observing many prices because they are more likely to observe one of the lower prices, while those observing only a few prices may not be that lucky. Moreover, since the frequency of high prices may increase with the number of firms some consumers, at least theoretically, may end up paying higher prices. It is therefore not obvious – in contrast to a full-information model – that all consumers benefit and, when they do, whether they benefit in the same way from increased competition. In this section we use the model presented in Section 2 to study and quantify the welfare gains from increased competition for different types of consumers.<sup>53</sup>

Denote the price paid by a consumer who observes  $s$  prices by  $y_s = \min \{p_1, p_2, \dots, p_s\}$  where  $p_1, p_2, \dots, p_s$  are i.i.d. random variables drawn from  $F_N(p)$ . For example, consumers that observe only one price pay, on average, the mean price  $E[p]$ . The distribution of  $y_s$  is  $G(y_s) = 1 - (1 - F_N(y_s))^s$ . As in Section 4, we can derive the inverse of the distribution of this statistic:

$$y_s(\boldsymbol{\mu}(N), v, c) = c + \frac{\mu_1(N)(v - c)}{\sum_{\ell=1}^N \ell \mu_\ell(N) (1 - G)^{\frac{\ell-1}{s}}},$$

---

<sup>52</sup>The estimated coefficients, in the order appearing in Table 4, are -0.712 (0.204), -0.099 (0.046), -0.0766 (0.033), -0.0641(0.028), 0.0900 (0.041), 0.1472 (0.059), 0.5473 (0.117) and are slightly more significant as those in panel D of Table 4.

<sup>53</sup>Note that we ignore non-price effects of competition such as better quality of service, shorter distances to gas stations, etc., so that “welfare effects” refers only to price effects.

where  $G$  takes values on  $[0, 1]$ .

In this section, we analyze the effect of increased competition on the average price paid by the consumer (the average transaction price). The expected price paid by a consumer who observes  $s$  prices is,

$$\begin{aligned} E[y_s(\boldsymbol{\mu}(N), v, c)] &= \int_{-\infty}^{\infty} sz(1 - F_N(z))^{s-1} f_N(z) dz \\ &= c + \int_0^1 \frac{\mu_1(N)(v - c)}{\sum_{\ell=1}^N \ell \mu_{\ell}(N)(1 - G)^{\frac{\ell-1}{s}}} dG \end{aligned}$$

As shown in Section 4, how  $E[y_s(\boldsymbol{\mu}(N), v, c)]$  changes with  $N$  will depend on how  $\boldsymbol{\mu}(N)$  changes into  $\boldsymbol{\mu}(N + 1)$ . We go back to our examples from Section 4.

**Example 3** (continuation of Example 2). Recall that when  $N = 2$  we assumed  $\boldsymbol{\mu}(2) = (0.5, 0.5)$  and when  $N = 3$  we had  $\boldsymbol{\mu}(3) = (0.45, 0.1, 0.45)$ . Let us suppose that  $N = 4$ , and we have  $\boldsymbol{\mu}(4) = (0.43, 0.05, 0.05, 0.47)$ . We numerically calculate the average price paid  $E[y_s(\boldsymbol{\mu}(N), v, c)]$  by the different types of consumers  $s = 1, 2, 3, 4$  and for  $N = 2, 3, 4$ . We obtain,

	Expected price paid		
	N=2	N=3	N=4
s=1	0.549	0.554	0.585
s=2	0.451	0.415	0.423
s=3	--	0.353	0.343
s=4	--	--	0.298

Two main points deserve our attention. First, the price paid by some consumers may increase, rather than decrease as  $N$  increases. This is the case for consumers who are exposed to one price, as well as for consumers exposed to two prices when  $N$  increases from 3 to 4. Second, the price paid by some consumers may be non-monotonic in  $N$  as in the case of consumers who observe two prices. In this example these consumers benefit when there is entry in a market with two firms, but they pay a higher price when there is entry in a market with three firms.

**Example 4** (continuation of Example 1). Under the assumption that  $\frac{\mu_\ell(N+1)}{\mu_1(N+1)} \geq \frac{\mu_\ell(N)}{\mu_1(N)}$  for all  $\ell$ ,  $F_N$  dominates  $F_{N+1}$  in a first-order stochastic sense and, by implication, so does  $G_N$  over  $G_{N+1}$ . As a result, the average price paid by all consumer types,  $E[y_s]$ , will decrease in  $N$ .

We estimate  $E[y_s]$  for each market as follows. We draw  $s$  residual prices with replacement from the sample of  $K_m$  residual prices observed in each municipality (pooled over gas stations and days).<sup>54</sup> We take the minimum of the  $s$  prices and store it. We repeat this 10,000 times and compute the average of the 10,000 stored minimum prices. This average is our estimate of  $E[y_s]$  for each  $s = 1, 2, \dots, N$  in each market (characterized by a given  $N$ ). That is, we obtain  $N$  estimates of  $E(y_s)$  in each market corresponding to the expected price paid by different consumers observing, respectively,  $s = 1, \dots, N$  prices.

There are two dimensions of these estimates that are of interest. First, as consumers observe more prices in a given market, the price they end up paying should be lower on average. That is, broader price exposure should result in lower prices paid. We clearly see this in Figure 10 where we plot the estimates of  $E[y_s]$  in each market against  $s$ , as well as the predicted value of a locally weighted regression of the estimate of  $E[y_s]$  on  $s$ . The gains from being better informed – the difference in expected price paid as  $s$  increases by 1 – are positive in 99.4 percent of the observations.

Although the path of expected prices paid in each market declines with  $s$ , Figure 10 points out that there is substantial heterogeneity in the price paid for given  $s$  across markets. This heterogeneity is a reflection of the different price distributions across markets having different number of stations. This is precisely the other issue of interest – and the focus of this Section – namely, the relationship between  $N$  and the price paid,  $E[y_s]$ , for given  $s$ .

To address this issue, we regress our estimate of  $E[y_s]$  on  $\ln(N)$  and on the other controls used in the previous regressions. We run a separate 2SLS regression for various

---

<sup>54</sup>We report estimates based on prices drawn with replacement. Since, in reality, consumers do not sample with replacement we also replicated our calculations when prices are drawn without replacement; the results are practically identical.

values of  $s$ . Note that when  $s = 1$  the estimate of  $E[y_1] = E[p]$  is the mean price in the market which is zero by construction. We therefore present, in Table 9, the estimated effect of  $\ln(N)$  on  $E[y_s]$  for  $s = 2, \dots, 7$ , i.e., for consumers that observe up to 7 prices. Note also that as  $s$  increases, the number of observations declines because there are fewer municipalities with  $N$  above  $s$  (see Table 2) and the parameters are not precisely estimated for  $s \geq 5$ .

Two results in Table 9 are noteworthy. First, the estimates are all negative. A negative coefficient means that the prices paid by consumers decrease as the number of competitors increases. All types of consumers benefit from increasing the number of gas stations. Second, the negative effect of  $N$  increases with the number of price observations ( $s$ ) up to, and including,  $s = 4$  but stabilizes thereafter. This means that the gains from increased competition – in terms of price reduction – are maximal for consumers observing 4 prices. Entry of additional stations does not result in additional gains for consumers who observe 5 or more prices. In sum, the magnitude of the welfare improvement depends on shopping behavior: it is about twice as large for consumers that observe 4-5 prices than for consumers that observe only 2 prices.

Finally, we note that the welfare analysis was based on residual prices. It is conceivable that the stations' characteristics and/or their productivity also respond to competition and that this response is reflected in the stations' mean prices. By focusing on the residual prices we may be missing these kind of effects. In our data, however, we may not be missing much because we did not find significant effects of the number of stations on the mean price of gasoline (bottom panel in Table 7).

## 9 Conclusions

In markets where the amount of price information varies across consumers, prices are typically dispersed in equilibrium. An increase in the number of firms usually affects each seller's pricing strategy and this in turn alters the entire distribution of equilibrium prices. Traditionally, empirical research has focused on estimating the impact of competition on

the mean and variance of prices.<sup>55</sup> Although this is certainly useful, these statistics are not sufficient to perform a detailed welfare analysis because competition can affect different parts of the price distribution in opposite directions.

This paper has tried to fill this gap. We examined how the distribution of gasoline prices in the Netherlands changes with the number of competitors in the market. We used population size and local taxes as instruments for the number of gas stations. We found that as competition – the number of gas stations – increases the distribution of prices spreads out, with the low prices going down and the high prices going up. Consequently, *competition has an asymmetric effect on prices.*

This result has important welfare implications because when some prices increase and others decline, the price actually paid by consumers will depend on their shopping behavior. All (hypothetical) consumers in our data, irrespective of whether they are informed about one or more prices, benefit from an increase in the number of stations. The magnitude of the welfare gain, however, is greater for those consumers that observe more prices. As a result, *an increase in competition has a positive but unequal effect on the welfare of consumers.*

Our empirical strategy and interpretation of the results are closely linked to the theoretical model presented in Section 2 which we believe to be appropriate to the retail gasoline market. The model makes clear that increased competition has an effect on prices only when it increases the amount of price information consumers have. The paper can therefore be also interpreted as examining the effect of changes in consumers' information on the distribution of prices. Thus, policies aimed at increasing the amount of price information (e.g., through mandatory price labels) or at lowering the costs of disseminating and gathering information (e.g., through the Internet) can affect the distribution of prices and welfare.

Since price dispersion is prevalent in many markets, we believe the paper has a general message that goes beyond the present application to the gasoline market in the

---

<sup>55</sup>In our data, the standard deviation of residual prices increases with the number of stores. The 2SLS coefficient of  $\ln(N)$  in a regression specification similar to those in panel D of Table 4 is 0.086 (s.e. 0.033).

Netherlands. The price effects of competition-enhancing policies (e.g., industry deregulation, trade liberalization, etc.) are not as straightforward as one may be led to believe based on standard oligopoly theory. As a result, welfare implications are not obvious either. In fact, we have shown, theoretically and empirically, that increased competition can have unequal effects among consumers; some consumers may even experience declines in their welfare as a result of higher prices.

In order to identify which consumers benefit more and which benefit less from increased competition in gasoline prices we would require a mapping between shopping behavior and socio-economic characteristics of interest. If such data were available we would be able to say something about how the distribution of the benefits from increased competition varies with income. For example, consumers that observe only a few prices may be high-income consumers (whose value of time is higher) and these consumers may benefit less from competition than low-income consumers. This, however, is beyond the scope of this paper and is left for future research. Moreover, a complete welfare analysis, should recognize the effect of increased competition on other dimensions of consumer welfare such as increased variety, quality and accessibility.

Lastly, although our empirical work is motivated by a particular theoretical framework, we think the empirical findings reported in the paper are of interest on their own right and, if verified in other data sets, they should be taken into account when formulating theoretical models of pricing in oligopolistic markets.

## References

- [1] Armstrong, Mark (2008), “Interactions between Competition and Consumer Policy,” *Competition Policy International* 4(1), 97-147.
- [2] Baye, Michael. R., John Morgan and Patrick Scholten (2004), “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.
- [3] Baye, Michael. R., John Morgan and Patrick Scholten (2006), “Information, Search, and Price Dispersion, chapter 6 in *Handbook on Economics and Information Systems Volume 1*, (T. Hendershott, Ed.), Amsterdam: Elsevier.
- [4] Bassett, Gilbert, Mo-Yin S. Tam and Keith Knight (2002), “Quantile Models and Estimators for Data Analysis”, *Metrika*, 55, 17-26.
- [5] Barron, John M., Taylor, Beck A. and John R. Umbeck (2004), “Number of sellers, average prices, and price dispersion”, *International Journal of Industrial Organization* 22(8-9), 1041-1066.
- [6] Borenstein, Severin and Nancy Rose (1994), “Competition and Price Dispersion in the U.S. Airline Industry,” *Journal of Political Economy*, 102, 653-683.
- [7] Bresnahan, Timothy F. and Peter Reiss (1991), “Entry and Competition in Concentrated Markets”, *Journal of Political Economy*, 99(5), 977-1009.
- [8] Chamberlain, Gary (1994), “Quantile regression, censoring and the structure of wages”, chapter 5 in *Advances in Econometrics, Sixth World Congress*, edited by Christopher Sims, Cambridge University Press.
- [9] Chandra, Ambarish and Mariano Tappata (2008), “Price Dispersion and Consumer Search in the Retail Gasoline Market”, <http://strategy.sauder.ubc.ca/tappata/research.htm>.

- [10] Janssen, Maarten and José Luis Moraga-González (2004), “Strategic Pricing, Consumer Search and the Number of Firms”, *Review of Economic Studies*, 71(4), 1089-1118.
- [11] Gerardi, Kristopher S. and Adam Hale Shapiro (2009), “Does Competition Reduce Price Dispersion: New Evidence from the Airline Industry”, *Journal of Political Economy*, 117(1), 1-37.
- [12] Guryan, Jonathan and Kerwin Charles (2008), “Prejudice and Wages: An Empirical Assessment of Becker’s The Economics of Discrimination”, *Journal of Political Economy*, October 2008, 116(5), 773-809.
- [13] Goldmanis, Maris, Hortaçsu, Ali, Chad Syverson and Onsel Emre (2009), “E-commerce and the Market Structure of Retail Industries”, *Economic Journal*, forthcoming.
- [14] Haile, Philip A., Han Hong, and Mathew Shum (2003), “Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions”, <http://www.econ.yale.edu/~pah29/npmcv.pdf>.
- [15] Hortaçsu, Ali and Chad Syverson (2004), “Product Differentiation, Search Costs and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds”, *Quarterly Journal of Economics*, 119, 403-456.
- [16] Hosken, Daniel S., Robert S. McMillan and Christopher T. Taylor (2008), “Retail Gasoline Pricing: What Do We Know?”, *International Journal of Industrial Organization*, 26(6), 1425-1436.
- [17] Knight, Keith (2002), “Comparing conditional quantile estimators: first and second order considerations”, <http://www.utstat.utoronto.ca/keith/home.html>.
- [18] Lach, Saul (2002), “Existence and Persistence of Price Dispersion: an Empirical Analysis”, *Review of Economics and Statistics*, August, 433-444.

- [19] Lewis, Matthew (2008), “Price Dispersion and Competition with Differentiated Sellers”, *Journal of Industrial Economics* 56 (3), 654-678.
- [20] NMa (2006): Benzinescan 2005/2006, De Nederlandse Mededingingsautoriteit.
- [21] Roberts, Mark and Dylan Supina (2000), “Output Price and Markup Dispersion in Producer Micro Data: The Roles of Producer Heterogeneity and Noise”, *Advances in Applied Microeconomics, Vol 9, Industrial Organization*, Michael Baye (ed.), JAI Press.
- [22] Stahl, Dale O. (1989), “Oligopolistic Pricing with Sequential Consumer Search”, *American-Economic-Review*, 79(4), 700-712.
- [23] Varian, Hal (1980), “A Model of Sales”, *American Economic Review*, 70, 651-659.
- [24] Waterson, Michael (2003), “The Role of Consumers in Competition and Competition Policy,” *International Journal of Industrial Organization* 21, 129-50, 2003.
- [25] Wildenbeest, Matthijs R. (2008), An empirical model of search with vertically differentiated products, mimeo.
- [26] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge.
- [27] Wu. X. and J. Perloff (2007), “Tax Incidence Varies Across the Price Distribution,” *Economics Letters*, 96, 93-99.

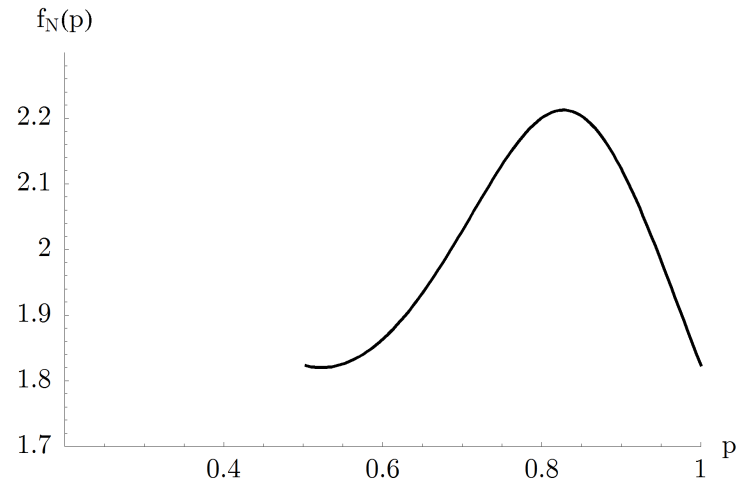


Figure 1: Density function for  $N = 5, \boldsymbol{\mu}(N) = (0.73, 0.2, 0.01, 0.01, 0.05), v = 1, c = 0$ .

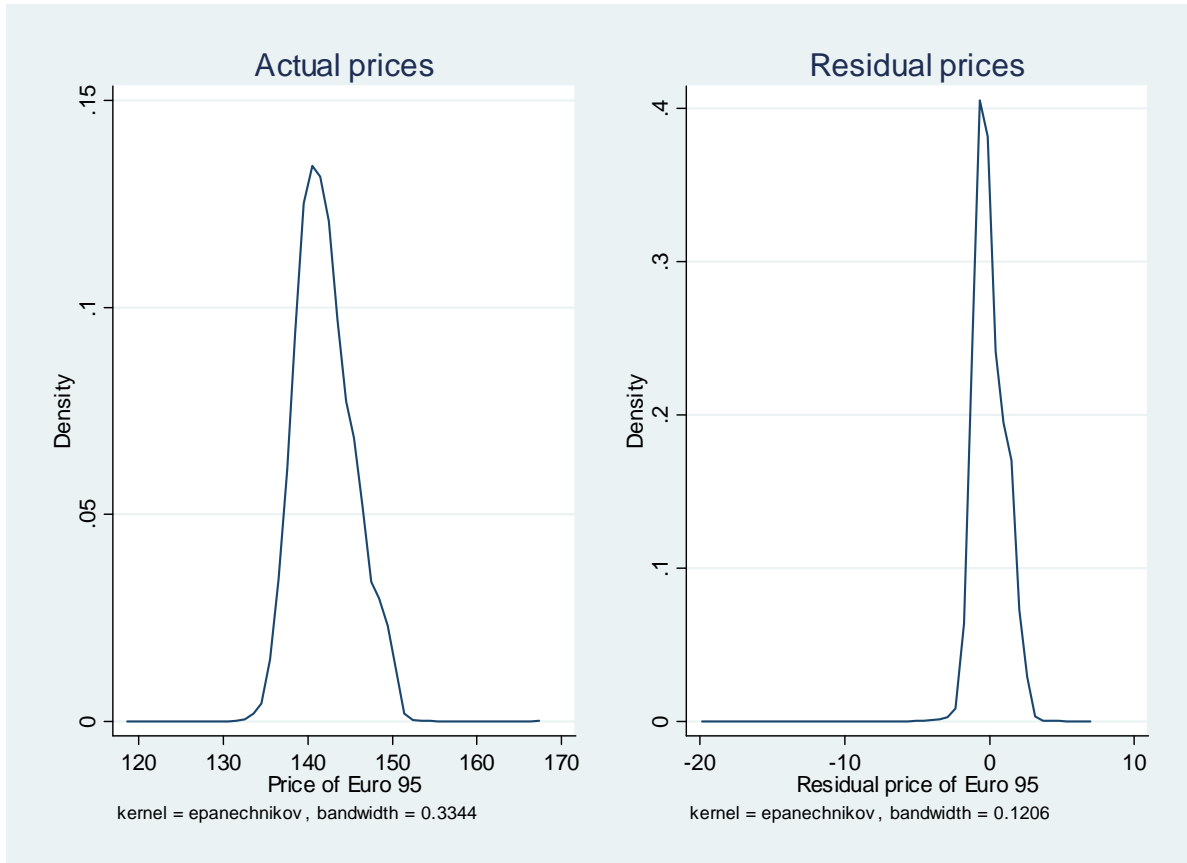


Figure 2: Density of Euro 95 raw and residual prices in the Netherlands

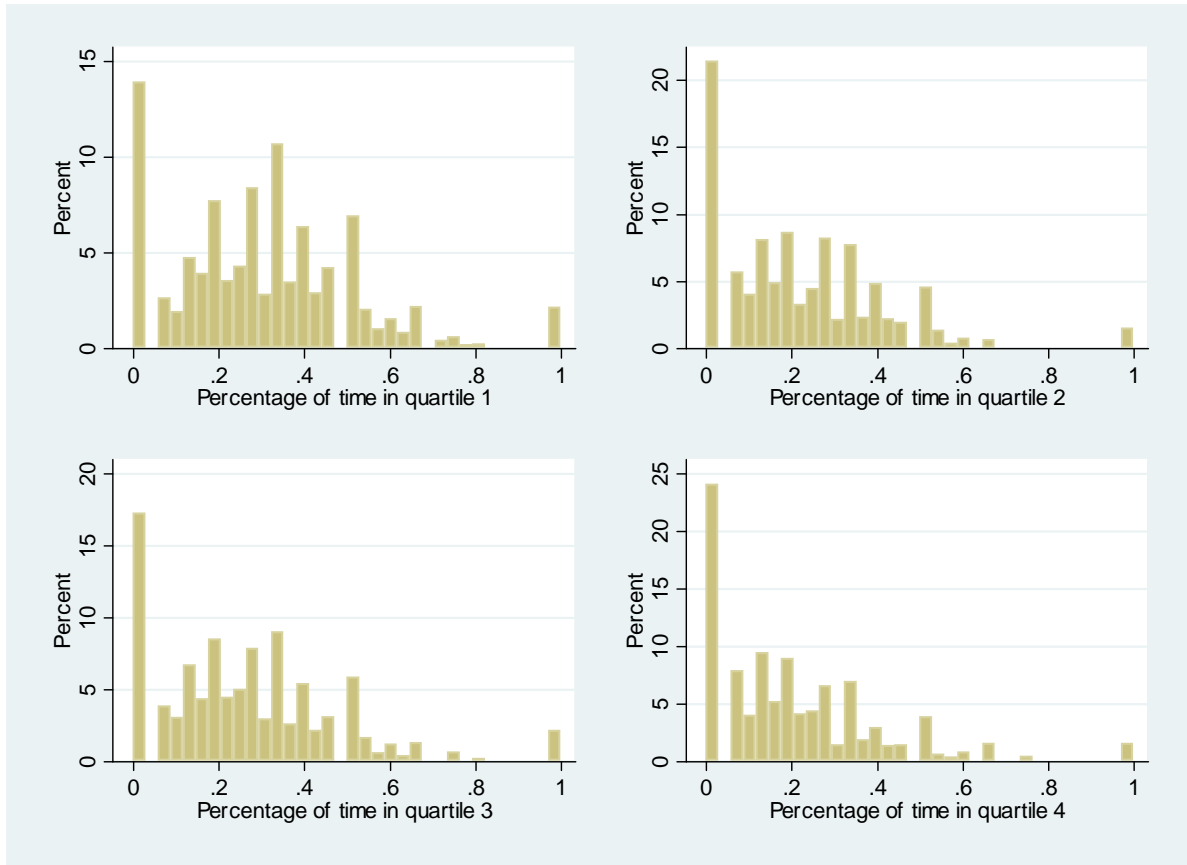


Figure 3: Time spent in each quartile of cross-sectional price distribution

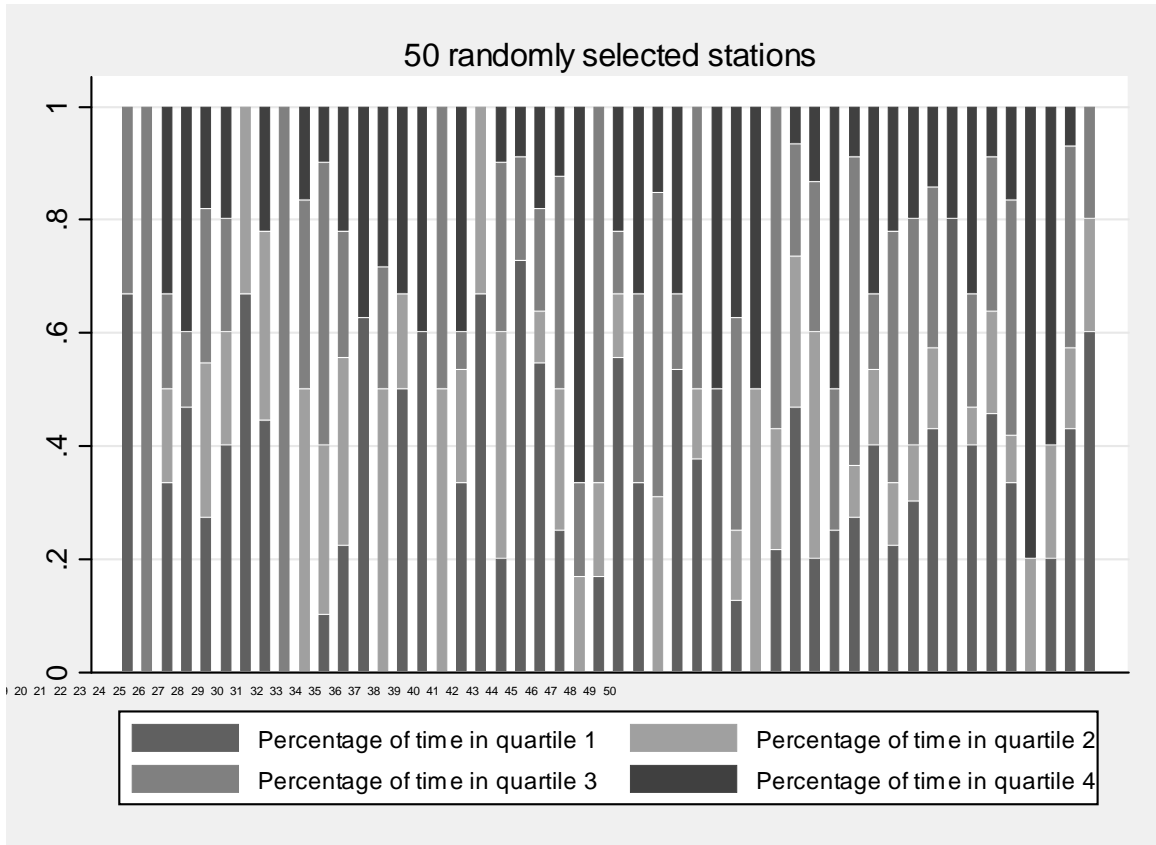


Figure 4: T1-T4 per gas station

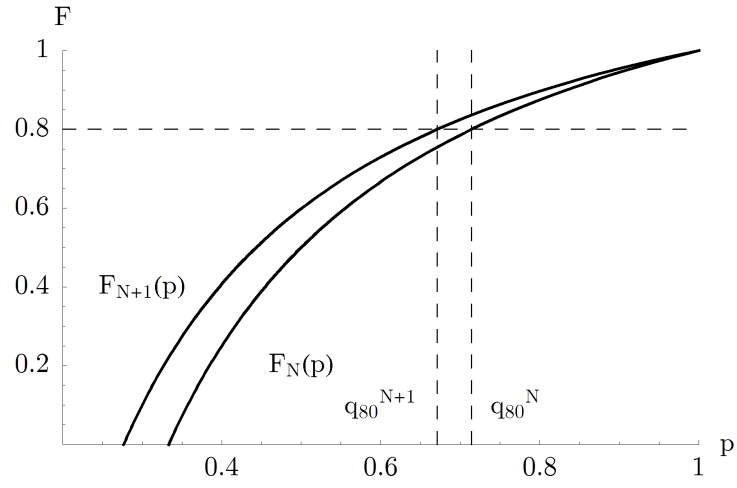


Figure 5:  $F_N(p)$  for  $N = 2$  and  $3$ ,  $\boldsymbol{\mu}(2) = (0.5, 0.5)$ ,  $\boldsymbol{\mu}(3) = (0.45, 0.52, 0.03)$

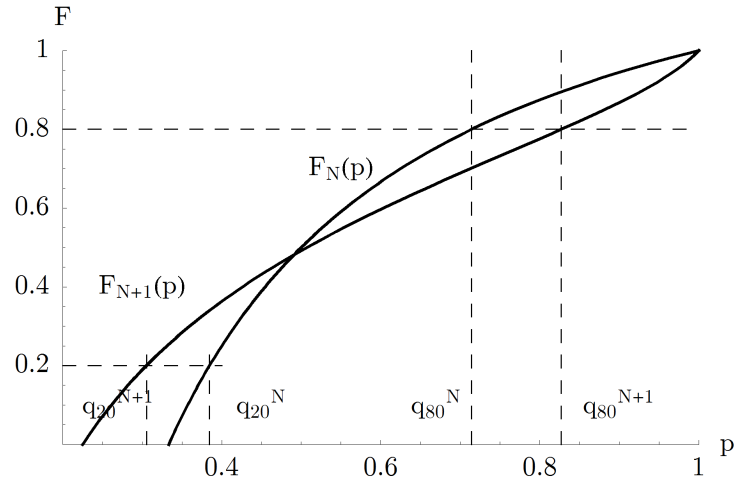


Figure 6:  $F_N(p)$  for  $N = 2$  and  $3$ ,  $\boldsymbol{\mu}(2) = (0.5, 0.5)$ ,  $\boldsymbol{\mu}(3) = (0.45, 0.1, 0.45)$

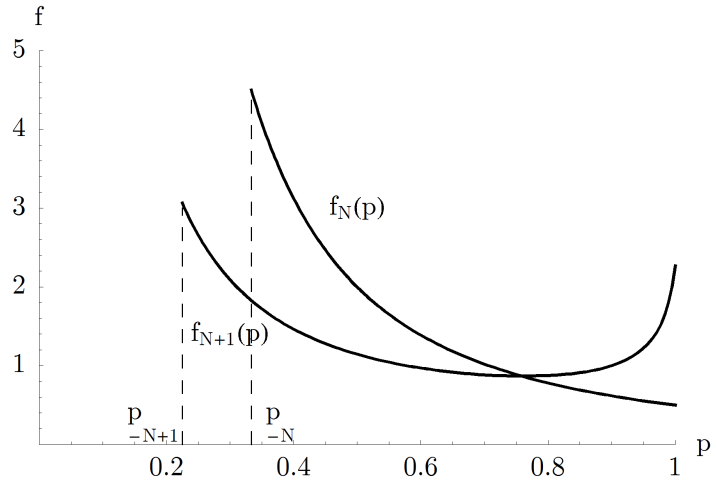


Figure 7:  $f_N(p)$  for  $N = 2$  and  $3$ ,  $\boldsymbol{\mu}(2) = (0.5, 0.5)$ ,  $\boldsymbol{\mu}(3) = (0.45, 0.1, 0.45)$

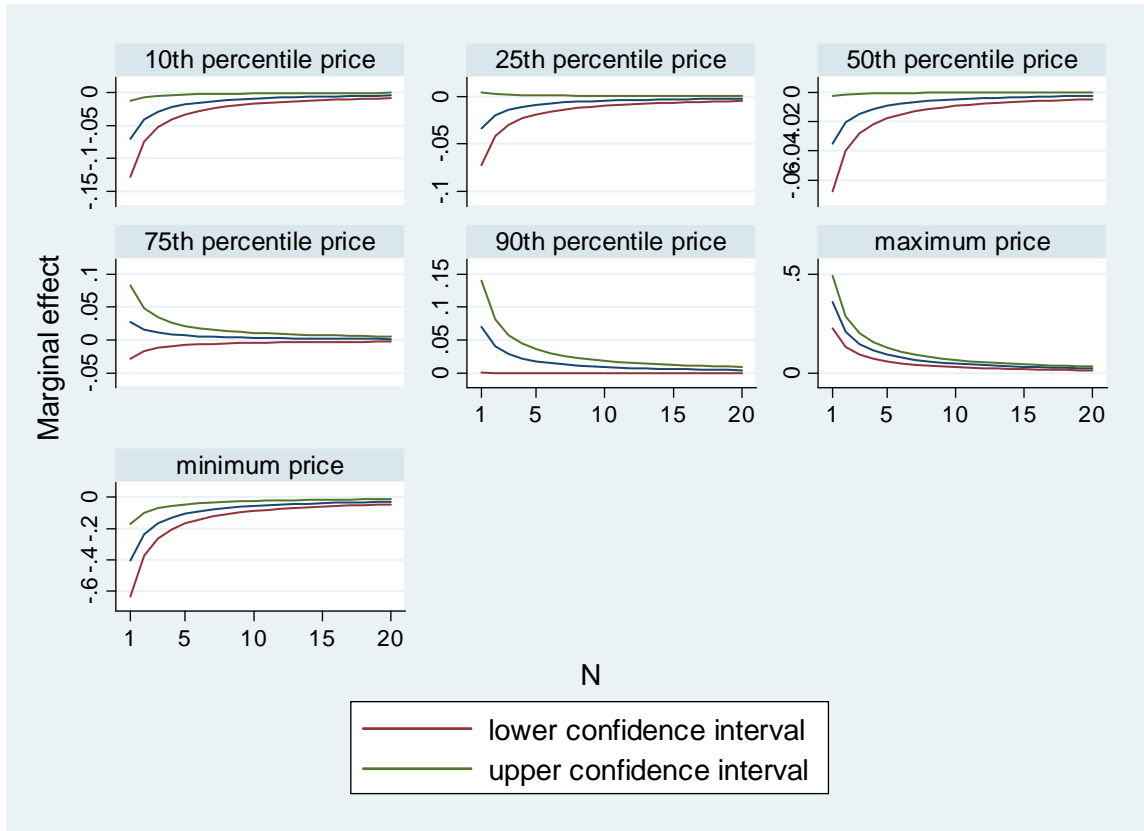


Figure 8: Marginal effect of the number of stations

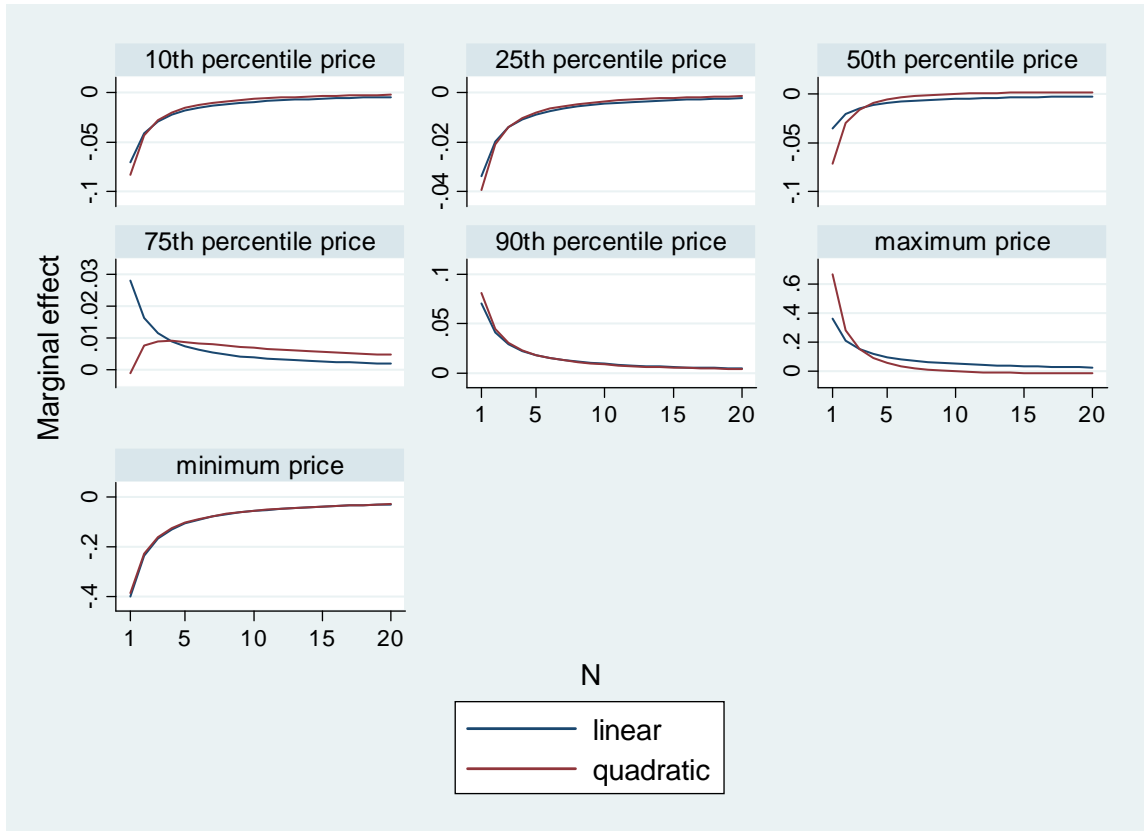


Figure 9: Marginal effects of squared and linear specifications

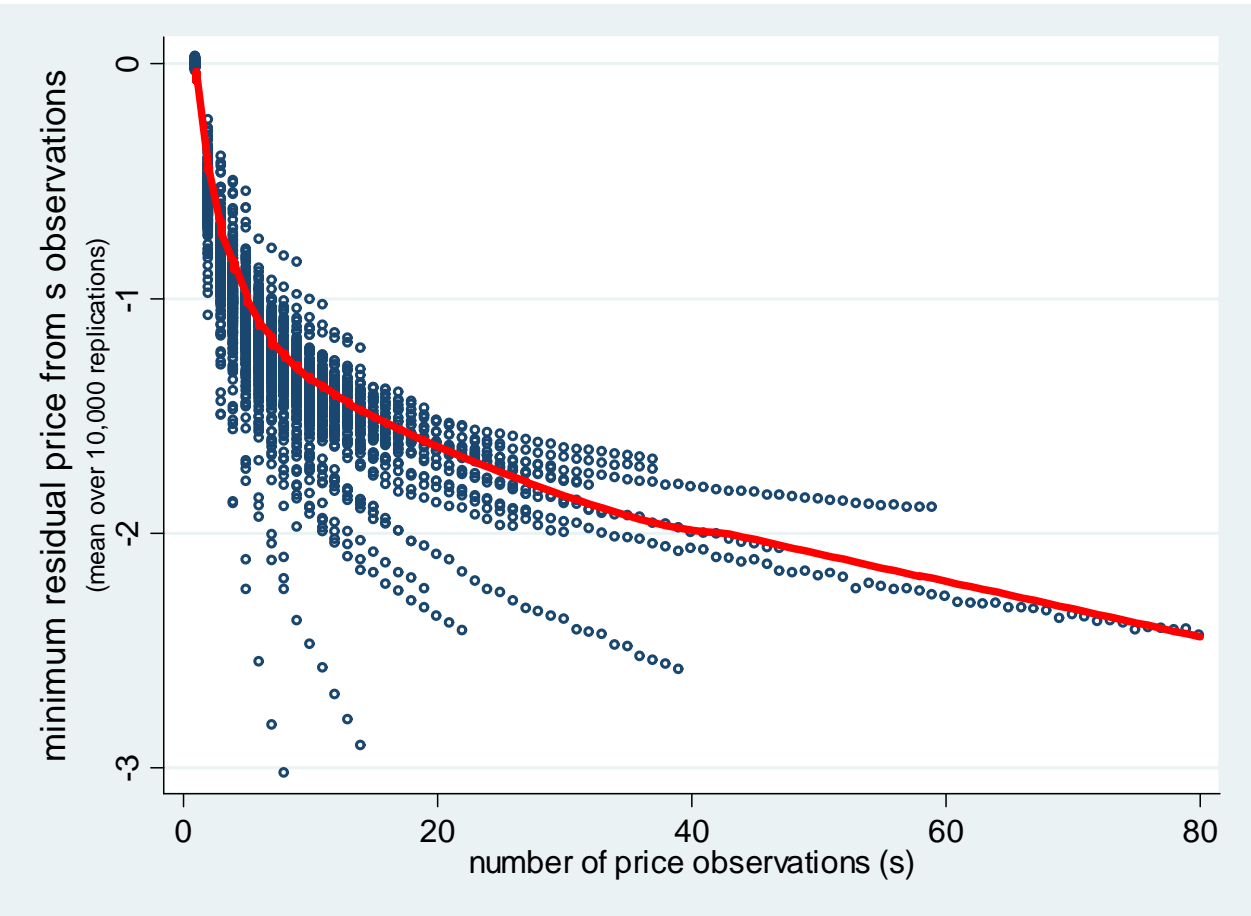


Figure 10: Expected price paid and number of price observations

**Table 1. One-week transition matrix (percentages)**

---

---

		t+7			
		q <sub>1</sub>	q <sub>2</sub>	q <sub>3</sub>	q <sub>4</sub>
t	q <sub>1</sub>	33	22	22	22
	q <sub>2</sub>	27	27	27	18
	q <sub>3</sub>	32	23	28	17
	q <sub>4</sub>	27	22	31	20

---

A station enters the calculations only when it has data at t and at t+7. Entries in each are weighted averages of the week-specific transition probabilities for each day t with weights equal to the share of observations in day t in the originating quartile out of the total number of observations for all days. Entries may not add up to 100 due to rounding.

**Table 2. Distribution of number of gas stations across markets**

Number of stores	Frequency	Percent	Cumulative
1	16	3.6	3.6
2	35	8.0	11.6
3	45	10.2	21.8
4	48	10.9	32.7
5	56	12.7	45.5
6	36	8.2	53.6
7	26	5.9	59.6
8	30	6.8	66.4
9	24	5.5	71.8
10	21	4.8	76.6
11	19	4.3	80.9
12	17	3.9	84.8
13	8	1.8	86.6
14	10	2.3	88.9
15	8	1.8	90.7
16	6	1.4	92.1
17	2	0.5	92.5
18	7	1.6	94.1
19	3	0.7	94.8
21	1	0.2	95.0
22	2	0.5	95.5
23	2	0.5	95.9
24	1	0.2	96.1
25	2	0.5	96.6
26	1	0.2	96.8
27	3	0.7	97.5
29	1	0.2	97.7
30	1	0.2	98.0
31	1	0.2	98.2
32	1	0.2	98.4
33	1	0.2	98.6
37	2	0.5	99.1
39	1	0.2	99.3
47	1	0.2	99.6
59	1	0.2	99.8
80	1	0.2	100.0
Total	440	100	

**Table 3. Residual price distribution by number of gas stations**

Number of stores	Minimum price	10 <sup>th</sup> percentile	25 <sup>th</sup> percentile	Median price	75 <sup>th</sup> percentile	90 <sup>th</sup> percentile	Maximum price
1	-1.54	-1.23	-0.74	-0.13	0.81	1.4	1.88
2	-1.45	-1.12	-0.72	-0.17	0.64	1.39	1.92
3	-1.9	-1.22	-0.76	-0.16	0.8	1.48	2.25
4	-1.73	-1.18	-0.75	-0.17	0.75	1.47	2.38
5	-2.02	-1.21	-0.76	-0.16	0.7	1.49	2.44
6	-1.85	-1.19	-0.76	-0.19	0.72	1.53	2.38
7	-1.94	-1.23	-0.76	-0.19	0.74	1.54	2.51
8	-2.83	-1.26	-0.79	-0.18	0.79	1.54	2.77
9	-2.27	-1.23	-0.74	-0.22	0.74	1.52	2.8
10	-2.27	-1.23	-0.79	-0.19	0.74	1.52	2.65
11	-2.27	-1.22	-0.79	-0.18	0.81	1.51	2.59
12	-2.37	-1.23	-0.78	-0.18	0.75	1.53	2.75
13	-2.12	-1.15	-0.72	-0.16	0.68	1.44	2.29
14	-2.57	-1.24	-0.76	-0.16	0.72	1.59	2.65
15	-2.68	-1.24	-0.75	-0.15	0.73	1.44	2.83
16	-2.99	-1.16	-0.7	-0.15	0.71	1.41	2.59
17	-2.18	-1.32	-0.87	-0.2	0.9	1.58	2.67
18	-2.68	-1.24	-0.76	-0.18	0.74	1.54	3.11
19	-2.89	-1.25	-0.75	-0.21	0.81	1.49	2.64
21	-2.06	-1.2	-0.74	-0.27	0.74	1.49	2.41
22	-3.02	-1.36	-0.78	-0.1	0.88	1.47	2.73
23	-2.88	-1.16	-0.73	-0.14	0.67	1.43	2.76
24	-1.95	-1.31	-0.93	-0.18	0.86	1.79	2.61
25	-2.43	-1.19	-0.71	-0.2	0.68	1.45	2.64
26	-3.37	-1.32	-0.82	-0.19	0.8	1.54	2.72
27	-2.94	-1.25	-0.83	-0.21	0.75	1.53	2.88
29	-2.05	-1.34	-0.84	-0.14	0.71	1.4	2.47
30	-4.04	-1.24	-0.81	-0.19	0.85	1.58	4.16
31	-2.64	-1.24	-0.77	-0.18	0.72	1.46	2.53
32	-3.34	-1.18	-0.74	-0.19	0.8	1.37	2.89
33	-3.7	-1.17	-0.8	-0.17	0.52	1.45	3.14
37	-2.31	-1.22	-0.76	-0.26	0.75	1.55	3.4
39	-4.07	-1.28	-0.76	-0.09	0.76	1.64	3.43
47	-4.66	-1.26	-0.76	-0.23	0.73	1.61	3.23
59	-2.5	-1.25	-0.79	-0.18	0.78	1.45	2.8
80	-5.52	-1.28	-0.74	-0.25	0.75	1.54	3.32

Entries are weighted averages of each residual price statistic over all markets (municipalities) where the weights are the municipality's share of observations. Unweighted results are very similar.

**Table 4. Effect of number of gas stations on the price distribution**

	minimum price	10th percentile	25th percentile	median price	75th percentile	90th percentile	maximum price
<b>Panel A: OLS without controls</b>							
Log (number of stations)	-0.508*** (0.0523)	-0.0266 (0.0175)	-0.0163* (0.00958)	-0.0161* (0.00897)	0.00715 (0.0138)	0.0438** (0.0173)	0.368*** (0.0341)
R-squared	0.11	0.01	0.01	0.01	0.00	0.01	0.16
<b>Panel B: OLS with 39 provincial dummies</b>							
Log (number of stations)	-0.549*** (0.0606)	-0.0285 (0.0194)	-0.0209** (0.0106)	-0.0239** (0.00960)	0.0165 (0.0164)	0.0513*** (0.0196)	0.356*** (0.0408)
R-squared	0.17	0.09	0.11	0.17	0.12	0.14	0.25
<b>Panel C: OLS with 39 provincial dummies and other controls</b>							
Log (number of stations)	-0.305*** (0.0886)	-0.0168 (0.0337)	-0.0248 (0.0196)	-0.0227 (0.0189)	-0.000228 (0.0312)	0.0428 (0.0373)	0.320*** (0.0640)
Tests:							
Other controls zero (p-value)	0.00	0.98	0.65	0.33	0.95	0.14	0.64
R-squared	0.19	0.09	0.13	0.20	0.13	0.16	0.27
<b>Panel D: 2SLS with 39 provincial dummies and other controls</b>							
Log (number of stations)	-0.580*** (0.171)	-0.101** (0.0423)	-0.0487* (0.0280)	-0.0507** (0.0241)	0.0403 (0.0406)	0.102** (0.0510)	0.518*** (0.0970)
Tests:							
Other controls zero (p-value)	0.37	0.77	0.44	0.13	0.95	0.04	0.46
J-test (p-value)	0.99	0.90	0.70	0.40	0.60	0.29	0.59
R-squared	0.18	0.07	0.13	0.20	0.12	0.16	0.26

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market.

The top two panels are based on 439 observations. The bottom two panels are based on 423 observations; we lose 16 observations because of missing municipality-level data on the other controls. The instruments in the 2SLS panel are population size and local tax rates, both in logs.

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 5. Determinants of the number of gas stations**

	Dep. var.: log(number of stations)			
	(1)	(2)	(3)	(4)
Log(population)	0.821*** (0.0255)	0.805*** (0.0251)	0.760*** (0.0633)	0.758*** (0.0637)
Log(municipal tax)	-0.189*** (0.0663)	-0.176*** (0.0668)	-0.0850 (0.0659)	-0.0811 (0.0666)
Average income per hh	--	--	0.00162 (0.0112)	0.00121 (0.0112)
Share of business cars	--	--	1.515*** (0.543)	1.514*** (0.543)
Area	--	--	0.0000664 (0.000384)	0.0000700 (0.000386)
Land area	--	--	0.00365*** (0.000710)	0.00366*** (0.000711)
Urbanized land share	--	--	-0.00284 (0.00292)	-0.00286 (0.00292)
Agrarian land share	--	--	-0.00201 (0.00165)	-0.00203 (0.00165)
Road length (km)	--	--	-0.00128*** (0.000265)	-0.00128*** (0.000265)
Sample size	--	--	0.00300*** (0.000695)	0.00301*** (0.000696)
F-test for significance of IV's	580.1	602.9	72.2	72.2
Provincial Effects	No	Yes	Yes	Yes
Observations	440	440	424	423
R-squared	0.72	0.81	0.84	0.84

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 6. Effect of number of gas stations on the price distribution -- additional quantiles (2SLS)**

	<b>minimum price</b>	<b>1th %</b>	<b>2nd %</b>	<b>3rd %</b>	<b>4th %</b>	<b>5th %</b>	<b>10th %</b>	<b>25th %</b>	
Inst	-0.580*** (0.171)	-0.5873*** (.1559)	-0.3237*** (.108)	-0.1871** (.0894)	-0.1513* (.0852)	-0.0794 (.0701)	-0.101** (0.0423)	-0.0487* (0.0280)	
Tests:									
Other controls zero (p-value)	0.37	0.04	0.13	0.38	0.45	0.54	0.77	0.44	
J-test (p-value)	0.99	0.36	0.31	0.81	0.74	0.6	0.9	0.7	
R-squared	0.18	0.09	0.06	0.07	0.07	0.07	0.07	0.12	
	<b>50th %</b>	<b>75th %</b>	<b>90th %</b>	<b>95th %</b>	<b>96th %</b>	<b>97th %</b>	<b>98th %</b>	<b>99th %</b>	<b>maximum price</b>
Inst	-0.0507** (0.0241)	0.0403 (0.0406)	0.102** (0.0510)	0.1727*** (.0635)	0.1865*** (.0682)	0.2493*** (.0736)	0.31737*** (.0803)	0.4464*** (.0877)	0.518*** (0.0970)
Tests:									
Other controls zero (p-value)	0.13	0.95	0.04	0.68	0.48	0.51	0.18	0.01	0.46
J-test (p-value)	0.4	0.6	0.29	0.52	0.1	0.09	0.46	0.7	0.59
R-squared	0.19	0.12	0.16	0.11	0.12	0.10	0.14	0.22	0.26

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market.

Based on 423 observations. The instruments are population size and local tax rates, both in logs.

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 7. Robustness checks I**

<b>4 Size Groups: Two-stage 2SLS with 39 provincial dummies and other controls</b>							
	<b>minimum price</b>	<b>10th percentile</b>	<b>25th percentile</b>	<b>median price</b>	<b>75th percentile</b>	<b>90th percentile</b>	<b>maximum price</b>
3 - 6 Stations	-0.4246*** (.1649)	-0.1440** (.0671)	-0.0939** (.0435)	-0.0629* (.0366)	0.0801 (.0619)	0.2209*** (.0792)	0.6927*** (.1223)
7 - 10 Stations	-0.4154 (.3464)	-0.0370 (.0378)	-0.0098 (.0265)	-0.0501** (.0234)	0.0106 (.0388)	0.0179 (.0519)	0.2410** (.1146)
11+ Stations	-0.3321 (.3229)	-0.0036 (.0404)	-0.0296 (.0279)	0.0193 (.0233)	0.0617 (.0396)	0.0779 (.0569)	0.0355 (.1584)
Tests:							
Other controls zero (p-value)	0.03	0.99	0.49	0.32	0.88	0.08	0.50
R-squared	0.19	0.08	0.12	0.20	0.13	0.14	0.25

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market.

The instruments in the 2SLS regression are the predicted probabilities of belonging to a group size. The equation is just identified. These predictions were obtained from a first-stage probit regression of each size dummy on population size and local tax rates (both in logs) as well as on the other controls and provincial dummies.

The number of observations in each regression is 423 municipalities. Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<b>Raw prices: 2SLS with 39 provincial dummies and other controls</b>								
	<b>minimum price</b>	<b>10th percentile</b>	<b>25th percentile</b>	<b>mean price</b>	<b>median price</b>	<b>75th percentile</b>	<b>90th percentile</b>	<b>maximum price</b>
Log (number of stations)	-1.423*** (.376)	-1.434*** (.373)	-0.818** (.387)	-0.184 (.384)	-0.251 (.421)	0.402 (.432)	1.336*** (.415)	1.354*** (.414)
Tests:								
Other controls zero (p-value)	0.00	0.00	0.02	0.05	0.02	0.07	0.01	0.00
J-test (p-value)	0.28	0.47	0.52	0.25	0.33	0.23	0.06	0.18
R-squared	0.47	0.45	0.41	0.44	0.42	0.38	0.38	0.41

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled stations in each market-day. The instruments are population size and local tax rates, both in logs.

The number of market-day observations in each regression is 7091. Standard errors clustered at the municipality level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 8. Robustness checks II**

<b>Adding Neighbours : 2SLS with 39 provincial dummies and other controls</b>							
	<b>minimum price</b>	<b>10th percentile</b>	<b>25th percentile</b>	<b>median price</b>	<b>75th percentile</b>	<b>90th percentile</b>	<b>maximum price</b>
Log (number of stations)	-0.567*** (0.176)	-0.0950** (0.0430)	-0.0422 (0.0292)	-0.0493** (0.0235)	0.0272 (0.0414)	0.0886* (0.0520)	0.502*** (0.101)
Log (number of neighbouring stations)	0.0188 (0.0773)	-0.00210 (0.0221)	-0.00829 (0.0158)	0.00587 (0.0134)	0.0214 (0.0223)	0.00466 (0.0293)	0.00684 (0.0493)
Tests:							
Other controls zero (p-value)	0.42	0.45	0.55	0.44	0.81	0.12	0.48
J-test (p-value)	0.99	0.90	0.70	0.40	0.60	0.29	0.59
R-squared	0.18	0.08	0.13	0.14	0.14	0.15	0.25

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market. The instruments are population size and local tax rates, both in logs.

Based on 420 observations corresponding to municipalities with non-zero number of neighbours.

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<b>Diesel: 2SLS with 39 provincial dummies and other controls</b>							
	<b>minimum price</b>	<b>10th percentile</b>	<b>25th percentile</b>	<b>median price</b>	<b>75th percentile</b>	<b>90th percentile</b>	<b>maximum price</b>
Log (number of stations)	-0.474** (0.214)	-0.0791 (0.0553)	-0.0736*** (0.0274)	-0.0146 (0.0202)	0.0738** (0.0317)	0.0910** (0.0462)	0.502*** (0.101)
Tests:							
Other controls zero (p-value)	0.20	0.16	0.18	0.23	0.41	0.44	0.21
J-test (p-value)	0.81	0.15	0.45	0.06	0.13	0.41	0.17
R-squared	0.16	0.16	0.17	0.10	0.14	0.12	0.29

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market. The instruments are population size and local tax rates, both in logs.

Based on 424 observations. Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 9. Expected price paid and the number of gas stations**

	Dep. Var: Expected price paid when observing s prices, $E[\text{Min}\{p_1, p_2, \dots, p_s\}]$					
Number of observations (s)	<b>s = 2</b>	<b>s = 3</b>	<b>s = 4</b>	<b>s = 5</b>	<b>s = 6</b>	<b>s = 7</b>
Log (number of stations)	-0.0689*** (0.0182)	-0.0924** (0.0375)	-0.163*** (0.0564)	-0.133* (0.0760)	-0.136 (0.0975)	-0.115 (0.127)
Tests:						
Other controls zero (p-value)	0.08	0.15	0.15	0.14	0.22	0.15
J-test (p-value)	0.93	0.62	0.59	0.78	0.27	0.26
Number of observations	407	373	329	282	227	192
R-squared	0.041	0.012	0.007	0.021	0.058	0.068

Other controls include: average income per household, share of business cars, area (km<sup>2</sup>), land area, urbanized and agrarian land shares, road length (km) and the number of sampled observations by market.

The instruments are population size and local tax rates, both in logs.

Robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$