

Price dispersion on the French retail gasoline market

(Very preliminary)

Etienne Chamayou*

CREST and Department of Economics, Ecole Polytechnique

April 16, 2014

Abstract: Abstract to be added here.

Keywords: Industrial Organization, Search

JEL Classification Numbers: XXX

*e-mail: *etienne.chamayou@ensae.fr*

1 Abstract

This article studies price dispersion on the French retail gasoline market, following and discussing the methodology employed by Chandra and Tappata (2011) with US data to test a standard prediction of consumer search models: randomized pricing. Even though the presence of differentiation and shocks affecting competition in French data are shown to require careful treatment, findings are consistent with Chandra and Tappata (2011). Price rankings in local markets exhibit significant variability over time, with variability increasing in distance between stations. This arguably supports the connection made by Stigler (1961) between price dispersion and the "ignorance in the market". Price dispersion furthermore decreases with cost and increases with the number of firms, in accordance with predictions from a model a la Varian (1980).

2 Introduction

The seminal paper of the consumer search literature, Stigler (1961), advocated the idea that a key element of competition was the relation between the cost of searching for prices by consumers and price dispersion i.e. the lasting existence of different prices for a homogeneous product. While the paper only provides evidence of static price dispersion (e.g. retailers posting different prices for a given car model in a suburb), thus leaving open discussions about differentiation, the empirical literature would soon strive to make the case stronger by considering temporal price dispersion, namely sellers' price moving within the market distribution over time. This was long done essentially by considering panel data of standard goods sold at various stores, introducing fixed effects to account for potential differentiation, and producing statistics to account for changes of rank in the distribution of prices (e.g. quantile to quantile changes). Shortcomings typically included the fact

that price distributions included sellers that were not necessarily operating in the same environment.

As argued by Chandra and Tappata (2011), the retail gasoline market appears to be particularly relevant to a discussion about the role of consumer search. The presence of sustained retail price variations due to oil cost fluctuations generates an information issue that consumers cannot ignore. High product homogeneity offers the opportunity to work with raw prices rather than prices which may have been cleaned inadequately. And most importantly, current data allow to compare prices posted by sellers which are actually competing within the same market.

Competition in the retail gasoline market has attracted a lot of attention in France over last years. In 2006, a governmental website was launched by the Ministry of Economics with a view to increase transparency in the market. All stations having sold over 500m3 of gasoline over the previous year are since then required to keep prices posted on the website. In 2011, a large inquiry was ordered by the government as retail price adjustments to oil cost variations were suspected to be faster upwards than downwards (the "Rocket and feathers" hypothesis first investigated by ? in the UK). Little evidence was yet found and the report rather emphasized signs pointing toward a significant degree of competitiveness. Finally, in August 2013, following an election promise, the gouvernement of the then newly elected French President François Hollande has implemented a 3c/l tax cut for 3 month, asking gasoline retailers on the market to produce the same effort.

The effect of the creation of the website has not been evaluated, probably for lack of data before 2006. The 2011 report yet called for further development of the website, and Germany has taken a similar initiative in 2012. More generally, in a context where dynamic prices generate heightened debates (e.g. Amazon), the idea to monitor competition by constraining firms to increase transparency enjoys strong popularity among politicians. The

theory of consumer search however tells us that measures affecting information are far from innocuous. On the one hand, in a situation of strong competition, aligned prices provide little incentive for consumers to search, thus undermining the stability of such an equilibrium. Contrarily, collusion appears quite sustainable in a context where a reliable website makes monitoring costless and, again, aligned prices discourage search by consumers.

TODO: contribute to the understanding of the relation between dispersion and search, taking advantage of the heterogeneity in local contexts of competition.

3 Literature

Whereas price dispersion had already been obtained in some settings, Varian (1980) innovated by modeling price dispersion as a result of mixed pricing strategies. According to the paper, price dispersion thereby obtained could be interpreted as "temporal" price dispersion, typically in the form of "sales". This provided a rationale for rank reversals i.e. competing sellers being successively cheaper/more expensive one vs. another.

The dynamic interpretation of the model is not unambiguous however. In the model, firms are ex-ante indifferent between all prices in the support of the equilibrium price distribution (also holds in terms of randomization over utilities). Ex-post, indifference obviously no more exists. The cheapest firm attracts shoppers but would be better off increasing its price to match the second cheapest price. Other firms may consider either increasing price to consumers' reservation price, or trying to undercut the cheapest firm. Consequently, price rigidity should lead to carefully interpret results regarding rank reversals.

Chandra and Tappata (2011) take advantage of a data set including daily prices of gas stations in the US over one year and a half. Chandra and Tappata (2011) argue that gas stations located across the same corner should display less rank reversal vs. those a bit further as consumers' information on prices should decrease with distance. This is shown

to be true (use of the Kolmogorov-Smirnov test to prove that rank reversals are drawn from different distributions for gas stations respectively 1-mile and 2-mile away from each other and of quantile regressions of each pair of stations' standard deviation on distance between stations) and all the more convincing as close gas stations display a lower average spread. Regressions of price dispersion (measured in the various ways described previously) on marginal costs and the number of firms in the market within 1-mile yields results consistent with the extension of Varian (1980) proposed by the paper (resp. negative and positive impacts).

4 The French retail gasoline market

4.1 Data

Since 2006, French law makes it mandatory for any gas station having sold over 500m³ gas the previous year to keep prices posted on a web site (www.prix-carburants.economie.gouv.fr). This legal obligation was motivated by the French Ministry of Finance's will to increase transparency on the retail gas market. Gas stations (in France and many countries) have indeed long been suspected of colluding and taking advantage of oil price variations to maintain high margins (through asymmetries in retail price responses to wholesale cost increases and decreases).

As it turns out, the French website only enjoys a moderate success (c. 200k-300k visitors per month).

Since September 2011, retail gas station prices were collected on a daily basis. Data downloaded from the web site include over c. 10,000 gas stations' addresses, services (presence of a shop, car wash etc.) and daily price of gas. I have used Google's geocoding service to obtain gas stations' gps coordinates and computed the distance separating gas stations.

Information on gas stations' competitive environment could probably be further refined using their address and observing strategic price interactions (e.g. a gas stations located on a "Route Nationale/Departementale", the price of which would be higher/uncorrelated with other gas stations in the vicinity). Rotterdam wholesale gas price is used as the best proxy for gas stations' variable costs.

Essentially data from gouvernement website prix-carburants.gouv.fr, from website zagaz.fr and INSEE data: prices of gas (sp95-sp95/e10?) and diesel on a daily basis, gas station information: opening hours, services, gps coordinates (special treatment), INSEE data: information about local markets (rural/center/suburb, population etc.). Price collected with date at which it was set hence possibility to correct when data is missing and check that no loss of information. Generally no intra day changes...

Regarding location (...) . Information provided by user is typically of better quality than information resulting from geocoding. Such information was collected from the website www.zagaz.com.

4.2 Context

The period covered is marked by two significant events of different natures. One is a change of pricing strategy by a leading retail brand affecting the whole country and the other is a governmental intervention.

The first event is the progressive conversion by the historical incumbent "Total" of a number of its gas stations to low cost gas stations branded "Total Access" i.e. gas stations whose prices are aligned with supermarket gas stations. The French retail gasoline market is indeed characterized by the penetration of supermarkets, which currently account for over 50% of total sales. Though these conversions are obviously not heterogeneous shocks to the market, they yet involve a significant number of gas stations lowering their price by

c. 0.10 €/per liter virtually overnight.

The second event is of political nature. On August 29, 2012, a decrease in price of 6c per litre was announced by the government, following an election promise made by François Hollande. This decrease was (to be) achieved by a decrease of tax of 3c per litre and an equivalent "effort" by gas station operators. Also, it must be noted that a small portion of the tax on gasoline is set by regions on an annual basis.

4.3 Descriptive statistics

Descriptive statistics are provided for a period of 640 days starting September 9, 2011 and ending May 6, 2013. Four subperiods of data are missing whose respective day lengths are 11 (2012/07/08-2012/07/18), 10 (2012/08/13-2012/08/22), 49 (2012/12/04-2013/01/21) and 6 (2013/04/04-2013/04/09). The end of the period studied has been determined quite arbitrarily. It essentially avoids including another missing subperiod and keeps the database to a reasonable size.

Data include prices at 10,054 gas stations, of which 421 (20 missing) are identified as highway gas stations and thus treated as a specific market.

- Price rigidity: price durations and variations (sales?)
- Aggregate price vs. Rotterdam price
- Prices by brand / region (taxes)
- Distance and competition

Static price dispersion in French Data is easily seen in any single day price cross section. The resulting price distribution is bimodal, largely reflecting the presence of two dominant types of actors: supermarkets and oil companies. Supermarkets justify their relatively low prices by the fact that they use gasoline to attract consumers, while oil companies mention

the quality of the service to justify higher prices and sell to business clients for which the price paid is not (necessarily) the price displayed at the gas station.

In comparison, data used in Chandra and Tappata (2011) include prices of gas (regular, mid-grade, premium) and diesel at gas stations in California, Florida, Texas and New Jersey from January 2006 to May 2007. The average price of regular gas fluctuates quite a bit over time (from c.\$2.5/gal to c.\$3.5/gal for California) and from one state to the other (average price in Texas over all stations and periods is \$2.576/gal vs. \$3.019/gal in California), largely due to difference in taxes. The number of rivals within 2 miles (c.3.2km) of each station is rather similar in all states: from 13.38 in New Jersey to 14.22 in Texas (resp. 4.25 to 4.74 with 1 mile). The average distance to the closest rival ranges from 0.38 to 0.49 miles (c. 600 to 800m).

5 Pair price dispersion

A simple way to measure temporal price dispersion between two stations with daily observations is to consider the probability that the station which is in general cheaper (in terms of day count) turns out to be more expensive. Formally, considering the prices p_{it} and p_{jt} of two stations i and j over T_{ij} days, such that $p_{it} \geq p_{jt}$ is observed most of the time, the rank reversals statistic writes:

$$r_{ij} = \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} \mathbb{1}_{p_{jt} > p_{it}}$$

Though the computation and intuition of rank reversals is fairly straightforward, its actual use involves dealing with two fundamental practical questions. Though rank reversals at nearby competitors certainly provide more convincing evidence of temporal price

dispersion linked to consumer information than variations in the distribution of prices at a municipal/regional/national level, it raises the issue of actual market definition. Also, as static price dispersion leads to a reduction in temporal price dispersion *ceteris paribus* (in the case of randomization over utilities, there may be dispersion without rank reversals at all), one may need to control for it. As the issue of static price dispersion gets more severe, it can become unclear to which extent nearby sellers actually compete for the same customers, potentially connecting the two questions.

Statistics provided in Chandra and Tappata (2011) to describe prices and mark up suggest that the issue of static price dispersion is not severe. In all pairs obtained with a 1 mile maximum distance between stations, depending on fuel type, average spread amounts to 1% to 2% of average price, over 90% of pairs exhibit positive rank reversals, and the average pair rank reversals varies between c.12% and c.15%. Results do not vary significantly with a 2 mile maximum distance. Static price dispersion can therefore arguably be neglected, thus avoiding the shortcomings associated with price cleaning.

Using maximum 1km and 2km allows to build databases containing respectively X and X pairs of gas stations. As expected, rank reversals between two stations decreases in the average spread between their prices, with a large number exhibiting no rank reversals. In order to reduce the influence of static dispersion, the sample is subset based on the value of the average spread between prices. Results are presented for a maximum average spread of 0.02c/l i.e. a 1% to 2% average difference in prices (NB: X% of stations include 3 digits in price display). Finally, a last element is taken in consideration: the conversion of XXX gas stations by the brand Total to low cost gas stations, which generates a large number of spurious rank reversals.

INSERT GRAPHS + STATS DES

The percentage of reversed pairs per day varies between X and X etc.

TODO: test with distance vs. rank reversals (control for spread) considering cleaned prices (compare to Lach and Hosken (relevant?) and raw prices... Noise? + Discuss duration of rank reversals and deterministic moves. Insist on economic significance. (+ dispersion vs. price rigidity?)

6 Market price dispersion

While the emphasis has so far been laid on examining empirical evidence about the link between price dispersion and consumer information, the following section seeks to compare predictions from a model a la Varian (1980) with French retail gasoline market data.

The methodology employed by Chandra and Tappata (2011) consists in considering each gas station successively as the center of a market delimited by a circle of a given radius (robustness of results to variations in radius size is checked). Various statistics accounting for price dispersion can then be computed for each market and period. If one observes N stations close to each other (e.g. with a limit radius of 2km: if no two stations are separated by a distance of more than 2 km) and with prices posted over T periods, this will result in $N*T$ observations as each station is successively considered and, for each station, a statistic representing price dispersion is computed at each period. An expected consequence is that it tends to over represent markets with higher gas station density (typically cities).

Regressions are then run using raw prices and cleaned prices. Also, in order to address the issue of overlapping markets, regression are performed on dispersion data generated in a way that prevents market overlap (the precise methodology is not given?). Descriptive statistics performed on raw prices are provided.

Standard market dispersion statistics such as range and standard deviation are strongly influenced by static price dispersion. As a consequence, contrarily to the previous section where descriptive statistics performed on raw prices arguably provided a most conservative

Figure 1: Overview of market price dispersion

	Radius 5km	Radius 3km	No overlap
Nb of markets	6794	5614	386
Avg nb competitors	11.2	6.7	3.9
Raw prices			
Market price range	0.116	0.105	0.089
Market gain from search	0.046	0.044	0.036
Market price std	0.046	0.045	0.045
Clean prices			
Market price range	0.033	0.028	0.020
Market gain from search	0.017	0.014	0.010
Market price std	0.011	0.011	0.010

Each price dispersion statistic corresponds to the average across markets of average price dispersion over time. Markets with less than 3 competitors are dropped. At the market level, price dispersion has been computed only for days when more than two thirds of competitors' prices are available. Only markets with more than 50 days of data are then kept for the overall average.

first idea of price dispersion, clean prices are now to be preferred despite their shortcomings. Results are reported with clean prices, raw prices for markets obtained with radiuses of respectively 3 and 5km. Additionnally, results are reported for non overlapping markets built in a way that is detailed in appendix.

References

- Chandra, A. and M. Tappata (2011). Consumer search and dynamic price dispersion: an application to gasoline markets. *RAND Journal of Economics* 42(4), 681–704.
- Stigler, G. (1961). The economics of information. *The Journal of Political Economy* 69(3), 213–25.
- Varian, H. (1980). A model of sales. *American Economic review* 70(4), 651–659.

A Data processing

Price and station information: Duplicate detection. Not trivial: typically occurs when brand changes... new record with new address/brand... use price information Investigation of brand changes more detailed in descriptive statistics.

Location: same corner gas station... crucial for consumer information / advantage in location Duplicates: picture of competition, potentially interesting if stations closed can be found in data. Brand changes: Connected to duplicates... important policy implications Highway gas stations: different market

The database which lies behind the website is first essentially recreated, using the same gas station identification numbers as it is contained in webpages (zip code + 3 digits which appear to merely reflect the registration order of gas stations with the same zip code). Two observations follow: there appears to be a significant turnover in the data, some prices appear to reflect mistakes by managers (extraordinary variations in price).

Since there are very few gas stations which are newly opened each year, it must be that gas stations either stop providing information because they are not/no more committed to do so or create a new account and are then registered under a different identification number... hence the need to reconcile them.

Treatment of prices:

- Daily files are merged recursively on station ids which leads to create a database of 10,433 individuals over 640 days, including unobserved periods which result in missing values within series. Since dates of price changes are known, missing periods can yet be filled in many cases (e.g: if price on day 2 is missing, it can be checked in day 3 that price hasn't changed since day 1 and if price on days 10-15 are missing, it can be seen on day 16 that the last change was made on day 13: prices for 13-15 are then input backward and forward for 10-12). TODO: stats

- Abnormal values are then looked for based on prior observation. This allows to observe that some prices are input as 0.156 instead of 1.56 and correct rather than transform in missing values. TODO: stats
- Abnormal price variations are then looked for based on observation. Quite often it appears that the variation must have been the consequence of a mistake given the size and the following reverse change. Such changes are eliminated. When no converse changes can be found due to missing data, suspect periods are also switched to missing.
- Abnormal price durations are finally examined based on prior observation. As mentioned in XXX, a duration of 1 month is considered suspect, 1 month and 1/2 highly suspect, 2 months almost certainly an error.

Station information:

- Matching of stations with INSEE codes: Zip codes: problem of cedex and changing zip codes. City name matching implies the generic problem of string comparison, not to mention the fact that the same name can be used in different regions and that city names sometimes change (small municipalities are regrouped). Approach: matching on zip then city name
- Matching of databases: address standardization but still remains a big issues as quite different addresses can be provided, a piece of info can be up to date in one database while a bit old in the other. It requires a multicriteria approach.
- Geocoding: address standardization
- Highway gas stations