A network model of price dispersion

Giacomo Pasini[*] Paolo Pin[†] Simon Weidenholzer[‡]

January 2008

Abstract

We analyze a model of price competition *à la* Bertrand in a network environment. Firms only have limited information on the structure of the network: they know the number of potential customers they can attract and the degree distribution of customers. This incomplete information framework stimulates the use of Bayesian-Nash equilibrium. We find that, if there are customers only linked to one firm, but not all of them are, then an equilibrium in pure strategies fails to exist. Instead, we find a symmetric equilibrium in randomized strategies. Finally, we test results on US data, finding empirical evidence supporting our model.

1 Introduction

What we observe in everyday life is a stark dispersion of prices for otherwise homogeneous products. In addition there are even fluctuations of prices across time for the same product offered by the same seller. The typical explanation of this phenomena is that goods are state contingent: an ice-cream in the desert or in summer is not the same product as an ice-cream in Siberia or in wintertime. This does however only marginally explain much of the variability observed in the data. For instance, gasoline sold across US cities (see below for a description of the dataset we analyze), or chicken sold

[*]Dipartimento di Scienze Economiche, Università di Venezia. Email: giacomo.pasini@unive.it

[†]Abdus Salam International Centre for Theoretical Physics, Trieste and Università di Venezia. Email: pin@unive.it

[‡]Institut fuer Volkswirtschaftslehre, Universität Wien. Email: simon.weidenholzer@univie.ac.at

in different supermarkets in The Netherlands, or plane tickets sold online, display unexplainable price dispersion both along time and across locations.[1]

One of the ways that classical economic theory explains the formation of prices is the Bertrand competition model. Bertrand competition, however, imposes strict restrictions: agents are homogeneous and fully informed while firms compete against each others on prices. In the case of constant marginal costs, firms will undercut each other, thereby gaining all of the demand, and pushing the equilibrium price down to marginal costs.

Previous works (Salop and Stiglitz (1977, 1982), Wilde and Schwartz (1979), Varian (1980)) relaxed the full information assumption, dividing agents into informed and uninformed customers. The informed would compare prices and buy from the firm with the lowest price while the uninformed would just sample one firm. In this setup it can be shown that the equilibrium is such that firms randomize on prices with identical probability distributions. Baye and Morgan (2001) analyze advertising and consider a gatekeeper who charges advertising fees. Interestingly, they obtain similar equilibrium properties as the models mentioned above.

However, note that this setup does note well suit the case, where some firms have greater local market power while others are facing much more competitive environments, e.g. the only gas station in a small town vs. the agglomerations of gas stations around big cities.

A different type of models (e.g. Kranton and Minehart (2001) or Corominas–Bosch (2004)) analyze bargaining situations on exogenously given non–regular networks. However, these bargaining get soon very complex as the size of the network grows. Hence, these models restrict their analysis to simple networks. On the contrary, we keep the transaction between buyers and sellers simple, which allows to generalize to any topology of relations between buyers and sellers.

We model our market through a network. This allows customers to be more heterogeneous in their ability to compare prices: they can be not only either uninformed or fully informed, but also partially informed. Both customers and firms are nodes of this network. A firm can only attract those customers it is linked to and customers can only compare prices of the firms they are linked to.

We assume that firms only have information about the number of potential customers they could attract and on the probability distribution characterizing the number of links each customer has (i.e. the number of firms whose

[1]Supermarket prices of homogeneous goods in the Netherlands have been analyzed empirically by Wildenbeest (2007). Clemons et al (1999) analyzed prices of plane tickets, sold by online travel agencies.

prices she can compare and whose goods she can buy). We are excluding any other knowledge by the firms about the full topology of the network. This assumption seems to be realistic since real world wholesaler normally do not know the shopping habits of each individual consumer, but rely their strategy on survey data representing the shopping habits of average customers. This incomplete information framework motivates the use of the Bayesian-Nash equilibrium concept, as e.g. in Jackson and Yariv (2007) or Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2007).

Within this setup, if there are completely uninformed customers, an equilibrium in pure strategies fails to exist. Instead, firms will randomize on prices. We identify two interesting properties of this equilibrium price distribution using stochastic dominance criteria. First, if we add links to a customer the average price will decline. Second, if we add a link to a customer with few links this will have a larger effect on the average price than if we add the link to a customer with already many links.

Finally, in order to restore the balance between theory and application, we test our theoretical model on a gasoline prices dataset, making reasonable assumptions on the shape of the networks under simple characteristics such as town size or population density. We find that much of the variability in prices can be explained by these simple characteristics.

Section 2 describes the model and the analytical results. Section 3 analyzes the data, while Section 4 concludes.

2 Model description

Formally our environment is characterized by an exogenously given bipartite undirected network, that is a network where nodes are of two distinct types (firms and customers) and each link can only be between a firm and a customer.[2] We assume that N potential buyers and H firms are located in this bipartite network.

Customers are assumed to need at most one unit each of a homogeneous good produced by firms. Customers compare all prices of the firms they are linked with and buy from the firm offering the lowest price, provided that this price does not exceed a reservation price r. In case of a tie, demand is assumed to be equally randomly split.

We assume that firms know the probability distribution characterizing the number of links each customer has (i.e. the number of firms whose prices she can compare and whose goods she can buy). So, firms know the probability

[2] We do not model informational links between customers, since this case can be replicated connecting them exactly to the same firms.

that a given customer has only one link (call this probability q_1), two links (q_2), three links (q_3), and so forth, yielding a probability distribution $\vec{q}$. We call $\vec{q}$ the *degree distribution of customers*. In this incomplete information environment firms consider customers to be a priori homogeneous, even if the number of connections they have will vary across the network. In addition, we assume that each firm i knows the potential number of customers that it can attract, i.e. it knows its own degree d_i.

We assume that firms produce at constant marginal costs (which we set to 0 without loss of generality). In this sense, our model is a good approximation of the behavior of shops and wholesalers. Under this assumption the private information of firms (i.e. the number of customers each firm can potentially attract) does not affect the optimal strategy of each firm.

So, the private information of firm i is the number d_i of its own links. On the contrary, the public information of firms is given by the reservation price r of buyers, and the vector $\vec{q}$ of probabilities (such that $\sum_{i=1}^{\infty} q_i = 1$).

The strategy of each firm i is to fix a price p_i , or more generally, as we will see, a distribution $f_i(p)$ of prices.

The number of expected links in the network is $L = N \cdot \sum_{i=1}^{H} iq_i$. The number of expected links for a single firm is then $\frac{L}{H} = \frac{N}{H} \cdot \sum_{i=1}^{H} iq_i$.

2.1 Formal analysis

It is easy to show that if $q_1 = 0$ (i.e. all consumers compare at least two prices), then we obtain the same result as in Bertrand competition. Formally:

Lemma 1 (Bertrand competition) *If $q_1 = 0$, then the only equilibrium is such that each firm plays $p = 0$.*

Proof First note that no firm will charge a price above r and below 0. In the former case it will attract no customers at all whereas in the latter it will make losses. So, consider a situation where all firms charge the same price $0 \leq p \leq r$. A given firm i with degree d_i expects profits of

$$\pi_i = d_i \sum_{j=1}^{k} \frac{q_j}{j} p \quad . \tag{1}$$

A deviant firm could now charge a slightly lower price $p - \epsilon$ and thereby attract all potential customers. This would yield a profit of $d_i(p - \epsilon)$ which is, for small enough $\epsilon > 0$, higher than the profit specified in (1). If $q_1 = 0$ firms keep on undercutting until we reach the point where $p = 0$ and $\pi = 0$. ∎

It is also trivial to check that, if $q_1 = 1$, i.e. all customers go to exactly one shop, all firms will act as monopolists. Formally:

Lemma 2 (Monopoly) *If $q_1 = 1$, then the only equilibrium is such that each firm plays $p = r$.*

In what follows we address the remaining interesting case where $q_1 \in (0,1)$. In particular, we show that if $0 < q_1 < 1$ the only symmetric equilibrium is such that all firms will randomize on price. This result is related to previous results, as Salop and Stiglitz (1977, 1982), Wilde and Schwartz (1979) and Varian (1980).

We stress that, even if the game we consider is a one-shot game, the equilibrium we expect is not in pure strategies, but rather an equilibrium where each firm randomizes strategies. This motivates our search for empirical evidence in the time-series data discussed in Section 3.

We start by noting that under the assumption of constant marginal costs, every potential customer can be considered as an independent game between all the firms connected to her. In this sense, firms are homogeneous and can only be distinguished by the number of games they play (the number of customers they are connected to, which they know). Each game is played by j different players (with probability q_j), but this number is unknown to the players. In the next proposition we will consider the symmetric strategies when there is a positive non–trivial probability that a customer is connected only to one firm ($q_1 \in (0,1)$).

Proposition 3 *If $q_1 \in (0,1)$, then there is no symmetric equilibrium in pure strategies, but there exist a symmetric equilibrium in randomized strategies, given by the distribution function $f(p)$. The support of $f(p)$ is $[q_1 r, r]$, on this support its cumulative distribution $F(p) = \int_{q_1 r}^{r} f(p)dp$ is characterized by the implicit equation*

$$\sum_{j=1}^{H} q_j \left[1 - F(p)\right]^{j-1} = \frac{q_1 r}{p} \tag{2}$$

$$\Phi_{\vec{q}}\left[1 - F(p)\right] = \frac{q_1 r}{p}\left[1 - F(p)\right] \tag{3}$$

where $\Phi_{\vec{q}}$ is the generating function of the probability vector $\vec{q}$.[3]

[3]I.e.

$$\Phi_{\vec{q}}(x) \equiv \sum_{i=1}^{\infty} q_i x^i \quad ,$$

it works for any $x \in [0,1]$.

Proof If $q_1 \in (0,1)$ a firm i can charge a price of r and assure itself a profit of $d_i q_1 r$. Note that this implies that no firm will charge a price lower than $q_1 r$.

Suppose that all firm charge the same price $q_1 r$. They will make an expected profit, on each customer, of $\sum_{j=1}^{k} \frac{q_j}{j} r q_1 < r q_1$, and hence could improve by setting a price of r.

Suppose now that the symmetric equilibrium has a point of mass on a given price $p > q_1 r$. In this case a profitable deviation would be to shift this point of mass to $p - \epsilon$, for some low enough $\epsilon > 0$, as discussed in the proof of Lemma 1.

Suppose finally that the support of $f(p)$ is exactly $[q_1 r, r]$ (i.e. $f(p) > 0$ for any $p \in [q_1 r, r]$). If we show that this assumption can be consistent with the well known requirement of a mixed–strategies equilibrium, that every strategy in the support yields the same expected payoff, then this equilibrium exists.

We suppose that each firm plays a random strategy $f(p)$, where the support of $f(p)$ is $[q_1 r, r]$, and there are no points of mass.

Let $F(p) = \int_{q_1 r}^{r} f(p) dp$ be the symmetric cumulative distribution function of prices. The probability that a customer with j links buys a product from a given firm is $[1 - F(p)]^{j-1}$. This means that the profit to a firm i by setting price p is given by

$$\pi_i(p) = d_i \int_{q_1 r}^{r} \left(\sum_{j=1}^{H} q_j \left[1 - F(p)\right]^{j-1} \right) p \ f(p) \ dp \quad . \tag{4}$$

Since firms randomize on prices that guarantee the same expected profit, we define $\pi_i \equiv \pi_i(p)$ for all p with $f(p) > 0$, and then $\pi_i = \pi_i(r) = d_i r q_1$. Equation (2), which is independent from d_i, comes directly from the previous point and from (4).

We have shown that a symmetric equilibrium in mixed strategies exists, which has price–support on $[q_1 r, r]$, and satisfy the property that each price on the support guarantees the same expected payoff. ∎

If $q_i = 0$ for any $i > H$, then the definition is still valid, but one can truncate the sum at H. We assume $0^0 = 1$, so that $\Phi_{\vec{q}}(0) = q_1$ and then, in our case, also $\Phi_{\vec{q}}[1 - F(r)] = q_1$. Since $\sum_{i=1}^{\infty} q_i = 1$: $\Phi_{\vec{q}} : [0,1] \to [q_1, 1]$. It is also easy to check that $\Phi_{\vec{q}}$ is increasing and, if $\vec{q'}$ first order stochastically dominates (FOSD) $\vec{q}$, then $q'_1 \le q_1$, and

$$\Phi_{\vec{q'}}(x) < \Phi_{\vec{q}}(x)$$

for any $x \in (0,1)$.

We now provide two examples on how the probability distribution $f(p)$ can be inferred from different degree distributions.

Example 1 (Power law network) *Suppose that probabilities q_i fall down with a power law, so that $q_i \propto \alpha^i$, with $\alpha \in (0,1)$.*
In order to get the normalization $\sum_{i=1}^H q_i = 1$, since $\sum_{i=1}^H \alpha^i = \alpha \frac{1-\alpha^H}{1-\alpha}$ we need

$$q_i = \alpha^{i-1} \frac{1-\alpha}{1-\alpha^H} \quad .$$

Equation (3) reads

$$\begin{aligned} \sum_{i=1}^{H} \alpha^{i-1} \frac{1-\alpha}{1-\alpha^H} \left[1-F(p)\right]^{i-1} &= q_1 \frac{r}{p} \\ \frac{1-\left(\alpha\left[1-F(p)\right]\right)^H}{1-\alpha\left[1-F(p)\right]} &= \frac{r}{p} \end{aligned} \tag{5}$$

for $p \in \left[\frac{1-\alpha}{1-\alpha^H} r, r\right]$.
If $H \to \infty$ then (5) becomes

$$\begin{aligned} \frac{1}{1-\alpha\left[1-F(p)\right]} &= \frac{r}{p} \\ F(p) &= \frac{p}{\alpha r} - \frac{1-\alpha}{\alpha} \end{aligned} \tag{6}$$

for $p \in [(1-\alpha)r, r]$, so that $f(p) = \frac{1}{\alpha r}$ on this support (uniform probabilities), and is 0 otherwise. The expected price is $E[f(p)] = r - \frac{\alpha}{2}$.
If α increases, also the probabilities that customers have more connections increase, and expected prices decrease. ∎

Example 2 (Random Network) *Suppose that H and N are fixed, and any link between a customer and a firm has probability λ. In this case, by the binomial distribution,*

$$q_i = \binom{H}{i} \lambda^i (1-\lambda)^{H-i} \quad .$$

Note that there is also a positive probability for $q_0 = (1-\lambda)^H$. The expected number of potential customers is actually $\left(1-(1-\lambda)^H\right) N < N$.

We can divide every q_i, with $i > 0$ by $(1-q_0)$ to re–normalize things, obtaining

$$q_i = \binom{H}{i} \lambda^i \frac{(1-\lambda)^{H-i}}{1-(1-\lambda)^H} \quad .$$

Noting that $q_1 = H\lambda\frac{(1-\lambda)^{H-1}}{1-(1-\lambda)^H}$ *equation (3) becomes now*

$$\begin{aligned}\sum_{i=1}^{H}\binom{H}{i}\lambda^i\frac{(1-\lambda)^{H-i}}{1-(1-\lambda)^H}\left[1-F(p)\right]^{i-1} &= \frac{q_1 r}{p}\\ \frac{(1-\lambda)^{H-1}}{1-(1-\lambda)^H}\left[\left(1+\lambda\left[1-F(p)\right]\right)^H-1\right] &= H\lambda\frac{(1-\lambda)^{H-1}}{1-(1-\lambda)^H}\frac{r}{p}\left[1-F(p)\right]\\ \left(1+\lambda\left[1-F(p)\right]\right)^H &= 1+H\lambda\frac{r}{p}\left[1-F(p)\right] \qquad (7)\end{aligned}$$

for $p \in [q_1 r, r]$, $F(p) = 0$ *for* $p < q_1 r$ *and* $F(p) = 1$ *for* $p > r$. *It is not possible to compute analytically* $F(p)$ *and hence* $f(p)$.

There are however two things that it is possible to infer from (7). Note first that if we call $\xi(p) \equiv \lambda\left[1-F(p)\right]$, *then (7) can be written as*

$$\left(1+\xi(p)\right)^H = 1+H\frac{r}{p}\xi(p)$$

1. *As H (keeping fixed* λ*) grows we approximate (by the law of large numbers) a regular network with* $H\lambda$ *links per customer, moreover* $q_1 \to 0$. *(7) tells us that in this case, for any* $p > q_1 r \to 0$*:* $1-F(p) \to 0$. *This implies that we approximate Bertrand Competition.*

2. *As* λ *grows (keeping fixed H)* $\xi(p)$ *is fixed,* $1-F(p)$ *decreases, and so* $F(p)$ *increases.*

Finally note that an increase in λ *implies higher chances for the customers to be connected to more firms, and that an increase in* $F(p)$ *means that the expected price decreases.* ∎

The popular case of a scale–free network ($q_i \propto i^{-\gamma}$, with $\gamma > 0$) could also be treated analytically, but the series could not be solved for explicitly and the result would just be a rephrasing of equation (2).[4]

2.2 Comparative statics

Although, we can not explicitly solve for the equilibrium pricing strategies we can nevertheless perform some comparative statics using the implicit form

[4]Suppose that $q_i \propto i^{-\gamma}$. Under the additional assumption that $H \to \infty$, (2) becomes:

$$Li_\gamma[1-F(p)] = \frac{r}{p}[1-F(p)] \quad ,$$

for any $p \in [\frac{r}{\zeta(\gamma)}, r]$. Here $Li_\gamma(x) \equiv \sum_{i=1}^{\infty} x^i i^{-\gamma}$ is the *polylogarithm*, and $\zeta(\gamma) \equiv \sum_{i=1}^{\infty} i^{-\gamma}$ is the *Riemann zeta function*.

provided in Proposition 3. In particular, we can show that an increase in the number of links of the network will reduce the average price. Formally, if we assume some degree distribution $\vec{q'}$ first order stochastically dominates (FOSD) another degree distribution $\vec{q}$, then the resulting probability distribution of prices under q will FOSD the resulting price distribution under q'.

Proposition 4 *Consider two customer degree distributions, $\vec{q}$ and $\vec{q'}$, with $q_1 > 0$ and $q'_1 > 0$. Call $f(p)_{\vec{q}}$ and $f(p)_{\vec{q'}}$ the resulting symmetric equilibria (as defined in (2)). If $\vec{q'}$ FOSD $\vec{q}$, then $f_{\vec{q}}(p)$ FOSD $f_{\vec{q'}}(p)$.*

Proof Remember equation (3) and define

$$\Psi_{\vec{q}}(x) \equiv \frac{\Phi_{\vec{q}}(x)}{x} = \sum_{i=1}^{\infty} q_i x^{i-1} \quad ,$$

so that

$$\Psi_{\vec{q}}[1 - F_{\vec{q}}(p)] \equiv \frac{\Phi_{\vec{q}}[1 - F_{\vec{q}}(p)]}{1 - F_{\vec{q}}(p)} = \frac{q_1 r}{p} \quad . \tag{8}$$

for $p \in [q_1 r, r]$.
Note that for any $\vec{q}$, both $\Psi_{\vec{q}}$ and $\Psi_{\vec{q}}^{-1}$ are increasing functions.[5] Equation (8) can be written as

$$F_{\vec{q}}(p) = 1 - \Psi^{-1}\left(\frac{q_1 r}{p}\right) \quad , \tag{9}$$

for $p \in [q_1 r, r]$, $F(p) = 0$ for $p < q_1 r$ and $F(p) = 1$ for $p > r$. Given the definition of generating functions, if $\vec{q'}$ FOSD $\vec{q}$, then $q'_1 \leq q_1$, and

$$\Psi_{\vec{q'}}(x) < \Psi_{\vec{q}}(x)$$

for any $x \in (0, 1)$.

Hence,

$$\Psi_{\vec{q'}}^{-1}\left(\frac{q'_1 r}{p}\right) < \Psi_{\vec{q}}^{-1}\left(\frac{q_1 r}{p}\right)$$

for any $p \in [q'_1 r, r]$. From (9) we have that $F_{\vec{q'}}(p) > F_{\vec{q}}(p)$ for any q_1 and $p \in [q_1 r, r]$. This implies that $f_{\vec{q}}(p)$ FOSD $f_{\vec{q'}}(p)$. ∎

Note that Proposition 4 implies the following: if we increase the number of shops each customers visits then the equilibrium price distribution will put

[5] $\Psi_{\vec{q}}^{-1}$ is well defined, given the monotonicity of $\Psi_{\vec{q}}$.

more weight on lower and less weight on higher prices. Moreover, this implies that the average price will decline if we add links to an existing network.

Furthermore, note that the result of Proposition 4 is well in accordance with the final considerations provided in Examples 1 and 2.

Likewise, we find that the marginal contribution, in terms of FOSD of price distributions, of adding a new link to a customer is higher the less original links she has.

Proposition 5 *Suppose the network is fixed, not complete, every customer has at least one link and there is at least a customer with only one link. We add a link to a customer j, where j stands for the number of links she originally has. If $\vec{q}$ is the degree distribution obtained by adding it to a customer k, and $\vec{q'}$ is the degree distribution obtained by adding it to a customer $k' > k$, then $f_{\vec{q'}}(p)$ FOSD $f_{\vec{q}}(p)$.*

Proof Note that $\vec{q}-\vec{q'} = \left(0, \ldots, \frac{1}{L+1}, \ldots, -\frac{1}{L+1}, \ldots, 0\right)$, where L is the total number of links in the original network, $\frac{1}{L+1}$ is in k^{th} position and $-\frac{1}{L+1}$ is in $(k+1)^{th}$ position.

The proof goes exactly as for Proposition 4. If $k = 1$, then $q'_1 < q_1$, otherwise $q'_1 = q_1$. Given the definition of generating functions,

$$\Psi_{\vec{q'}}(x) < \Psi_{\vec{q}}(x)$$

for any $x \in (0, 1)$. ∎

Proposition 5 allows us to draw an important conclusion for competition: the return of adding links (in terms of the reduction of the average price) is larger for customers with few links, than for customers with already many links.

Finally, we do not perform any global welfare analysis because, in our settings, it would be meaningless. Given that, in equilibrium, every customer will buy one and only one unit of the good, the surplus will be either on the side of firms or on the side of customers, but its aggregate value will remain constant. In this sense, our model is a zero–sum game.

3 Empirical analysis

Price dispersion is a well-known issue in the empirical literature: price dispersion over time is at the basis of time series econometrics, but same goods also have different prices in cross–sectional samples. The empirical implication of our model is that the "law of one price" is not the only equilibrium outcome: given the network structure the model suggests firms will randomize on prices. Thus price dispersion turns out to be an equilibrium. Hence, given a network which is ex–ante heterogeneous, i.e. an environment where the number of potential clients around each firm is not fixed, we expect to find evidence of persistent price dispersion along time. Further, the comparative statics exercise of the previous section shows that a network with a higher number of potential links (i.e. a big city compared to a small village) implies a lower average price.

Wildenbeest (2007) provides empirical evidence of persistent price dispersion in groceries' goods using data from www.supers.nl, a Dutch website that publishes daily prices once a month for a set of groceries goods sold in 15 different supermarket chains in the Netherlands. The website is freely accessible, thus in principle clients could compare prices every time they need to shop and behave consequently. Nevertheless, price dispersion is present and does not fall over time. Baye et al (2004) find the same kind of evidence on thousands of consumer goods whose prices are daily compared on a website (Shopper.com)[6].

The idea behind using prices taken from the Internet to provide evidence of price dispersion can be thought to as a "worst case scenario": the Internet should reduce search costs to zero and thus ex-ante information heterogeneity in the network. So, if there is price dispersion in this case, it can only be worse in a "real world environment". Nevertheless, there may be a potential self selection of customers in this setting: those who compare prices on the Internet even before going out to the supermarket, are individuals somewhat more "realsophisticated" than the average.

Since both Wildenbeest and Baye et al provide evidence in support of our model, we look for price dispersion on gasoline prices in the US. While there may be self selection on the market for different types of cars, regardless of the vehicle, everyone needs to buy gasoline. We use prices collected monthly by the US bureau of labor statistics representatives across US to compute the Consumer Price Indexes, thus we have the "primary source": customers in our network check prices when they go to the gas station, and that is where our data is collected. We have average prices on 5 different types of gasoline, 28 urban areas and 12 region/size class groupings, i.e. average prices in three different city size classes and in the four US macro regions. Depending on the area and gasoline type the time series may vary in length spanning the period between January 1978 and October 2007, thus covering several

[6]Further empirical evidence on price dispersion can be found in Baye et al (2006).

business cycles (see table 2 and 3 in the appendix for details on regular grade gasoline series).

As in Baye et al (2004), the first part of our analysis is based on the graphical analysis of the time series of three measures of price dispersion: the first one, standard deviation, is the most straightforward. Since price dispersion is likely to depend on the price level, we computed the coefficient of variation, which is the ratio of standard deviation over the relative mean. The last measure, the interquartile range, is obtained in order to reduce the impact of potential outliers. These measures are computed for each gasoline type and for each month aggregating data at the macro–regional level: North East, Midwest, South and West. While these areas are quite wide, each of them cover both a number of urban areas and suburban ones, thus firms (i.e. gas stations) with a higher and lower number of potential customers. Furthermore customers could potentially live in one area and buy gasoline in the neighboring one, implying that each area is not isolated from each other.

The time series of each of these measures for leaded and unleaded gasoline are reported graphically in table 4 and 5 in the appendix. Results, regardless of the measure chosen, support the presence and persistence of price dispersion: this phenomenon is not going to decline, since there are no clear down sloping trends. Moreover, the magnitude of such a variability is substantial: for both leaded and unleaded gasoline the mean coefficient of variation lies between 0.02 and 0.05, meaning that the price in different areas within the same region deviateson average between 2% and 5% from the regional mean, with a peak at 13.5% (leaded gasoline in the West). Dispersion is not constant over regions: North East, which is the region characterized by the presence of many urban area, is in general the lower–variability region among the four.

The second part of our analysis is based on regression results. The aim is to test the implication that big cities - which are network with a high number of potential links - have a lower average price than non urban areas. To do so we compute differences of (log)prices between a urban area and a sub–urban one. Then, we use it as the dependent variable Δp_t in a regression over a intercept and its autoregressive lag:

$$\Delta p_t = \alpha_0 + \alpha_1 \Delta p_{t-1} + u_t \tag{10}$$

Where u_t is supposed to have zero mean and σ_t^2 unknown variance. The two original time series are likely to be integrated, i.e. they are likely to be correlated to an underlying oil prices data generating process. The fact that we are considering the difference allows us to control for this common factor and for any other common exogenous shock. The autoregressive term account for inertia in price adjustment and persistent shocks. Therefore, this regression provide us a direct test of our hypothesis: under the null that prices in urban areas are on average lower than in non–urban areas, the intercept term should be significantly different from zero and negative, since it represents Δp_t mean after controlling for its lag. A second, indirect evidence is given by the residuals skewness: regression (10) residuals should be normally distributed and thus symmetric around zero. An excessive left skewness would suggest that not only prices are on average lower on big cities, but that the whole distribution of big cities prices lies on the left of non–urban areas prices. Results are reported in table 1 for 6 Δp_ts: we consider two metropolitan areas in Southern U.S. (Houston–Galveston–Brazoria, and Dallas-Fort Worth, both in Texas and both with more than two million inhabitants) and B,C,D area/size average prices for Southern US as well. The Bureau of Labor statistics define four area/size aggregations: A is the average over census metropolitan areas with more than two million inhabitants, D is the average over areas with less than 50.000 inhabitants, B and C are intermediate classes. Results do not vary significantly across size, while if the difference is taken between the two metropolitan areas, the test is rejected at 99% level. We ran the same regressions on the other regions and gasoline specification, with no significant differences.[7]

4 Conclusion

Our work aims at explaining some of the price dispersion observed in real data both along time and across locations, using combined tools from network theory and standard consumer theory. The network models the heterogeneity in information and purchasing possibility of the consumers at one side, and the variability in the number of potential customers on the side of firms. We maintain the assumption that firms' have constant marginal costs, which is reasonable when considering shops or gas stations. We show that there is no symmetric equilibrium in pure strategies, but there is one in randomized strategies. Empirical evidence support our model: we analyzed twenty years of monthly prices on five different types of gasoline sold in United States. Whatever the time series or the area is chosen, prices exhibit substantial and persistent dispersion, confirming the fact that such an empirical evidence is an equilibrium feature and not a temporary state of the market. Moreover, regression analysis ran on the same data goes in favor of the model as well: metropolitan areas - i.e. networks with a high number of potential links - have

[7]The data is not reported but is available upon request.

Table 1: Regression results

Coefficient	Estimate	Std. dev.	t–stat	Skewness
Δp : **Houston–Galveston–Brazoria minus B–size average**				
α_0	-0.0482	0.0099	-4.8623	-0.3188
α_1	0.8959	0.0441	20.3216	
Δp : **Houston–Galveston–Brazoria minus C–size average**				
α_0	-0.0409	0.0152	-2.6884	-0.3310
α_1	0.9255	0.0292	31.6625	
Δp : **Houston–Galveston–Brazoria minus D–size average**				
α_0	-0.0383	0.0083	-4.6162	-0.5097
α_1	0.8648	0.0261	33.1395	
Δp : **Houston–Galveston–Brazoria minus Dallas–Fort Worth**				
α_0	-0.0082	0.0039	-2.1338	-0.3070
α_1	0.8064	0.0366	22.0557	
Δp : **Dallas–Fort Worth minus B–size average**				
α_0	-0.0379	0.0082	-4.6013	-0.3302
α_1	0.8821	0.0380	23.2047	
Δp : **Dallas–Fort Worth minus C–size average**				
α_0	-0.0305	0.0147	-2.0788	-0.5795
α_1	0.9231	0.0275	33.6055	
Δp : **Dallas–Fort Worth minus D–size average**				
α_0	-0.0293	0.0086	-3.4057	-0.2527
α_1	0.8676	0.0266	32.6393	

lower average prices and price distribution tilted to the left with respect to sub-urban and rural areas.

References

[1] Baye, M. R., and J. Morgan (2001) “Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets”, The American Economic Review, Vol. 91, No. 3., pp. 454-474.

[2] Baye, M. R., J. Morgan and P. Scholten (2004) "Price Dispersion in the

Small and in the Large: Evidence from an Internet Price Comparison Site" , Journal of Industrial Economics, Vol. 52, No. 4, pp. 463–496.

[3] Baye, M. R., J. Morgan and P. Scholten (2006) "Information, Search and Price Dispersion" in: Terry Hendershott (Editor), *Handbook on Economics and Information Systems*, Elsevier, forthcoming

[4] Clemons, E. K., I.–H. Hann, and L. M. Hitt (1999) “The Nature of Competition in Electronic Markets: An Empirical Investigation of On-line Travel Agent Offerings”, Ninth Workshop on Information Systems and Economics.

[5] Corominas–Bosch, M., (2007) “Bargaining in a network of buyers and sellers”, Journal of Economic Theory, Vol. 115, pp. 35–77.

[6] Galeotti, A., S. Goyal, M. Jackson, F. Vega–Redondo and L. Yariv (2007) “Network Games”, mimeo.

[7] Kranton, R.E., and D. F. Minehart (2001) “A Theory of Buyer-Seller Networks”, The American Economic Review, Vol. 91, No. 3., pp. 485-508.

[8] Jackson, M.O., and L. Yariv (2007) "Diffusion of Behavior and Equilibrium Properties in Network Games" , American Economic Review (Papers and Proceedings).

[9] Salop, S., and J. Stiglitz (1977) "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion", The Review of Economic Studies.

[10] Salop, S., and J. Stiglitz (1982) "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents", The American Economic Review.

[11] Varian, H.R., (1980) "A Model of Sales", The American Economic Review.

[12] Wilde, L.L., and A. Schwartz (1979) "Equilibrium Comparison Shopping", The Review of Economic Studies.

[13] Wildenbeest, M. (2006) "An Empirical Model of Search with Vertically Differentiated Products", mimeo.

Appendix: Data description and figures

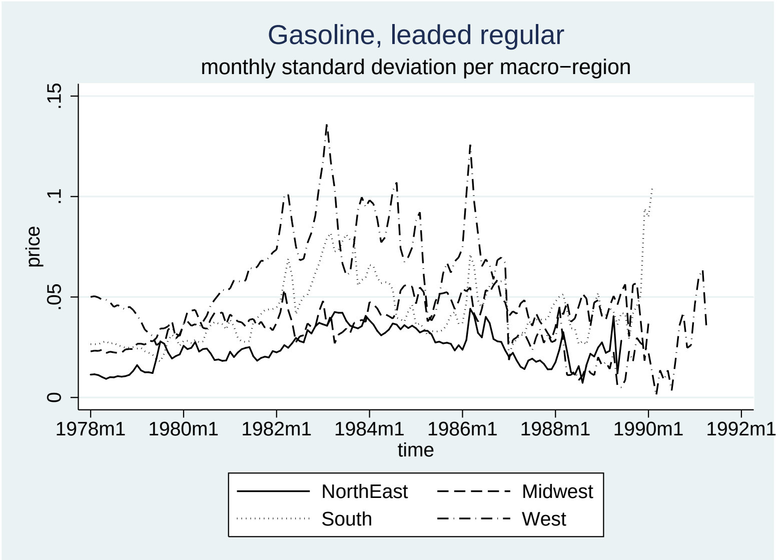
Table 2: Gasoline, leaded regular

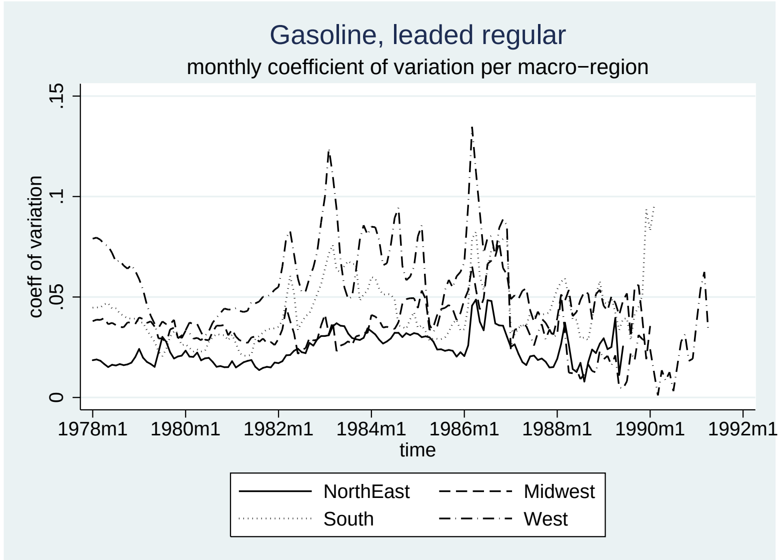
series id	area	begin	end
APUA10174712	New York-Northern New Jersey Long Island, NY-NJ-CT-PA	1978, Jan	1989, Dec
APUA10274712	Philadelphia-Wilmington-Atlantic City PA-NJ-DE-MD	1978, Jan	1989, Dec
APUA10374712	Boston-Brockton-Nashua, MA-NH-ME-CT	1978, Jan	1989, Dec
APUA10474712	Pittsburgh, PA	1978, Jan	1989, Dec
APUA10574712		1978, Jan	1986, Dec
APUA10674712		1978, Jan	1986, Dec
APUA20774712	Chicago-Gary-Kenosha, IL-IN-WI	1978, Jan	1989, Dec
APUA20874712	Detroit-Ann Arbor-Flint, MI	1978, Jan	1989, Dec
APUA20974712	St. Louis, MO-IL	1978, Jan	1990, Dec
APUA21074712	Cleveland-Akron, OH	1978, Jan	1988, Dec
APUA21174712	Minneapolis-St. Paul, MN-WI	1978, Jan	1986, Dec
APUA21274712	Milwaukee-Racine, WI	1978, Jan	1986, Dec
APUA21374712	Cincinnati-Hamilton, OH-KY-IN	1978, Jan	1986, Dec
APUA21474712	Kansas City, MO-KS	1978, Jan	1986, Dec
APUA31574712		1978, Jan	1988, Dec
APUA31674712	Dallas-Fort Worth, TX	1978, Jan	1989, Dec
APUA31774712		1978, Jan	1988, Dec
APUA31874712	Houston-Galveston-Brazoria, TX	1978, Jan	1989, Dec
APUA31974712	Atlanta, GA	1978, Jan	1986, Dec
APUA32074712	Miami-Fort Lauderdale, FL	1978, Jan	1987, Dec
APUA42174712	Los Angeles-Riverside-Orange County, CA	1978, Jan	1991, Dec
APUA42274712	San Francisco-Oakland-San Jose, CA	1978, Jan	1991, Dec
APUA42374712	Seattle-Tacoma-Bremerton, WA	1978, Jan	1986, Dec
APUA42474712	San Diego, CA	1978, Jan	1986, Dec
APUA42574712	Portland-Salem, OR-WA	1978, Jan	1986, Dec
APUA42674712	Honolulu, HI	1978, Jan	1986, Dec
APUA42774712	Anchorage, AK	1978, Jan	1986, Dec
APUA43374712	Denver-Boulder-Greeley, CO	1978, Jan	1986, Dec
APUB10074712	Northeast Size B	1978, Jan	1989, Dec
APUB20074712	Midwest Size B	1978, Jan	1989, Dec
APUB30074712	South Size B	1978, Jan	1990, Dec
APUB40074712	West Size B	1978, Jan	1988, Dec
APUC10074712	Northeast Size C	1978, Jan	1989, Dec
APUC20074712	Midwest Size C	1978, Jan	1991, Dec
APUC30074712	South Size C	1978, Jan	1990, Dec
APUC40074712	West Size C	1978, Jan	1991, Dec
APUD10074712	Northeast Size D	1978, Jan	1986, Dec
APUD20074712	Midwest Size D	1978, Jan	1990, Dec
APUD30074712	South Size D	1978, Jan	1989, Dec
APUD40074712	West Size D	1978, Jan	1986, Dec

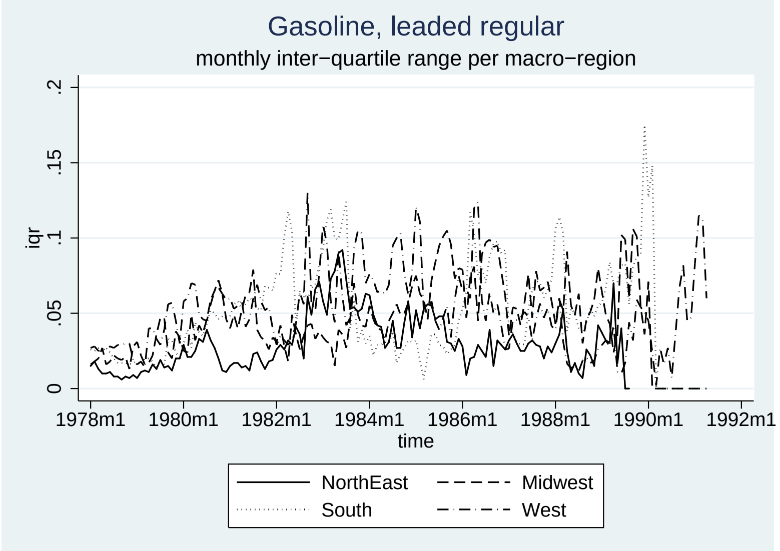
Table 3: Gasoline, unleaded regular

series id	area	begin	end
APUA10174714	New York-Northern New Jersey Long Island, NY-NJ-CT-PA	1978, Jan	2007, Oct
APUA10274714	Philadelphia-Wilmington-Atlantic City PA-NJ-DE-MD	1978, Jan	2007, Oct
APUA10374714	Boston-Brockton-Nashua, MA-NH-ME-CT	1978, Jan	2007, Oct
APUA10474714	Pittsburgh, PA	1978, Jan	1997, Dec
APUA10574714		1978, Jan	1986, Dec
APUA10674714		1978, Jan	1986, Dec
APUA20774714	Chicago-Gary-Kenosha, IL-IN-WI	1978, Jan	2007, Oct
APUA20874714	Detroit-Ann Arbor-Flint, MI	1978, Jan	2007, Oct
APUA20974714	St. Louis, MO-IL	1978, Jan	1997, Dec
APUA21074714	Cleveland-Akron, OH	1978, Jan	2007, Oct
APUA21174714	Minneapolis-St. Paul, MN-WI	1978, Jan	1986, Dec
APUA21274714	Milwaukee-Racine, WI	1978, Jan	1986, Dec
APUA21374714	Cincinnati-Hamilton, OH-KY-IN	1978, Jan	1986, Dec
APUA21474714	Kansas City, MO-KS	1978, Jan	1986, Dec
APUA31174714	Washington-Baltimore, DC-MD-VA-WV	1998, Jan	2007, Oct
APUA31574714		1978, Jan	1997, Dec
APUA31674714	Dallas-Fort Worth, TX	1978, Jan	2007, Oct
APUA31774714		1978, Jan	1997, Dec
APUA31874714	Houston-Galveston-Brazoria, TX	1978, Jan	2007, Oct
APUA31974714	Atlanta, GA	1978, Jan	2007, Oct
APUA32074714	Miami-Fort Lauderdale, FL	1978, Jan	2007, Oct
APUA42174714	Los Angeles-Riverside-Orange County, CA	1978, Jan	2007, Oct
APUA42274714	San Francisco-Oakland-San Jose, CA	1978, Jan	2007, Oct
APUA42374714	Seattle-Tacoma-Bremerton, WA	1978, Jan	2007, Oct
APUA42474714	San Diego, CA	1978, Jan	1986, Dec
APUA42574714	Portland-Salem, OR-WA	1978, Jan	1986, Dec
APUA42674714	Honolulu, HI	1978, Jan	1986, Dec
APUA42774714	Anchorage, AK	1978, Jan	1986, Dec
APUA43374714	Denver-Boulder-Greeley, CO	1978, Jan	1986, Dec
APUB10074714	Northeast Size B	1978, Jan	1997, Dec
APUB20074714	Midwest Size B	1978, Jan	1997, Dec
APUB30074714	South Size B	1978, Jan	1997, Dec
APUB40074714	West Size B	1978, Jan	1988, Dec
APUC10074714	Northeast Size C	1978, Jan	1997, Dec
APUC20074714	Midwest Size C	1978, Jan	1997, Dec
APUC30074714	South Size C	1978, Jan	1997, Dec
APUC40074714	West Size C	1978, Jan	1997, Dec
APUD10074714	Northeast Size D	1978, Jan	1986, Dec
APUD20074714	Midwest Size D	1978, Jan	2007, Oct
APUD30074714	South Size D	1978, Jan	2007, Oct
APUD40074714	West Size D	1978, Jan	1986, Dec

Table 4: Regular leaded gasoline

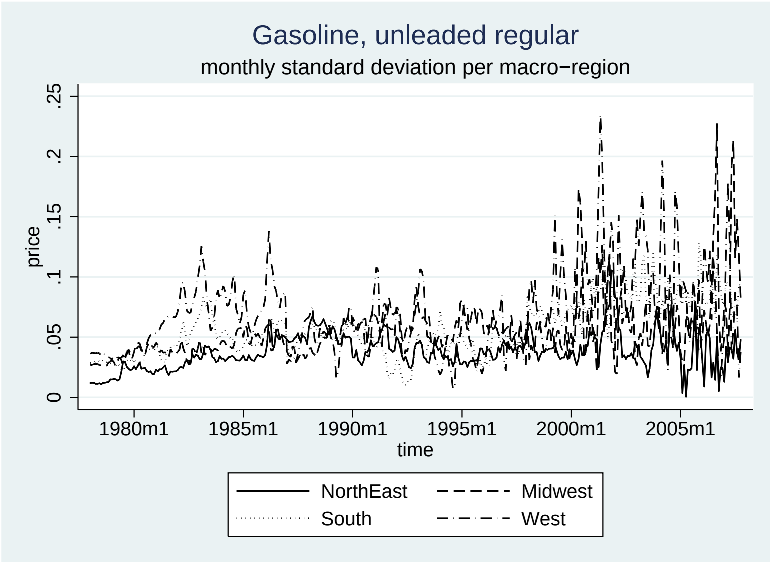


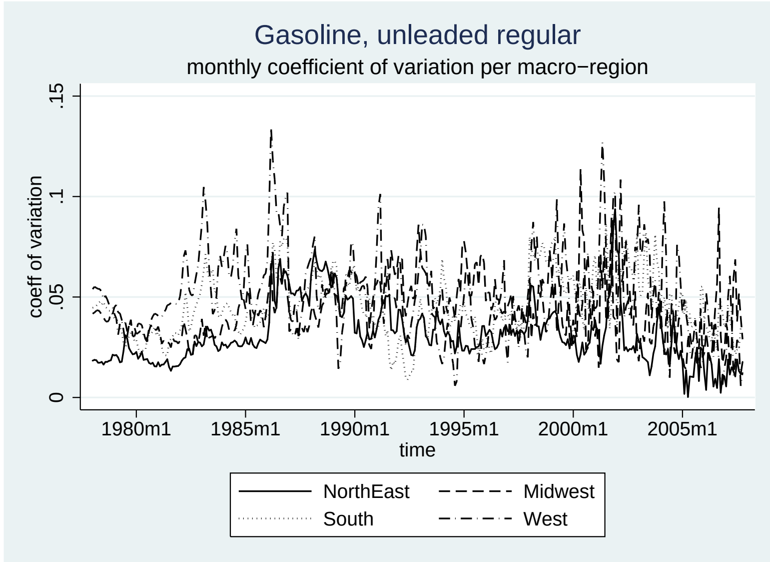


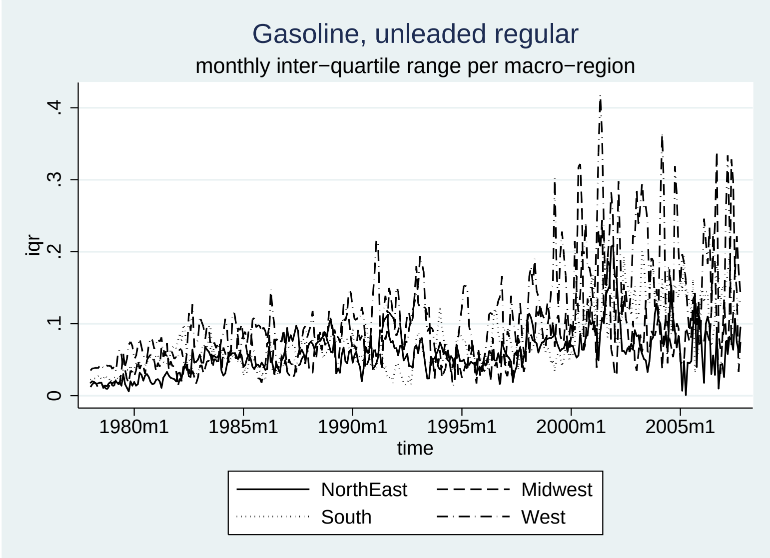


Variable	Mean	Std. Dev.	Min.	Max.	N
North East					
standard deviation	0.025	0.009	0.007	0.044	138
coeff. variation	0.024	0.008	0.008	0.049	138
interquartile range	0.03	0.018	0	0.092	140
Midwest					
standard deviation	0.04	0.009	0.018	0.058	145
coeff. variation	0.04	0.01	0.019	0.077	145
interquartile range	0.047	0.026	0	0.106	156
South					
standard deviation	0.043	0.017	0.018	0.105	146
coeff. variation	0.043	0.016	0.02	0.095	146
interquartile range	0.053	0.033	0	0.175	151
West					
standard deviation	0.052	0.029	0.001	0.137	160
coeff. variation	0.049	0.027	0.001	0.135	160
interquartile range	0.051	0.026	0.002	0.13	160

Table 5: Regular unleaded gasoline







Variable	Mean	Std. Dev.	Min.	Max.	N
North East					
standard deviation	0.038	0.015	0.001	0.105	358
coeff. variation	0.031	0.015	0	0.094	358
interquartile range	0.059	0.03	0.001	0.209	358
Midwest					
standard deviation	0.056	0.026	0.019	0.227	358
coeff. variation	0.044	0.015	0.014	0.114	358
interquartile range	0.075	0.051	0.009	0.339	358
South					
standard deviation	0.057	0.022	0.01	0.128	358
coeff. variation	0.046	0.016	0.009	0.088	358
interquartile range	0.075	0.042	0.015	0.23	358
West					
standard deviation	0.068	0.034	0.007	0.235	358
coeff. variation	0.05	0.022	0.006	0.134	358
interquartile range	0.11	0.068	0.013	0.418	358