Search and Equilibrium Prices:
Theory and Evidence from Retail Diesel

Luís Cabral
New York University and *CEPR*

Dominik Schober
University of Mannheim and *CEER (Zentrum für Europäische Wirtschaftsforschung)*

Oliver Woll
CEER (Zentrum für Europäische Wirtschaftsforschung)

March 2018

Abstract. We examine the relation between consumer search and equilibrium prices when collusion in endogenously determined. We develop a theoretical model and show that average price is a U-shaped function of the measure of searchers: prices are highest when there are no searchers (local monopoly power) or when there are many searchers (and sellers opt to collude). We test this prediction with diesel retail prices in Dortmund, Germany. We estimate a U-shaped relation with statistical precision and a €.025/liter price variation due to the variation in the measure of searchers.

Cabral: Paganelli-Bull Professor of Economics, Stern School of Business, New York University; and Research Fellow, CEPR; luis.cabral@nyu.edu. Schober: schober@zew.de. Woll: woll@zew.de. We are grateful to John Asker and to seminar participants at IIOC 2017 and APIOC 2017 for comments and suggestions. The usual disclaimer applies.

1. Introduction

Car driving is an integral part of many people's daily lives; understandably, gasoline prices are a concern for many drivers. A common complaint, especially in Europe, is that prices are too high. In addition to taxes, two common "culprits" for high prices are collusion and imperfect information.

In some cases, collusion has taken the form of explicit cartel agreements. For example, in 2011 the Brazilian competition policy authority (CADE) investigated various regional gasoline cartels. Several gas stations were indicted for secret agreements to maintain high prices.

In other cases, while there is no evidence of an explicit cartel, there are reasons to believe tacit collusion to be in place: For example, in 2011 Bundeskartellamt, Germany's Cartel Office, conducted an investigation regarding collusive behavior in the German retail gasoline and diesel market. It concluded that "retail prices of the majority of off-motorway petrol stations are higher than they would have been if effective competition had been in place" (sector inquiry, p. 19). Nevertheless, the Cartel Office abstained from direct intervention, arguing that "direct measures by the authority to reduce prices will have little hope of success" (ibid).

Regarding imperfect information, it is often said that there is considerable price dispersion, which makes it difficult for a driver to find the lowest-price pump. For example, at 5pm on May 1st, 2017, diesel prices in Dortmund, Germany, ranged from 49.5 to 60.4 cents per liter (net of taxes). In fact, imperfect information concerns have led competition authorities such as the Bundeskartellamt to collect and divulge price information.

In this paper, we look at the *combined* effect of search and collusion in determining retail gasoline prices. We derive a theoretical model of pricing and price search in the tradition of Varian (1980). We show that, the greater the measure of searchers, the lower the average competitive equilibrium price. This is not entirely surprising and is in line with previous theoretical results.

Next we consider the possibility of collusion. Whereas much of the previous literature on collusion has focused on the *feasibility* of collusion, our primary focus is on the expected *profitability* of collusion. In other words, whereas the previous literature has focused primarily on the incentive constraint (no-deviation from collusive agreement), we focus on participation constraint (collusion vs static Nash equilibrium). Our point is relatively simple: aside from product market revenues and costs, engaging in collusion has a positive expected cost. This includes the cost of reaching an agreement (explicit collusion), finding a focal point (tacit collusion), as well as the fines or other antitrust penalties in case the firm is indicted for collusion. In this context, an increase in the extent of search, by lowering equilibrium profits in the no-collusion equilibrium, increases the relative benefit from collusion, and thus the probability that collusion takes place.

Taken together, the static and collusion effects imply the prediction of a non-monotonic relation between the extent of search and expected price, specifically, a relation that is U-shaped: If there is very little search, then there is no collusion but equilibrium price is high since, absent the disciplining device of consumer search, sellers enjoy local monopoly power. By contrast, if many consumers are searchers then, absent collusion, equilibrium price is very low. This implies that the gains from collusion are high, the likelihood collusion is high, and so is average equilibrium price. In sum, theory predicts that average price is high if the measure of searchers is either very low or very high.

We test our theoretical predictions with retail diesel fuel data from Germany. Our basic regression has price as a dependent variable (at the gas station level) and measures of market search as independent variables. Specifically, we proxy for the extent of search by measuring the percentage of young inhabitants in the relevant area: work by Germany's Bundeskartellamt indicates that young people are disproportionately more likely to use a price-comparison app that greatly helps the process of finding low gasoline prices. (Below we discuss the validity of this proxy for the extent of consumer search.)

We also consider a series of other controls, including in particular the degree of competition in each station's neighborhood. Following Schaumans and Verboven (2015), we instrument for the number of competitors by measuring the number of competitors in the outer ring of each gas station's market.

Our results are broadly consistent with the theoretical prediction of a U-shaped relation between the degree of search and price. Considering the sample range of the variable "share of young people," we estimate (with statistical precision) a price range (highest estimated average price minus lowest estimated average price) of about 2.5 cents of Euro per liter, a value that, extrapolated to the German gasoline market, would correspond to €1.7 billion per year.

■ **Related literature.** The effect of search costs on the nature of oligopoly competition has been a topic of research interest at least since Stigler (1964). From a formal, game-theoretic standpoint, two important articles are Varian (1980) and Stahl (1989), who develop models where consumers can be divided into searchers and non-searchers; Varian (1980) considers the case when search is simultaneous, whereas Stahl (1989) assumes sequential search. Both of these models imply that average price is decreasing in the fraction of searchers (that is, consumer search increases the level of market competitiveness). Moreover, both models imply price dispersion and that the variance of the price distribution is a non-monotonic (inverted-U) function of the degree of searchers: price dispersion is minimal when the fraction of searchers is 0 or 100%.

When it comes to collusion, the literature on the effects of consumer search is surprisingly scarce. (There is an extensive literature on transparency, but it refers to transparency among sellers.) Building on the work of Nilsson (1999) and Schultz (2005), Petrikaitė (2014) develops a sequential-search model in the tradition of Stahl (1989) and shows that the critical discount factor above which collusion is sustainable is a non-monotonic function of the share of shoppers: first decreasing and then increasing.[1] This results from the different rates at which deviation profits and punishment profits vary as a function of the measure of searchers. To the extent that more favorable collusion stability conditions are associated with higher prices, Petrikaitė's (2014) result suggests an inverted U-shaped relation between the measure of searchers and equilibrium price (in homogenous-product markets): collusion is most likely (and prices highest) when the measure of searchers is neither too low nor too high.

Differently from Petrikaitė (2014), our theoretical model assumes simultaneous search, in the tradition of Varian (1980), an approach we believe is more appropriate for the empirical case we consider. More important, rather than stressing the stability of collusion we focus on the participation constraint (is collusion worthwhile?). This results in a prediction

1. This result is derived analytically for the two-seller case and numerically for higher values of n.

regarding the relation between the fraction of searchers and average price which is the opposite of Petrikaitė's (2014). Our paper also includes an empirical test which vindicates our theoretical prediction.

From an empirical point of view, Pennerstorfer et al. (2015) is closest to our paper.[2] They find evidence of an inverted-U relationship between the measure of searchers and the degree of price dispersion, as well as a monotonically decreasing relationship between the measure of searchers and average price. Both of these results are consistent with Varian (1980) and Stahl (1989). Our results differ in that, for high values of the measure of searchers, average price is increasing in the measure of searchers. At a theoretical level, this result is consistent with a framework where we add the possibility of collusion to Varian's (1980) static model; at an empirical level, we identify with statistical precision an effect of significant economic impact (2.5 cents of Euro per liter).

■ **Road map.** The rest of the paper is organized as follows. In Section 2 we lay down our theoretical model. First, we present the static game, which follows Varian (1980). Second, we consider the possibility of collusion with the assumption that firms collude if and only if the gains from collusion exceed collusion costs. Section 2 includes our main theoretical result (average price is a U-shaped function of the measure of searchers) and concludes with a preliminary regression on pseudo-data generated by the theoretical model.

Section 3 presents the data we use for empirical analysis, in particular our choices of variables to measure price and the extent of search. Section 4 presents the results of our basic regressions of price on the extent of search, as well as a series of robustness checks.

Section 7 concludes the paper.

2. Theoretical analysis

This section lays out the theoretical framework which we then use in Section 4 to test specific predictions. We first develop our basic model of demand and competition in retail. We then consider two possible games (and equilibria) resulting from the basic model: the equilibrium of the static game and the collusive equilibrium of an infinitely-repeated version of the static game.

■ **Model.** Consider a market with n firms and a measure 1 of consumers. Each consumer has a valuation r (reservation level) for at most one unit supplied by one of the firms. We assume that each consumer is initially "attached" to a firm, each firm with equal probability. Following Varian (1980), a fraction ϕ of these consumers are searchers, meaning they purchase from the lowest-price seller independently of the firm they are attached to.[3] The other consumers, a fraction $1-\phi$, are loyal consumers, which means they only purchase from the firm they are attached to, if at all.

2. Pennerstorfer et al. (2015) also cite a series of studies on the relation between the measure of searchers and price dispersion based on Internet search data: Brown and Goolsbee (2002), Tang et al. (2010), Sengupta and Wiggins (2014). Following Ellison and Ellison (2005) and Ellison and Ellison (2009), they question the validity of online data for a "clean" analysis of the relation between the measure of search and market outcomes.

3. If more than one firm sets the same lowest price, we assume consumers are equally likely to choose any of these; unless one of the firms is the "home" firm, in which case the consumer chooses the home firm. This assumption is not critical for our results but simplifies the analysis.

■ **Static and collusive equilibria.** The static equilibrium of this game corresponds to Varian (1980). Firms mix between r and lower values of p, all the way down to $\underline{p}$. Each firm is indifferent between any of these prices: a higher price implies a higher margin but also a lower probability of attracting any searchers (who purchase from the firm setting the lowest price). Given each firm's indifference, equilibrium profits are easy to determine: each firm is as well as setting $p = r$, which leads to profits

$$\widehat{\pi} = (1 - \phi)\, r/n \tag{1}$$

where the hat over π denotes static Nash equilibrium.

Consider now the infinite repetition of the static game, assuming that each player discounts the future according the discount factor δ. It is well known that, if δ is close enough to 1, then there exists a collusive equilibrium such that firms set $p = r$ in each period. In fact, there exist many such equilibria. A particularly simple one corresponds to grim strategies: each firm sets $p = r$ if, in the past, all firms set $p = r$; and all firms revert to playing the static Nash equilibrium forever if ever any firm deviates from $p = r$. Since all firms set the same price, all consumers purchase from the seller they are assigned to. It follows that firm profit is given by

$$\pi^* = r/n \tag{2}$$

regardless of firm type. Note that, as expected, $\pi^* > \widehat{\pi}_i$.

■ **The determinants of collusion.** Grout and Sonderegger (2005) aptly summarize the theoretical literature on collusion by stating that it is "primarily concerned with the compliance of independent firms with agreements that reduce competition within a market." Specifically, much of the extant theoretical work is based on the repeated-game framework; and typically assumes firms play grim-strategies (set monopoly prices and, if a firm deviates from the prescribed equilibrium, revert to the static equilibrium forever).

A common result in this literature is that, if the discount factor is greater than some critical threshold δ', then grim-strategy collusion is feasible (see, e.g., Friedman, 1971). The literature then goes about deriving comparative statics results with respect to various exogenous parameters: that is, studying how each of these exogenous parameters affects the critical value δ'. As Harrington (2015) put it, "the focus of economic theory has been on characterizing the market conditions conducive to satisfying the stability condition." In the context of search and collusion, a particularly important reference is Petrikaitė (2014).

While obviously there is value in this approach, we believe it misses an important issue: by stressing whether collusion is *feasible*, it largely ignores the issue of whether collusion in *profitable*. To quote Harrington (2015),

> When there exists a stable collusive arrangement, when is it that firms want to replace competition with collusion? That is, when collusion is feasible, when is it desirable?

Accordingly, we follow a route different from most of the previous literature. We assume that the discount factor δ is sufficiently high so that colluding by setting $p = r$ is part of a repeated-game Nash equilibrium.[4] We then go back to Becker's (1968) classic approach

4. This is consistent with Harrington's (2015) claim that "many industries which could sustain a collusive arrangement, do not; and there are many instances of cartels which could have effectively operated prior to when the cartel was formed but did not." This suggests that $\delta > \delta'$ is not the binding constraint.

and ask the question: when does crime (in this case collusion) pay?[5]

In answering this question, we make an important assumption: each period that firms collude, they must pay a collusion cost c. The idea is that, by engaging in tacit or explicit collusion, firms create a liability for themselves: the possibility that an investigation be initiated that might lead to a conviction or, at the very least, a compliance cost with competition authorities, as well a cost in terms of public relations.

Within this framework, the question we address is whether collusion is worthwhile once the collusion cost is taken into consideration. The answer is quite straightforward: collusion pays if and only if

$$\pi^* - c > \widehat{\pi} \tag{3}$$

which is equivalent to

$$\phi \geq \phi' \equiv \frac{n\,c}{r} \tag{4}$$

■ **Comparative statics.** Equation (3) corresponds to the core of our analysis: the cost-benefit analysis of collusion. Whereas most of the prior literature focuses on the stability of collusion, we ask the question of when firms prefer collusion to the alternative of static Nash competition. Similar to the prior theoretical literature on collusion, we ask how exogenous parameters impinge on whether (3) is or is not satisfied. This brings us to our central result:

Proposition 1. *For low values of ϕ (resp. high values of ϕ), the price distribution is decreasing (resp. increasing) in ϕ in the sense of first-order stochastic dominance. The same is true for the minimum value of the support of the price distribution.*

Proof: Consider first the static equilibrium (no collusion). By setting monopoly price a seller expects to sell only to loyal consumers, that is, the non-searching consumers who are attached to the seller. This leads to a profit of

$$(1-\phi)\,r/n$$

By selling at a lower price, a seller expects to sell to non-searchers plus to all searchers if its price p is lower than the prices set by all other sellers. This corresponds to an expected profit of

$$p\left(\frac{1-\phi}{n} + \phi\left(1-F(p)\right)^{n-1}\right)$$

Mixing implies indifference between any p in the interval $[\underline{p}, r]$, that is, implies the equality of the above two profit expressions:

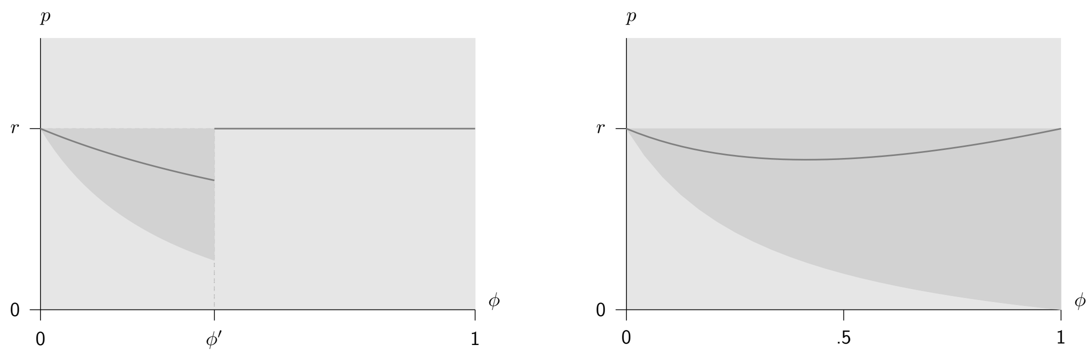
$$p\left(\frac{1-\phi}{n} + \phi\left(1-F(p)\right)^{n-1}\right) = \frac{1-\phi}{n}\,r \tag{5}$$

for all $p \in [\underline{p}, r]$, where $F(p)$ is the seller's price mixed strategy. olving for $F(p)$, we get

$$F(p) = \sqrt[n-1]{1 - \frac{1-\phi}{\phi\,n}\,(r-p)} \tag{6}$$

5. See also Landes (1983).

Figure 1
Proposition 1 illustrated



which is increasing in ϕ, which implies the distribution of p is decreasing in p in the sense of first-order stochastic dominance.

From (4), the above pricing pattern prevails if and only if $\phi < \phi' \equiv n\,c/r$. For $\phi > \phi'$, firms prefer to engage in collusion and set $p = r$. Together with the previous analysis, we conclude that: (a) If ϕ is sufficiently low, then $\phi < n\,c/r$ and price is decreasing in the sense of first-order stochastic dominance. (b) If ϕ is sufficiently high, then $\phi > n\,c/r$ and $p = r$, which implies that price increases (weakly) in ϕ.

Solving (5) for $F(\underline{p}) = 0$, we get

$$\underline{p} = \frac{(1-\phi)\,r}{1+\phi\,(n-1)} \tag{7}$$

from which we get

$$\frac{d\,\underline{p}}{d\phi} = -\frac{n\,r}{\big(1+\phi\,(n-1)\big)^2} < 0$$

and the result for $\underline{p}$ follows in the same manner as the ordering in term of first-order stochastic dominance. ■

In words, if the measure of searchers is small then the competitive solution (no collusion) prevails, and price is decreasing as the measure of searchers increases. If however the measure of searchers increases beyond the c/r threshold, then prices increase to monopoly price, whereas further increases in ϕ lead to price constant at r. Together, these results generate the two branches of the U-shaped relation between ϕ and p.

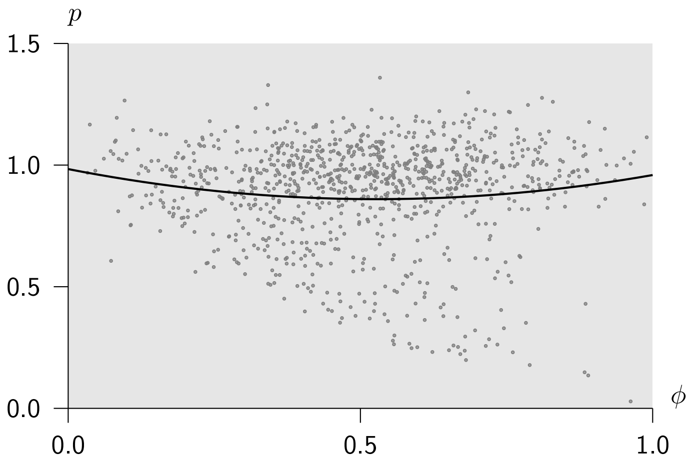
■ **Proposition 1 illustrated.** Suppose for simplicity that $n = 2$. Then (6) implies that p is uniformly distributed between $\underline{p}$ and r. Moreover, from (7) we conclude that $\underline{p} = (1-\phi)/(1+\phi)\,r$. It follows that average price is given by

$$\overline{p} = \tfrac{1}{2}(r+\underline{p}) = \tfrac{1}{2}\left(r + \frac{1-\phi}{1+\phi}\,r\right) = \frac{r}{1+\phi}$$

If $\phi \geq \phi'$, however, then collusion takes place and $p = \underline{p} = r$.

Combining the two cases, we obtain the values of p as a function of ϕ. This is shown in the left panel of Figure 1, which in turn illustrates Proposition 1. Note that, for low values of ϕ, average price is decreasing in ϕ. (Price is distributed in the shaded region.) However,

Figure 2
Model-generated pseudo data (and estimated OLS regression)



as ϕ increases from a value lower than ϕ' to a value greater than ϕ', then average price increases. Together, these observations illustrate the U-shaped relation between the extent of search and average price.

The discontinuity implied by the switch from no-collusion to collusion leads to a relation between p and ϕ which is non-monotonic but not very similar to a U. Partly, this results from our stark assumption of a fixed value of c, which in turn leads to collusion with probability 0 or 100%. Suppose however that the value of c is randomly distributed. For example, suppose that c is uniformly distributed between 0 and $\overline{c}$. Collusion profit is given by $\pi^* = \frac{1}{2}\, r$, whereas static Nash equilibrium profits are given by $\frac{1}{2}\,(1-\phi)\, r$. It follows that collusion takes place iff and only if $c \leq \frac{1}{2}\, r - \frac{1}{2}\,(1-\phi)\, r = \frac{1}{2}\,\phi\, r$, which happens with probability $\frac{1}{2}\,\phi\, r / \overline{s}$. By contrast, with probability $1 - \frac{1}{2}\,\phi\, r/\overline{s}$ average price is given by $r/(1+\phi)$ (as before). Overall, average price is now given by

$$\overline{p} = \tfrac{1}{2}\,\phi\,\frac{r}{\overline{s}}\, r + \left(1 - \tfrac{1}{2}\,\phi\,\frac{r}{\overline{s}}\right)\,\frac{r}{1+\phi}$$

For simplicity, suppose that $r = 1$ and that $\overline{s} = \frac{1}{2}$. Then the above reduces to $(1+\phi^2)/(1+\phi)$, which is shown on the right-hand panel of Figure 1. Now, the "forces" of competition and collusion work in a continuous way, resulting in a U-shaped relation between the measure of searchers and average price. For low values of ϕ, an increase in ϕ has a strong negative effect on average price. Moreover, the possible switch to collusion (which takes place in the unlikely event that collusion cost is very small) does not have a big effect on average price because average price itself is very high. As ϕ continues to increase, the difference between static Nash profits and collusion profits increases. As a result, the likelihood that collusion cost is lower than the gain from collusion increase. Moreover, the gap between static equilibrium price and collusion price in greater, which implies that the switch from no collusion to collusion has a greater impact on price. This in turn explains why the slope of the relation between ϕ and p becomes positive.

■ **Pseudo data.** If we were to take Proposition 1 literally, that is, for a specific set of parameter values n, r, c and changing the value of ϕ, then we would expect a cloud of points (ϕ, p) for $\phi < \phi'$ and $p = r$ for $\phi > \phi'$. In reality, in particular if we were to obtain cross-section data, we would expect the values of n, r, c also to vary.

Specifically, in order to get an better feel for the nature of Proposition 1, we next use our theoretical model, augmented by simple distribution assumptions on its parameters, to generate a pseudo data set of individual gas station prices. Specifically, we assume that $c \sim N(.1, .05)$, $r \sim N(1, .1)$, $\phi \sim N(.5, .2)$ and $n \in \{2, ..., 6\}$ (the latter with all values equally likely).

Figure 2 shows a scattered plot of the data generated as well as the results of an OLS regression of price on ϕ, ϕ^2 and n. Although the R^2 is rather low (not surprising, given that firms play mixed-strategies), the prediction of Proposition 1 is borne out by the data. In particular, we notice a U-shaped relation between ϕ and average price. In the next sections, we seek to test the predictions of Proposition 1 with actual data.

■ **Testable predictions.** The property that the price distribution is decreasing (or increasing) in ϕ (in the sense of first-order stochastic dominance) implies a series of testable predictions for gas-station-level regressions. For low levels of ϕ, (a) price is decreasing (in the sense of first-order stochastic dominance) in ϕ; (b) the lowest value of the price distribution is decreasing in ϕ; and (c) the mean value of the price distribution is decreasing in ϕ. For high values of ϕ, the opposite is true.

3. Data

Our empirical test consists of regressing measures of seller pricing behavior on measures of consumer search behavior. We divide our data description accordingly.

■ **Prices.** Since 2013, the German Cartel Office (Bundeskartellamt) — specifically, its Market Transparency Unit for Fuels — has been collecting detailed retail fuel prices in an effort to improve its ability to oppose illegal market practices. Companies which operate gas stations are obliged to report price changes for the most commonly used types of fuel — Super E5, Super E10 and Diesel — in real time.

We obtained from the Market Transparency Unit for Fuels data on real-time prices at the pump in a variety of German cities during the month of May 2015. So as to work with a manageable-sized dataset, we use 15-minute time intervals. Moreover, given limited availability of crucial RHS-variable data, we focus our analysis in the city of Dortmund, a city with close to 600,000 inhabitants and 86 gas stations (as of May 2015).

We complement station-level price data with additional station-level information. Following the above observation regarding the use of price-search apps, we proxy for the fraction of searchers by measuring the share of young inhabitants (ages 16 to 29) in each of the local areas Dusseldorf is divided into.[6]

Regarding a station's local market competition, we consider two variables. First, the fraction of neighboring stations that are not branded. To the extent that no-brand stations are more competitive, this variable measures one dimension of market competitiveness. According to a German Cartel Office 2011 investigation, the branded stations belong to one

6. Note that we distinguish between local markets (defined as the market faced by each gas station) and local areas (defined as the areas of the municipality of Dortmund for which average age is available).

Figure 3
Average price by 15 minute daily intervals

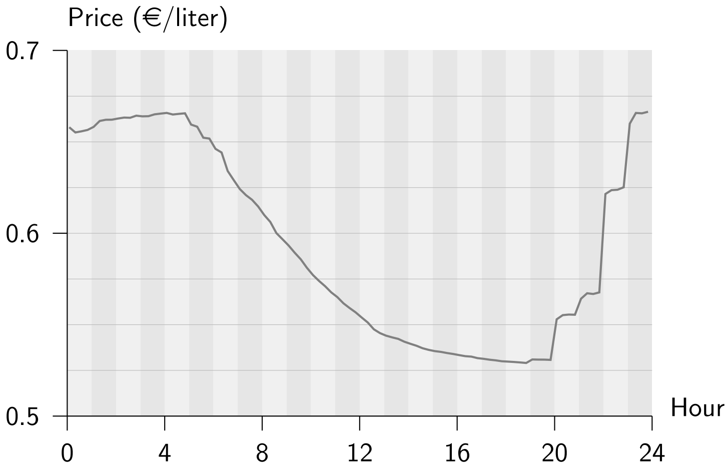
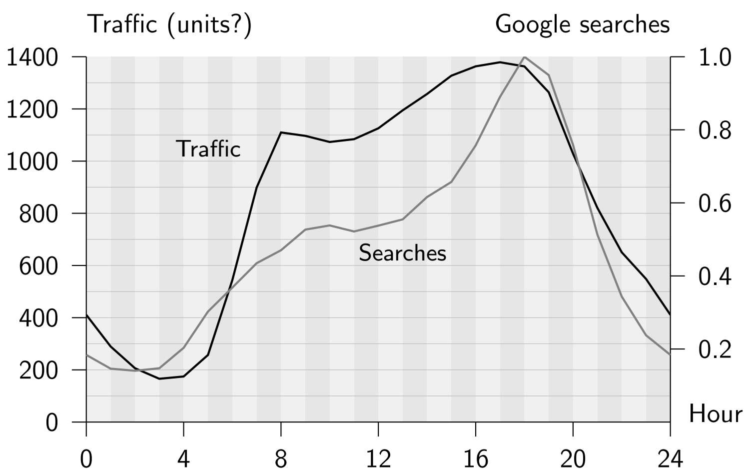


Figure 4
Traffic and search intensity



of the following four networks: Aral, Esso, Shell and Total. They represent 50.49% of all gas stations in our dataset.

Additionally, we measure the number of competitors as a proxy for the intensity of competition, which should have a substantial influence on price levels. Specifically, we consider the number of competitors within a 1.25-mile radius. Since the number of competitors depends on price itself, the number of stations in the greater environment of 3 to 12 miles (divided by population) is used as an instrument for market potential (Schaumans and Verboven, 2015).

We considered additional local-market variables, e.g., number of cars and household income. These variables did not seem to have any significant correlation with prices and were thus excluded from our analysis.

As Figure 3 shows, there is a clear daily price-setting pattern, with higher prices at night gradually declining throughout the day and reaching lower levels during rush hour.

Figure 5
Diesel price distribution

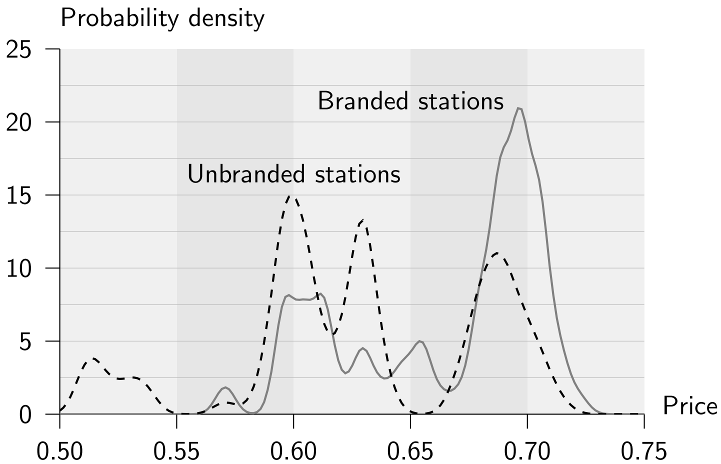


Figure 4 shows the traffic and search patterns by time of day. As can be seen the 5-6pm period is both the the period of highest traffic and the period of most intense search (as measured by Google searches). Considering the problem we are interested in analyzing (the effect of search on market competition), we focus our analysis during rush-hour market competition, that is, the four 15-minute slots beginning at 5pm. Further analysis suggests that the qualitative results are robust to changes in the time frame considered.

Figure 5 shows the kernel density of prices, with branded and non-branded gas stations classified separately. A striking feature of this figure is the significant difference between prices by branded and non-branded stations. This difference justifies their different treatment as well as the use of the fraction of non-branded stations as a measure of market competitiveness. Additionally, we observe significant variation in prices, which is consistent with a central feature of our theoretical model (which in turn follows Varian, 1980, to a great extent).

■ **Search.** We are unable to observe consumer search directly. Instead, we create an indirect measure of the extent of consumer search. The Market Transparency Unit for Fuels — which we referred to at the beginning of the section — does not offer price information directly to the public. However, a variety of private consumer information service providers have access to it and consumers in turn can search prices by accessing these services.

> Via the internet, a smartphone or navigation system, motorists will be able to gain information on the current fuel prices and find the cheapest petrol station in their vicinity or along a specific route. This will allow for a better overview of prices and an informed choice which will increase competition.[7]

(Note that, in this setting, our assumption regarding search — searchers obtain information on all prices at once, rather than sequentially — seems justified.) By the Transparency Unit's own account,

7. http://www.bundeskartellamt.de/, visited July 2017.

Table 1
App downloads and age distribution

Dependent variable:	App use		
Age 18 to 30	.6064*** (.1110)		
Age 18 to 50		.5209*** (.0991)	
Age 50 to 65			-.8216*** (.1925)
Constant	-4.3977*** (1.3601)	-16.7038*** (3.7347)	22.3748*** (4.4778)
R^2	.0912	.1218	.1003
N	512	512	512

Notes: Standard errors in parentheses. Star levels: 10, 5 and 1%.

> The gasoline price app is very popular with consumers: 24 percent of German car drivers have already used the offer since its introduction in September and have compared gas prices over the Internet or smartphone apps; another 61 percent have heard of the possibility but did not use it yet. The response to the gasoline price app is particularly high for men and for younger age groups. With regard to gender, 30 percent of men and 18 percent of women have already compared prices. With regard to age, 39 per cent of the drivers in the 16–29 age bracket have used the app, as opposed to only 14 percent of the 60+ age bracket.[8]

Consistently with this evidence, we propose as a proxy for the extent of search the fraction of local population in the 16–29 age bracket ("share of young", or simply SOY). Specifically, the municipality of Dortmund is divided into 62 areas. For each gas station (each observation) we use the value of SOY in the area where the gas station is located as a measure of the extent of search the gas station's local market is subject to.

Our empirical analysis hinges on our measure of the extent of search. For this reason, some additional discussion of our measure is in order. Two conditions must be satisfied for SOY to be a good proxy. First, it must be the case that age is a good proxy for the use of the app (or apps) that provide information about prices. Second, it must be that the set of residents in a station's local area is a good proxy for that station's potential demand.

Regarding the first condition, a reality check is provided by data on app downloads and app use. We obtained data on the number of search queries entered in one of the app providers in Germany, *Tank-Navigator* (by Mammuth Applications).[9] Table 1 displays the results of simple OLS regression of app use on an age dummy. The unit of observation is a municipality in Germany ($N = 512$).[10] The dependent variable is total number of

8. Our translation from the German.
9. See https://www.tank-navigator.de/. Mammuth Applications is ranked among the bigger price comparison website providers: it belongs to the 100–500k app downloads category.
10. Age distribution is obtained from http://www.inkar.de/

app queries per municipality per capita by municipality; the independent variable is the percentage of municipality population of age 18–30 (first equation) or age 18–50 (second equation). The results clearly indicate that the presence of younger age groups has a strong, positive impact on the frequency of search queries.

Regarding the second condition (place of residence and place of purchase), one would expect each driver's consideration set to include stations near home, near the workplace, or along the daily commute. Our restriction to the area of residence is clearly a restriction to a subset of the real consideration set. However, to the extent that there is a positive correlation between a driver's home location and the probability that local gas stations belong to their consideration set, SOY works as a proxy (though possibly a noisy proxy). Moreover, we considered a number of alternative definitions of the age-by-location variable (the results of which are included in the robustness checks section) and found that the results are remarkably stable.

Notwithstanding the above arguments, the question may be raised as to how important search is in determining prices: is the measure of searchers that significant? Is there sufficient variation in this measure? The Germany Cartel Office's website provides a list of price-search apps as well as an estimate of the number of downloads per app.[11] A total of 57 apps are listed, of which download estimates are available for 25. For each of these, we obtain a lower bound and an upper bound of the number of downloads. The sum of the lower bounds is given by 4.56 million, whereas the sum of the upper bounds is given by 19.2 million. Considering that we only have data for 25 out of 57 apps, we might extrapolate these bounds by multiplying by the factor 57/25. This gives 10.4 million (lower bound) and 43.8 million. In order to get an idea of scale, note that the total number of registered cars in Germany is given by 45.8 million.[12]

There are reasons to believe the upper bound 43.8 million overestimates the total number of downloads. For example, apps with missing data are likely to be smaller than apps for which data is available. Moreover, some users my have downloaded more than one app, so the number of downloads may overestimate the number of users who have downloaded apps. Against that, we should also consider the fact that car usage is not uniform; and that app downloaders are likely to have purchased diesel more frequently than average.

The mere fact a user downloads an app does not mean he or she is a searcher: a searcher is someone who downloads an app and uses it. As mentioned earlier, we have data on usage for a particular app, *Tank-Navigator*. From June 18, 2014 to October 16, 2016 (2.33 years) we observe 1,914,096 queries, or 821,500 per year. Considering that the number of app downloads varies from 100 and 500 thousand, this corresponds to 8.21 or 1.64 queries per user per year.

Taking the simple average of the above two numbers, we get an estimate of 4.9 queries per user per year. The average number of visits to the pump is given by 26.9 times per year.[13] Assuming that search implies one query per refueling event, we estimate that about 18% of refueling events are associated with a price query.

Due to limited data availability, these calculations are very gross. That said, considering

11. (http://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/MTU-Fuels/mtufuels_node.html

12. www.kba.de/DE/Statistik/Fahrzeuge/Bestand/bestand_node.html#rechts

13. This data was obtained from "Gesellschaft für Konsumforschung (GfK)" (2012), a report by GfK Tankstellenpanel Deutschland, a major consumer research corporation. See www.gfkps.com/imperia/md/content/ps_de/tankstellenpanel/gfk_tankstellenpanel_vortrag_29_februar_2012.pdf.

Table 2
Descriptive statistics

Variable	# obs	Mean	St.Dev.	Min	Max
Price (net, Euro)	180,855	0.57	0.06	0.50	0.80
Share of young (age 18-29)	180,855	0.14	0.03	0.09	0.16
Share of young (age 18-65)	180,855	0.64	0.01	0.60	0.66
Number of competitors	180,855	6	3	1	16
Share of branded stations	180,855	0.50	0.30	0	1
Branded station	180,855	0.50	0.50	0	1
Stations / population	180,855	0.014	0.0059	0.0051	0.0311

the order of magnitude of our estimates of searchers; considering the relation between search activity and user age; and considering the geographical variation in age, we believe that the effect identified by our theoretical model plausibly is also present in the data and is identified by the cross-sectional variation in our age variable.

■ **Summary statistics.** Table 2 presents summary statistics of the data we use for our regressions. Notice in particular that the variable "Share of young," which plays a central role in our analysis of the relation between search and price, varies from 8% and 20%, with a standard deviation of 3%.

4. Results

Proposition 1 implies a U-shaped relation between measures of price (price, average price, minimum price) and the extent of market search. As a proxy for the latter, we measure the fraction of young consumers in the gas-station's neighborhood (SOY). Therefore, a natural way of testing Proposition 1 is to regress measures of prices on a quadratic function of the share of young people; that is, to use SOY and SOY^2 as explanatory variables. Proposition 1 predicts a negative coefficient on the squared term.

Our observations are at the gas station level. As mentioned earlier, we consider rush-hour 15 minute time intervals in Dortmund, a total of 12,398 observations. Our basic regression model takes the form

$$p_{it} = \beta_0 + \beta_1 \, SOY_i + \beta_2 \, SOY_i^2 + \beta_3 \, \widehat{NOC}_i + \beta_4 \, SNB_i + \beta_5 \, BRA_i + \epsilon_{it}$$

where i indexes the station, t the time period,

- p: price measure
- SOY: share of 16-19 year old population in station i's Dortmund area
- NOC: number of gas stations in station i's neighborhood (1.25-mile-radius circle)
- SNB: share of non-branded gas stations in station i's neighborhood
- BRA: dummay variable equal to 1 iff station i is branded (Aral, Esso, Shell or Total)

Table 3
Parametric regression of price equations

Dependent variable:	Price	Mean Price	Min Price
Share of young consumers	-2.4633*** (0.9430)	-2.1744*** (0.7231)	-2.9960*** (1.3602)
Share of young consumers squared	10.2963*** (3.6037)	9.1199*** (2.7585)	11.1640** (5.0516)
Number of Competitors	-0.0011 (0.0016)	-0.0011 (0.0010)	0.0018 (0.0012)
Share of branded stations	0.0036 (0.0070)	0.021** (0.0049)	0.0368*** (0.0082)
Branded station	0.0141*** (0.0023)	-0.0018 (0.0015)	-0.0060** (0.0027)
Cars per 10^9 people	.0949* (.0561)	.0849** (.0363)	.0100 (.0288)
Constant	0.6278*** (0.0638)	0.6161*** (0.0482)	0.7155*** (0.0918)
Number of observations	12,398	12,398	12,398
R^2	0.36	0.43	0.36

ϵ is an error term and β_k $(k = 0, ..., 5)$ are regression parameters.

Competitive intensity in a local market (here measured by NOC typically has an effect on price level. However, the number of competitors is determined endogenously by entry decisions, which in turn are based on local market price. To circumvent potential endogeneity issues, following Schaumans and Verboven (2015) we instrument for the endogenous number of local market participants by measuring the number of stations in the greater environment. Specifically, we regress p on $\widehat{NOC}$, the predicted value from the regression

$$NOC_i = \alpha_0 + \alpha_1\, SGE_i + \nu_i$$

where SGE_i is the number of competitors in the greater environment around station i, that is, the "donut" centered around station i (a 12-mile-radius circle that excludes the 3-mile-radius inner circle, both centered on station i's location). X_i is the usual second stage variables used for the IV regression.

We consider three different price measures:

- Price. Station i's price.
- Mean price. Average of all price set by stations in the neighborhood of station i, that is, within a 1.25 mile radius circle.
- Min price. Minimum price among all stations in station i's neighborhood, that is, within a 1.25 mile radius circle.

The results of our regressions are shown in Table 3. From a theoretical point of view, the coefficients of interests are the coefficients associated with the squared variables. As the

Figure 6
Estimated price regression (Table 3)

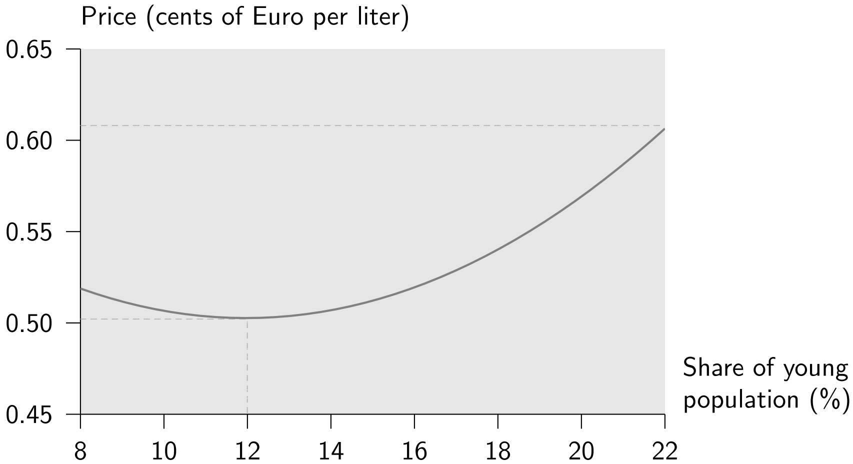


table shows, the coefficients on SOY^2 are all positive, as predicted by Proposition 1, and statistically significant (at the 5% significance level for price, at the 1% level for mean and min price).

In order to get an idea of the economic significance of these estimates, Figure 6 plots estimated price (in cents of Euro per liter) as a function of the share of 18-29 year olds in a station's local area. Although SOY only varies from 8 to 22% (cf Table 2), SOY's range implies a price variation of 2.5 cents, that is, about 5% of price. In a market with very low margins, we believe this is a significant range.

Figure 6 also allows us to restate our main theoretical point, this time with actual industry estimates: Gas stations in areas with very few searchers (low values of SOY) have relatively high market power (regardless of whether they are branded or not branded). For this reason, sellers find it optimal not to collude: the benefits from collusion do not compensate the potential costs. In this context, an increase in SOY, which proxies for an increase in the fraction of searchers, leads to a more competitive market. Given that firms do not collude, this corresponds to a lower price, as shown by the decreasing portion of the graph in Figure 6.

As SOY continues to increase, and price decreasing, seller profits become so low that it pays to collude: the cost is the same, but the benefit — the difference between collusion and competition profits — increases. As a result, an increase in SOY leads to an increase in price, due to the switch from competitive to collusive pricing.

5. Robustness checks

We performed a number of robustness checks, the results of which can be found in Table 4. In all regressions the dependent variable is gas station price. For reference, column (1) reproduces the results from the price equation in Table 3.

■ **Market definition.** We considered different radii in our definition of a local market as well as the neighboring market: 3 km instead of 2 km for the inner circle and 15 km instead of 12.5 for the outer circle. Column (2) in Table 4 displays the results of this alternative regression. As can be seen, the core coefficients estimates remain very similar in size and statistical significance.

■ **Share of young consumers.** We considered different definitions of the *SOY* variable. First, instead of focusing on the district where a gas station is located, we also include neighboring districts. The idea is that, since most drivers are commuters, they may fill up at a pump close to the workplace rather than the district of residence. Column (3) in Table 4 shows that, again, the value and the precision of the core coefficients remains virtually unchanged.

A different variation in the *SOY* variable consists in changing the age range. Column (4) corresponds to measuring *SOY* as the share of inhabitants with ages 30 to 65 (that is, the combination of the 18–29 and 30–65 brackets). Naturally, the independent variable's mean value is greater under this alternative definition: from .14 to .64. In order to make the models comparable, we compute the product of the estimated coefficient and the mean value of the dependent variable. Regarding SOY, we get $-2.4633 \times .14^2 - 2.4633 \times .14 = .4847$ in the base model and $20.1623 \times .64^2 - 25.2195 \times .64 = .4787$ in model (4). In other words, the relevant coefficient is remarkably similar.

Finally, column (5) combines the independent-variable changes in columns (3) and (4). The coefficient estimates change very marginally with respect to model (4), though their statistical significance is lower.

■ **Different functional forms.** Although a quadratic functional form is a common approach to estimate non-linear effects, it is not the only one. As an alternative, we considered a series of quantile regressions. Specifically, we divided our observations in an odd number of bins according to the *SOY* variable and estimate dummy variables for each quantile leaving the central quantile as the omitted variable.

Table 5 reports the results for terciles. As predicted by theory — and consistent with our quadratic regression approach — the coefficients for the first and third tercile are positive and statistically significant. The coefficient estimate of about 1 cent is lower than the 2.5 cent range we estimated based on our quadratic regression. However, we must take into account the fact that terciles correspond to a significant level of aggregation. Our results for quintiles and septiles are consistent with this interpretation. Specifically, the estimated coefficient of the price regression for the fifth quintile is .0217 (p value less than 1%); and the estimated coefficient for the seventh septile is .0222 (p value less than 3%), both of which are remarkably consistent with our 2.5 cent euro range estimate.

■ **Alternative dataset.** A particularly important robustness test is to repeat the analysis on a different dataset. Although the price data is available for all of Germany, other variables must be obtained on a municipality-by-municipality basis; and the way the data is organized is not uniform. For this reason, extending the analysis to other German cities is not a trivial task. That said, we were able to obtain similar *SOY* data (though not directly comparable) for Mannheim, Germany. The results of our regressions of price on the measure of searchers show a similar U-shaped pattern.

Table 4
Revised price equations with different independent variable definitions:
(1) Base model (price regression from Table 3)
(2) Different definitions of local market and neigbhorhood market
(3) SOY includes neighboring districts
(4) SOY includes ages 16–65
(5) Combination of (3) and (4)

Dependent variable: price:	(1)	(2)	(3)	(4)	(5)
Share of young consumers	-2.4633*** (0.9430)	-2.4091*** (0.9036)	-2.4769** (0.9953)	-22.1281** (9.6413)	-20.7947 (9.1514)
Share of young consumers squared	10.2963*** (3.6037)	10.1743*** (3.5175)	10.4910*** (3.9202)	17.7220** (7.6173)	16.6456** (7.2349)
Number of Competitors	-0.0011 (0.0016)	-0.0011* (0.0006)	-0.0018 (0.0018)	-0.0016 (0.0016)	0.0012 (0.0034)
Share of branded stations	0.0036 (0.0070)	0.0034 (0.0068)	0.0035 (0.0069)	0.0034 (0.0068)	0.0044 (0.0076)
Branded station	0.0141*** (0.0023)	0.0148*** (0.0023)	0.0136*** (0.0025)	0.0136*** (0.0023)	0.0145*** (0.0029)
Cars per 10^9 people	0.0949* (0.0561)	0.0847 (0.0574)	0.1015* (0.0563)	0.1065** (0.0460)	0.1086** (0.0537)
Constant	0.6278*** (0.0638)	0.6344*** (0.0597)	0.6273*** (0.0706)	7.3864 (3.0290)	6.9506** (2.8565)
Number of observations	12,398	12,398	12,398	12,398	12,398
R^2	0.36	0.36	0.34	0.39	0.34

Table 5
Regression results for terciles

Dependent variable:	Price	Mean Price	Min Price
Tercile t1	0.0080* (-0.0045)	0.0082*** (-0.0032)	0.0166*** (-0.0037)
Tercile t3	0.0062* (-0.0036)	0.0063** (-0.0026)	0.0035 (-0.0032)
Number of Competitors	-0.0021 (-0.0019)	-0.0020* (-0.0012)	-0.002 (-0.0014)
Share of branded stations	0.0039 (-0.007)	0.0206*** (-0.0051)	0.0357*** (-0.0075)
Branded station	0.0128*** (-0.0026)	-0.0031* (-0.0018)	-0.0063** (-0.0025)
Cars per 10^9 people	0.0266 (-0.0454)	0.0195 (-0.0279)	-0.0411* (-0.0212)
Constant	0.5216*** (-0.025)	0.5248*** (-0.0161)	0.5397*** (-0.0149)
Number of observations	12,398	12,398	12,398
R^2	0.31	0.37	0.38

6. Model calibration and estimation on pseudo data

One possible criticism of our analysis — and of much of reduced-form regression analysis — is that there are alternative narratives consistent with the empirical correlation we find. In this section we present one additional argument in favor of our interpretation: we calibrate our theoretical model and generate a pseudo data set. We then repeat our empirical analysis on this pseudo data set and obtain similar results to those coming from the actual dataset.

Our theoretical model implies that, taking into account the possibility of collusion, price is determined as follows:

$$p = \begin{cases} r & \phi \geq \phi' \\ \frac{(1-\phi)}{(1-\phi)+n\,\phi\,\xi^{n-1}}\; r & \phi < \phi' \end{cases}$$

where

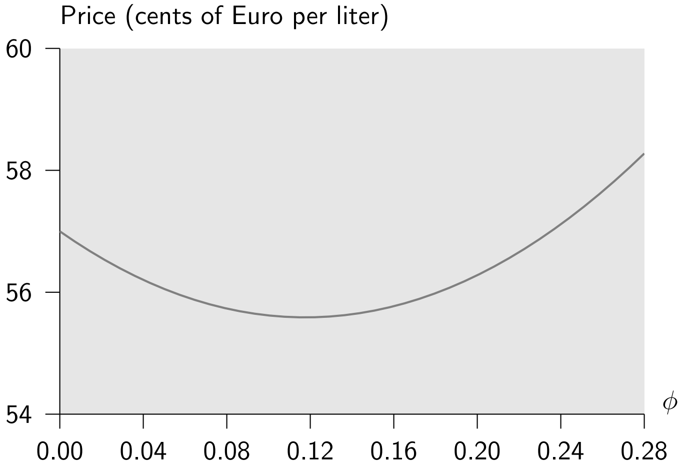
$$\phi' \equiv \frac{n\,c}{r}$$

and ξ is uniformly distributed in [0,1]. The above equation is obtained by solving (6) with respect to p and recalling that, by definition, the actual values of $F(p)$ are uniformly distributed in [0,1].[14]

Naturally, all markets are not equal in terms of the relevant exogenous parameters. Accordingly, we generate values of ϕ, n, r and c from random distributions with parameter values that calibrate the model to our Dortmund dataset. Specifically, we assume

14. That is, if we pick p at random according to $F(p)$, then the corresponding values of $F(p)$ are distributed according to $U[0,1]$.

Figure 7
Estimated regression with pseudo data



- $n \sim N(7, 3)$. These parameter values are obtained directly from Table 2
- $r \sim N(60, 6)$. The value of σ_r is chosen based on Figure 5, whereas the value of μ_r is chosen so as to replicate the mean price in the data.[15]
- $\phi \sim N(.14, .06)$. We assume that ϕ is equal to the measure of searchers. The parameter values are then obtained directly from Table 2
- $c \sim N(1.4, 1.4)$. We assume a unit coefficient of variation and calibrate μ_c so that collusion takes place in about one half of all of the observations.

We generate 12,398 observations (the same sample size as our actual data) and perform the same regressions as we do in our actual data set. The pseudo-data regression has $R^2 = .011$. We estimate the coefficients on the measure of searchers with very good statistical precision: the t statistics corresponding to the ϕ and ϕ^2 coefficients are typically between 3 and 10, depending on the particular random dataset generated.

Figure 7 shows the predicted relation between ϕ and price that is implied by the pseudo data set. Although the variables SOY and ϕ are not directly comparable (the former is a proxy of the latter), we note that the relation between the measure of searchers and the amplitude of predicted price is fairly similar across the two figures.

In sum, we conclude that, for reasonable parameter values, our theoretical model — a combination of the Varian model and the firms' decision to collude — replicates the U-shaped relationship found in the data. In other words, Proposition 1 is not just a possibility result, it is a reasonable possibility result.

7. Conclusion

This paper makes both a conceptual and an empirical contribution. At the conceptual level, we propose a new approach to the problem of tacit collusion. The vast majority of the IO literature has focused on the issue of sustainability of collusion. One criticism

15. See Table 2. Note that μ_r is the mean reservation price, not mean price.

of this approach is that, in the words of Harrington (2015), “many industries which could sustain a collusive arrangement, do not; and there are many instances of cartels which could have effectively operated prior to when the cartel was formed but did not.” Motivated by this criticism, we propose a different approach to tacit collusion, one that focuses on the participation constraint rather than the incentive (no-deviation) constraint. An application of this approach to the problem of search suggest a U-shaped relation between the measure of searchers and equilibrium price.

The second contribution of the paper is to present empirical evidence of a U-shaped relationship between the measure of searchers and price. We do so with retail diesel data from Dortmund gas stations. Our empirical results are consistent with theory: we estimate a U-shaped relationship with statistical precision. Moreover, we show that the range of variation in customer search leads is associated with a price variation of about 2.5 cents per liter (about 5% of sale price).

References

Becker, G. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, Vol. 76 (1968), pp. 1–54.

Brown, J.R. and Goolsbee, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy*, Vol. 110 (2002), pp. 481–507.

Ellison, G. and Ellison, S.F. "Lessons About Markets from the Internet." *Journal of Economic Perspectives*, Vol. 19 (2005), pp. 139–158.

—. "Search, Obfuscation, and Price elasticities on the Internet." *Econometrica*, Vol. 77 (2009), pp. 427–452.

Friedman, J.W. "A Non-cooperative Equilibrium for Supergames." *Review of Economic Studies*, Vol. 38 (1971), pp. 1–12.

Grout, P.A. and Sonderegger, S. "Predicting Cartels." Working Paper, Office of Fair Trading (2005).

Harrington, J. "Thoughts on Why Certain Markets are More Susceptible to Collusion and Some Policy Suggestions for Dealing with Them." Background paper, OECD Global Forum on Competition (2015).

Landes, W.M. "Optimal Sanctions for Antitrust Violations." *University of Chicago Law Review*, Vol. 50 (1983), pp. 652–678.

Nilsson, A. "Transparency and Competition." Technical report, Stockholm School of Economics Working Paper No. 298 (December) (1999).

Pennerstorfer, D., Schmidt-Dengler, P., Schutz, N., Weiss, C., and Yontcheva, B. "Information and Price Dispersion: Theory and Evidence." Technical report, University of Vienna Working Paper (2015).

Petrikaitè, V. "Collusion with Costly Consumer Search." *International Journal of Industrial Organization*, Vol. 44 (2014), pp. 1–10.

Schaumans, C. and Verboven, F. "Entry and Competition in Differentiated Products Markets." *The Review of Economics and Statistics*, Vol. 97 (2015), pp. 195–209.

Schultz, C. "Transparency on the Consumer Side and Tacit Collusion." *European Economic Review*, Vol. 49 (2005), pp. 279–297.

Sengupta, A. and Wiggins, S.N. "Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet." *American Economic Journal: Economic Policy*, Vol. 6 (2014), pp. 272–307.

Stahl, D.O. "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, Vol. 79 (1989), pp. 700–712.

Stigler, G. "A Theory of Oligopoly." *Journal of Political Economy*, Vol. 72 (1964), pp. 44–61.

Tang, Z., Smith, M.D., and Montgomery, A. "The Impact of Shopbot Use on Prices and Price Dispersion: Evidence from Online Book Retailing." *International Journal of Industrial Organization*, Vol. 28 (2010), pp. 579–590.

Varian, H. "A Model of Sales." *American Economic Review*, Vol. 70 (1980), pp. 651–659.