// NO.19-018 | 05/2019

DISCUSSION PAPER

// LUÍS CABRAL, DOMINIK SCHOBER, AND OLIVER WOLL

Search and Equilibrium Prices: Theory and Evidence from Retail Diesel

Leibniz Association

ZEW

Search and Equilibrium Prices: Theory and Evidence from Retail Diesel

Luís Cabral
New York University and *CEPR*

Dominik Schober
CEER (Zentrum für Europäische Wirtschaftsforschung) and *KU Eichstätt-Ingolstadt*

Oliver Woll
CEER (Zentrum für Europäische Wirtschaftsforschung)

First Version: March 16, 2018

This Version: March 27, 2019

Abstract. We examine the relation between consumer search and equilibrium prices when collusion is 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.

JEL L1, L4, L5, L9

Keywords Collusion, Cartelization, Fuel Retailing, Search, Competitive Intensity

---

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, APIOC 2017, and EARIE 2018 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 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 the 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 of 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 drivers among each gas station's customers: 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, which we measure with the number of competing stations. 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.[1]

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 (2016) 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.[2] 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 (2016) 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 (2016), our theoretical model assumes simultaneous search, in

1. See Bresnahan and Reiss (1991), Nevo (2001), Schaumans and Verboven (2015) for similar approaches.
2. This result is derived analytically for the two-seller case and numerically for higher values of n.

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 regarding the relation between the fraction of searchers and average price which is the opposite of Petrikaitė's (2016). Our paper also includes an empirical test which vindicates our theoretical prediction.

From an empirical point of view, Pennerstorfer et al. (2015) and Chandra and Tappata (2011) are closest to our paper.[3] 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).

A recent empirical paper which is close to our is Luco (2018), who considers the Chilean gasoline market. He makes use of the sequential implementation of an online price-disclosure policy to perform a differences-in-differences analysis. The question he addresses is different from the question we address in our theoretical model and empirical analysis. That said, there is a relation between the result of the two papers, an issue to which we return in Section 4.

■ **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 while subsequent Section 5 contains a series of robustness checks. In Section 6 we present additional evidence regarding our main narrative, namely that the incentives to engage in collusive practices is related to the propensity of consumers to search. 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

3. 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.

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.[4] 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.

■ **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 (2016).

4. 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.

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 is *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.[5] We then go back to Becker's (1968) classic approach and ask the question: when does crime (in this case collusion) pay?[6]

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.

As a first step in our analysis we make a series of simplifying assumptions. First, that collusion is a binary state: either firms collude or they do not. Second, the cost of collusion, c, is the same for all firms and known with certainty. Third, the cost of collusion is invariant with respect to price. These simplifying assumptions regarding collusion allow us to focus on the central focus of the paper: the relation between the propensity to search, collusion, and equilibrium prices. Later in this section we consider various extensions.

The question we address is whether collusion is worthwhile once the collusion cost is taken into consideration. In the simplified version of the model we consider, 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.*

5. 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.
6. See also Landes (1983).

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. Solving for $F(p)$, we get

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

The derivative of the root argument with respect to ϕ is given by $-(r-p)/n/p/\phi^2$, which is negative. It follows that the $F(p)$ is increasing in ϕ, which implies the distribution of p is decreasing in ϕ 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 ϕ 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

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

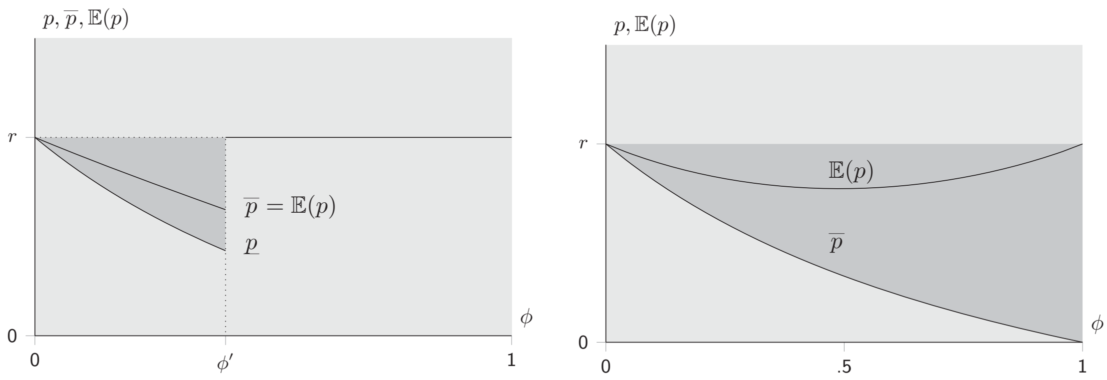
from which we get

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

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 $n\, c/r$ 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.

Figure 1
Proposition 1 illustrated



■ **Proposition 1 illustrated.** Suppose for simplicity that $n = 2$. From (6), we get

$$f(p) = \frac{(1-\phi)\,r}{2\,\phi\,p^2}$$

From (7) we conclude that $\underline{p} = (1-\phi)/(1+\phi)\,r$. It follows that average price (under no collusion) is given by

$$\overline{p} = \int_{\underline{p}}^{r} x\,\frac{(1-\phi)\,r}{2\,\phi\,x^2}\,dx = \frac{(1-\phi)\,r}{2\,\phi}\,\ln\!\left(\frac{1+\phi}{1-\phi}\right) \tag{8}$$

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 illustrates Proposition 1. Note that, for low values of ϕ, average price is decreasing in ϕ. (Price is distributed in the shaded region.) However, 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.

■ **Partial collusion.** So far we have assumed that collusion is a binary outcome which implies a cost c. However, one might expect the cost of engaging in collusion to be an increasing function of the collusion price level. For example, a higher price might lead to closer scrutiny by of antitrust authorities, which in turn increases the likelihood of an investigation and the expected cost from collusion.

Specifically, suppose that the cost from collusion is a function of the collusion price p, that is, $c = C(p)$. Then firms would choose to collude at a price level that maximizes $p/n - C(p)$. (Recall that we assume zero marginal cost and a measure one of consumers, each buying one unit.) Let p^* be the price that maximizes net profits from collusion. Notice that, since collusion profits are independent of the measure of searchers, the value of p^* is itself independent of ϕ.

The decision of whether to collude is now determined by the inequality

$$p^*/n - c > \widehat{\pi} = (1-\phi)\,r/n$$

which is equivalent to

$$\phi \geq \phi'' \equiv \frac{n\,c + r - p^*}{r}$$

If $p^* = r$, then we are back to the previous result. More generally, if $p^* < r$, then we obtain a higher value of the critical ϕ. This is intuitive: if the expected benefit from collusion is lower, then it takes a higher value of the measure of searchers to justify the switch to collusion.

In sum, despite the change in the precise threshold separating the competitive and the collusive outcome, as well as the price set in case of collusion, the nature of the result remains the same as before.

■ **Randomly distributed collusion cost.** 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 profit is 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 $\frac{1}{2}\,\phi\,r/\overline{c}$. By contrast, with probability $1 - \frac{1}{2}\,\phi\,r/\overline{c}$ average price is given by $r/(1+\phi)$ (as before). Overall, average price is now given by

$$\mathbb{E}(p) = \tfrac{1}{2}\,\phi\,\frac{r}{\overline{c}}\,r + \left(1 - \tfrac{1}{2}\,\phi\,\frac{r}{\overline{c}}\right)\,\frac{(1-\phi)\,r}{2\,\phi}\,\ln\!\left(\frac{1+\phi}{1-\phi}\right)$$

(We distinguish $\overline{p}$, equilibrium price under the Varian (1980) equilibrium, from $\mathbb{E}(p)$, the overall equilibrium price which includes the possibility of collusion.) The right-hand panel of Figure 1 plots the value of (overall) average price as a function of ϕ for the case when $r = 1$ and $\overline{c} = \frac{1}{2}$. 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 profit and collusion profit increases. As a result, the likelihood that collusion cost is lower than the gain from collusion increases. Moreover, the gap between static equilibrium price and collusion price is 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.

■ **Endogenous search.** The model considered so far assumes that consumers' search cost is either zero or infinity, so they either always search or never search. In a sense, we take buyer search as an exogenous phenomenon: to the extent that some buyers have zero search cost and others infinite search cost, the choice of whether or not to search is trivial.

A more general case assumes that consumer search cost is distributed according to the continuously differentiable cdf $G(s)$. Suppose also that, in the eyes of consumers, the value of c, the sellers' collusion cost, is distributed according to the continuously differentiable cdf $H(c)$.

Specifically, suppose that the distributions of c and s are common knowledge; that sellers and buyers simultaneously choose their collusion/pricing and search strategies; and, as before, that if no collusion takes place then pricing follows the Varian (1980) model.

In this extended model, an equilibrium is defined by a critical value of collusion cost, c^*, below which sellers collude; a critical value of search cost, s^*, below which buyers search; a price distribution conditional on collusion not taking place; and the conditions that sellers' collusion and pricing strategies, as well as buyers' search strategy, are optimal given the other agents' choices.

For simplicity, we continue to consider the case when $n = 2$. From (4), we conclude that collusion is a best response if and only if

$$c \leq \tfrac{1}{2}\,\phi\,r$$

Absent collusion, sellers play a Varian (1980) game and the price distribution is characterized by the cdf given in (6), where we note that ϕ, the measure of searchers, is now an endogenous variable. By contrast, if sellers collude then price is equal to r.

Under collusion, the gains from search are zero (all firms set the same price). Under Varian pricing, expected gain from search, v, equals the difference between expected price (what a buyer expects to pay from visiting a random store) and expected minimum price $q = \min_i p_i$ (what a buyer expects to pay from observing all prices):

$$v = \big(1 - H(c^*)\big)\,\big(\mathbb{E}(p) - \mathbb{E}(q)\big)$$

where q corresponds to the first order statistic of (p_1, p_2). If $n = 2$, $\mathbb{E}(p)$ is given by (8). As to q, its cdf is given by $K(q)$, which, for $n = 2$, is given by

$$K(q) = 1 - \big(1 - F(q)\big)^2 = 1 - \left(\frac{(1-\phi)\,(r-q)}{2\,\phi\,q}\right)^2$$

It follows that the corresponding density is given by

$$k(q) = \frac{(1-\phi)^2\,(r-q)\,r}{2\,\phi^2\,q^3}$$

Therefore

$$\mathbb{E}(q) = \int_{\underline{p}}^{r} p\,\frac{(1-\phi)^2\,(r-x)\,r}{2\,\phi^2\,x^3}\,dx = \frac{(1-\phi)\,r}{\phi} + \frac{(1-\phi)^2\,r}{2\,\phi^2}\,\ln\!\left(\frac{1-\phi}{1+\phi}\right) \tag{9}$$

Together, (8) and (9) imply that the value from search is given by

$$\mathbb{E}(p) - \mathbb{E}(q) = \frac{(1-\phi)\,r}{2\,\phi}\left(\frac{1}{\phi}\,\ln\!\left(\frac{1+\phi}{1-\phi}\right) - 2\right)$$

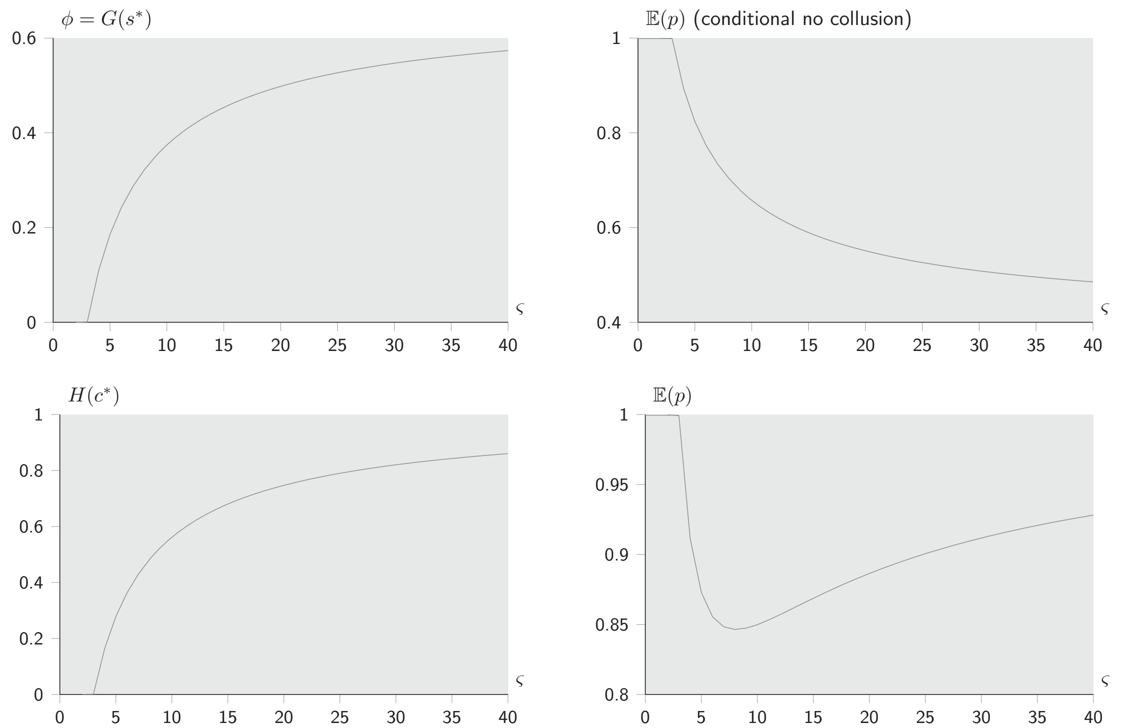
A first equilibrium condition is that, for the indifferent searcher, search cost is equal to expected gain from search: $s^* = v$, which implies

$$s^* = \big(1 - H(c^*)\big)\,\big(\mathbb{E}(p) - \mathbb{E}(q)\big) \tag{10}$$

A second equilibrium condition is that, for the indifferent colluding firms, the cost from collusion is equal to the expected gain from collusion. This implies

$$c^* = \tfrac{1}{2}\,G(s^*)\,r \tag{11}$$

Figure 2
Search propensity (ς) and equilibrium values



For simplicity, we continue to use ϕ to denote the measure of searchers, although we now explicitly take into account that ϕ is endogenously determined, specifically,

$$\phi = G(s^*) \tag{12}$$

Putting (10)–(12) together, we get the following equation in ϕ

$$G^{-1}(\phi) = \left(1 - H\left(\tfrac{1}{2}\,\phi\, r\right)\right)\;\frac{(1-\phi)\,r}{2\,\phi}\;\left(\frac{1}{\phi}\;\ln\left(\frac{1+\phi}{1-\phi}\right) - 2\right) \tag{13}$$

Given an equilibrium value of ϕ, s^* and c^* are uniquely determined by (10) and (11).

Equation (13) cannot be solved in closed form. Numerical computations show that it admits a unique interior solution. As an illustration, consider the case when c and s are uniformly distributed: $H(c) = \gamma\, c$ and $G(s) = \varsigma\, s$ (that is, $\gamma = 1/\overline{c}$ and $\varsigma = 1/\overline{s}$). Without loss of generality, assume $r = 1$.

Figure 2 displays the main equilibrium values as a function of ς, a measure of the propensity to search. The top left panel depicts the equilibrium measure of searchers. For a distribution of high search costs (low ς), we get a corner solution where no buyers search and sellers price at r, whereby there are no gains from search.[7] As the distribution of search costs becomes more favorable (higher ς, that is, lower search costs), a unique interior solution is obtained, with a cutoff value s^* that falls strictly between 0 (minimum search cost) and $1/\varsigma$ (maximum search cost).

7. This situation is reminiscent of the equilibrium of Diamond's (1971) model.

The top right panel shows average price *conditional on no collusion taking place.* The greater the value of ς, the greater the measure of searchers and the lower the price distribution (in the sense of first-order stochastic dominance).

A lower price distribution is also associated with lower equilibrium profits. As the bottom left panel shows, this corresponds to a probability of collusion, $H(c^*)$, which is increasing in ς (except for very low values of ς, when the measure of searchers is zero and so is the probability of collusion).

Finally, the bottom right panel shows the overall equilibrium price taking into account both endogenous search and endogenous collusion. For very low values of ς (high search costs) we get the corner solution with no search and price at the monopoly level. As ς increases and the measure of searchers increases, the first effect on price is that the Varian (1980) equilibrium price decreases, causing a decline in average price. However, as ς further increases, the effect of increased probability of collusion eventually dominates (notice how average price as a function of ς is a convex function, implying that the static competitive effect of an increase in ς becomes marginally lower). As a result, equilibrium price eventually becomes an increasing function of ς.

Figure 2 also suggests that a good proxy for ς is also a good proxy for ϕ. We will use this idea next, when we consider testable empirical implications.

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

Regarding price dispersion, our model has the same implication as static models in the Varian (1980) and Stahl (1989) tradition, namely, the relation between price dispersion and the propensity to search follows an inverse-U pattern: for low values of ς, an increase in ς leads to an increase in dispersion; and for high values of ς the opposite is true. An alternative second moment of the price distribution is price range. Unlike price dispersion, the prediction is that an increase in ς leads to an increase in price range for all values of ς. The idea is that price range under the Varian (1980) is increasing in ϕ; and as long as the probability of collusion is not equal to 1, the price range in the static equilibrium is the dominant effect.

Finally, to the extent that we can find direct evidence of tacit collusion, an additional testable implication is that the probability of collusion is increasing in ς.

3. Data

Our empirical tests consist 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 easily-accessible (but crucial) RHS-variable data, we focus our analysis on the city of Dortmund, a city with close to 600,000 inhabitants and 86 gas stations (as of May 2015). Finally, we measure prices at the pump net of taxes.

We complement station-level price data with additional station-level information. Specifically, we proxy for the fraction of searchers among each station's customers by measuring the share of young inhabitants (ages 18 to 29) in each of the local areas Dortmund is divided into.[8] Since this corresponds to an important part of our empirical strategy, later in this section we take a closer look at its validity.

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 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.[9]

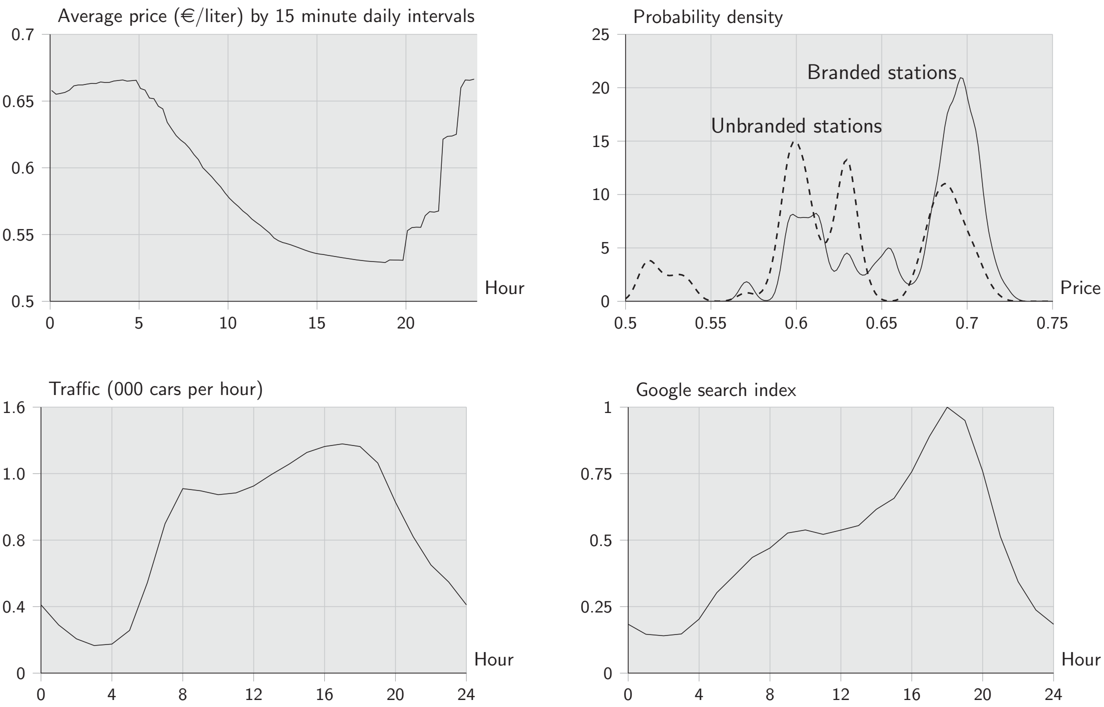
We consider additional local-market variables, for instance the number of cars, household income and population density. Except for the number of cars, these variables do not seem to have any significant correlation with prices and are thus excluded from our analysis.

Figure 3 shows various features of the data. From the top left panel we see that 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. The top right panel shows the kernel density of prices at 5pm, with branded and non-branded gas stations classified separately. Two striking features of this figure is the considerable variability in prices, which is consistent with the theoretical model; and the significant difference between prices by branded and non-branded stations. This latter difference justifies the different treatment of branded and non-branded stations, as well as the use of the fraction of non-branded stations as a measure of local market competitiveness. The bottom panels show the traffic and Google search patterns by time of day. As can be seen, the 5–6pm period is 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 period, that is, the five 15-minute slots beginning at 5pm. Further analysis suggests that the qualitative results are robust to changes in the

8. 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).

9. See Bresnahan and Reiss, 1991;Nevo, 2001; Schaumans and Verboven, 2015. Population and population density are more frequently used as instruments. Our preference for the number of competitors is justified by the similarity of building restrictions in neighboring areas, thus allowing us to capture correlations not present by population-based variables.

Figure 3
Data description



time frame considered.

■ **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.[10]

(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,

> 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.

10. See http://www.bundeskartellamt.de/, visited July 2017.

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 18–29 age bracket have used the app, as opposed to only 14 percent of the 60+ age bracket.[11]

Consistently with this evidence, we propose as a proxy for the propensity to search a variable reflecting the fraction of population in the 18–29 age bracket ("share of young", or simply SOY) among each station's customer base.

Specifically, we construct our station-level SOY variable following a gravity-equation approach.[12] We determine the distance from each of the 86 stations to the GPS center of each of the 62 administrative areas of Dortmund (for which we have SOY data).[13] We then compute each station's SOY value as a weighted average of all 62 values, with weights calculated according to the population in the administrative area divided by the square of the Euclidian distance from the respective station to each of the area centers.[14]

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.

Moreover, we also considered alternative definitions of station-level SOY that take into account commuting patterns. Specifically, we use data on commuting patterns provided by the North Rhine-Westphalia Statistical Office to create SOY weights that take into account inner city commuters driving from other city districts as well as from outside of the city.[15] We perform a number of robustness checks by changing the relative weight given to commuters, as shown at the end of this section.

Notwithstanding the multiple robustness checks we preform, our empirical analysis hinges on our measure of the propensity to search, that is, the use of the SOY variable. One way to test the validity of our proxy is to test whether the use of the app (or apps) that provide information about prices is related to SOY. We obtained data on the number of search queries directed to one of the app providers in Germany, *Tank-Navigator* (by Mammuth Applications).[16] 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 ($N = 512$).[17] The dependent variable is total number of app queries per municipality per capita by municipal-

11. Our translation from the German.
12. See Allain et al. (2017) for a similar construction in a study of the French supermarket sector.
13. See *Dortmunder Stadtatlas 2015*, https://www.dortmund.de/ media/p/statistik_3/ statistik/veroeffentlichungen/jahresberichte/Statistikatlas_2015.pdf. The 586,181 inhabitants of Dortmund then break down to an average of 9,455 inhabitants per area.
14. Estimating regression models with unweighted SOY does change the coefficients quantitatively.
15. See https://www.it.nrw.de/statistik/b/daten/eckdaten/r543Pendler.html, https://www.lwl.org/LWL/Kultur/Westfalen_Regional/Verkehr/Berufspendler_2011, last visited October 2017. See also Spiekermann and Wegener (1995).
16. 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.
17. Age distribution is obtained from http://www.inkar.de/.

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%.

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	0.01	0.11	0.16
Share of young (age 18-65)	180,855	0.64	0.02	0.61	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
Cars per 10^9 people	180,855	0.50	0.06	0.36	0.65
Stations / population	180,855	0.014	0.0059	0.0051	0.0311

ity; 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.

■ **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 11% and 16%, with a standard deviation of 1%. Although the support of the SOY variable is relatively small with an interval between minimum and maximum values measuring only about 6%, this is in line with other research investigating the impact of search on market outcomes. We therefore believe this to be sufficient to influence retailers' behavior and market outcome.[18]

18. See Brown (2018) for a recent study of search on markups in the health care sector.

We return to this issue at the end of Section 6.

4. Results

Proposition 1 implies a U-shaped relation between a variety of price measures (price level, average price, minimum price) and the propensity to consumer search. As a proxy for the latter, we estimate each station's fraction of young consumers (SOY) among its customer base. Therefore, a natural way of testing Proposition 1 is to regress measures of prices on a quadratic polynomial of SOY; that is, to use SOY and SOY^2 as explanatory variables. Proposition 1 predicts a negative coefficient on the linear term and a positive 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

$$p_{it} = \beta_0 + \beta_1\, SOY_i + \beta_2\, SOY_i^2 + \beta_3\, \widehat{NOC}_i + \beta_4\, SBR_i + \beta_5\, BRA_i + \beta_6\, CAR_i + \sum_{\tau=1}^{T} \beta_\tau\, I_{\tau,t} + \epsilon_{it}$$

where i indexes the station, t the time period, I_t denotes day fixed effects (controlling for, e.g., aggregate input price or demand shocks), ϵ is an error term, β_k $(k = 0, ..., T+6)$ the regression parameters, and the independent variables are defined as follows:[19]

- p: price measure
- SOY: estimate of share of 18–29 year old population shopping at station i
- NOC: number of gas stations in station i's neighborhood (1.25-mile-radius circle)
- SBR: share of branded gas stations in station i's neighborhood
- BRA: dummy variable equal to 1 iff station i is branded (Aral, Esso, Shell or Total)
- CAR: number of cars per inhabitant in station i's neighborhood

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, we instrument for the number of local market participants by measuring the number of stations in the greater environment around station i. Specifically, we regress p on $\widehat{NOC}$, the predicted value from the regression

$$NOC_i = \alpha_0 + \alpha_1\, SGE_i + \alpha_2\, X_i + \nu_i$$

where SGE_i is the number of competitors in the wider area around station i, that is, the "donut" centered on station i (a 12-mile-radius circle that excludes the 3-mile-radius

19. We cluster standard errors at the station level or use the White correction for standard errors. The results only differ marginally; we report the results obtained by the White correction.

Table 3
Parametric regression of price equations

Dependent variable:	Price	Mean Price	Min Price
Share of young consumers	-5.9284** (2.9651)	-6.8834*** (2.0216)	-8.0028*** (2.2951)
Share of young consumers squared	23.0416** (11.125)	26.5829*** (7.5637)	29.9891*** (8.4662)
Number of Competitors	-0.0016*** (0.0006)	-0.0016*** (0.0004)	-0.0016*** (0.0005)
Share of branded stations	0.0044 (0.0069)	0.0216*** (0.0053)	0.0363*** (0.0077)
Branded station	0.0157*** (0.0025)	0.0000 (0.0016)	-0.0031 (0.0024)
Cars per 10^9 people	0.0601 (0.0378)	0.0535** (0.022)	0.0151 (0.0177)
Constant	0.8900*** (0.1883)	0.9570*** (0.1308)	1.0493*** (0.1539)
Number of observations	12,398	12,398	12,398
R^2	0.34	0.44	0.41

inner circle, both centered on station i's location); and X_i includes the usual second-stage variables used for the IV regression. [20]

Regarding the regressions's dependent variable, we consider three different price measures:

- Price level. 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 interest are the coefficients associated with the squared variables.[21] As the 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).

20. See Bresnahan and Reiss (1991), Nevo (2001), Schaumans and Verboven (2015) for similar approaches to instrumenting for competition variables. Population and population density are more frequently used as instruments. Our preference for the number of competitors is justified by the similarity of building restrictions in neighboring areas, thus allowing us to capture correlations not present by population-based variables. That said, we add that using population variables does not change our results significantly.
21. It is clear that the coefficient of the linear variable has to be statistically significantly negative not to describe a progressively increasing or degressively decreasing function instead of a U-shape.

In order to get an idea of the economic significance of these estimates, Figure 4 plots estimated price (in cents of Euro per liter) as a function of each station's share of 18-29 year olds. Although SOY only varies from 9 to 16% (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 4 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 4.

As SOY continues increasing, 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.

As mentioned in Section 1, a recent paper by Luco (2018) looks at the effects making price information available in the Chilean gasoline market. He concludes that price disclosure leads to an increase in prices, and that the price increase is greater in areas where search costs are lowest. Our research question is different from Luco's (2018). However, we can use our results to perform the conceptual experiment of comparing the observed price levels to the price levels that would persist if price information were not available (and thus app-based search were impossible). Our model would predict that all prices would be at the reservation level (as there would be no search). Given that, the effect of introducing search is to decrease prices in areas with moderate levels of search but not in areas with high levels of search. This differs from Luco (2018) in two different ways. First, he observes an increase in prices, whereas we would predict a decrease in prices. This is because we assume that, other than the cost of engaging in collusion, sellers have no informational difficulty in engaging in collusion. Second, Luco (2018) observe no non-monotonicity in the effect of search costs. In terms of our theoretical model, his results are consistent with the left branch of our U-shaped relation between search and prices.

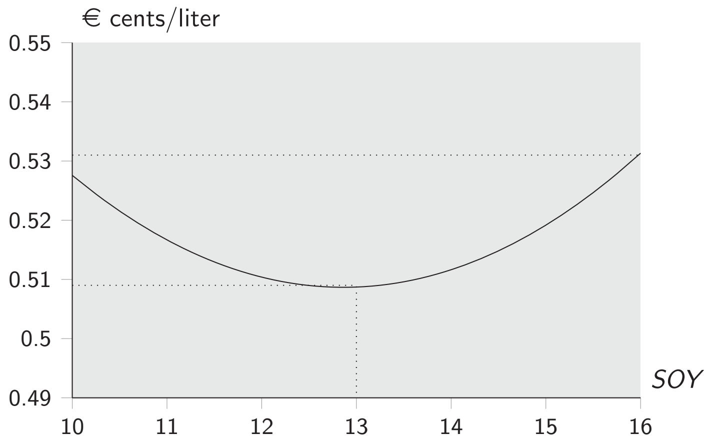
5. Robustness checks

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: 2 miles instead of 1.25 miles for the inner circle and 15 miles instead of 12 miles 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.

Figure 4
Estimated price regression (Table 3)



First, we also include each station's 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 18 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 $23.0416 \times .14^2 - 5.9284 \times .14 + 0.89 = .512$ in the base model and $44.6679 \times .64^2 - 56.4875 \times .64 + 18.3492 = .493$ 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 into an odd number of bins according to the SOY variable and estimated dummy variable coefficients 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.

■ **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

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 18–65
(5) Combination of (3) and (4)

Dependent variable: price:	(1)	(2)	(3)	(4)	(5)
Share of young consumers	-5.9284** (2.9651)	-5.0858* (2.9678)	-6.8946** (3.38)	-56.4875* (32.5124)	-18.6928* (10.0372)
Share of young consumers squared	23.0416** (11.125)	19.8391* (11.0509)	26.5247** (12.6618)	44.6679* (25.5706)	14.9195* (7.8876)
Number of Competitors	-0.0016*** (0.0006)	-0.0024* (0.0013)	-0.0017*** (0.0006)	-0.0006 (0.0008)	-0.0014** (0.0007)
Share of branded stations	0.0044 (0.0069)	0.0041 (0.007)	0.0045 (0.0069)	0.0054 (0.0067)	0.0034 (0.0065)
Branded station	0.0157*** (0.0025)	0.0139*** (0.0025)	0.0159*** (0.0025)	0.0167*** (0.0024)	0.0150*** (0.0025)
Cars per 10^9 people	0.0601 (0.0378)	0.0756* (0.0427)	0.0545 (0.0371)	0.0661* (0.034)	0.0573* (0.0317)
Constant	0.8900*** (0.1883)	0.8237*** (0.1885)	0.9596*** (0.2171)	18.3492* (10.3157)	6.3645** (3.1844)
Number of observations	12,398	12,398	12,398	12,398	12,398
R^2	0.34	0.31	0.34	0.37	0.36

Table 5
Regression results for terciles

Dependent variable:	Price	Mean Price	Min Price
Tercile t1	0.0091* (0.0052)	0.0088** (0.0035)	0.0161*** (0.0047)
Tercile t3	0.0078** (0.0035)	0.0078*** (0.0026)	0.0056** (0.0023)
Number of Competitors	-0.0022** (0.0009)	-0.0020*** (0.0006)	-0.0016** (0.0007)
Share of branded stations	0.0053 (0.0073)	0.0225*** (0.0059)	0.0391*** (0.0086)
Branded station	0.0161*** (0.003)	0 (0.0022)	-0.0034 (0.0027)
Cars per 10^9 people	0.0134 (0.0358)	0.0112 (0.0223)	-0.0203 (0.0195)
Constant	0.5411*** (0.0203)	0.5408*** (0.0154)	0.5351*** (0.0164)
Number of observations	12,398	12,398	12,398
R^2	0.29	0.37	0.41

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.

■ **Other robustness checks.** Our base regressions include as a dependent variable prices a 15 minute intervals during rush hour. One possible additional robustness check is to restrict our regressions to one single minute observation during rush hour. The results from these regressions are essentially identical to those when we pool observations, both in terms of coefficient size and in terms of statistical significance.

We also censored the support of the SOY variable as a final sanity check. We cut off 10% of the SOY variable support at both the lower and upper tails of the distribution. The U-shaped relationship between price measures and SOY remains stable.

6. Additional empirical evidence

In this section, we provide two additional sources of evidence, which, together with the price regressions presented in Section 4, strengthen our case for a U-shaped relationship between the propensity to search and price levels, as well as the theoretical basis for this relationship.

Our argument for a U-shaped relationship is based on the composition of two effects: (a) increased competition when there is no collusion; and (b) increased likelihood of collusive behavior. One way to provide additional evidence for our theory is therefore to test directly for each of the above components. First, we test whether the range of gas station prices increases when the propensity to search increases. The idea is that the price range is determined by the static equilibrium, where the upper bound is given by monopoly price and the lower bound is decreasing in the measure of searchers. Second, we directly test whether the propensity to collude is increasing in the propensity to search.

Both of these tests are consistent with our theoretical model, as we show below. Our last step is to show that our theoretical model reasonably explains the magnitude of the effects we observed. Specifically, we argue that the variation in propensity to search, once fed into our theoretical model, implies a reasonable variation in expected price as well as the estimated measure of searchers.

■ **Search and price range.** Since the price range under no-collusion is wider than the price range under collusion, as long as the event of no collusion takes place with positive probability we expect the price range to be increasing in the measure of searchers. In other words, as far as the price range is concerned, the effect of an increase in the measure of searchers corresponds to the effect under no collusion. And, as it is well known from the Varian (1980) model, such range increases in the measure of searchers.

We estimated an OLS regression of price range on SOY as well as on other independent variables considered before. Consistent with theory, we obtain a positive and statistically significant SOY coefficient: 0.4466. Considering the range of SOY, from 0.108 to 0.163 (a range of .055), this corresponds to an increase in the price range of about 2.4 cents, which is admittedly on the low end. In sum, our results regarding price range are consistent with theory.

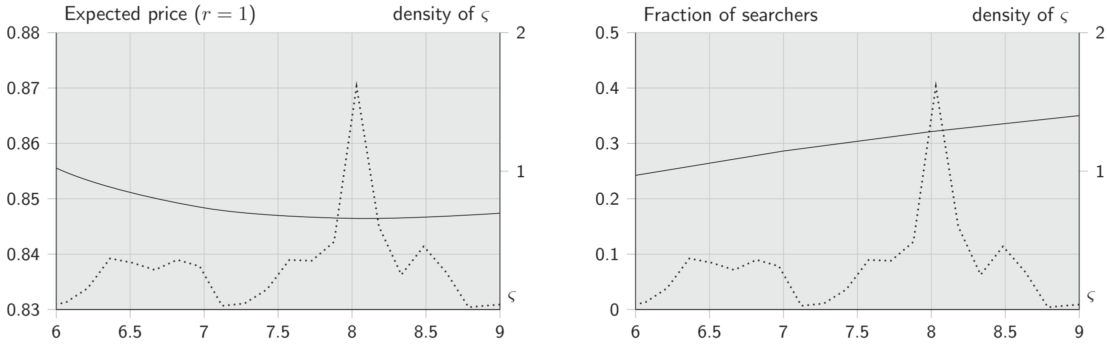
■ **Search and propensity to collude.** Subsequently to our sample period, an important event took place in the German retail fuel market: the Shell network of gas stations introduced a price-matching guarantee (PMG) scheme to its cardholders.

It has been argued (both theoretically and empirically) that PMGs can be used as mechanisms to facilitate collusion. In fact, starting in June 2015 we observe a series of attempts by gas stations to increase prices. Although there is no direct evidence of collusion — that is, no "smoking gun" — there are various indications that firms attempted to tacitly collude following Shell's PMG offer. In particular, whereas in our sample prices were uniformly declining during the day, we now observe a pattern of midday price jumps. Moreover, a large fraction of these price jumps (about 90%) amount to exactly 3 cents; and a large fraction takes place at exactly 12:01 or 12:02 (given the way the data is collected, it may be difficult to distinguish these times from noon). In other words, the behavior of gas stations following Shell's PMG exhibits many of the features typical of focal-point tacit-collusion equilibria.

In a different paper (Cabral et al., 2018) we examine in greater detail the causes of these midday price jump patterns. For the present purpose, we take advantage of the considerable spatial variation in their prevalence to examine their correlation with search intensity.

Our theoretical model predicts that the greater the propensity to search, the greater the propensity to collude. The idea is that more searchers imply fiercer market competition, which in turn increases the appeal of tacit collusion. In order to test this prediction, we define a dependent dummy variable (at the gas-station-level) that equals 1 if the seller increases price at noon. When we run a probit regression on SOY and the other independent

Figure 5
Search propensity (ς) and equilibrium values



variables used in our base regressions, we obtain a positive and statistically significant SOY coefficient estimate: 20.8768 (standard deviation 9.1595). In terms of economic significance, as we consider the range of variation in SOY (from 0.108 to 0.163), this translates into a 0.72 cent price increase. In sum, we find evidence consistent with the positive effect of the measure of searchers on the likelihood of collusion.

■ **Back to the theoretical model.** Our empirical section tests a qualitative implication of the theoretical model, namely that the relation between a proxy for the propensity to search (namely SOY as a proxy for ς) is nonmonotonic: for low values of ς, expected price is decreasing, whereas for high values of ς, expected price is increasing in ς.

From Table 2, we see that our proxy for the propensity of search, SOY, ranges from 0.108 to 0.163, with a relatively low standard deviation of 0.0132. Since our analysis relies heavily on the spatial variation of this variable, it is reasonable to ask whether the distribution of SOY, once fed into the theoretical model, is consistent (in terms of orders of magnitude) with our empirical results.

Suppose that the propensity to search by a given gas stations' customer base (ς in our theoretical model) is proportional to the measure of young inhabitants, SOY. The best case for a U-shaped relation between SOY and average price results from a mapping from SOY to the coefficient on propensity to search which places the distribution of ς (see the lower right panel of Figure 2) around the minimum of the U-shaped relation between ς and $\mathbb{E}(p)$.

Figure 5 reports the results of this exercise. Suppose that $\varsigma = SOY/1.8$. Then the distribution of ς is approximated by the kernel density distribution found in both panels of Figure 5. For the relevant range of ς, we obtain a range of predicted expected prices given by the U-shaped curve on the left-hand panel. This is taken from Figure 2, which in turn refers to the equilibrium of the endogenous-search model with $n = 2$ and r normalized to 1.

The implied range of price variation is about one percent of r, which in turn corresponds to two thirds of 1 cent of an euro. This value is lower than the value implied by our reduced-form regressions, but of the same order of magnitude.

An additional sanity check is to evaluate the measure of searchers implied by the proposed values of ς. This is shown on the right-hand panel of Figure 5. As can be seen, the measure of searchers varies from about 25 to about 35 percent.

Is this a reasonable range for the fraction of searchers in equilibrium? As mentioned earlier, the German Cartel Office's website provides a list of price-search apps as well as an estimate of the number of downloads per app.[22] 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.[23]

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 may 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.[24] 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 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.

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 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

22. See http://www.bundeskartellamt.de/EN/Economicsectors/MineralOil/MTU-Fuels/mtufuels_node.html.
23. See www.kba.de/DE/Statistik/Fahrzeuge/Bestand/bestand_node.html#rechts.
24. 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.

participation constraint rather than the incentive (no-deviation) constraint. An application of this approach to the problem of search suggests 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 propensity to search 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 is associated with a price variation of about 2.5 cents per liter (about 5% of sale price). Adding to this, we find evidence for an increasing probability of collusion when search intensity increases, which furthermore leads to increasing price dispersion.

References

ALLAIN, M.L., CHAMBOLLE, C., TUROLLA, S., AND VILLAS-BOAS, S. "Retail Mergers and Food Prices: Evidence from France." *Journal of Industrial Economics*, Vol. 65 (2017), pp. 469–509.

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

BRESNAHAN, T.F. AND REISS, P.C. "Entry and Competition in Concentrated Markets." *Journal of Political Economy*, Vol. 99 (1991), pp. 977–1009.

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.

BROWN, Z. "An Empirical Model of Price Transparency and Markups in Health Care." *National Bureau of Economic Research and University of MichiganWorking Paper*, (2018).

CABRAL, L., DÜRR, N., SCHOBER, D., AND WOLL, O. "Price Matching Guarantees and Collusion: Theory and Evidence from Germany." Working Paper, New York University (2018).

CHANDRA, A. AND TAPPATA, M. "Consumer Search and Dynamic Price Dispersion: an Application to Gasoline Market." *The Rand Journal of Economics*, Vol. 42 (2011), pp. 681–704.

DIAMOND, P.A. "A Model of Price Adjustment." *Journal of Economic Theory*, Vol. 3 (1971), pp. 156–168.

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.

LUCO, F. "Who Benefits from Information Disclosure? The Case of Retail Gasoline." Technical report, forthcoming in *American Economic Journal: Microeconomics* (2018).

Nevo, A. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, Vol. 69 (2001), pp. 307–342.

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).

Petrikaitė, V. "Collusion with Costly Consumer Search." *International Journal of Industrial Organization*, Vol. 44 (2016), 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.

Spiekermann, K. and Wegener, M. "Car Pooling and Energy Consumption." *Proceedings, 7th World Conference on Transport Research, Sydney*, (1995).

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.

// IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de