Markets for Data

Dirk Bergemann[*] Alessandro Bonatti[†]

First Version: January 27, 2012
Current Version: February 13, 2012

Very Preliminary and Incomplete Version.
Do not Circulate.

Abstract

We develop a model of data provision and data pricing in an environment with strategically interacting firms. The demand for data is generated by firms which seek to tailor their product positioning, or price, to either the individual or the aggregate demand. In turn, the data provider determines the amount of information released to the individual firms and the price to access it. We derive the optimal information and pricing policy of the data provider, under either individual or aggregate tailoring by the firms.

We show that frequently the optimal information policy is to provide only partial and noisy information to the competing firms. In addition, the revenue of the data provider is commonly maximized by asymmetric or even exclusive information policies.

Keywords: Data Markets, Data Providers, Data Brokers, Online Data, Information, Media Markets, Advertising.

JEL Classification: D44, D82, D83.

[*]Department of Economics, Yale University, New Haven, CT 06520, dirk.bergemann@yale.edu
[†]MIT Sloan School of Management, Cambridge, MA 02142, bonatti@mit.edu

1 Introduction

As the cost of collecting, aggregating and storing data has dramatically decreased over the past decade, new markets for data have emerged in many forms and varieties. Individual firms amass both internal data and external data. The former is proprietary data, which is generated by the transactions of the firm. The latter is made available by data providers, and enables firms to improve their competitive position, vis-à-vis potential customers and other firms. Data vendors offering access to databases and data analytics provide information with respect to virtually any economic transaction. Financial data providers, such as Bloomberg and Thomson Reuters, provide real-time and historical financial data. Credit rating agencies, whether for individuals, such as Equifax and Transunion, or for business and governmental agencies, such as Moody's and Standard & Poor's, provide detailed financial information. Data brokers, such as LexisNexis and Acxiom, maintain and grow large databases on individual consumers. Online aggregators such as Spokeo and Intelius mine publicly available data to compile personal data profiles.

Whatever the source of the data, these vendors can sell access to their database and to their data analytics to downstream firms. Firms may use the information they acquire in order to improve their product positioning, in terms of the specifics of the product offered and the conditions at which the product is offered.

This is especially relevant in the case of display advertising markets. Consider platforms such as BlueKai, that promote online marketplaces for the exchange of user data. The goal of these platforms is to monetize the match-making potential by linking websites that monitor traffic to advertisers who are interested in information about individual users. This is possible in part because of the current practice, common among several advertising networks and exchanges, to allow buyers (advertisers) to use their own data when selecting (i.e. cherry-picking) which impressions to acquire. This in turn enables advertisers to better target their messages on other websites as well, by tracking or following their desired users across different pages. While this has led to severe concerns over users' privacy on the web, it also creates the potential for a data provider (the platform) to appropriate a large fraction of the surplus resulting from an improvement in the targeting technology.

In the present paper, we shall investigate the role of data providers for competition among firms and for consumers' welfare. We shall distinguish between individual (or consumer) tailoring and aggregate (or market) tailoring. In the case of individual tailoring, the individual firm can improve its product positioning vis-à-vis the individual consumer. In the case of market tailoring (market research, location and supply decision), the individual firm can improve its positioning vis-à-vis the aggregate market condition. We shall pursue the role of the individual and the aggregate tailoring in two distinct frameworks.

In each framework, we compare the firms' equilibrium profits in the strategic tailoring game under different information structures. This allows us to investigate the equilibrium value of acquiring private information. We interpret the private information about consumers or markets as being induced by the equilibrium purchases of data from a fourth party (i.e., neither a competing firm nor the consumer). Specifically, firms compete in tailoring to consumers (or markets) under an information structure that has been determined "upstream," in the market for data.

We then assess the profitability of selling data that induces different given information structures and the properties of strategic information acquisition to the equilibrium price of data. We summarize our framework for the market for data in the following diagram.

Data Provider

/ \

Firm i Firm j

\ /

Consumer/Market

Figure 1: Upstream and Downstream Interaction

More specifically, in the case of individual tailoring (e.g., internet advertising, mail advertising, private offers), we consider the a model of horizontal differentiation with competition between two firms. The identity of the consumer is the location of the consumer on the unit interval. In the absence of additional information, the two firms compete against each other with a simple pricing policy, and the privately informed consumers sort themselves across competitors along their preferences, i.e. their location. We then consider the case when the individual firms may acquire information about the individual consumer, which will allow to offer a contingent pricing policy, and hence a tailored pricing policy, to attract the consumer in question. Finally, ask whether firms have an incentive to acquire information, and what price the data provider can charge to grant each firm access to the database.

For the case of the aggregate tailoring, we consider a standard Cournot oligopoly with uncertain aggregate demand. Here, individual firms share a common prior about the realized level of demand. By acquiring additional information from the data provider, each firm can improve its forecast of the realized demand and hence can more closely tailor the quantity it supplies to the prevailing market condition.

The two models, the Hotelling model of horizontal differentiation, and the Cournot oligopoly model emphasize different aspects of tailoring. In the Hotelling model, the values of the firms (i.e. the location of the consumer) are negatively correlated across firms, and firms tailor the positioning of their product in terms of the price at which the product is offered. In the Cournot model, the values of the firms, the state of demand, are positively correlated, and the firms tailor the positioning of their product in terms of the quantity in which the product is offered. Nonetheless, the value of information across these different models of competition, and the optimal data provision policy display many interesting common features.

2 Literature

The literature on the optimal choice of information structures is recent. Bergemann and Pesendorfer (2007) consider the design of optimal information structures within the context of an optimal auction. Thus, the principal simultaneously controls the design of the information and the design of the allocation rule. More recently, Kamenica and Gentzkow (2011) consider the design of the information structure by the principal when the agent will take an independent action on the basis of the received information. In an extension, Gentzkow and Kamenica (2011) consider the design of the information structure in the context of a competitive environment. In either case, the principal can influence the eventual decision (of the agents) in his favor by an appropriate choice of the information structure, but they do not discuss the price of information.

In our earlier work, Bergemann and Bonatti (2011), we analyzed the role of (exogenous) information structures on the competition for advertising space. In this strategic environment, more precise information about consumers' locations allows for better targeting of advertisement messages, and improves firms' revenues. It may, however, be detrimental to the seller of the advertising space. Thus, we are motivated by this result to investigate further the role of information on industry profits, and the scope for profitable information intermediaries. We depart from the model in Bergemann and Bonatti (2011) along two major directions: first, we endogenize the information structure and investigate the price of the information itself; second, we consider firms operating on interdependent product markets. In particular, we relate the mode of product-market competition to the profitability of different information provision policies. The role of information in a specific model of competition is also analyzed in Ganuza (2004), while associated notions of ranking of information structures are developed in Ganuza and Penalva (2010).

The basic model of noisy information in an oligopoly environment has been extensively studied in the context of information sharing by oligopolists, see Raith (1996) for a general model and a summary of the results in the literature. Yet, to the best of our knowledge, the question as to what happens if information can be made available (at a price) by a third party, has not been studied in the literature at all. To the extent that the database provides the competitor with a certain information structure, we might say that the database, by choice of the information structure, induces a specific game of incomplete information among the competitors (while being consistent with the prior information of the agents). The associated class of games is analyzed in Bergemann and Morris (2011b) and Bergemann and Morris (2011a) as a problem of robust prediction and an associated equilibrium concept, referred to as Bayes correlated equilibrium.

The implications of specific information structures in auctions are analyzed in recent work by Abraham, Athey, Babaioff, and Grubb (2011) and Celis, Lewis, Mobius, and Nazerzadeh (2011). Both papers are motivated by asymmetries in bidders' ability to access additional information about the object for sale. Consequently, they examine the role of the irregular distributions of values resulting from the private acquisition of data by a single bidder. In particular, Abraham, Athey, Babaioff, and Grubb (2011) focus on a common value environment with binary valuation, and a given mechanism, while Celis, Lewis, Mobius, and Nazerzadeh (2011) propose an (approximately) optimal mechanism in a private values model.

3 Consumer Tailoring

We consider the classic Hotelling problem of horizontal differentiation. There are two firms located at $\{-1,+1\}$ on the real line. The location $\theta \in [-1,1]$ of the buyer determines his preferences θ, namely

$$u(\theta, p_-, p_+) = \begin{cases} v+\theta-p_+, & \text{if} \quad \theta \text{ buys from } +1; \\ v-\theta-p_-, & \text{if} \quad \theta \text{ buys from } -1. \end{cases}$$

The market demand is given by a distribution $F(\theta)$ of buyers' locations θ over the interval $[-1,1]$. Each firm has only limited information about each individual consumer, and an important, but special case is the situation when each firm has zero additional information about the consumer, hence all consumers look identical to the firms, and the only information that each firm has is the common prior, namely the distribution $F(\theta)$.

3.1 Information

We want to think about a third party (the data provider) providing additional information (the database) about each consumer, and charging firms a fixed price. We may initially assume that the database is sufficiently rich and well informed that it can identify each consumer and his willingness to pay. The optimal data provision policy is to determine the nature of the information provided to the firms and the price to be charged. In the environment of all possible information policies, three information policies seem relevant and interesting benchmarks.

1. The database could provide symmetric and complete information about the customers to each firm. In consequence, the firms would compete for each customer individually with a personalized (tailored) price. We would then evaluate the expected firm profit under complete information. The resulting allocation would be efficient, and the competition would seemingly be fierce, but for all consumer beyond the marginal consumer, the winning firm would be able to extract the entire additional willingness-to-pay.

2. The database could provide the complete information only to a single firm, and not inform the other firm at all. The informed firm would be at an informational advantage over the uninformed firm, the allocation would presumably not be efficient, but the informed firm might be willing to pay a substantial premium for the access to the database.

3. The database could provide a symmetric, but noisy information structure to the competing firms. In particular, it could inform each firm exactly about the consumers close to its location, but provide only noisy information, namely the prior, on the consumers located far away from it. This bipartite information structure would allow the winning firm to price exactly up to the willingness to pay, and hence extract all the surplus, when the uninformed firm is only a weak competitor as it acts under very coarse information.

3.2 Special Information Structures

By comparing the profit realized under each information structure, we receive information about the surplus that the monopolist database could extract from the seller through its information policy.

Before analyzing the equilibrium on the market for data, we examine a few special information structures in detail, starting with the benchmark case of no information.

3.2.1 Prior Information

In the benchmark case of no information, we have a standard Hotelling duopoly game. Given the two prices, p_+ and p_-, the location of the indifferent consumer is denoted by

$$\hat{\theta} = \frac{p_+ - p_-}{2}.$$

If the distribution $F(\theta)$ is symmetric about the median $\theta = 0$, then at a symmetric equilibrium with full market coverage we have

$$p^* = 1/f(0).$$

Clearly, this is the case only if v is high enough. In particular, as long as

$$\frac{1}{2} - f(0)\, v \leq 0,$$

the market is covered and we have

$$p^* = \min\{v, 1/f(0)\}.$$

Thus profits are given by

$$\Pi^* = \frac{1}{2} p^*.$$

When v is too low, firms operate in local monopolies, and the optimal price is given by (for firm +1)

$$\begin{aligned} p^* &= \arg\max_p \{p(1 - F(p - v))\}. \\ \Pi^* &= p^*(1 - F(p^* - v)) \in [v/2, p^*/2]. \end{aligned}$$

In particular, under the Uniform distribution, $p > v$ for all $v > 2$ and conversely, for $v < 1$ not all consumers participate, and firms operate in a local monopolies regime.

3.2.2 Complete Information

In the full-information benchmark, both firms can tailor their prices to the consumer. We adopt the natural tie-breaking rule for Bertrand competition that assigns the consumer to the firm with the higher willingness to pay. Under full information, the allocation of consumers to firms is then efficient. Firm $+1$ sells to all consumers $\theta \geq 0$ at a price of $p_+(\theta) = 2\theta$. Similarly, firm -1 sells to all consumers $\theta < 0$ at a price of -2θ. The resulting profits are given by

$$\Pi_+ \;=\; 2\int_0^1 \theta \mathrm{d}F(\theta) \tag{1}$$

$$\Pi_- \;=\; 2\int_0^{-1} \theta \mathrm{d}F(\theta). \tag{2}$$

Because the incomplete-information price was determined by a local condition, while profits under complete information depend on the whole distribution, it is not hard to find examples in which the ranking of these two structures is strict in either direction. We focus on the case of the uniform distribution when we turn to the market for the database. Under the Uniform distribution, one can show that providing no information dominates full information (in terms of joint profits) for all values of v for which the market is covered. Conversely, for all $v \in [0,1]$, information about consumers' location allows firms to reach the entire market, and the resulting profits under complete information exceed those under incomplete information.

3.2.3 Asymmetric Information

We now examine the role of asymmetric information structures. We start from the simplest such structure. Suppose firm $+1$ is perfectly informed about the consumer's location, while firm -1 only knows the prior. To best understand the strategic effects that arise under this structure, assume (for now) that v is large, so that the market is covered, and look for a pure-strategy equilibrium. We now show that such an equilibrium does not exist. Towards a contradiction, let firm -1 choose price $p > 0$ in equilibrium. The best response of firm $+1$ is then to charge each consumer θ a price of

$$p_+^*(p, \theta) = \max\{p + 2\theta, 0\}. \tag{3}$$

Notice that the best response of firm -1 would then be to undercut firm $+1$ and win the fraction of the market for which $p_+^* > 0$. In addition, charging $p_- = 0$ is not a best response to the informed firm's strategy in 3 which leaves a local monopoly region to the uninformed firm. If we restrict attention to weakly undominated strategies (in particular, non-negative prices), there are no pure-strategy equilibria.[1]

We now specialize the model and consider a uniform distribution of types $\theta \in [-1, 1]$. We characterize a mixed strategy equilibrium that does not involve weakly dominated strategies for either firm. Clearly, in such an equilibrium, the informed firm cannot randomize over an open set of θ, because the uninformed firm cannot adjust its strategy in response to the unobserved type, and

[1] There are, indeed, pure strategy equilibria in which the informed firm sets a price of $p_+(\theta) = 2\theta$ and the uninformed firm sets $p_- = 0$. These equilibria are analogous to zero-profit equilibria in the complete-information benchmark, in $p_-(\theta) = \min\{-2\theta, 0\}$ and $p_+(\theta) = \min\{0, 2\theta\}$. We rule them out because they rely very heavily on the tie-breaking rule, and they lead to less interesting profit comparisons.

the informed firm cannot be indifferent for two values of the gap (2θ) in the consumer's willingness to pay. We therefore look for an equilibrium in which only the uninformed firm randomizes.

Given the realization p_- of the uninformed firm's action, the marginal consumer is denoted by

$$\hat{\theta}\,(p_-) : p_+(\hat{\theta}) = p_- + 2\hat{\theta},$$

where $p_+\,(\theta)$ is the informed firm's (pure) strategy. We conjecture and verify that $p'_+\,(\theta) < 2$, implying that if θ buys from $+1$ so do all $\theta' > \theta$.

In order for firm -1 to be indifferent, it must then be the case that

$$\mathbb{E}\Pi_-\,(p_-) = p_- F(\hat{\theta}\,(p_-)) = p_- \frac{1 + \hat{\theta}\,(p_-)}{2}$$

is independent of p_-, i.e.,

$$\mathbb{E}\Pi_-(\hat{\theta}) = (p_+(\hat{\theta}) - 2\hat{\theta})\frac{1 + \hat{\theta}}{2} = k \ \forall \hat{\theta}.$$

therefore, we need

$$p_+(\hat{\theta}) = \max\left\{2\hat{\theta} + \frac{2k}{1 + \hat{\theta}}, 0\right\}, \tag{4}$$

with $k > 0$. In particular, if we let $k = -t\,(1 + t)$ we have $p_1\,(t) = 0$. That is, type $\theta = t < 0$ is the farthest type receiving a positive price offer by firm $+1$. In order for firm -1 to be willing to serve all types $\theta \leq t$ we need to insure that it does not profit by raising prices above $-2t$. In other words, we need

$$\Pi_-\,(\theta) = -2\theta\frac{1 + \theta}{2} \text{ increasing at } \theta = t,$$

which corresponds to $t \leq -1/2$. Under this condition, firm -1 serves all types that receive a zero-price offer from firm $+1$. However, firm -1 can always guarantee a profit level of $1/4$ by best-responding to a price of zero by firm $+1$ and only making sales to types $\theta \in [-1, -1/2]$. Therefore, we must have $k = 1/4$ and $t = 1/2$.

We now ask which distribution of prices $G\,(p_-)$ induces firm $+1$ to price according to $p_+(\hat{\theta})$, defined in (4). Conditional on signal θ, the expected profits of firm $+1$ are given by

$$\mathbb{E}\Pi_+\,(\theta) = \max_p\,[p\,(1 - G\,(p - 2\theta))]\,.$$

Therefore, letting $p_- = p_+ - 2\theta$ and using (4) we have the following differential equation:

$$\frac{g\,(p_-)}{1 - G\,(p_-)} = \frac{1}{p_- + \frac{4k}{p_-} - 2},$$

with the two boundary conditions $G\,(k) = 0$ (notice that when $\theta = 1$ buys from -1 we must have $p_- = k$ from (4)), and $G\,(-2t) = 1$. These conditions determine the values of c and k. The equilibrium strategy is therefore given by

$$G(p_-) = 1 - \frac{e^{4/3}}{4/3} \frac{e^{-\frac{1}{1-p_-}}}{1-p_-}, \; p_- \in \left[\frac{1}{4}, 1\right].$$

The optimal price of the informed firm is given by (4) and therefore by

$$p_+(\theta) = 2\theta + \frac{1}{2(1+\theta)}, \; \theta \in \left[-\frac{1}{2}, 1\right].$$

its expected profits as a function of θ are given by

$$\mathbb{E}\Pi_+(\theta) = \int_{-1/2}^{1} p_+(\theta)(1 - G(p_+(\theta) - 2\theta)) \frac{1}{2} d\theta.$$

We can show that not only profits, but also prices are (pointwise) higher than under complete information (i.e. the price $p_+(\theta)$ charged is higher than 2θ for all $\theta > 0$). Notice that the equilibrium is inefficient, and that all consumers participate as long as $v \geq 5/4$.

When v is not high enough, the informed firm enjoys a local monopoly region. In particular, as long as $v \geq 1/2$, the informed firm is able to extract the whole surplus from consumers $\theta \in [t(v), 1]$. The uninformed firm's profits are still given by $k = 1/4$ and the informed firm's strategy $p_+(\theta)$ is unchanged for $\theta \in [-1/2, t(v)]$. The distribution of prices $G(p_-)$ is modified so that $G(v - t(v)) = 0$. We summarize our results in the next proposition.

Proposition 1 (Asymmetric Information Structure)
For any $v \geq 1/2$ the unique mixed strategy equilibrium is given by

$$\begin{aligned}
G(p_-) &= 1 - \frac{1 - c(v)}{1 - p_-} e^{\frac{1}{1-c(v)} - \frac{1}{1-p_-}}, \quad p_- \in [c(v), 1], \\
c(v) &= \max\left\{\frac{1}{2}\left(v + 1 - \sqrt{v^2 + 2v - 1}\right), \frac{1}{4}\right\}, \\
p_+(\theta) &= \begin{cases} v + \theta & \textit{for} \quad \theta > t(v) \\ 2\theta + \frac{1}{2(1+\theta)} & \textit{for} \quad \theta \in [-1/2, t(v)] \\ 0 & \textit{for} \quad \theta \in [-1, -1/2], \end{cases} \\
t(v) &= \min\{v - c(v), 1\}.
\end{aligned}$$

For all v in this range, the profits of the uninformed firm are given by $\Pi_- = 1/4$. In contrast, the profits of the informed firm are given by

$$\Pi_+(v) = \int_{-1/2}^{1} p(\theta)(1 - G(p(\theta) - 2\theta)) \frac{1}{2} d\theta.$$

Using our expression for $G(\cdot)$ we obtain

$$\Pi_+(v) = \frac{3}{32}\operatorname{Ei}\left(\frac{1}{3}\right)e^{\frac{1}{3}} + \frac{9}{16} \approx 0.67, \text{ for } v \geq 5/4, \tag{5}$$

while for the case of $v < 5/4$,

$$\begin{aligned}
\Pi_+(v) &= (1-c(v))\, e^{\frac{1}{1-c(v)}} \left(\frac{1}{2e} \int_{-1/2}^{v-c} (2\theta+1)\, e^{-\frac{1}{2\theta+1}} d\theta \right) + \int_{v-c}^{1} (v+\theta) \frac{1}{2} d\theta \\
&= (1-c(v))\, e^{\frac{1}{1-c(v)}} \left(\frac{1}{4e} \int_{\frac{1}{2(v-c)+1}}^{\infty} \frac{1}{y^3} e^{-y} dy \right) + \int_{v-c}^{1} (v+\theta) \frac{1}{2} d\theta. \qquad (6)
\end{aligned}$$

This expression is single-peaked in v. This suggests that when the consumers' base values are lower, and hence the competition-intensifying concern is diminished, information about consumer location becomes even more valuable. Clearly, profits must be increasing in v over some range, due to the first-order demand intercept effect.

3.2.4 Summary

We now summarize the profit levels under the three information structures analyzed so far in three graphs. The range of v is $[1/2, 3/2]$ and blue corresponds to no information, red to symmetric complete information and green to asymmetric information. Figure 1 shows profit levels for each individual firm.

Figure 1: Individual Firm Profits

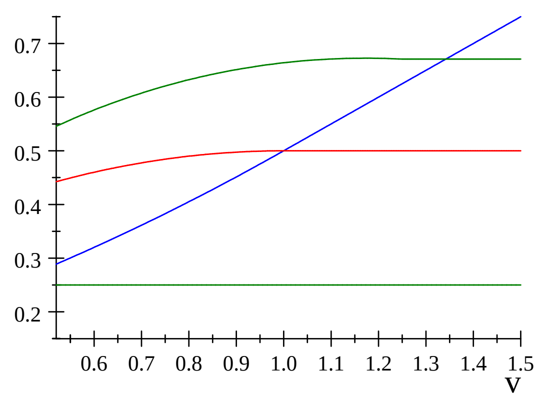


Figure 2 shows industry-level profits under each information structure.

And finally, Figure 3 shows the value of information for each firm, as a function of v, given that its competitor has acquired information (red) or it has not (blue).

Figure 3 allows us to characterize the price of data under each scenario.

Figure 2: Total Profits

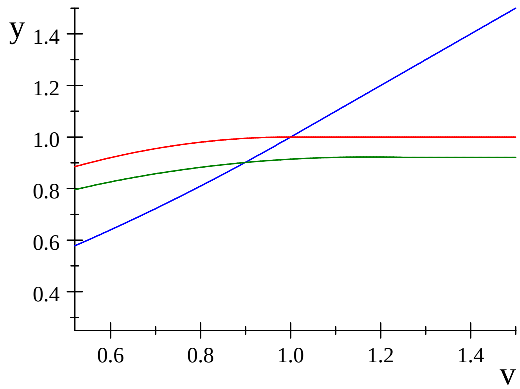
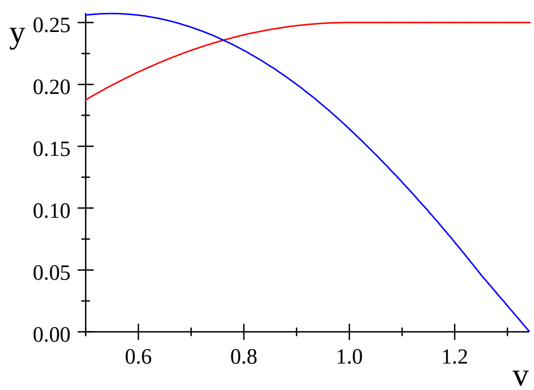


Figure 3: Individual Gains from Information



3.3 Equilibrium Price of Data

We now relate our findings about special information structures to the profitability of selling consumer-level information to firms competing in a differentiated-products oligopoly. Depending on parameters, most importantly the degree of product differentiation, information may or may not be profitable in our strategic setting. In particular, uncertainty increases firms' profits when consumers' brand preferences are weak, as more information would make price competition fiercer. Conversely, when firms can enjoy local monopoly consumer segments, information helps tailor the price to the most valuable customers.

We tease out these effects in the following game: a single data provider offers a database for sale to each firm at a price p_D; firms simultaneously choose whether to buy the data; after observing the information purchasing decisions, firms compete in prices. Suppose now that the database reveals the exact location θ of the consumer. The three special information structures described in Section 3.2 correspond to the three subgames following the data purchases. Throughout this section, we maintain the uniform distribution assumption.

Quite simply, the normal form of the game played in the first stage is given by

$$\begin{array}{c|ll|} & \text{buy} & \text{not} \\ \cline{2-3} \text{buy} & \Pi_F - p_D,\ \Pi_F - p_D & \Pi_I - p_D,\ \Pi_U \\ \text{not} & \Pi_U,\ \Pi_I - p_D & \Pi_H,\ \Pi_H \\ \cline{2-3} \end{array}$$

In this matrix, Π_H denotes the Hotelling payoff, Π_I and Π_U denote the informed and uninformed payoffs in the asymmetric information structure, and Π_F denotes the complete information payoff.

Under the uniform distribution, based on the previous section's results, the incomplete- and complete-information profits are given by

$$\Pi_H = \begin{cases} 1 & \text{if} \quad v \geq 2 \\ \frac{v}{2} & \text{if} \quad v \in [1, 2] \\ \frac{1}{2}\left(\frac{1+v}{2}\right)^2 & \text{if} \quad v \leq 1, \end{cases}$$

and

$$\Pi_F = \begin{cases} \frac{1}{2} & \text{if} \quad v \geq 1 \\ \frac{1}{2}v\left(1 - \frac{1}{2}v\right) + \frac{1}{4} & \text{if} \quad v < 1. \end{cases}$$

Assuming the market is covered when only one firm is fully informed (i.e, $v \geq 1/2$), the asymmetric information profits are given by $\Pi_U = 1/4$ for the uninformed firm and by $\Pi_I = \max\{(6),(5)\}$ depending on whether $v \gtrless 5/4$.

We now focus on the complete market coverage and examine equilibria in the information acquisition stage. Let $\hat{v}$ be the critical level for which (see Figure 3) $\Pi_F - \Pi_U = \Pi_I - \Pi_H$. We have the following result.

Proposition 2 (Data Acquisition Game)

1. *Assume $v \geq \hat{v}$. Then for all $p_D \in [\Pi_I - \Pi_H, \Pi_F - \Pi_U]$, there are two pure-strategy equilibria, in which both firms, or neither, buy the dataset.*

2. *Assume $v \geq \hat{v}$. Then for all $p_D \in [\Pi_F - \Pi_U, \Pi_I - \Pi_H]$, there are two pure-strategy equilibria, in which each firm buys the dataset.*

Note that the gains from information for a firm are higher when its competitor is informed ($\Pi_F - \Pi_U > \Pi_I - \Pi_H$). In other words, this result confirms the intuition that information acquisition decisions are strategic complements. Therefore, in the best equilibrium (for the data company), the price charged for the data corresponds to the entire value of information when a competitor is fully informed. Clearly, the comparison of the no-information and complete information payoffs was irrelevant for this calculation. In addition, it is not hard to see that the data company could exploit the strategic complementarity by making sequential offers.

Proposition 3 (Optimal Prices of Data)

1. *Consider the mixed strategy equilibrium of the simultaneous-moves game: it is always more profitable for the data provider to sell information to both firms, i.e.* $p_D^* = \Pi_F - \Pi_U$.

2. *If the data provide makes sequential offers and* $v < 2\Pi_I$*, then in the unique subgame perfect equilibrium both firms buy the dataset at a price of* $p_D = p_D^*$.

If the data company can make sequential pricing offers, and the first firm rejects the price p_D, the second firm receives an offer of $p_D = \Pi_I - \Pi_H$. As long as this is positive, the second firm will purchase. The first firm is therefore willing to pay up to $\Pi_F - \Pi_U = 1/4$ to avoid the resulting informational disadvantage. However, firms realize that information may have a negative equilibrium value, and in particular if $v \geq 2\Pi_I$ no firm will purchase the data (even if $p_D = 0$). This occurs because no firm wishes to sing, were one firm to

To summarize, we have identified two key elements that affect the value of information in a differentiated products oligopoly: (a) the demand intercept v, which determines whether the market is covered under incomplete information, and highlights the potential efficiency gains in information revelation (additional consumers may be served when firms can tailor their prices); and (b) the local degree of competitiveness, measured by the value of the density at the median $f(0)$.

For the first criterion (a), we have shown that a low demand intercept enhances the value of information, while for the second criterion (b), we remark for now that the Beta family of distributions (restricted to be symmetric about the median)

$$f(\theta) = \frac{((1+\theta)(1-\theta))^{a-1}}{\int_{-1}^{1} ((1+x)(1-x))^{a-1}\,\mathrm{d}x}, \quad a > 0$$

provides a useful comparison tool.

We now move in a different direction and compare the (asymmetric but) full information revelation policy with a noisier one, that restores some elements of Hotelling competition.

3.4 Partial Information Structures

We analyze the asymmetric information structure in which firm -1 is perfectly informed about the consumer's location, and firm $+1$ is uninformed, if and only if $\theta \leq 0$. The opposite is true for firm $\theta \geq 0$. Under this structure, the following strategy for each firm gives us the pure-strategy equilibrium:

$$\begin{aligned} p(\theta) &= 2|\theta| \\ p(\emptyset) &= 0. \end{aligned}$$

Again, the equilibrium profits are given by (1) and (2) for the two firms. This shows that, for general distributions, the data provider can create the same industry-wide value of information by limiting the total amount of information collected. We will come back to this theme when we introduce a cost of information gathering.

4 Market Tailoring

We consider the notion of market tailoring in the oligopoly environment with normally distributed uncertainty about a cost or a demand shock. The basic problem is phrased for a common value environment with uncertainty about the nature of the demand, represented by θ.

4.1 Information Structures

The database, in the baseline model, is assumed, to have complete information about θ. It can provide the information to each of the competing firms, possibly in a different way. The underlying parameter θ is distributed according to a common prior given by a normal distribution F, but the firms do not know the exact realization of the state of the world. In the baseline model, the firms do not have any information beyond the prior F.

We are interested in a number of information policies:

1. The database could provide a single noisy, but common signal, to all the firms in the market.
2. The database could provide a single noisy signal to each individual firm, but the signals could be conditionally independent across the firms in the market.
3. The database could provide some of the market participant with privileged information, and not provide the other firms with any information. The privileged firms could receive noiseless or noisy (and correlated, or conditionally independent) information.

We know from recent work, see Bergemann and Morris (2011a), that a complete and symmetric information provision policy does not maximize the aggregate industry profits across all information structures. In fact, some degree of idiosyncratic information is necessary to maintain a moderate level of profits for the firms. This suggests that the optimal information structure is one that will use less than complete information, and will offer personalized, randomized information to each of the firms.

4.2 Oligopoly with Common Demand Uncertainty

We consider an oligopoly model with a finite number of symmetric firms $i \in \{1, ..., I\}$ and uncertainty about the state of the common demand θ. If the state of the world θ were known, then the individual firm i receives profits

$$q_i \left(\theta - b \sum_{j=1}^{I} q_j \right) - \frac{1}{2} c q_i^2 .$$

The inverse demand function is assumed to be linear

$$\theta - b \sum_{j=1}^{I} q_j,$$

and the cost function of the individual firm is strictly concave, in fact quadratic. The first order condition under complete information about the state of demand θ are hence given by:

$$\theta - b \left(\sum_{j \neq i} q_j + 2q_i \right) - cq_i = 0,$$

and so the best response function is given by

$$q_i = \frac{\theta - b \sum_{j \neq i} q_j}{2b + c} = \frac{\theta - b (I - 1) q_{-i}}{2b + c}, \tag{7}$$

where the average action q_{-i} of the opponents of firm i are given by:

$$q_{-i} = \frac{1}{I - 1} \sum_{j \neq i} q_j.$$

The unique complete information equilibrium action profile is therefore the solution to:

$$q = \frac{\theta - b (I - 1) q}{2b + c} \Rightarrow q^* = \frac{\theta}{b + c + bI}.$$

Now, in a world with uncertain demand, the true state of demand, θ, is given by a normal distribution with positive mean and variance, and so:

$$\theta \sim N \left(\mu_\theta, \sigma_\theta^2 \right).$$

The present setting is therefore described by quadratic payoff functions and normally distributed uncertainty.

Interestingly, in the context of the linear oligopoly under uncertainty, an extensive literature has examined the incentives of competing firms to share cost and demand information. The problem of information sharing among firms was pioneered in work by Novshek and Sonnenschein (1982), Clarke (1983) and Vives (1984), who examined to what extent competing firms have an incentive to share information in an uncertain environment. In this strand of literature, which is surveyed in Vives (1990) and embedded in a very general framework by Raith (1996), each firm receives a private signal about a source of uncertainty, say a demand or cost shock. The central question then is under which conditions the firms have an incentive to commit ex-ante to an agreement to share information in some form. A striking result by Clarke (1983) was the finding that in a

Cournot oligopoly with uncertainty about a common parameter of demand, the firms will never find it optimally to share information. The complete lack of information sharing, independent of the precision of the private signal and the number of competing firms, is surprising. After all, it would be socially optimal to reduce the uncertainty about demand and a reasonable conjecture would be that the firms could at least partially appropriate the social gains of information.

But, perhaps curiously, the literature on information sharing never asked under which condition a data provider would be able to enter the market and provide information about the common value to some or all of the competing firms. This is the question that is of central interest for us.

We allow each firm to purchase and hence to receive a two-dimensional signal. In the first dimension, the signal is privately observed and idiosyncratic to the firm, whereas in the second dimension, the signal is publicly observed and common to all the agents. In either dimension, the signal is normally distributed and centered around the true state of the world θ. Accordingly, we consider the following bivariate normal information structure. Each firm i is observing a private and a public noisy signal of the true state of the world θ. The private signal x_i, observed only by firm i, is defined by:

$$x_i = \theta + \varepsilon_i, \tag{8}$$

and the public signal, common and commonly observed by all the firms is defined by:

$$y = \theta + \varepsilon. \tag{9}$$

The random variables ε_i and ε are normally distributed with zero mean and variance given by σ_x^2 and σ_y^2, respectively; moreover ε_i and ε are independently distributed, with respect to each other and the state θ. This model of bivariate normally distributed signals appears frequently in games of incomplete information, see Morris and Shin (2002) and Angeletos and Pavan (2007) among many others. It is at times convenient to express the variance of the random variables in terms of the precision:

$$\tau_x \triangleq \sigma_x^{-2}, \tau_y \triangleq \sigma_y^{-2},\ \tau_\theta \triangleq \sigma_\theta^{-2} \quad \text{and} \quad \tau \triangleq \sigma_\theta^{-2} + \sigma_x^{-2} + \sigma_y^{-2};$$

we refer to the vector (τ_x, τ_y) as the information structure of the game.

A special case of the noisy environment is the environment with zero noise. In this environment, the complete information environment, each agent observes the state of the world θ without noise.

In the above cited work on information sharing, the individual firms receive a private, idiosyncratic and noisy signal x_i about the state of demand θ. Each firm can commit to transmit the information, noisy or noiseless, to an intermediary, such as a trade association, which aggregates the information. The intermediary then discloses the aggregate information to the firms. Importantly, while the literature did consider the possibility of noisy *or* noiseless transmission of the private information, it a priori restricted the disclosure policy to be noiseless, which implicitly restrict the information policy to disclose the same, common signal to all the firms. An information policy is then a pair a information transmission and information disclosure policies.

We shall now ask how much information the data provider will disclose and how the data provider will price the information. To this end, we first have to describe, the expected value of the Bayes Nash equilibrium that is played by the firms subsequent to the information transmission. For given variances, we then analyze the Bayes Nash equilibrium of the basic game.

Proposition 4 (Linear Bayes Nash Equilibrium)
The unique Bayes Nash equilibrium is a linear equilibrium: $q(x,y) = q_0^* + q_x^* x + q_y^* y$.

The derivation of the linear equilibrium strategy already appeared in many contexts, e.g., in Morris and Shin (2002) for the beauty contest model, and for the present general environment, in Angeletos and Pavan (2007).

Now, if we ask how the joint distribution of the Bayes Nash equilibrium varies with the information structure, then given the normality of the equilibrium distribution, it follows that it is sufficient to consider the second moments, that is the variance-covariance matrix. The variance-covariance matrix of the equilibrium joint distribution over individual actions q_i, q_j, and state θ is given by:

$$\Sigma_{q_i,q_j,\theta} = \begin{bmatrix} \sigma_q^2 & \rho_q \sigma_q^2 & \rho_{a\theta}\sigma_q\sigma_\theta \\ \rho_q\sigma_q^2 & \sigma_q^2 & \rho_{a\theta}\sigma_q\sigma_\theta \\ \rho_{a\theta}\sigma_q\sigma_\theta & \rho_{a\theta}\sigma_q\sigma_\theta & \sigma_\theta^2 \end{bmatrix}. \tag{10}$$

We denote the correlation coefficient between action q_i and q_j shorthand by ρ_q rather than ρ_{qq}. Given the structure of the variance-covariance matrix, we can express the equilibrium coefficients q_x^* and q_y^* in terms of the variance and covariance terms that they generate.

Since the correlation coefficient of the actions has to be nonnegative, the above representation suggest that as long as the correlation coefficient $(\rho_a, \rho_{a\theta})$ satisfy:

$$0 \leq \rho_a \leq 1, \text{ and } \ \rho_a - \rho_{a\theta}^2 \geq 0, \tag{11}$$

we can find information structures (τ_x, τ_y) such the resulting coefficients are indeed the equilibrium coefficients of the associated Bayes Nash equilibrium strategy.

Proposition 5 (Information and Correlation)
For every $(\rho_a, \rho_{a\theta})$ *such that* $0 \leq \rho_a \leq 1$, *and* $\rho_a - \rho_{a\theta}^2 \geq 0$, *there exists a unique information structure* (τ_x, τ_y) *such that the associated Bayes Nash equilibrium displays the correlation coefficients* $(\rho_a, \rho_{a\theta})$*:*

$$\tau_x = \frac{(1-\rho_a)\,\rho_{a\theta}^2}{\left((1-\rho_a) + (1-r)\left(\rho_a - \rho_{a\theta}^2\right)\right)^2 \sigma_\theta^2},$$

and

$$\tau_y = \frac{\left(\rho_a - \rho_{a\theta}^2\right)\rho_{a\theta}^2\,(1-r)^2}{\left((1-\rho_a) + (1-r)\left(\rho_a - \rho_{a\theta}^2\right)\right)^2 \sigma_\theta^2}.$$

4.3 Value of Information

We next consider the value of information for the oligopolist with respect to the uncertain state of the world θ. We can now compute the expected value of the equilibrium. The payoff function is given by:

$$\mathbb{E}\left[q_i\left(\theta - b\sum_{j=1}^{I} q_j\right) - \frac{1}{2}cq_i^2\right] =$$

In general, the quadratic form is given by:

$$\mathbb{E}\left[\varepsilon\Lambda\varepsilon\right] = tr\left[\Lambda\Sigma\right] + \mu^T\Lambda\mu, \tag{12}$$

and here the quadratic form is given by:

$$\Lambda = \begin{bmatrix} -b - \frac{1}{2}c & -\frac{1}{2}b & \cdots & \frac{1}{2} \\ -\frac{1}{2}b & 0 & & 0 \\ \vdots & & \ddots & \vdots \\ \frac{1}{2} & 0 & \cdots & 0 \end{bmatrix}$$

and the variance matrix is given by

$$\Sigma = \begin{bmatrix} \sigma_a^2 & \rho_a\sigma_a^2 & \cdots & \rho_{a\theta}\sigma_a\sigma_\theta \\ \rho_a\sigma_a^2 & \ddots & & \vdots \\ \vdots & & \sigma_a^2 & \rho_{a\theta}\sigma_a\sigma_\theta \\ \rho_{a\theta}\sigma_a\sigma_\theta & \cdots & \rho_{a\theta}\sigma_a\sigma_\theta & \sigma_\theta^2 \end{bmatrix}.$$

The value of the equilibrium outcome hence depends only on the trace, and is given by:

$$\begin{aligned} \Lambda\Sigma &= \begin{bmatrix} -b - \frac{1}{2}c & -\frac{1}{2}b & \cdots & \frac{1}{2} \\ -\frac{1}{2}b & 0 & & 0 \\ \vdots & & \ddots & \vdots \\ \frac{1}{2} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \sigma_a^2 & \rho_a\sigma_a^2 & \cdots & \rho_{a\theta}\sigma_a\sigma_\theta \\ \rho_a\sigma_a^2 & \ddots & & \vdots \\ \vdots & & \sigma_a^2 & \rho_{a\theta}\sigma_a\sigma_\theta \\ \rho_{a\theta}\sigma_a\sigma_\theta & \cdots & \rho_{a\theta}\sigma_a\sigma_\theta & \sigma_\theta^2 \end{bmatrix} \\ &= \begin{bmatrix} \left(-b - \frac{1}{2}c\right)\sigma_q^2 - \frac{1}{2}\left(I-1\right)b\left(\rho_q\sigma_q^2\right) + \frac{1}{2}\rho_{q\theta}\sigma_q\sigma_\theta & \cdots & \cdots \\ \vdots & -\frac{1}{2}b\rho_q\sigma_q^2 & 0 \\ & \ddots & \vdots \\ & \cdots & \frac{1}{2}\rho_{q\theta}\sigma_q\sigma_\theta \end{bmatrix}, \end{aligned}$$

and hence the trace

$$\begin{aligned} tr\left[\Lambda\Sigma\right] &= \left(-b-\frac{1}{2}c\right)\sigma_q^2-(I-1)\,b\left(\rho_q\sigma_q^2\right)+\rho_{q\theta}\sigma_q\sigma_\theta \\ &= -\frac{1}{2}c\sigma_q^2-Ib\left(\rho_q\sigma_q^2\right)+\rho_{q\theta}\sigma_q\sigma_\theta . \end{aligned}$$

Now using the firm's best response function, we get a simple expression for:

$$\begin{aligned} & -\frac{1}{2}c\left(\frac{\sigma_\theta\rho_{q\theta}}{2b+c+b\rho_q\,(I-1)}\right)^2-Ib\left(\rho_q\left(\frac{\sigma_\theta\rho_{q\theta}}{2b+c+b\rho_q\,(I-1)}\right)^2\right)+\rho_{q\theta}\left(\frac{\sigma_\theta\rho_{q\theta}}{2b+c+b\rho_q\,(I-1)}\right)\sigma_\theta \\ = \ & \frac{1}{2}\sigma_\theta^2\rho_{q\theta}^2\frac{4b+c-2b\rho_q}{\left(2b+c-b\rho_q+b\rho_q I\right)^2} \end{aligned}$$

It therefore follows that we will always choose $\rho_{q\theta}$ as large as possible, or using the inequalities in (11), we get:

$$\rho_q \geq -\frac{1}{I-1}, \tag{13}$$

and

$$\rho_q+\frac{1}{I-1}\geq \rho_{q\theta}^2, \tag{14}$$

we obtain, now in terms of ρ_q alone:

$$f(\rho)=\left(\rho+\frac{1}{I-1}\right)\frac{4b+c-2b\rho}{(2b+c+(I-1)\,b\rho)^2}$$

and hence, where we take $2b+c\triangleq c$; $I-1\triangleq I$:

$$f(\rho)=\left(\rho+\frac{1}{I}\right)\frac{c+2b\,(1-\rho)}{(c+Ib\rho)^2}$$

which gives us an interior solution

$$\rho=\frac{\left((2b+c)^2-4b^2\right)I-2b\,(2b+c)}{\left(2b^2+b\,(2b+c)\right)I^2-2b^2I+4b\,(2b+c)\,I}\in\left(-\frac{1}{I-1},1\right),$$

after observing that with inserting $2b+c\triangleq c$, we get

$$\rho=\frac{1}{bI}\frac{\left(c^2+4bc\right)I-2bc-4b^2}{6b+4c+4bI+cI}\in\left(-\frac{1}{I},1\right), \tag{15}$$

with the important property that as I increases ρ_q decreases, that is the value of information, but more importantly even the optimal size of information, the correlation ρ_q, is decreasing in the number of competitors.

We therefore have the following comparative static result.

Proposition 6 (Demand and Correlation)
The privately optimal (symmetric) level of correlation is increasing in c, decreasing in b, and initially increasing, but eventually decreasing in I.

We can compute the value of the Bayes Nash equilibrium. We can then ask for the participation conditions. To Be Completed.

5 Discussion

We consider the optimal information policy of a data provider in two different environment, which we described as individual and aggregate tailoring. We focused on the optimal decision of single data provider, a monopolist, in these two environments.

The findings across these two different environments have important common features. The firms purchase the information to tailor their decision more closely to the market conditions, whether it is to the individual consumer, or to the aggregate consumer, whether it is through a price or through a quantity/quality decision. Yet, while more information enhanced the efficiency of the (equilibrium) decisions of the firms, an increase in information inevitably leads to more symmetry in the competition and an ensuing decrease in the rents that the firms were able to capture in the market. For this reason, the optimal information policy of the data provider was never to provide and disclose all the information it had available but rather leave the competing firms with some uncertainty. The exact extent and form of the residual uncertainty is largely determined by a desire to leave the competing firms with as large a rent as possible.

In the case of individual tailoring this lead to a symmetric, but biased information policy, in the sense that the local firm received much better information about the local customers than about those further away. The complete information about the local customers, those over which the local firm, has some local monopoly power allowed it to extract all the surplus, while the competitive strategy by its adversary becomes limited through limited information.

In the case of market tailoring, the common value, lead the database to pursue a symmetric policy across the firms, but inject some amount of idiosyncrasy, even at the cost of introducing additional uncertainty to allow each firm to pursue an independent, or at least partially independent policy, even in the presence of a common demand shock.

5.1 Future Developments

The current analysis emphasizes the role of the database to influence the nature of the competition and the information rents among the firms. There are a number of further issues that we should think about.

5.1.1 Advertising Exchanges and other Information Intermediaries

Advertising exchanges, Google, and many other firms both buy (or at least collect information) and then sell the information. Similarly for credit report agencies, rating agencies (and in fact we should try to find out what the payment arrangement or information exchange arrangement presently are in these markets.

We now add a third party, the data company, which buys information from one party and sells it to another one. We notice the analogy with the online advertising market, and we explore the phenomenon of cherry picking that arises as a consequence of contingent pricing of consumers (signals). Furthermore, we consider the possible asymmetry in provision and use of information on a given site. In particular, site S might sell information about its users' behavior, but prohibit advertisers from filtering its content with their own information. We examine the resulting free-riding or coordination problem that arises among sites.

5.1.2 Buying and Selling Information

We now characterize the price charged by firms to report information back to the data company and the equilibrium level of precision for the dataset. We consider the steady state equilibrium of a dynamic model in which each consumer is being tracked over time, and may enter and exit the population at a fixed rate. A candidate for such a model is one of Brownian learning (two-point prior) for which the steady-state distribution (given for example the uniform prior) is a Beta.

Notice that our linear-quadratic utility model only requires knowledge of the mean and variance of the distribution of a consumer's type θ given the signal from the dataset. In particular, we can look at the case of several isolated or competing advertisers, and establish that the dataset's precision will be maximal, with the price of (buying and selling) information adjusting to changes in the market power and mode of competition.

References

Abraham, I., S. Athey, M. Babaioff, and M. Grubb (2011): "Peaches, Lemons, and Cookies: Designing Auction Markets with Dispersed Information," Discussion paper, Microsoft Research.

Angeletos, G., and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75, 1103–1142.

Bergemann, D., and A. Bonatti (2011): "Targeting in Advertising Markets: Implications for Offline Vs. Online Media," *Rand Journal of Economics*, 42(3), 417–443.

Bergemann, D., and S. Morris (2011a): "Correlated Equilibrium in Games of Incomplete Information," Discussion paper, Cowles Foundation for Research in Economics, Yale University and Princeton University.

——— (2011b): "Robust Predictions in Games of Incomplete Information," Discussion paper, Cowles Foundation for Research in Economics, Yale University.

Bergemann, D., and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Celis, L., G. Lewis, M. Mobius, and H. Nazerzadeh (2011): "Buy-it-now or Take-a-chance: A simple sequential screening mechanism," in *Proceedings of the 20th international conference on World wide web*, pp. 147–156. ACM.

Clarke, R. (1983): "Collusion and the Incentives for Information Sharing," *Bell Journal of Economics*, 14, 383–394.

Ganuza, J., and J. Penalva (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78, 1007–1030.

Ganuza, J.-J. (2004): "Ignorance Promotes Competition: An Auction Model of Endogeneous Private Valuation," *RAND Journal of Economics*, 35.

Gentzkow, M., and E. Kamenica (2011): "Competition in Persuasion," Discussion paper, University of Chicago.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Morris, S., and H. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92, 1521–1534.

Novshek, W., and H. Sonnenschein (1982): "Fulfilled Expectations Cournot Duopoly with Information Acquisition and Release," *Bell Journal of Economics*, 13, 214–218.

RAITH, M. (1996): "A General Model of Information Sharing in Oligopoly," *Journal of Economic Theory*, 71, 260–288.

VIVES, X. (1984): "Duopoly Information Equilibrium: Cournot and Bertrand," *Journal of Economic Theory*, 34, 71–94.

——— (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305–321.