Discussion Paper Series – CRC TR 224

Discussion Paper No. 383
Project B 03

Search, Data, and Market Power

Carl-Christian Groh [1]

November 2022

[1] Department of Economics, University of Mannheim, Email: cgroh@mail.uni-mannheim.de

Support by the CRC

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Search, Data, and Market Power*

Carl-Christian Groh†

November 13, 2022

Link to most recent version

Abstract

I study the relationship between data and market power in a duopoly model of price discrimination with search frictions. One firm receives a signal about the valuation of any arriving consumer while its rival receives no information. A share of consumers, referred to as searchers, have equal valuation for the good of either firm and optimally choose which firms to visit. The remaining consumers are captive. In equilibrium, a large majority of searchers will only visit the firm with data. The market share of the firm with data converges to one as the share of searchers in the market goes to one, regardless of the signal structure. Reductions of search frictions induce higher market concentration. The establishment of a right to data portability can address the competitive imbalances caused by data advantages.

Keywords: search, price discrimination, imperfect customer recognition, market power
JEL Classification: D43, D83, L13, L15

---

*I thank Dirk Bergemann, Nicolas Bonneton, Bernhard Ganglmair, Federico Innocenti, Marina Halac, Tsung-Hsien Li, Volker Nocke, Francisco Poggi, Nicolas Schutz, Fiona Scott Morton, Marcel Preuss, Thomas Tröger, Ernst-Ludwig von Thadden, and Jidong Zhou for insightful comments. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project B03) is gratefully acknowledged. This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences. Declarations of interest: none.

†Groh: Department of Economics, University of Mannheim, L7, 3–5, 68161 Mannheim, Germany (email: carl-christian.groh@uni-mannheim.de).

1 Introduction

This paper studies the relationship between data and market power. Data is becoming increasingly relevant in the digital age and is accumulating unevenly — some firms are building up significant advantages in terms of the scope and precision of the data they possess.[1] In order to ensure the proper functioning of digital markets, it is hence imperative to understand how such data advantages will translate into competitive advantages and foster market dominance. This question has gathered significant attention by policymakers (European Commission, 2020) and researchers (Kirpalani & Philippon, 2021; Bergemann & Bonatti, 2022; Eeckhout & Veldkamp, 2022) alike. I study said relationship in a theoretical model of price discrimination with search frictions, in which individual-level consumer data is used to personalize prices and consumers optimally choose which firms to visit.[2]

I show that consumers' search choices substantially amplify the transmission of data advantages into competitive advantages. Even arbitrarily small data advantages can make it optimal for nearly all consumers to only visit the firm with a data advantage, thus granting this firm market shares close to one. This result underscores the importance of regulating digital markets. This is because such extreme forms of market dominance will reduce consumer welfare, for example by deterring entry or by reducing the incentives of firms to innovate. To guide policy, I study the optimal regulation in such contexts. Whereas reductions of search frictions can only exacerbate market dominance, the establishment of a right to data portability (as defined in the EU GDPR and the DMA) is an effective way of correcting the competitive imbalances caused by data advantages.[3]

I consider a duopoly model of a final goods market with search frictions. Every consumer can costlessly visit one firm, but has to pay a search cost to visit another firm after the first. Some consumers are *searchers*: They have equal valuation for the good of either firm and want to buy the good at the lowest possible price. The remaining consumers are *captive consumers*, who can only buy the good at the firm they are captive to. The valuation of any consumer is uniformly drawn from the unit interval and is private information to the consumer.

The two firms have different degrees of information about consumers' valuations. One firm in the market, referred to as the *firm with data*, exogenously receives a private signal about the valuation of every consumer who visits it. This signal can take on two realizations:

[1] See, for example, Statista (2021) and Statista (2022).

[2] There is mounting empirical evidence for price discrimination in online markets — see Hannak et al. (2014), Larson et al. (2015), and Escobari et al. (2019). Regulatory bodies around the world are becoming concerned about this business practice — see OECD Secretariat (2016) and European Commission (2019).

[3] For details, see article 20 of the European Union General Data Protection Regulation (GDPR) and article 6 of the EU Digital Markets Act (DMA).

low or high. The high signal realization becomes more likely to occur when a consumer's valuation rises. Using this signal, the firm with data will price discriminate: It will offer a relatively low price (the low signal price) to all consumers who arrive and generate the low signal and a higher price (the high signal price) to all other arriving consumers. The other firm, referred to as the *firm without data*, receives no information about any consumer and will thus offer the same price to all arriving consumers.[4]

As a benchmark, I solve a variant of the above model in which every consumer can only visit one firm in Section 4.1. Then, the decision problem of any searcher boils down to choosing which firm to visit. Because the firm with data price discriminates, searchers with low valuations prefer to visit the firm with data, while searchers with high valuations prefer to visit the firm without data. This is because consumers with low (high) valuations are likely to be identified as such and receive a comparatively low (high) price at the firm with data. Formally, the equilibrium strategy of searchers is thus a cutoff rule: Any searcher will visit the firm with data if her valuation is below a cutoff and the firm without data if her valuation is above the cutoff.

This search behaviour affects prices through a *selection effect*. Because searchers with low valuations visit the firm with data and vice versa, the average valuation of consumers who visit the firm without data is larger than the average valuation of consumers who visit the firm with data. Thus, these search patterns entail upward pressure on the uniform price of the firm without data and downward pressure on the prices of the firm with data.

A key message of this paper is that this selection effect amplifies the transmission of data advantages into competitive advantages. Simply put, this effect imposes a competitive externality on the firm without data: It pushes up the uniform price the firm without data would optimally set, which is to the benefit of the firm with data because it incentivizes searchers to visit this firm. In fact, a large majority of searchers will just visit the firm with data in equilibrium — only searchers with very high valuations will optimally visit the firm without data. Moreover, the market share of the firm with data converges to one as the share of searchers approaches one, regardless of the signal structure.

Why does the market only equilibrate when the firm without data is just visited by its captive consumers and searchers with very high valuations? Intuitively, the selection effect becomes weak enough to enable equilibrium existence. Because the mass of searchers who visit the firm without data is small, the distribution of consumer valuations is very similar at the two firms. As a consequence, the optimal uniform price of the firm without data will be between the prices set by the firm with data. However, the selection effect is still active,

[4]I focus on equilibria in which firms play pure strategies. In addition, I show that firms play pure strategies in any equilibrium in which prices are drawn from distributions with connected support.

which means that the optimal uniform price of the firm without data will lie just below the highest price of the firm with data (the high signal price), but significantly above the lowest price of this firm (the low signal price). These prices make it optimal for all searchers, except those with very high valuations, to visit the firm with data, because the potential benefit of receiving the low signal price at the firm with data is comparatively large.

In Section 4.2, I show that all previous insights go through when consumers can visit both firms, albeit under slightly stronger restrictions on the share of searchers. Formally, I solve the aforementioned model when the costs of visiting a second firm are arbitrary, while the analysis in Section 4.1 only considers the case in which these search costs are prohibitive.

To begin with, I show that no consumer will visit both firms in equilibrium if the share of searchers is not too low.[5] This result is based on two separate arguments: Firstly, any searcher who initially visits the *firm without data* in equilibrium would never continue searching, because the price this firm offers is non-stochastic.[6] Secondly, there exists no equilibrium in which searchers continue searching after visiting the *firm with data* if there are enough searchers in the market. This is because searchers who arrive at the firm without data after visiting its rival exert upward pressure on the uniform price of this firm.[7] When the share of searchers is large enough, the price the firm without data would set in such a hypothetical equilibrium is thus so high that it is not worthwhile for any consumer to pay a search cost in pursuit of this price.

In equilibrium, all consumers thus only visit one firm and all results that were derived within the baseline model extend verbatim. The firm with data price discriminates and hence, the selection effect is active. As before, a large majority of searchers will thus only visit the firm with data. Moreover, the market share of the firm with data approaches one as the share of searchers goes to one, regardless of the signal structure.

Reductions of search frictions can only exacerbate the dominant position of the firm with data. When search costs are above a certain threshold, the possibility of searching plays no role and changes in search costs do not affect the equilibrium outcomes. At sufficiently low search costs, reductions of search costs further increase the market share of the firm with data. Intuitively, searchers constrain the prices of the firm with data with the threat of searching when search costs are sufficiently small. By strengthening this threat, reductions of search costs will induce the firm with data to lower its prices. These reduced prices raise

[5] I define ρ as the share of searchers in the market. Assuming that $\rho \geq 0.2$ is sufficient for this result when restricting attention to linear signal distributions, independent of the exact level of search costs.

[6] Any searcher who finds it optimal to continue searching after visiting the firm without data would not initially visit this firm in equilibrium. She would be strictly better off by visiting the firm with data first and searching thereafter if and only if a high price is obtained, since this endows her with an option value.

[7] Note that the firm without data offers a uniform price and there are search costs. Thus, any searcher would only continue searching after visiting the firm with data if she would buy at the firm without data.

the incentives of searchers to visit the firm with data, thus granting this firm even higher market shares.

In Section 5, I argue that the market dominance that arises from data advantages within my framework creates a need for regulatory interventions. In short, this is because the accompanying distortions can raise the average price level and will impede innovation and entry in a dynamic context.

Thereafter, I study the effects of two policies designed to curb data advantages — the establishment of a right to anonymity and a right to data portability. A right to anonymity allows consumers to ensure that the firm with data receives no signal about them. Conversely, a right to data portability enables consumers to transfer the information that the firm with data has about them to the firm without data. Whereas the former is inconsequential, the establishment of a right to data portability can be very effective. No consumer would exercise their right to anonymity, because this would be indicative of having a high valuation. By contrast, the incentives to exercise one's right to data portability are highest for low-valuation consumers. Through an unraveling effect, the establishment of a costless right to data portability can thus induce all searchers to visit the firm without data in equilibrium.

In Section 6, I consider various extensions of the baseline model. First, I solve a model in which the firm with data receives a continuous signal about the valuations of visiting consumers. The previous equilibrium predictions extend as long as the signal remains noisy, i.e. as long as the firm with data does not know the valuations of arriving consumers perfectly. Next, I show that the previous results also hold when both firms receive signals about the valuations of visiting consumers, but the signal of one firm is less precise, or when consumers' preferences admit quality differentiation as in Mussa & Rosen (1978).

The rest of the paper proceeds as follows: I offer a detailed literature review in Section 2. In Section 3, I set up the theoretical framework, which is solved in Section 4. Sections 5 and 6 contain the analysis of the aforementioned policy proposals and extensions. I conclude and argue why my insights apply more generally, for example in insurance markets, in Section 7.

2 Related Literature

The findings I establish are novel because all previous work on the competitive effects of data advantages does not consider heterogeneous search patterns in the analysis. In preceding papers, there are either no search frictions (e.g. Eeckhout & Veldkamp, 2022; Rhodes & Zhou, 2022), search is random (Freedman & Sagredo, 2022) or there is no consumer heterogeneity that affects search decisions (Kirpalani & Philippon, 2021). As a result, the selection effect that drives the strong relationship between data advantages and competitive advantages in my framework is absent in previous work.

Several recent papers study the competitive effects of data advantages. In Belleflamme et al. (2020), a firm probabilistically either knows a consumer's valuation perfectly or knows nothing about this consumer. Bounie et al. (2021), Gu et al. (2019), Garcia (2022), and Delbono et al. (2022) study models where firms receive non-stochastic information about consumer preferences and some firms receive more informative data (e.g. a finer partition of the Hotelling line).[8] Rhodes & Zhou (2022) consider a setting in which some firms conduct first-degree price discrimination, whereas their rivals can only offer uniform prices.[9] Eeckhout & Veldkamp (2022) study a model in which better data reduces demand risk, thus inducing firms with data advantages to invest more into reducing marginal costs and attaining scale. In contrast to my work, there are no search frictions in all the aforementioned contributions.

The papers that are closest to mine within this area are Kirpalani & Philippon (2021) and Freedman & Sagredo (2022), because these papers consider frameworks with search frictions. Freedman & Sagredo (2022) examine a model of quality differentiation in which a unit mass of sellers offer quality-price menus to consumers. The firms observe signals about consumers' tastes for quality and different firms have access to signals with varying precision levels. Consumers are randomly matched with either one or two sellers. The key distinction to my work thus lies in the fact that consumers' choice sets are unrelated to their preferences in Freedman & Sagredo (2022) — in their framework, consumers neither choose how many firms nor which kind of firms to visit. The heterogeneous consumer search patterns that are central in my model are thus absent in Freedman & Sagredo (2022).

In Kirpalani & Philippon (2021), consumers choose whether to search for a good on a platform or an outside market. The platform has access to better data, which allows firms on the platform to generate a match with a higher probability. In contrast to my work, there is no consumer heterogeneity in Kirpalani & Philippon (2021) that affects the relative utility of search on the platform vs. searching on the outside market. In equilibrium, all consumers must hence be indifferent between searching on the platform or on the outside market. Thus, the aforementioned separating search behavior of consumers in my model is also absent in Kirpalani & Philippon (2021). In addition, the prices that consumers pay on the platform and on the outside market are the same in Kirpalani & Philippon (2021), i.e. no seller can conduct finer price discrimination in this model.

My work also relates to the growing literature that studies price discrimination in search markets. Armstrong & Zhou (2016) and Preuss (2021) consider models where firms condition

[8] Clavorà Braulin (2021) considers a framework in which consumer preferences vary in two dimensions and firms may acquire different information about the components of a consumer's preferences.

[9] Guembel & Hege (2021) and Osório (2022) consider settings in which firms have different abilities to target their products to the individual preferences of consumers, but there is no price discrimination.

prices on a consumer's search history.[10] Fabra & Reguant (2020) study a simultaneous search setting where firms perfectly observe a consumer's desired quantity and price discriminate based on this information. Mauring (2021) and Atayev (2021) study a setting with shoppers and non-shoppers as defined in Burdett & Judd (1983) and Stahl (1989). Mauring (2021) and Atayev (2021) assume that firms receive imperfect information about the affiliation of a particular consumer to the groups of shoppers and non-shoppers. Marshall (2020) and Groh (2022) are the only papers which consider models of price discrimination based on information about valuations together with search, as this paper does. In all the listed contributions, consumers do not engage in directed search and no firm has a data advantage.[11]

Bergemann et al. (2021) study a homogenous goods model with search frictions in which competing firms receive information about the number of price offers a consumer obtains. In Bergemann et al. (2021), different firms may observe signals with varying levels of informativeness. In contrast to my work, all consumers have the same valuation in Bergemann et al. (2021) and consumers do not engage in directed search. Bergemann & Bonatti (2022) consider a model in which a platform uses data to match consumers and firms.

Ke et al. (2022) study the information design problem of an intermediary that connects sellers with consumers. In this model, every consumer just has a match at one seller. Ex ante, both the consumer and the sellers do not know with which seller the consumer has a match. By contrast, the intermediary perfectly knows said information and designs a *public* information structure about this. Consumers engage in directed search by visiting firms according to the intermediary's recommendations. However, all firms are ex ante symmetric in Ke et al. (2022) and the intermediary's signals are public, so no firm has an informational advantage and all firms obtain the same expected outcomes.

3 Theoretical framework

There is a unit mass of consumers, who each want to buy at most one unit of an indivisible good that is produced by two firms at zero marginal cost. Consumers can costlessly visit one firm, but visiting a second firm after the first incurs search costs $s > 0$. There are two different groups of consumers, namely *captive consumers* and *searchers*. Captive consumers can only buy at the firm they are captive to and have zero valuation for the good of the other firm. By contrast, searchers have equal valuation for the good of either firm. The valuation

[10] Garrett et al. (2019) consider a model of second-degree price discrimination in which consumers differ in their choice sets, but firms do not have information about consumers. Braghieri (2019) studies a search model in which consumers decide whether or not to reveal their horizontal characteristic to firms.

[11] My work is also related to Esteves (2014) and Peiseler et al. (2021), who study price discrimination based on imperfect information about preferences in competitive settings without search frictions.

of any consumer, which I call v, is drawn uniformly from the unit interval. Searchers make up a share $\rho \in (0,1)$ of the total mass of consumers, while a share $0.5(1-\rho)$ of consumers is captive to either firm. If a consumer with valuation v buys the good at price p, the utility of the consumer is:

$$u(v,p) = v - p \tag{1}$$

The two firms are heterogeneous in the information that is available to them for pricing. One firm, which I call the *firm with data*, exogenously receives a binary private signal $\tilde{v} \in \left\{\tilde{v}^L, \tilde{v}^H\right\}$ about the valuation of any consumer who visits it. I define the probability distribution of this signal, which only depends on the consumer's valuation v, as $Pr(\tilde{v}^H|v)$, where $Pr(\tilde{v}^L|v) := 1 - Pr(\tilde{v}^H|v)$. As I will formalize later, I restrict attention to probability distributions that are monotonic in v. I define the signal $\tilde{v}^H$, which becomes more likely to occur when a consumer's valuation increases, as the high signal. The other firm, which I name the *firm without data*, receives no signal about the valuations of arriving consumers.

Both firms can offer a different price to any consumer who visits. Thus, the game's timing is as follows: At the beginning, every consumer observes her valuation (and whether she is a searcher or captive to some firm) and optimally decides which firm to visit first. The firm that is visited first offers a price to the consumer. Based on her valuation and this price offer, the consumer then decides whether to visit the other firm at cost $s > 0$. If the consumer visits a second firm, this firm also offers the consumer a price upon arrival. Crucially, both firms receive no information about any consumer's search history (i.e. they do not know whether an arriving consumer visits them first or second) and do not know whether a consumer is captive or a searcher. This setup implies that, as in Diamond (1971), firms cannot induce more consumers to visit them via downward deviations from equilibrium prices.

I study perfect Bayesian equilibria. Throughout the analysis, I mainly focus on equilibria in which firms play pure strategies. A pure strategy of the firm without data is a uniform price, which I call p^{nd}. A pure strategy of the firm with data is a price tuple (p^L, p^H). This firm offers the price p^L (p^H) to all consumers that visit it and generate the low (high) signal.[12] The strategy of a searcher must define which firm to visit first, based on her valuation. This decision is captured by a measurable function $d : [0,1] \to [0,1]$, where $d(v)$ is the probability that a searcher with valuation v visits the firm with data first. Moreover, the strategy of a searcher must also codify after which initial price offers they would continue searching, conditional on the firm that is visited first. Captive consumers always visit the firm they are

[12]The assumption that $\rho < 1$, i.e. that every firm has captive consumers, ensures that all information sets of both firms are on the equilibrium path, which rules out the existence of perfect Bayesian equilibria that are sustained by implausible off-path punishments.

captive to and do not search thereafter.

In the model, consumers know which firm has a data advantage. This assumption can be motivated along two dimensions. Firstly, knowledge of this fact can arise through learning. Over time, consumers can communicate with their peers and learn which firm sets stochastic prices and which firm sets a uniform price, allowing them to infer which firm uses data to personalize prices. Secondly, such awareness might result from regulation. The European Union, for example, has recently implemented regulation that mandates firms which engage in personalized pricing to inform any visiting consumer about this fact.[13] The benefits of measures that increase consumer awareness of personalized pricing have also been stressed by the OECD's competition committee.[14]

Consider the monopoly benchmark. I define $\Pi^M(p_j|\tilde{v}^k)$ as the profit a monopolist with access to the aforementioned information structure makes when offering the price p_j to consumers who generate the signal $\tilde{v}^k$, with global maximizers $\{p^{k,M}\}_{k\in\{L,H\}}$ given by:

$$p^{k,M} = arg\max_{p_j} \underbrace{p_j \int_{p_j}^{1} Pr(\tilde{v}^k|v)dv}_{:=\Pi^M(p_j|\tilde{v}^k)}, \quad k \in \{L,H\} \tag{2}$$

In the analysis that follows, I impose the following assumptions on $Pr(\tilde{v}^H|v)$:

Assumption 1 *The function $Pr(\tilde{v}^H|v)$ is strictly increasing, continuous, and satisfies $Pr(\tilde{v}^H|v) \in (0,1)$ for all $v \in [0,1]$. Moreover, $\Pi^M(p_j|\tilde{v}^L)$ and $\Pi^M(p_j|\tilde{v}^H)$ are strictly concave in p_j.*

Under this assumption, $p^{L,M} < 0.5 < p^{H,M}$ holds: When observing the low (high) signal, a monopolist will set a lower (higher) price than the price he would set when he has no information about a consumer, namely 0.5. This holds because the average valuation of consumers who generate the low (high) signal is relatively low (high).

I place no functional form restrictions on $Pr(\tilde{v}^H|v)$. Thus, my analysis also covers cases in which the signal $\tilde{v}$ is almost uninformative. Moreover, it is also possible that the firm with data receives a signal which induces it to set higher average prices. I will reference linear signal distributions (which all satisfy assumption 1) with the parameter α as an example throughout the analysis when illustrating the connection between assumptions and primitives and when visualizing results. A linear signal distribution $Pr(\tilde{v}^H|v)$ is given by:

$$Pr(\tilde{v}^H|v) = 0.5 + \alpha(v - 0.5) \tag{3}$$

[13]For details, please examine Directive 2019/2161 of the European Commission.
[14]See article 5 in OECD Secretariat (2016).

In addition, I impose the following tie-breaking rule:

Assumption 2 *Searchers visit either firm with equal probability if* $p^L = p^H = p^{nd}$.

In section 4.1, I solve the specified model under the restriction that $s \to \infty$. In section 4.2, I solve this model for arbitrary $s > 0$. I call the former framework the *baseline model* and the latter the *sequential search framework*.

4 Equilibrium analysis

4.1 Baseline model

Consider first the baseline model, in which it is prohibitively costly for searchers to visit a second firm ($s \to \infty$). In this framework, the only relevant choice that searchers have to make is which firm to visit. If firms play pure strategies, a searcher with valuation v prefers to visit the firm with data if and only if:

$$Pr(\tilde{v}^L|v)\max\{v - p^L, 0\} + Pr(\tilde{v}^H|v)\max\{v - p^H, 0\} \geq \max\{v - p^{nd}, 0\} \tag{4}$$

The strategy of searchers is represented by a function $d(v)$, where $d(v)$ is the probability that a searcher with valuation v visits the firm with data. Given the searchers' behaviour, the firm with data maximizes the following profit function through choice of the price p_j when observing the signal $\tilde{v}^k$, with $k \in \{L, H\}$:

$$\Pi^k(p_j; d(v)) = p_j\left[\underbrace{\rho\int_{p_j}^1 d(v)Pr(\tilde{v}^k|v)dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho)\int_{p_j}^1 Pr(\tilde{v}^k|v)dv}_{\text{captive consumer demand}}\right] \tag{5}$$

Analogously, the firm without data maximizes the following profit function:

$$\Pi^{nd}(p_j; d(v)) = p_j\left[\underbrace{\rho\int_{p_j}^1 (1-d(v))dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho)\int_{p_j}^1 (1)dv}_{\text{captive consumer demand}}\right] \tag{6}$$

I begin by characterizing equilibria in which firms play pure strategies. In such equilibria, the uniform price of the firm without data must lie between the prices of the firm with data. Moreover, the strategy of searchers is described by a cutoff rule:

Lemma 1 (Equilibrium search patterns)
In an equilibrium in which firms play pure strategies:

- *The ordering* $p^L < p^{nd} < p^H$ *must hold.*
- *There exists a* $\bar{v} > p^L$ *such that all searchers with* $v \in (p^L, \bar{v})$ *visit the firm with data and all searchers with* $v \in (\bar{v}, 1]$ *visit the firm without data.*

Simply put, the first result holds because the optimal prices of the firms satisfy the ordering $p^L < p^{nd} < p^H$ if the valuations of consumers who visit either firm follow the same distribution. This holds, for example, under random search or if all searchers visit a given firm. In general, setting a price p^H that is strictly below p^L can never be optimal for the firm with data.[15] Thus, we can restrict attention to equilibrium candidates in which $p^L \leq p^H$. The only candidate for an equilibrium in which $p^L = p^H$ holds is an equilibrium in which all firms set the same uniform price. But then, all searchers visit both firms with equal probability (by the tie-breaking rule described in assumption 2), and the optimal prices of the firms would satisfy $p^L < p^{nd} < p^H$, a contradiction. Thus, $p^L < p^H$ must hold.

Similar arguments establish that $p^{nd} \in (p^L, p^H)$ must hold in equilibrium. For example, suppose that $p^{nd} \leq p^L$. Then, all searchers with $v > p^{nd}$ visit the firm without data, implying that $p^{nd} \geq 0.5$ must hold. But then, the firm with data has a profitable downward deviation from p^L to $p^{L,M}$, since it only sells to captive consumers at p^L and $p^L \geq 0.5 > p^{L,M}$.

When deciding which firm to visit, any searcher thus faces a tradeoff: By visiting the firm with data, she will attain the lowest price p^L with probability $Pr(\tilde{v}^L|v)$, but she may also obtain an unfavorable outcome if she generates the high signal and is thus offered p^H. Because the probability of receiving p^L is strictly falling in v, it becomes strictly less favorable to visit the firm with data instead of its rival as a searcher's valuation increases. This implies the existence of a cutoff $\bar{v} > p^L$ that defines the optimal search behaviour.

The equilibrium search behaviour established above will affect the optimal prices (and their ordering) through a *selection effect*: Searchers visit the firm without data if their valuation is comparatively high and vice versa. Thus, the average valuation of consumers who visit the firm without data is *higher* than the average valuation of consumers who visit the firm with data. This effect entails upward pressure on the uniform price of the firm without data and downward pressure on the prices of the firm with data.

An equilibrium in which firms play pure strategies is described by a vector $(p^L, p^H, p^{nd}, \bar{v})$. The firms optimally set prices, given the search behaviour represented by $\bar{v}$. Before characterizing such equilibria, it is instructive to consider the best response functions of firms. To fix ideas, suppose that all searchers with $v < \bar{v}$ visit the firm with data and that searchers

[15] If $p^H < p^L$, there would either be a downward deviation from p^L to p^H when observing $\tilde{v}^L$ or vice versa.

with valuation $v > \bar{v}$ visit the firm without data. Then, the firm with data maximizes the following objective through choice of p_j when observing the signal $\tilde{v}^k$, with $k \in \{L, H\}$:

$$\Pi^k(p_j; \bar{v}) = p_j \Bigg[\underbrace{\rho \mathbb{1}[p_j \leq \bar{v}] \int_{p_j}^{\bar{v}} Pr(\tilde{v}^k|v)dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho) \int_{p_j}^{1} Pr(\tilde{v}^k|v)dv}_{\text{captive consumer demand}} \Bigg] \tag{7}$$

The firm without data maximizes the following objective function:

$$\Pi^{nd}(p_j; \bar{v}) = p_j \Bigg[\underbrace{\rho \int_{\bar{v}}^{1} \mathbb{1}[p_j \leq v]dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho) \int_{0}^{1} \mathbb{1}[p_j \leq v]dv}_{\text{captive consumer demand}} \Bigg] \tag{8}$$

I define the optimal prices of the firm with data as $p^{L,*}(\bar{v}) = arg \max_{p_j \in [0,1]} \Pi^L(p_j; \bar{v})$ and $p^{H,*}(\bar{v}) = arg \max_{p_j \in [0,1]} \Pi^H(p_j; \bar{v})$. Similarly, I define $p^{nd,*}(\bar{v}) = arg \max_{p_j \in [0,1]} \Pi^{nd}(p_j; \bar{v})$.

In the following two graphs, I visualize these best response functions for a given parametric example in which $\rho = 0.5$ and $Pr(\tilde{v}^H|v) = 0.5 + 0.7(v - 0.5)$. The functions $p^{L,*}(\bar{v})$, $p^{H,*}(\bar{v})$, and $p^{nd,*}(\bar{v})$ are plotted in blue, red, and yellow, respectively:

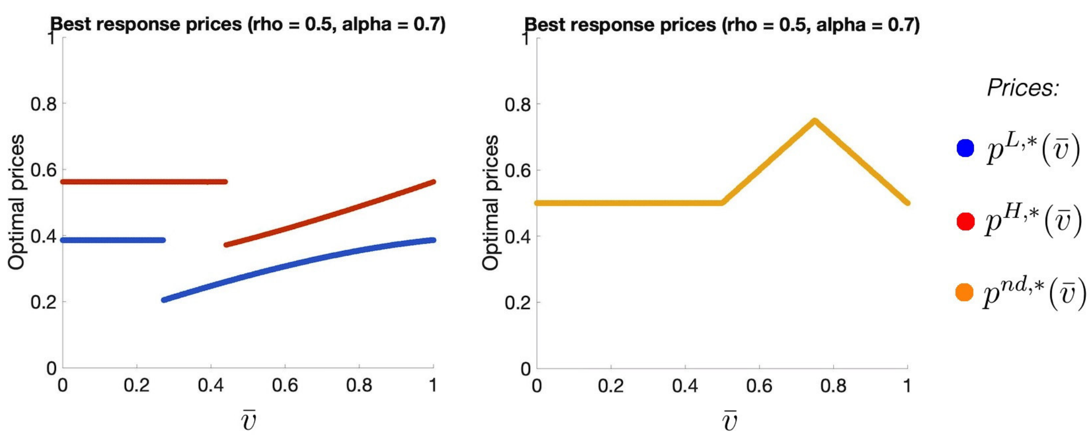

Figure 1: Best response functions

Consider first the optimal prices of the firm with data and recall that this firm is visited by searchers with valuation in $[0, \bar{v}]$. For low values of $\bar{v}$, this firm can only sell to searchers by setting very low prices, which yields low total profits. When $\bar{v}$ is low, it is hence optimal to forego these consumers entirely and to set prices that maximize the profits that accrue from captive consumers, namely $p^{L,M}$ and $p^{H,M}$, respectively. As $\bar{v}$ increases, it becomes optimal to set a price strictly below $\bar{v}$, thereby making the sale to some searchers. For such $\bar{v}$, the optimal prices of the firm with data are rising in $\bar{v}$, because the average valuation of consumers who visit the firm with data is rising in $\bar{v}$.

Now consider the optimal uniform price of the firm without data, which also responds non-monotonically to changes in $\bar{v}$. Recall that this firm is visited by searchers with valuations in the interval $[\bar{v}, 1]$. In general, the profits this firm attains from its captive consumers are maximized by setting the price 0.5. When $\bar{v} \leq 0.5$, setting the price 0.5 also maximizes the profits that accrue from searchers. When $\bar{v} > 0.5$, the profits generated by searchers are maximized by setting the price $p_j = \bar{v}$. Thus, the optimal price $p^{nd,*}(\bar{v})$ is equal to 0.5 when $\bar{v} < 0.5$ and will lie in between 0.5 and $\bar{v}$ for $\bar{v} \in [0.5, 1]$.

When $\bar{v} \in [0.5, 1]$, the optimal price of the firm without data depends on the mass of searchers who arrive at this firm and the corresponding strength of the selection effect. Given that these consumers entail upward pressure on the uniform price of this firm, this price will be comparatively low (high) when the mass of arriving searchers is small (large). When $\bar{v} \in [0.5, 0.5(1+\rho)]$, the mass of searchers who arrive at the firm without data is *large*, which implies that $p^{nd,*}(\bar{v})$ will be equal to $\bar{v}$. For $\bar{v} \in (0.5(1+\rho), 1]$, the mass of searchers who arrive at the firm without data becomes *small*, which means that the optimal price $p^{nd,*}(\bar{v})$ will be strictly below $\bar{v}$. Moreover, $p^{nd,*}(\bar{v})$ is now falling in $\bar{v}$, because the average valuation of consumers who visit the firm without data is *falling* in $\bar{v}$ in this interval.[16]

The presence of the selection effect implies that a majority of searchers must visit the firm with data in equilibrium:

Proposition 1 (Competitive advantages)
In an equilibrium in which firms play pure strategies, the cutoff $\bar{v}$ must satisfy $\bar{v} \geq 0.5(1+\rho)$.

Intuitively, any hypothetical equilibrium in which $\bar{v} < 0.5(1+\rho)$ holds is ruled out by an incompatibility between optimal search behavior and optimal pricing by the firm without data. To see this, note firstly that optimality of the searchers' choices requires that $p^{nd} < \bar{v}$ must hold in equilibrium. This is because any searcher with valuation just above p^{nd} would strictly prefer to visit the *firm with data* (since $p^L < p^{nd}$ must be true in an equilibrium by lemma 1). Thus, the ordering $p^L < p^{nd} < \bar{v}$ must be satisfied in an equilibrium in which firms play pure strategies.

However, previous results have established that setting a price $p^{nd} \in (p^L, \bar{v})$ cannot be optimal for the firm without data when $\bar{v} < 0.5(1+\rho)$. For any $\bar{v}$ and any p^L, the profits of this firm are equal to $\Pi^{nd}(p_j; \bar{v})$ when $p_j \in (p^L, \bar{v})$. If $\bar{v} < 0.5(1+\rho)$, the profits of this firm are thus strictly increasing in p_j at any possible equilibrium $p^{nd} \in (p^L, \bar{v})$, because the upward pricing pressure created by the large mass of arriving searchers is strong, a contradiction.

[16] In general, the average valuation of searchers who arrive at the firm without data is rising in $\bar{v}$, while their mass is falling in $\bar{v}$. Thus, increases in $\bar{v}$ entail opposing effects on the average valuation of all consumers who visit the firm without data. The latter effect dominates for $\bar{v} \in [0.5(1+\rho), 1]$.

Having defined the key properties of any equilibrium in which firms play pure strategies, I now establish the existence of such an equilibrium.

Proposition 2 (Equilibrium existence)
There exists an equilibrium in which firms play pure strategies.

The proof of proposition 2 is by construction. I show that there always exists a $\bar{v}^* \in [0.5(1+\rho), 1]$ that induces optimal prices (given by $p^{L,*}(\bar{v}^*)$, $p^{H,*}(\bar{v}^*)$, and $p^{nd,*}(\bar{v}^*)$) which, in turn, make it optimal for searchers to visit the firm without data if and only if their valuation is above $\bar{v}^*$. I will find such a $\bar{v}^*$ using a fixed point approach.

Continuity of the firms' best response functions plays an important role in the proof of proposition 2. Without further assumptions, the functions $p^{L,*}(\bar{v})$ and $p^{nd,*}(\bar{v})$ will both be continuous on the interval $\bar{v} \in [0.5(1+\rho), 1]$. However, the function $p^{H,*}(\bar{v})$ is not necessarily continuous for these $\bar{v}$ if $\Pi^H(0.5; 0.5 + 0.5\rho) \leq 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$ i.e. when the share of searchers (ρ) is too small. This represents the main technical challenge in proving this proposition. I relegate the formal arguments which show existence of an equilibrium in these constellations to the appendix and focus on the case in which $\Pi^H(0.5; 0.5 + 0.5\rho) > 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$ holds in the following discussion.[17]

Under this assumption, the optimal $p^{H,*}(\bar{v})$ will lie strictly below $\bar{v}$ for any $\bar{v} \in [0.5(1+\rho), 1]$.[18] Thus, the optimal price must satisfy a first-order condition, which guarantees continuity of the function $p^{H,*}(\bar{v})$ on $[0.5(1+\rho), 1]$.

To characterize the optimal search behavior of consumers, I define the following function:

$$\hat{v}(p^L, p^H, p^{nd}) := \sup\left\{v \in [0,1] : \underbrace{Pr(\tilde{v}^L|v)p^L + Pr(\tilde{v}^H|v)p^H}_{\text{exp. price at firm with data}} < p^{nd}\right\} \tag{9}$$

Conditional on (p^L, p^H, p^{nd}), all searchers will obtain a lower expected price at the firm with data if and only if their valuation is below $\hat{v}(p^L, p^H, p^{nd})$. Plugging in the best-response price functions into $\hat{v}(p^L, p^H, p^{nd})$ yields:

$$\hat{v}^B(\bar{v}) := \hat{v}\big(p^{L,*}(\bar{v}), p^{H,*}(\bar{v}), p^{nd,*}(\bar{v})\big) \tag{10}$$

A value $\bar{v}^* \geq 0.5(1+\rho)$ at which $\bar{v}^B(\bar{v}^*) = \bar{v}^*$, together with the implied optimal prices, constitutes an equilibrium. To see this, suppose that searchers visit the firm without data if $v > \bar{v}^*$ and the firm with data if $v < \bar{v}^*$, where $\hat{v}^B(\bar{v}^*) = \bar{v}^*$ and $\bar{v}^* \geq 0.5(1+\rho)$.

[17] This property holds for any linear signal distribution if $\rho \geq 0.13$.

[18] Consider any $\bar{v} \geq 0.5(1+\rho)$. The high signal profits from any price $p_j \geq \bar{v}$ are bounded from above by $0.5(1-\rho)\Pi^{H,M}(p^{H,M})$. By setting a price $p_j < \bar{v}$, e.g. $p_j = 0.5$, when observing $\tilde{v}^H$, the firm can attain higher profits, because $\Pi^H(0.5; \bar{v}) \geq \Pi^H(0.5; 0.5 + 0.5\rho)$ holds for any $\bar{v} \geq 0.5(1+\rho)$.

Given the search behavior represented by $\bar{v}^*$, the firm without data optimally sets the price $p^{nd,*}(\bar{v}^*)$. The optimal prices of the firm with data are $p^{L,*}(\bar{v}^*)$ and $p^{H,*}(\bar{v}^*)$. Given these prices, searchers optimally visit the firm where they receive the lower expected price (conditional on their valuation v). Thus, it is optimal for searchers to visit firms according to the cutoff rule implied by $\bar{v}^*$, because $\bar{v}^* = \hat{v}^B(\bar{v}^*)$. This means that the combination $(p^{L,*}(\bar{v}^*), p^{H,*}(\bar{v}^*), p^{nd,*}(\bar{v}^*), \bar{v}^*)$ constitutes an equilibrium.

Thus, proving that an equilibrium in pure strategies exists amounts to establishing the existence of a solution to the equation $\hat{v}^B(\bar{v}) - \bar{v} = 0$ in the interval $[0.5(1+\rho), 1]$. The existence of an appropriate fixed point can be verified by applying the intermediate value theorem to this equation, together with the boundary conditions (i) $\hat{v}^B(0.5+0.5\rho) > 0.5(1+\rho)$ and (ii) $\hat{v}^B(1) \leq 1$. At $\bar{v} = 0.5(1+\rho)$, $p^{nd,*}(0.5(1+\rho)) = 0.5(1+\rho)$ holds, while both optimal prices of the firm with data are strictly below $\bar{v}$. This establishes that $\hat{v}^B(0.5(1+\rho)) = 1$. The second boundary condition, namely $\hat{v}^B(1) \leq 1$, holds because $\hat{v}^B(\bar{v})$ is the supremum of a set with elements that cannot be larger than 1. Moreover, $\hat{v}^B(\bar{v})$ is continuous on $\bar{v} \in [0.5(1+\rho), 1]$ because all price functions are continuous in $\bar{v}$. Thus, a solution to $\hat{v}^B(\bar{v}) - \bar{v} = 0$ exists in the interval $[0.5(1+\rho), 1]$.

To build further intuition, I present a numerical example. Consider a linear signal distribution with $\alpha = 0.7$ and suppose that $\rho = 0.5$. For all possible equilibrium values of $\bar{v}$ on the x-axis[19], I have plotted the resulting $p^{L,*}(\bar{v})$ in blue, $p^{H,*}(\bar{v})$ in red, and $p^{nd,*}(\bar{v})$ in yellow, respectively, in the following graph. The function $\hat{v}^B(\bar{v})$ is plotted in green:

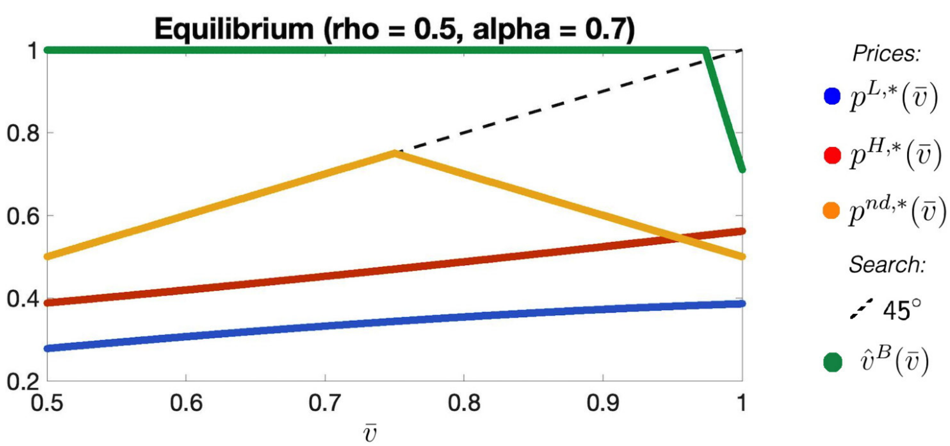

Figure 2: Visualization - equilibrium existence

The point $\bar{v}$ at which $\hat{v}^B(\bar{v})$ crosses the 45-degree line constitutes an equilibrium. The presence of the selection effect is central to the properties and the existence of an equilibrium.

[19] As argued previously, we can directly exclude equilibrium candidates in which $\bar{v} < 0.5$.

When $\bar{v} \leq 0.5(1+\rho)$, the selection effect is too strong to sustain an equilibrium. This manifests in the fact that the optimal uniform price of the firm without data lies above both prices the firm with data would set, so all searchers would prefer to visit the firm with data (i.e. $\hat{v}^B(\bar{v}) = 1$).

As $\bar{v}$ moves closer to 1, the selection effect becomes progressively weaker, i.e. the average valuations of consumers who visit either firm converge to each other.[20] This is accompanied by increases in the optimal prices of the firm with data and decreases in the optimal uniform price of the firm without data. These price changes will induce more searchers to visit the firm without data, which is represented by a falling $\hat{v}^B(\bar{v})$. When $\bar{v} \approx 1$, the optimal p^{nd} will lie just below the optimal p^H, while the optimal p^L lies substantially below these two prices. These prices, in turn, make the search behaviour represented by such high levels of $\bar{v}$ optimal. Only consumers with very high valuations, who are very likely to receive the high price at the firm with data, will optimally visit the firm without data.

Note that there may potentially exist multiple equilibria in which firms play pure strategies. This is because the search behavior of consumers with $v < p^L$ is not pinned down in equilibrium, which means that there might be equilibria in which consumers with $v < p^L$ do not visit the firm with data. This creates an issue of multiplicity, because the derivative of $\Pi(p_j; \tilde{v}^L)$ jumps down at $p_j = p^L$ in such equilibrium candidates.

However, this multiplicity is largely inconsequential for the analysis of market concentration, because $\bar{v} \geq 0.5(1+\rho)$ holds true in any equilibrium in which firms play pure strategies. Moreover, the issue of multiplicity is easily solved by imposing the refinement that searchers with a valuation in an open interval below the lowest equilibrium price visit the firm that offers this price. This is automatic if consumers face a tiny degree of uncertainty regarding their valuation v or the exact equilibrium prices.

This completes the characterization of equilibria in which firms play pure strategies. Now, I consider equilibria in which at least one firm plays a mixed strategy. Before moving forward, I impose a tie-breaking rule on the behavior of searchers with a valuation above the lowest equilibrium price, which I name $\underline{p}$, in such equilibria:

Assumption 3 *When firms play mixed strategies, any searcher with $v > \underline{p}$ who obtains equal expected utility by visiting either firm first visits both firms with equal probability.*

I restrict attention to equilibria in which firms draw prices from distributions with connected support.[21] As defined in Burdett & Judd (1983), a distribution $H(p)$ has connected

[20] To see this, consider the case where $\bar{v} = 1$. Then, the firm without data is only visited by its captive consumers, whose valuations are uniformly drawn from $[0, 1]$, as is the case for the searchers who all visit the firm with data. Thus, the valuation distribution of consumers who visit either firm is exactly equal.

[21] This restriction applies to any information set separately. To clarify this restriction, note that an

support if $H(p_1) \neq H(p_2)$ holds for any distinct prices p_1, p_2 in its support. There exists no such equilibrium in which firms mix.

Proposition 3 (No mixing)
Restrict attention to equilibria in which firms draw prices from distributions with connected support. In any such equilibrium, all firms play pure strategies.

This result is based on the following logic: I define the lowest price set by the firm without data and the firm with data as $\underline{p}^{nd}$ and $\underline{p}^{d}$, respectively. In an equilibrium in which firms mix, $\underline{p}^{d} = \underline{p}^{nd}$ must hold. Under our tie-breaking rule, there exists an interval of prices above this lowest price for which the profit functions of both firms are strictly concave. Thus, this lowest price $\underline{p}^{nd}$ must be offered with probability 1 by the firm without data. But then, the firm with data would also not mix. This is because it only sells to its captive consumers for any price above $\underline{p}^{nd}$, which means that the profits it makes are equal to $\Pi^M(p_j|\tilde{v}^k)$ for any price p_j it offers, which is a strictly concave function for either signal $\tilde{v}^k$.

Thus, we can restrict attention to equilibria in which firms play pure strategies, which I have characterized. In such equilibria, the firm with data has significant competitive advantages, as reiterated by the following corollary:

Corollary 1 (Market dominance)
The equilibrium market share of the firm with data approaches 1 as $\rho \to 1$.

Recall that ρ is the share of searchers in the market. As $\rho \to 1$, the measure of captive consumers approaches 0. In equilibrium, the measure of searchers who buy at the firm without data also approaches 0, because $\bar{v} \geq 0.5(1+\rho)$ and this lower bound converges to 1 as $\rho \to 1$. This is true even when there are multiple equilibria in which firms play pure strategies, because $\bar{v} \geq 0.5(1+\rho)$ holds in such any equilibrium. Thus, the equilibrium demand received by the firm without data approaches 0 as $\rho \to 1$, which implies that the market share of the firm with data approaches 1.

To build further intution for this result, I now visualize the equilibrium prices and search cutoffs for different values of ρ. A given graph corresponds to a fixed linear signal distribution, with $\alpha \in \{0.25, 0.6, 0.95\}$, while different levels of ρ are plotted on the x-axis of each graph. The color scheme of prices is as before, and the equilibrium levels of $\bar{v}$ are plotted in lilac.

---

equilibrium in which the firm with data draws prices from the interval $[0.3, 0.4]$ when observing $\tilde{v}^L$ and draws prices from $[0.5, 0.6]$ when observing $\tilde{v}^H$ is admissible. However, an equilibrium in which this firm draws prices from a distribution with support $[0.3, 0.4] \cup [0.5, 0.6]$ when observing $\tilde{v}^L$ is inadmissible.

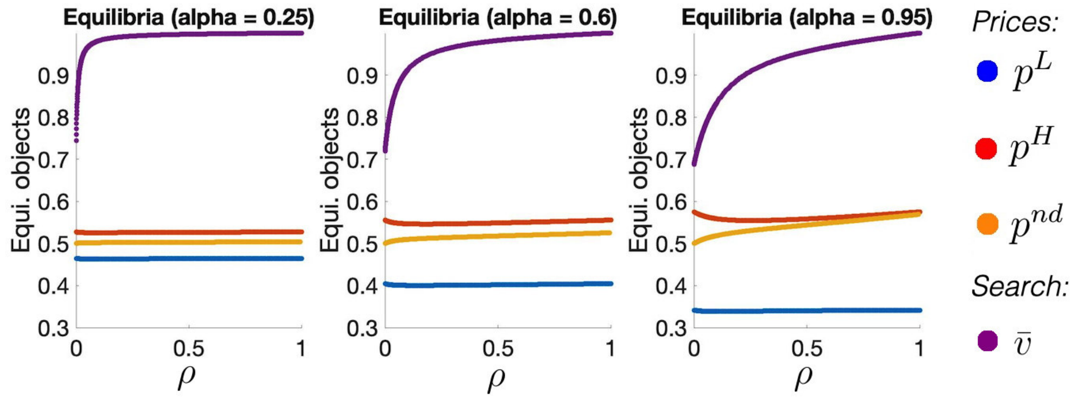

Figure 3: Baseline model - comparative statics (ρ)

When $\rho \to 1$, corollary 1 has established that $\bar{v} \to 1$. As a result, the uniform price of the firm without data approaches $Pr(\tilde{v}^L|1)p^{L,M} + Pr(\tilde{v}^H|1)p^{H,M}$, which is the expected price a searcher with valuation 1 would receive at a monopolist with access to data. To see this, note that the optimal low and high signal prices of the firm with data converge to $p^{L,M}$ and $p^{H,M}$ as $\bar{v}$ approaches 1, respectively. In order for the search behaviour represented by such a high level of $\bar{v}$ to be optimal, the uniform price of the firm without data has to be above the expected price at the firm with data (conditional on the valuation) for almost all searchers. This is guaranteed when the uniform price of this firm approaches $Pr(\tilde{v}^L|1)p^{L,M} + Pr(\tilde{v}^H|1)p^{H,M}$. Such a price is optimal for the firm without data because the slope of $p^{nd,*}(\bar{v})$ on $\bar{v} \in [0.5(1+\rho), 1]$ becomes very large as $\rho \to 1$.

As the signal of the firm with data becomes more informative, the degree of market dominance enjoyed by this firm falls. Formally, the equilibrium levels of $\bar{v}$ are falling in α. This holds by the following logic: When the precision of the signal (α) rises, searchers with high valuations are more likely to be recognized by the firm with data, in which case they receive an unfavorably high price. This reduces their incentives to visit this firm, which, in equilibrium, induces more searchers with high valuations to visit the firm without data.

I have established that arbitrarily small data advantages translate into substantial competitive advantages through directed consumer search. This result is underscored by considering what happens when no firm receives an informative signal. In this benchmark, both firms set the same uniform price in equilibrium and will thus receive exactly half of the market under the tie-breaking rule defined in assumption 2.

4.2 Sequential search framework

In this section, I show that all the results from the baseline model go through even if searchers can visit a second firm, albeit under slightly stronger restrictions on ρ. Formally, I no longer assume that s is prohibitively high, but consider an arbitrary $s > 0$. In terms of policy, these results show that the problem of market dominance cannot be solved by reducing search costs to negligibly small levels.

I begin the analysis by characterizing equilibria in which firms play pure strategies. As before, such an equilibrium needs to define the low signal and high signal price (p^L and p^H, respectively) of the firm with data, as well as the uniform price of the firm without data (p^{nd}). The strategy of searchers now specifies, for a given v, (i) which firm to visit first (captured by a function $d(v)$, as in the baseline model), (ii) after what price offers to search after visiting the firm with data first, and (iii) after what price offers to search after visiting the firm without data first.[22] Because searchers are forward-looking, they take into account under what conditions they would continue searching after sampling the first firm when deciding which firm to initially visit.

To express whether there is search on the equilibrium path, I define the probability with which a searcher with valuation v visits both firms in an equilibrium as $b(v)$. Consider the set $\{v \in [0,1] : b(v) > 0\}$. I say that there is search on the equilibrium path if and only if this set has strictly positive measure. When the share of searchers (ρ) is sufficiently large, there will be no search on the equilibrium path, independent of the exact value of search costs s. This is formalized by the following assumption and accompanying proposition:

Assumption 4 *Suppose that $p^{nd,s} + s > p^{H,M}$, where $p^{nd,s}$ solves the following:*

$$\left[\rho \int_{p^{nd,s}+s}^{1} Pr(\tilde{v}^H|v)dv + 0.5(1-\rho)\int_{p^{nd,s}}^{1} dv\right] = 0.5(1-\rho)p^{nd,s} \tag{11}$$

Proposition 4 (No search beyond the first firm)
Consider the sequential search framework and suppose that assumption 4 holds. There exists no equilibrium in which firms play pure strategies and there is search on the equilibrium path.

Assumption 4 requires that enough consumers engage in directed search, i.e. that ρ is high enough:

Remark 1 *For any linear $Pr(\tilde{v}^H|v)$, assumption 4 is satisfied if $\rho \geq 0.2$.*

[22]Off-path beliefs play no role in the analysis. All information sets of the firms are on the equilibrium path. Any searcher is only uncertain which node the game has reached when visiting the firm with data — then, she does not know which signal was generated. However, this does affect her incentives to continue searching, since these are fully pinned down by the initial price offer and the equilibrium price p^{nd}.

The proof of proposition 4 consists of three steps: Firstly, $p^L < p^H$ must hold in any equilibrium with on-path search. If $p^L = p^H$, any searcher would directly visit the firm which offers the lower uniform price and there would be no reason to search thereafter. If $p^H < p^L$, the firm with data would not be optimizing. Thus, the following arguments consider equilibria with $p^L < p^H$ and show that (i) no searcher who initially visits the firm without data in equilibrium would continue searching and (ii) that, under assumption 4, there exists no equilibrium in which searchers would continue searching after initially visiting the firm with data.

Result (i) follows from a contrapositive argument and requires no assumptions — any searcher who finds it weakly optimal to continue searching after visiting the firm without data (and receiving p^{nd}) would never optimally visit this firm first. This holds by the following logic: By visiting the firm without data first and searching thereafter, the best price this consumer will have in hand after search is p^L with probability $Pr(\tilde{v}^L|v)$ and p^{nd} with probability $Pr(\tilde{v}^H|v)$, while the search cost $s > 0$ is surely paid. Alternatively, the consumer could visit the firm with data first and continue searching if and only if p^H is received. The latter approach would achieve strictly higher expected utility than visiting the firm without data first and searching thereafter, because it yields the same distribution of prices, but saves search costs. By contraposition, no consumer would continue searching after visiting the firm without data in equilibrium — because any such consumer would have never optimally visited this firm first.

Now consider point (ii). The existence of equilibria in which a positive measure of searchers visit the firm without data second is ruled out under assumption 4 by the following logic: Firstly, searchers who visit the firm with data first would only find it optimal to continue searching if the initially received price is weakly above $p^{nd}+s$. Thus, $p^H \geq p^{nd}+s$ must hold in such an equilibrium. There exists no equilibrium in which $p^H = p^{nd} + s$ and a positive measure of consumers search after visiting the firm with data. Under this specification, a marginal downward deviation from p^H would be optimal for the firm with data, since this is sufficient to prevent search by all consumers.

Thus, it only remains to consider equilibria in which $p^{nd} + s < p^H$ (and $p^L < p^H$). In such equilibrium candidates, all searchers with $v > p^H$ who visit the firm with data first and receive p^H would continue searching and never return. Thus, the firm with data only makes the sale to captive consumers for prices in an open ball around p^H, which implies that $p^H = p^{H,M}$ must hold. The optimal price the firm without data would set in such an equilibrium is bounded from below by $p^{nd,s}$ (which is defined in the statement of assumption 4), given that all searchers who arrive at the firm without data surely buy for prices around p^{nd}. Under assumption 4, such an equilibrium cannot exist, because $p^{nd} + s > p^H$ would hold, which contradicts the premise of this equilibrium.

Thus, equilibria in which some consumers search after receiving p^H cannot exist when ρ is high enough. Intuitively, this is based on the following logic: Searchers who arrive at the firm without data second put upward pressure on p^{nd}. This is because visiting this firm second (i.e. paying the search cost $s > 0$) is only optimal for consumers who would buy at p^{nd}. When the share of searchers (ρ) is high, the upward pressure these consumers exert on p^{nd} is strong. Then, p^{nd} would be very high in such a hypothetical equilibrium —so high, in fact, that no searcher would find it optimal to pay a search cost in pursuit of this price.

Now, I turn my attention to equilibria without on-path search. Under an assumption on ρ, all the results established for the baseline model go through verbatim for these equilibria:

Assumption 5 *Assume that* $\Pi^H(0.5; 0.5 + 0.5\rho) > 0.5(1 - \rho)\Pi^{H,M}(p^{H,M})$.

Proposition 5 (Sequential search framework: equilibrium characterization) *Consider the sequential search framework. In an equilibrium in which firms play pure strategies and there is no search on the equilibrium path:*

- *There exists a* $\bar{v} > p^L$ *such that all searchers with* $v \in (p^L, \bar{v})$ *visit the firm with data first and all searchers with* $v \in (\bar{v}, 1]$ *visit the firm without data first.*
- *The cutoff* $\bar{v}$ *must satisfy* $\bar{v} \geq 0.5(1 + \rho)$.

Under assumption 5, such an equilibrium exists.

Remark 2 *For any linear* $Pr(\tilde{v}^H|v)$, *assumption 5 is satisfied if* $\rho \geq 0.13$.

Consider an equilibrium in which firms play pure strategies and there is no search on the equilibrium path. As before, the equilibrium prices must satisfy the ordering $p^L < p^{nd} < p^H$. Thus, the strategy of searchers must be a cutoff rule, because the distribution of prices at the firm with data becomes strictly less favorable as a consumer's valuation rises. Moreover, there exists no such equilibrium in which $\bar{v} < 0.5(1 + \rho)$ holds, because the firm without data would optimally set p^{nd} weakly above $\bar{v}$ in such a hypothetical equilibrium. This is inconsistent with optimal search behavior, because searchers with valuation just above p^{nd} would strictly prefer to visit the firm with data, given that $p^L < p^{nd}$ must hold.

The proof that an equilibrium without on-path search exists for any $s > 0$ under assumption 5 is by construction. First, consider the equilibrium derived for the baseline model (in which s was prohibitively high). I define the components of this equilibrium as $(p^{L,1}, p^{H,1}, p^{nd,1}, \bar{v}^1)$, where $\bar{v}^1 = \hat{v}^B(\bar{v}^1)$, $p^{L,1} = p^{L,*}(\bar{v}^1)$, $p^{H,1} = p^{H,*}(\bar{v}^1)$, and $p^{nd,1} = p^{nd,*}(\bar{v}^1)$. The arguments pertaining to proposition 2 establish that such a combination exists.

If search costs are so high that $p^{H,1} \leq p^{nd,1} + s$, this combination of prices and $\bar{v}$ remains an equilibrium. Then, searchers would never find it optimal to search after visiting the first firm, which implies that it is optimal to visit firms according to the search rule implied by $\bar{v}^1$. Given this search behaviour, firms will find it optimal to set the prices $p^{L,1}, p^{H,1}$ and $p^{nd,1}$, respectively, establishing that this vector of prices and $\bar{v}^1$ constitutes an equilibrium.

Thus, it only remains to establish that an equilibrium of the desired form exists when $p^{H,1} > p^{nd,1} + s$. Consider an equilibrium candidate $(p^{L,2}, p^{H,2}, p^{nd,2}, \bar{v}^2)$, in which $p^{L,2} = p^{L,*}(\bar{v}^2)$, $p^{nd,2} = p^{nd,*}(\bar{v}^2)$, $p^{H,2} = p^{nd,2} + s$, and $\bar{v}^2$ is a solution to the following equation:

$$\bar{v}^2 - \underbrace{\hat{v}\big(p^{L,*}(\bar{v}^2), p^{nd,*}(\bar{v}^2) + s, p^{nd,*}(\bar{v}^2)\big)}_{:=\hat{v}^S(\bar{v}^2)} = 0 \tag{12}$$

There exists a value $\bar{v}^2 \in [0.5(1+\rho), 1]$ that solves this equation. To see why it constitutes an equilibrium, consider the search behaviour of searchers: As before, searchers will maximize their expected utility by initially visiting the firm that offers them (based on their valuation) the lower expected price. Because $p^{H,2} = p^{nd,2} + s$, it is weakly optimal to refrain from searching after visiting the firm with data. Moreover, one can show that searchers with $v > \bar{v}^2$ would not search after visiting the firm without data. Thus, it is optimal for searchers to visit firms according to the cutoff rule implied by $\bar{v}^2$ and to refrain from searching thereafter.

It remains to show that the prices $(p^{L,2}, p^{H,2}, p^{nd,2})$ are optimal for firms if searchers visit firms according to the rule implied by $\bar{v}^2$. There will be no profitable deviations from $p^{L,2}$ and $p^{nd,2}$, because these prices are global maximizers of the respective profit functions when no consumer would ever leave to search, which are weakly above true profits for any price.

There will be no profitable deviations from $p^{H,2}$ under assumption 5 by the following logic: Because search costs are so low that $p^{H,1} > p^{nd,1} + s$, the ordering $p^{H,2} < p^{H,*}(\bar{v}^2)$ will hold. Intuitively, this represents the notion that searchers push down the high signal price of the firm with data using the threat of searching when s is sufficiently low. By strict concavity of the respective profit function, there will not be any profitable downward deviations from $p^{H,2}$. Moreover, assumption 5 guarantees that there will not be any profitable upward deviations (for which the firm with data would only sell to captive consumers). This is because equilibrium profits are bounded from *below* by $\Pi^H(0.5, 0.5(1+\rho))$, while the profits from any deviation above $p^{H,2}$ are bounded from *above* by $0.5(1-\rho)\Pi^{H,M}(p^{H,M})$.

As before, one can rule out the existence of equilibria in which firms mix (within the set of equilibria in which firms draw prices from distributions with connected support):

Proposition 6 (Sequential search framework: no mixing)
Consider the sequential search framework. Suppose that searchers who receive the same price

at both firms buy at each firm with equal probability. In any equilibrium in which firms draw prices from distributions with connected support, all firms play pure strategies.

Summing up, the key results from the baseline model are retained. A large majority of searchers only visit the firm with data. Moreover, the market share of the firm with data approaches 1, independent of the signal distribution, as $\rho \to 1$.

Corollary 2 (Sequential search framework: market dominance)
Consider the sequential search framework. The equilibrium market share of the firm with data approaches 1 as $\rho \to 1$.

When ρ (the share of seachers) approaches 1, both assumptions 4 and 5 will hold, independent of the signal distribution. Thus, an equilibrium will exist. In equilibrium, all consumers just visit one firm and $\bar{v} \geq 0.5(1+\rho)$ must hold, which implies the result.

It remains to study the effects of search cost reductions on the equilibrium objects to understand whether reductions of search frictions can at least mitigate the problem of market dominance. Within the equilibrium established for the baseline model, search cost reductions play no role. When s becomes sufficiently small, reductions of search costs exacerbate market dominance:

Corollary 3 (Reductions of search costs)
Suppose $s < p^{H,1} - p^{nd,1}$ *and that assumptions 4 and 5 hold. Then, the equilibrium level of* $\bar{v}$ *is weakly decreasing in search costs.*

I visualize these effects in the following graph, in which I plot the equilibrium quantities for different levels of search costs (on the x-axis) and α-ρ combinations. The color scheme is as before.[23]

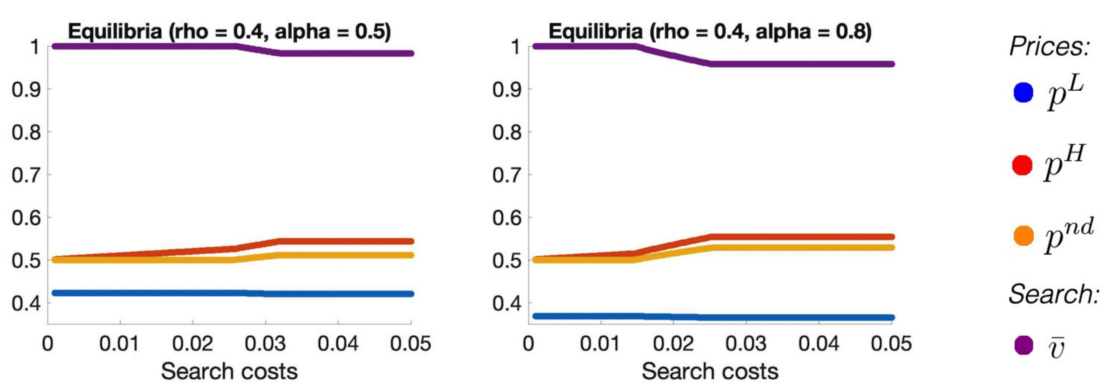

Figure 4: Comparative statics - search costs

[23]I assume that the equilibrium $(p^{L,1}, p^{H,1}, p^{nd,1}, \bar{v}^1)$ is played whenever it exists, i.e. when $p^{nd,1}+s \geq p^{H,1}$.

When search costs are sufficiently high (i.e., $s \geq 0.03$), the equilibrium $(p^{L,1}, p^{H,1}, p^{nd,1}, \bar{v}^1)$ from the baseline model is played, in which the possibility of searching is not relevant. When s becomes sufficiently small, the equilibrium quantities are given by $(p^{L,2}, p^{H,2}, p^{nd,2}, \bar{v}^2)$. Then, search cost reductions lead to lower price levels, but exacerbate the problem of market dominance. Intuitively, searchers are now able to constrain the high signal price of the firm with data with the threat of searching, which implies that this price will approach p^{nd} as search costs fall. This increases the incentives of searchers to visit the firm with data. In particular, all searchers will prefer to only visit the firm with data (i.e. $\bar{v} = 1$) when s becomes sufficiently small, because p^L always remains substantially below p^{nd} and p^H.

5 Welfare and policy recommendations

5.1 Data and consumer welfare

The effects of data advantages imply a need for regulatory interventions for two reasons. Firstly, personalized pricing by the firm with data may lead to higher average prices, thereby reducing consumer welfare. More importantly, the market dominance resulting from data advantages (no matter how small these data advantages are) can reduce consumer welfare by discouraging entry, distorting competition, and by reducing the incentives to innovate.

The personalized pricing that the firm with data implements can lower consumer welfare as such. For example, suppose that the firm with data receives a binary signal that is effective at identifying high-valuation consumers, where $Pr(\tilde{v}^H|v) = 0.5$ if $v < 0.6$ and $Pr(\tilde{v}^H|v) = 1$ if $v \geq 0.6$. If all searchers visit the firm with data, a searcher's ex ante expected utility is 0.1025, while it equals 0.125 when no firm has data. When ρ is high and $\bar{v}$ is thus close to 1, consumer welfare in the competitive equilibrium with data will hence be lower than when no firm has data. However, I note that the impact of personalized pricing on consumer welfare depends on the signal structure. This follows existing ideas, which are summarized by Acquisti et al. (2016).

The market dominance enabled by data advantages can deter entry. This is best conceptualized by augmenting the baseline model with an initial entry stage. There are two firms: the incumbent and the potential entrant, who has no data about consumers. Initially, the entrant has to decide whether or not to pay a fixed cost to enter the market, while the incumbent has to pay no such cost. After the entry decision, the product market competition game from the baseline model is played. If the incumbent has no data, both firms receive half of the market if the entrant enters. If the incumbent has a data advantage, the entrant is visited by a much lower mass of consumers, which makes entry less profitable. Thus, data advantages may discourage entry, which is to the detriment of consumers who have a strong preference for the entrant's product (e.g. the captive consumers in my model).

Empirical evidence by Li et al. (2021) shows that shielding firms from competitive pressures reduces their incentives to innovate. This effect can also be found in an extension of my framework, in which the product market competition described throughout the paper is modeled as a second stage that follows an innovation decision. Suppose that there is a high-quality and a low-quality firm. The valuations that searchers (and the corresponding captive consumers) have for the product of the high-quality firm, call these v, are uniformly distributed on $[0, 1]$. The valuation that any searcher has for the product of the low-quality firm is given by $v - \mu$. In accordance, the valuations that captive consumers have for the product of the low-quality firm are uniformly drawn from $[-\mu, 1 - \mu]$, where $\mu \geq 0$ captures the extent of the quality difference.

Suppose that no firm has data, but that $\mu > 0$. In a monopoly benchmark, the low-quality firm would set the price $0.5(1 - \mu)$, while the high-quality firm would set the price 0.5. In the competitive equilibrium, searchers thus only visit the high-quality firm.[24] Endowing the low-quality firm with data changes this prediction. To see this, define $p^{L,\mu}$ and $p^{H,\mu}$ as the prices this firm would set in the monopoly benchmark when receiving the low and high signal, respectively. If $p^{L,\mu} + \mu < 0.5 < p^{H,\mu} + \mu$ holds, the equilibrium predictions from the baseline model are retained — a large majority of searchers only visit the low-quality firm, because it has data.

This example shows how the presence of data can distort competition. In addition, these distortions reduce the incentives of the low-quality firm to reduce μ, e.g. by conducting product innovation. When this firm has no data, reducing μ to 0 will increase the market share of this firm from $0.5(1 - \rho)$ to 0.5, while the benefits of innovation are much smaller for this firm if it has a data advantage. This is to the detriment of consumers, who would benefit from innovation.

5.2 A right to anonymity

The first way of depriving the firm with data of its advantage is to endow consumers with a right to anonymity. I study the effects of such a policy by integrating this possibility into the baseline framework — now, any searcher can pay a cost $e \geq 0$ before obtaining a price quote at the firm with data to ensure that this firm receives no signal about them, i.e. to become anonymous. Everything else is as in the baseline model. Any searcher thus has three

[24] This is because any searcher will obtain the utility $\max\{v - 0.5, 0\}$ at the high-quality firm, which is strictly larger than the utility she would obtain at the low-quality firm, namely $\max\{v - 0.5(1 + \mu), 0\}$.

possible choices: (i) visit the firm without data, (ii) visit the firm with data and choose to become anonymous, or (iii) visit the firm with data and refrain from becoming anonymous.

The analysis requires a tie-breaking rule. I assume that whenever two of the approaches listed above entail the offering of an identical uniform price, both these choices will be selected by searchers with equal probability.

An equilibrium in this extension consists of a price p^a that the firm with data offers to all consumers who opt out of generating a signal, in addition to the prices (p^L, p^H, p^{nd}) introduced previously. The establishment of a right to anonymity will be inconsequential:

Proposition 7 (Ineffective anonymity)
Consider the baseline model, augmented with the right to anonymity. For any $e \geq 0$, the set of consumers who exercise this right has measure zero.

The intuition behind this result mirrors the insights of Belleflamme & Vergote (2016), who show a similar result in a monopoly setting. Only consumers with comparatively high valuations would ever want to exercise a right to anonymity — low valuation consumers, by contrast, benefit from the possibility that a firm profiles them. In equilibrium, firms will thus offer high prices to consumers who choose to become anonymous, which makes it detrimental for consumers to exercise this right.

5.3 A right to data portability

In this subsection, I integrate a right to data portability into the baseline model. Suppose that every searcher can, before obtaining a price quote anywhere, costlessly copy all the information the firm with data has about her and transfer this to the firm without data. A searcher now has three choices: As before, she can (i) visit the firm with data or (ii) obtain a price offer at the firm without data without porting her data. In addition, she can now (iii) obtain a price offer at the firm without data after porting her data. Formally, porting the data implies that the firm without data will, upon being visited, receive a signal about the consumer's valuation. The distribution of this signal is $Pr(\tilde{v}^H|v)$, just as for the firm with data.

A pure strategy of the firm with data remains a price tuple (p^L, p^H), while a pure strategy of the firm without data is now a vector $(p^{L,nd}, p^{H,nd}, p^{nd})$. This firm offers the price p^{nd} to all consumers who visit it but do not port their data and the prices $p^{L,nd}$ and $p^{H,nd}$ to all consumers who port their data and generate the low and high signal, respectively.

Endowing searchers with the ability to costlessly exercise their right to data portability can eliminate the advantage of the firm with data:

Proposition 8 (Data portability)
If searchers can costlessly exercise their right to data portability, there exists an equilibrium in which all *searchers visit the firm without data.*

This equilibrium has the following form: All searchers visit the firm without data. The firm with data is only visited by its captive consumers and will thus optimally set the monopoly prices, namely $p^{L,M}$ and $p^{H,M}$. Searchers exercise their right to data portability if and only if their valuation is below a cutoff v^t. If their valuation is above v^t, they visit the firm without data but don't port their data. This cutoff v^t solves:

$$v^t = \sup \left\{ v \in [0,1] : Pr(\tilde{v}^H|v)p^{H,nd} + Pr(\tilde{v}^L|v)p^{L,nd} - p^{nd} \leq 0 \right\} \tag{13}$$

Because of this separating behaviour, the prices that the firm without data would offer to consumers who port their data are lower than their monopoly counterparts, i.e. $p^{L,nd} \leq p^{L,M}$ and $p^{H,nd} \leq p^{H,M}$. This is because the average valuation of consumers who exercise their right to data portability is comparatively low. Because $p^{L,nd} \leq p^{L,M}$ and $p^{H,nd} \leq p^{H,M}$, it is optimal for all searchers to visit the firm with data. This is because visiting this firm and porting the data yields a higher expected utility than visiting the firm with data.

Calculating the equilibrium values of v^t shows that v^t is generally below 1. This is crucial, because it implies that the equilibrium prices satisfy $p^{L,nd} < p^L$ and $p^{H,nd} < p^H$, making it *strictly* optimal for searchers to visit the firm without data. This insight establishes that a right to data portability can effectively counteract the competitive effects of data advantages even when exercising this right is costly or generates a less informative signal.

6 Extensions

6.1 Continuous signals

The previous insights all go through even when the firm with data receives continuous signals about consumer valuations, so long as the information is not perfect. I show this in a model which retains all the specifications from the baseline model, with one exception: The firm with data now receives a signal $\tilde{v} = v + \epsilon$ about the valuation of any arriving consumer (v), where the noise term ϵ is uniformly distributed on the interval $[-\bar{\epsilon}, \bar{\epsilon}]$. I assume that $\bar{\epsilon} \in (0, 1/8)$. As $\bar{\epsilon} \to 0$, the signal becomes almost perfect. Just as in the baseline model, I assume that searchers can only visit one firm. I name this framework the *continuous signals framework*. In this and all the following extensions, I restrict attention to equilibria in which firms play pure strategies.

An equilibrium in which firms play pure strategies thus consists of (i) a uniform price (p^{nd}) of the firm without data, (ii) a function $p^d(\tilde{v})$ that defines what price the firm with data would offer after observing the signal $\tilde{v}$, and (iii) the search strategy of searchers, namely $d(v)$. In equilibrium, searchers separate exactly as before:

Lemma 2 (Continuous signals: search)
Consider the continuous signals framework. In equilibrium:

- *Any consumer with $v > 0$ obtains strictly positive expected utility at the firm with data.*
- *There exists a $\bar{v}$ such that all searchers with $v \in (0, \bar{v})$ visit the firm with data and all searchers with $v \in (\bar{v}, 1]$ visit the firm without data.*

Because $\bar{\epsilon} > 0$, the firm with data cannot perfectly price discriminate and any consumer with positive valuation can gain some utility (in expectation) by visiting this firm. Thus, all searchers with $v < p^{nd}$ will optimally visit the firm with data. However, the price distribution at the firm with data becomes strictly less favorable as a consumer's valuation increases. This implies that, as before, searchers with low valuations visit the firm with data and vice versa.

As before, this separating behavior will give rise to a selection effect. In equilibrium, a majority of searchers will thus visit the firm with data:

Proposition 9 (Continuous signals: equilibrium)
Consider the continuous signals framework. In equilibrium, $\bar{v} \geq 0.5(1+\rho)$ must hold and an equilibrium exists.

As before, any equilibrium candidate in which $\bar{v} < 0.5(1 + \rho)$ holds is ruled out by an incompatibility of optimal search and optimal pricing by the firm without data. Optimal behavior by searchers implies that $p^{nd} < \bar{v}$ must hold in equilibrium, since any consumer with positive valuation can attain strictly positive expected utility by visiting the firm with data. A searcher who is indifferent between both firms (i.e. a searcher with valuation $\bar{v}$) must hence receive strictly positive utility at the firm without data. However, the firm without data would optimally set a uniform price p^{nd} weakly above $\bar{v}$ when $\bar{v} < 0.5(1 + \rho)$, which rules out any such equilibria.

The optimization calculus of the firm without data is the same as in the baseline model. Thus, the uniform price it will set is given by $p^{nd,*}(\bar{v})$. The assumption that $\bar{\epsilon} < 1/8$ implies that the firm with data will price according to the following function for any $\bar{v} \geq 0.5(1+\rho)$:

$$p^{d,*}(\tilde{v}) = \begin{cases} 0.5(\tilde{v} + \bar{\epsilon}) & \tilde{v} \in [-\bar{\epsilon}, 3\bar{\epsilon}] \\ (\tilde{v} - \bar{\epsilon}) & \tilde{v} \in [3\bar{\epsilon}, 1 + \bar{\epsilon}] \end{cases} \tag{14}$$

The prices set by the firms respond continuously to changes in $\bar{v}$, which is sufficient to ensure equilibrium existence. It remains to study the properties of these equilibria in some more detail. To that end, I fix $\rho = 0.4$ and visualize the equilibrium values of $\bar{v}$ and p^{nd} for different levels of $\bar{\epsilon}$ (which are plotted on the x-axis) in the following graph:

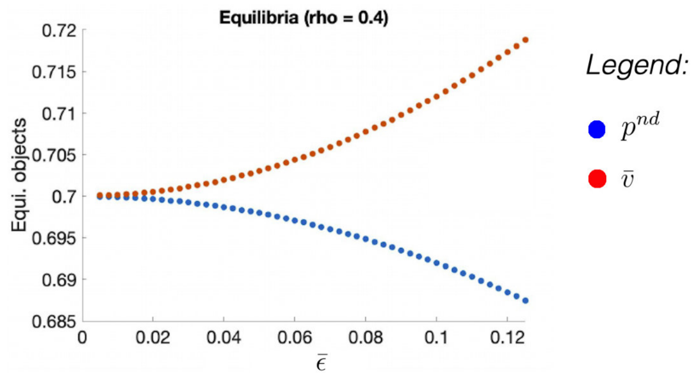

Figure 5: Equilibria under continuous signals

As $\bar{\epsilon} \to 0$, the cutoff $\bar{v}$ and the price p^{nd} both converge to $0.5(1+\rho)$. To see why, suppose that $\bar{\epsilon} \approx 0$ and that searchers visit firms according to a cutoff rule with $\bar{v} \approx 0.5(1+\rho)$. The firm without data will optimally set the price $p^{nd,*}(\bar{v})$, which is approximately equal to $0.5(1+\rho)$ for this search rule. Because the firm with data is able to almost perfectly price discriminate, every consumer will attain close to zero expected utility by visiting this firm. While consumers with $v < p^{nd} \approx 0.5(1+\rho)$ would still prefer to visit the firm with data, almost all consumers with valuation above p^{nd} would prefer to visit the firm without data, because the utility they attain there is linearly rising in v. Thus, the search behavior represented by the cutoff $\bar{v} \approx 0.5(1+\rho)$ is optimal, because $p^{nd} \approx 0.5(1+\rho)$.

6.2 Endowing both firms with data

In this section, I study a framework in which both firms receive binary signals about the valuation of any consumer who visits. I specify that there is one firm that has access to better data than the other. I define the two firms as the *firm with better data* and the *firm with worse data*. Everything else is as in the baseline model and all consumers can only visit one firm. I label the resulting framework the *dispersed data framework*.

The probability that a consumer with valuation v generates the high signal at the firm with better data is $Pr^{HB}(v)$ and $Pr^{HW}(v)$ at the firm with worse data. In the dispersed data framework, an equilibrium consists of a quadruple of prices $(p^{LB}, p^{HB}, p^{LW}, p^{HW})$ and the search strategy of searchers. The prices that the firm with better data offers to consumers that generate the low and high signal, respectively, are (p^{LB}, p^{HB}). The prices the firm with worse data offers in the respective information sets are (p^{LW}, p^{HW}).

To illustrate differences in signal precision, consider a monopolist with access to a signal with distribution $Pr^{HB}(v)$. I define the prices this firm would set after the low and high signal as $p^{LB,M}$ and $p^{HB,M}$, respectively. Analogously, the prices set by a monopolist who receives a signal with distribution $Pr^{HW}(v)$ in the respective information sets are defined as $p^{LW,M}$ and $p^{HW,M}$. In the analysis, I assume that the signal distributions are well behaved and that the firm with better data receives a more precise signal in the following sense:

Assumption 6 *Both functions $Pr^{HB}(v)$ and $Pr^{HW}(v)$ are strictly increasing in v, continuous, and map into $(0,1)$ for any v. The signal probability functions are such that:*

- *For any $v < 0.5$, $Pr^{HB}(v) < Pr^{HW}(v)$. For any $v > 0.5$, $Pr^{HB}(v) > Pr^{HW}(v)$.*
- *The prices that firms would set when all consumers randomize between firms satisfy the ordering $p^{LB,M} < p^{LW,M} < p^{HW,M} < p^{HB,M}$.*

In words, the signal which the firm with better data receives implies a higher chance of correctly recognizing whether a consumer's valuation is in the upper or lower half of the valuation interval. Moreover, the signal of the firm with better data is more informative in the sense that, when consumers randomly arrive at firms, this firm sets a lower price to consumers who generate the low signal and vice versa.

Moreover, I impose a tie-breaking rule concerning searchers with valuation below the lowest equilibrium price, which I call $p^{min} := \min\{p^{LW}, p^{HW}, p^{LB}, p^{HB}\}$.

Assumption 7 *Searchers with $v \leq p^{min}$ visit the firm that offers p^{min} with higher probability.*

The richness of the pricing possibilities enables the potential existence of equilibria with intractable search behaviour. To facilitate the analysis, I restrict attention to the following "simple" category of equilibria.

Definition 1 *An equilibrium is simple if and only if (i) all searchers visit each firm with probability 0.5 <u>or</u> (ii) there exists some $\bar{v}$ s.t. all searchers with valuation above $\bar{v}$ visit a particular firm and all searchers with valuation below $\bar{v}$ visit the other firm.*

The definition of a simple equilibrium does not impose restrictions on *which firm* consumers on either side of a cutoff $\bar{v}$ visit in an equilibrium where their strategy is a cutoff rule. Instead, I show that any simple equilibrium retains the structure of previous equilibria.[25] Moreover, the equilibrium $\bar{v}$ is bounded from below, as before.

[25] If both firms receive a signal with the same probability distribution, there is a unique simple equilibrium in which both firms follow the same pricing strategy and all searchers randomize between both firms.

Lemma 3 (Dispersed data: equilibrium characterization)
Consider the dispersed data framework. In a simple equilibrium:

- *The ordering* $p^{LB} \leq p^{LW}$ *holds. There exists a* $\bar{v}$ *such that searchers visit the firm with worse data if* $v > \bar{v}$ *and the firm with better data if* $v < \bar{v}$.
- $\bar{v} \geq \bar{v}^{LW}$ *must hold, where* $\bar{v}^{kW} \in (0,1)$ *solves the following for either* $k \in \{L, H\}$:

$$\rho \int_{\bar{v}^{kW}}^{1} Pr^{kW}(v)dv + 0.5(1-\rho)\left[\int_{\bar{v}^{kW}}^{1} Pr^{kW}(v)dv - \bar{v}^{kW} Pr^{kW}(\bar{v}^{kW})\right] = 0 \tag{15}$$

Low-valuation consumers prefer to visit the firm with better data, because they are more likely to generate the low signal, i.e. to be correctly identified, at this firm. As a result, consumers with low (high) valuations can expect a more favorable price distribution at the firm with better data (firm with worse data). The interpretation of $\bar{v}^{LW}$ is the same as in the baseline model:[26] For any $\bar{v} < \bar{v}^{LW}$, the firm with worse data would optimally set a low signal price (p^{LW}) that is weakly above $\bar{v}$. However, searchers with valuation just above p^{LW} would never visit the firm with worse data in a simple equilibrium, because the firm with better data will always offer the lowest equilibrium price. This means that $\bar{v} \leq p^{LW}$ cannot hold in a simple equilibrium, which rules out equilibria in which $\bar{v} < \bar{v}^{LW}$.

It remains to establish the existence of a simple equilibrium. For a given $\bar{v}$, the prices of the firm with better data need to maximize the following objective functions for the corresponding signal $\tilde{v}^k \in \{\tilde{v}^L, \tilde{v}^H\}$ in any such equilibrium:

$$\Pi^{kB}(p_j; \bar{v}) := p_j \Bigg[\underbrace{\rho \int_0^{\bar{v}} Pr^{kB}(v)\mathbb{1}[p_j \leq v]dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho)\int_0^1 Pr^{kB}(v)\mathbb{1}[p_j \leq v]dv}_{\text{captive consumer demand}}\Bigg] \tag{16}$$

The firm with worse data maximizes the following objective, given the signal $\tilde{v}^k \in \{\tilde{v}^L, \tilde{v}^H\}$:

$$\Pi^{kW}(p_j; \bar{v}) := p_j \Bigg[\underbrace{\rho \int_{\bar{v}}^{1} Pr^{kW}(v)\mathbb{1}[p_j \leq v]dv}_{\text{searcher demand}} + \underbrace{0.5(1-\rho)\int_0^1 Pr^{kW}(v)\mathbb{1}[p_j \leq v]dv}_{\text{captive consumer demand}}\Bigg] \tag{17}$$

I define the optimal prices of the firm with better data for a given $\bar{v}$ as $p^{LB,*}(\bar{v}) := arg\max_{p_j \in [0,1]} \Pi^{LB}(p_j; \bar{v})$ and $p^{HB,*}(\bar{v}) := arg\max_{p_j \in [0,1]} \Pi^{HB}(p_j; \bar{v})$. Analogously, I define the optimal prices of the firm with worse data as $p^{LW,*}(\bar{v}) := arg\max_{p_j \in [0,1]} \Pi^{LW}(p_j; \bar{v})$ and $p^{HW,*}(\bar{v}) := arg\max_{p_j \in [0,1]} \Pi^{HW}(p_j; \bar{v})$. In equilibrium, the search behavior of searchers will, as before, be

[26] In fact, this cutoff becomes $\bar{v}^{LW} = 0.5 + 0.5\rho$ when $Pr^{LW}(v) = Pr^{HW}(v)$, i.e. when the firm with worse data receives no informative signal, as in the baseline model.

determined by the expected prices they can anticipate at the two firms. For firm $j \in \{W, B\}$ and a fixed $\bar{v}$, these expected prices (conditional on v) are given by:

$$EP^j(v;\bar{v}) = Pr^{Lj}(v)p^{Lj,*}(\bar{v}) + Pr^{Hj}(v)p^{Hj,*}(\bar{v}) \tag{18}$$

For which searchers the expected price is lower at the firm with better data (given the equilibrium level of $\bar{v}$) is tracked by the following object:

$$\hat{v}^X(\bar{v}) = \sup\left\{v \in [0,1] : EP^B(v;\bar{v}) - EP^W(v;\bar{v}) < 0\right\} \tag{19}$$

Now, I define an assumption that ensures the existence of a viable candidate for a simple equilibrium. Afterwards, I establish when this candidate constitutes an equilibrium.

Assumption 8 *Define $\bar{v}^D = \max\{\bar{v}^{LW}, p^{HB,M}\}$. Suppose that $EP^B(\bar{v}^D;\bar{v}^D) < EP^W(\bar{v}^D;\bar{v}^D)$.*

Remark 3 *Assumption 8 is satisfied for any linear signal distribution.*

Proposition 10 (Dispersed data: equilibrium existence)
Consider the dispersed data framework. Under assumption 8, the following equation has a solution $\bar{v}^* \in [\bar{v}^{LW}, 1]$:

$$\bar{v}^* = \hat{v}^X(\bar{v}^*) \tag{20}$$

The combination $(p^{LB}(\bar{v}^*), p^{HB}(\bar{v}^*), p^{LW}(\bar{v}^*), p^{HW}(\bar{v}^*), \bar{v}^*)$ is an equilibrium if, given $(p^{LB}(\bar{v}^*), p^{HB}(\bar{v}^*), p^{LW}(\bar{v}^*), p^{HW}(\bar{v}^*))$, it is weakly optimal for searchers to visit the firm with better data if and only if $v \leq \bar{v}^*$.

In a hypothetical equilibrium of the above form, all prices of the firms are optimal by construction. Imposing optimality of the postulated search behavior is required, because the fact that $\bar{v} = \hat{v}^X(\bar{v})$ holds does not rule out the possibility that some consumers with $v < \bar{v}$ optimally visit the firm with worse data. This is because the search preferences of searchers have kinks at the equilibrium prices, which means that it may not necessarily be optimal for them to visit the firm where they receive the lower expected price.

However, numerical analysis shows that it is indeed optimal for searchers to visit the firm where they receive the lower expected price in equilibrium candidates of the above form, establishing that these combinations constitute equilibria. In appendix E.1, I study linear signal distributions as defined in equation (3), where the precision of the signal the firm with better data receives is α_b and the precision of the signal that its rival receives is α_w, with $\alpha_w < \alpha_b$. I consider $\rho \in \{0.05, 0.35, 0.65, 0.95\}$, $\alpha_w \in [0, 0.49]$ and $\alpha_b \in [0.5, 0.99]$ (with 25 grid points each). For different combinations of ρ, α_w, and α_b, I calculate the solution to equation (20). Given the implied prices, I then check whether it is optimal for all searchers with $v < \bar{v}^*$ to visit the firm with better data and vice versa for searchers with $v > \bar{v}^*$. I show that this requirement is met, i.e. that said combination of prices and $\bar{v}$ constitutes an equilibrium, for any of the parameter combinations listed above.

Finally, I visualize the comparative statics results of increases in α_w for different parameter combinations in the following figures:

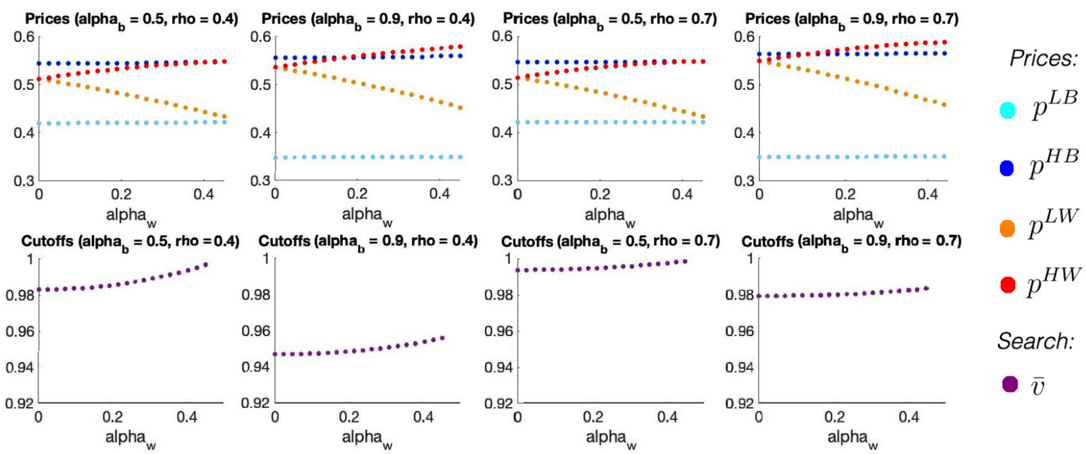

Figure 6: Equilibrium objects in the dispersed data framework

Summing up, the key prediction from the baseline model also holds true in the dispersed data framework when restricting attention to simple equilibria. Any simple equilibrium retains the property that $\bar{v}$ is bounded from below. The numerical simulations indicate that a simple equilibrium always exists and that $\bar{v} \to 1$ as $\rho \to 1$.

6.3 Quality differentiation

In this section, I integrate quality differentiation into the analysis by combining the previous search setup with the model of Mussa & Rosen (1978). The consumer's type is now denoted by $\theta \sim U[0, 1]$. The firms can offer different quality levels $q \in [0, 1]$. When paying the price p for a good with quality q, a consumer's net utility is:

$$u(q, p; \theta) = \theta q - p \tag{21}$$

There are two active firms, namely *the firm with data* and *the firm without data*. For any consumer who arrives, the firm with data receives a signal $\tilde{\theta} \in \{\tilde{\theta}^L, \tilde{\theta}^H\}$ about the consumer's type. The probability distribution of this signal is denoted by $Pr(\tilde{\theta} = \tilde{\theta}^H | \theta) := Pr^H(\theta)$, where $Pr^L(\theta) := 1 - Pr^H(\theta)$. The firm without data receives no information about any consumer. The provision of any quality level is costless. As in the baseline analysis, there are searchers and captive consumers, with shares $\rho \in (0,1)$ and $(1-\rho)$. Any consumer can only visit one firm. I label this model the *quality differentiation framework*.

By the revelation principle, it is without loss to restrict the strategy space of the firms to direct mechanisms. Thus, an equilibrium in this game consists of the following objects: (i) the search strategy of searchers, (ii) a quality-price menu $(q^{nd}(\theta), t^{nd}(\theta))$ offered by the firm without data, and two quality price menus $(q^L(\theta), t^L(\theta))$ and $(q^H(\theta), t^H(\theta))$ that the firm with data offers to consumers who generate the low signal and the high signal, respectively. I restrict the strategy space of the firms to menus in which the mapping from messages into qualities is a measurable function. I further restrict attention to *simple equilibria*:

Definition 2 *An equilibrium in the quality differentiation framework is simple if and only if (i) all searchers visit either firm with probability* 0.5 *<u>or</u> (ii) there exists a cutoff $\bar{\theta}$ such that all searchers with $\theta < \bar{\theta}$ visit a given firm and all searchers with $\theta > \bar{\theta}$ visit the other firm.*

Moreover, I impose some tie-breaking rules. I define the infimum of types that receive quality at the firm with data and at the firm without data as $\underline{\theta}^d$ and $\underline{\theta}^{nd}$, respectively.

Assumption 9 *If firms offer identical quality-price menus, searchers visit either firm with equal probability. For any other combination of menus, searchers with $\theta \leq \min\{\underline{\theta}^d, \underline{\theta}^{nd}\}$ visit the firm j with the lower $\underline{\theta}^j$. If $\underline{\theta}^{nd} = \underline{\theta}^d$, searchers with $\theta \leq \min\{\underline{\theta}^d, \underline{\theta}^{nd}\}$ visit the same firm.*

The main technical challenge in the following analysis stems from the fact that the density of types that arrive at either firm is not continuous — it has a jump discontinuity at $\bar{\theta}$. However, because this stark difference in the consumers' search choices can only occur at precisely one level of θ in a simple equilibrium, the types of consumers that arrive at the firms are still absolutely continuous random variables and admit a well-defined density.

To express these densities, I define the probability that a consumer arrives at the firm without data in equilibrium as $Pr(I^{nd})$ and the probability that a consumer arrives at the firm with data and generates the signal $\tilde{\theta}^k$ as $Pr(I^k)$. Because each firm has captive consumers, these probabilities are all strictly positive, i.e. all information sets of both firms are on the equilibrium path. To characterize the search behavior of consumers in a simple equilibrium, I define $g^H \in \{0, 0.5, 1\}$ and $g^L \in \{0, 0.5, 1\}$ as the probabilities with which a searcher with type $\theta > \bar{\theta}$ and $\theta < \bar{\theta}$ visits the firm with data, respectively. If all searchers visit either firm with equal probability, this is captured by setting $g^L = g^H = 0.5$ and choosing any $\bar{\theta}$. Defining $g := (g^L, g^H)$, the type of a consumer who visits the firm without data has the following probability density:

$$f^{nd}(\theta;\bar{\theta},g) = \begin{cases} (1/Pr(I^{nd}))(\rho(1-g^L)+0.5(1-\rho)) & \theta < \bar{\theta} \\ (1/Pr(I^{nd}))(\rho(1-g^H)+0.5(1-\rho)) & \theta > \bar{\theta} \end{cases} \tag{22}$$

The type of a consumer who visits the firm with data and generates the signal $\tilde{\theta}^k$ has an analogously defined probability density, which I call $f^k(\theta;\bar{\theta},g)$.

Given these densities, we can construct the virtual valuation functions. The virtual valuation function of consumers who visit the firm without data, which I call $J^{nd}(\theta;\bar{\theta},g)$, is $J^{nd}(\theta;\bar{\theta},g) = \theta - (1-F^{nd}(\theta;\bar{\theta},g))/f^{nd}(\theta;\bar{\theta},g)$. The virtual valuation function of consumers who visit the firm with data and generate the signal $\tilde{\theta}^k$ is $J^k(\theta;\bar{\theta},g) = \theta - (1 - F^k(\theta;\bar{\theta},g))/f^k(\theta;\bar{\theta},g)$. Note that $F^{nd}(\theta;\bar{\theta},g)$ and $F^k(\theta;\bar{\theta},g)$ are the cumulative density functions that accompany $f^{nd}(\theta;\bar{\theta},g)$ and $f^k(\theta;\bar{\theta},g)$, respectively. Moving forward, I impose the following assumptions:

Assumption 10 *The signal distribution $Pr^H(\theta)$ is continuous, strictly increasing, and maps into $(0,1)$ for any $\theta \in [0,1]$.*

Under these assumptions, the insights of Milgrom & Segal (2002) apply and the expected utility a consumer with type θ attains in an implementable mechanism can be expressed using the integrability condition. Thus, the expected revenue the firm without data obtains in an implementable mechanism is:

$$R^{nd}(q^{nd}(\theta);\bar{\theta},g) = -U^{nd}(0) + \int_0^1 q^{nd}(\theta)J^{nd}(\theta;\bar{\theta},g)f^{nd}(\theta;\bar{\theta},g)d\theta \tag{23}$$

The expected revenue the firm with data obtains from a consumer that generates $\tilde{\theta}^k$ has an analogous form. I have defined $U^{nd}(0)$ as the utility the lowest type would attain in a mechanism set by the firm without data. The set of consumer types for which the virtual valuations are positive are partially characterized by the cutoffs $\hat{\theta}^{nd}, \hat{\theta}^L$, and $\hat{\theta}^H$, which are defined as follows:

$$\hat{\theta}^k = \inf\left\{\theta : J^k(\theta;\bar{\theta},g) > 0\right\} \quad ; \quad \hat{\theta}^{nd} = \inf\left\{\theta : J^{nd}(\theta;\bar{\theta},g) > 0\right\} \tag{24}$$

Note that the virtual valuation functions can jump down at $\bar{\theta}$, which means that the virtual valuations are not necessarily positive for all θ above said cutoffs. In addition, while the functions $J^{nd}(\theta;\bar{\theta},g)$ and $J^H(\theta;\bar{\theta},g)$ are both piecewise strictly increasing in θ by construction, the low signal virtual valuation function $J^L(\theta;\bar{\theta},g)$ may be non-monotonic. To deal with the former problem in the equilibrium analysis, I set up the following assumption:

Assumption 11 *Fix $g_L = 0$ and $g_H = 1$. For any $\bar{\theta} \leq 0.5$, $J^L(\theta; \bar{\theta}, g) < 0 \ \forall \theta \leq \bar{\theta}$.*

Remark 4 *For any linear $Pr^H(v)$, assumption 11 is satisfied if $\rho \geq 0.34$.*

This assumption rules out equilibria in which searchers separate in a different way than previously, i.e. equilibria in which searchers with low θ visit the firm without data. Such equilibria could only be sustained if the firm without data implements an ironing mechanism, in which it starts providing quality at lower types than the firm with data (formally, $\hat{\theta}^{nd} < \hat{\theta}^L$). This assumption rules out such equilibria, given that any such equilibrium must feature $\hat{\theta}^L < \bar{\theta}$ and $\bar{\theta} \leq 0.5$, which is made impossible under said assumption. Moreover, there generally exists no equilibrium in which all searchers randomize between the firms. These notions are formalized in the following lemma:

Lemma 4 (Quality differentiation: search)
Consider the quality differentiation framework:

- *When all searchers visit firms randomly, the cutoffs satisfy $\hat{\theta}^L < \hat{\theta}^{nd} < \hat{\theta}^H$.*
- *Suppose assumption 11 holds as well. In a simple equilibrium, there exists a $\bar{\theta}$ such that searchers visit the firm with data if $\theta < \bar{\theta}$ and the firm without data if $\theta > \bar{\theta}$.*

The first result follows from the fact that consumers who generate the low signal have a distribution of types θ with less mass at high types and vice versa. This shifts up the distribution of virtual valuations. Intuitively, a firm would be more willing to offer positive quality to a consumer with a given θ when observing the low signal (rather than the high signal or no signal), because the mass of consumers with higher types for whom this decision would incur a revenue loss becomes smaller. The second result holds because the firm without data cannot attract low type consumers with an ironing mechanism in equilibrium.

The main result from the baseline analysis is retained in any simple equilibrium. This is formalized in the following proposition, together with the accompanying assumptions:

Assumption 12 *Fix $g^L = 1$ and $g^H = 0$.*

- *The function $J^L(\theta; \bar{\theta}, g)$ is piecewise strictly increasing in θ for $\theta \in [0, \bar{\theta})$ and $\theta \in (\bar{\theta}, 1]$.*
- *For any $\bar{\theta} \geq 0.5(1+\rho)$, $\lim_{\theta \downarrow \bar{\theta}} J^k(\theta; \bar{\theta}, g) > 0$ holds for both $k \in \{L, H\}$.*

Remark 5 *Assumption 12 is satisfied for any linear signal distribution.*

This assumption guarantees that, in equilibrium, all functions $J^x(\theta; \bar{\theta}, g)$ (with $x \in \{nd, L, H\}$) will be strictly negative if $\theta < \hat{\theta}^x$ and strictly positive if $\theta > \hat{\theta}^x$. Thus, the optimal $q^x(\theta)$ assigns quality 1 to all types above $\hat{\theta}^x$ and quality 0 to all types below this threshold. Such an equilibrium exists and retains the key property from the baseline analysis:

Proposition 11 (Quality differentiation: equilibrium)
Consider the quality differentiation framework. Under assumptions 11 and 12, $\bar{\theta} \geq 0.5(1+\rho)$ *must hold in any simple equilibrium, and a simple equilibrium always exists.*

Any value $\bar{\theta} < 0.5(1+\rho)$ cannot constitute an equilibrium, since $\bar{\theta} \leq \hat{\theta}^{nd}$ would hold for any such $\bar{\theta} < 0.5(1+\rho)$. This is not consistent with optimal consumer search behavior — because $\hat{\theta}^{L} < \hat{\theta}^{nd}$, searchers with θ just above $\hat{\theta}^{nd}$ would strictly prefer to visit the firm with data, but visit the firm without data in the supposed equilibrium. Thus, $\bar{\theta} \geq 0.5(1+\rho)$ must hold in a simple equilibrium, which exactly replicates the result from the baseline analysis. Such a simple equilibrium always exists under the stated assumptions. Thus, the main equilibrium result from the baseline model is retained in this extension, albeit under slightly stronger restrictions on ρ, as expressed in remark 4.

7 Conclusion

I have analyzed the relationship between data and market power in a duopoly model of directed search and personalized pricing. One of the firms in the market has a data advantage — in the baseline model, this firm receives a signal about the valuation of any consumer who visits it, while its rival receives no such information. Consumers can costlessly visit one firm, but have to pay a search cost to visit a second firm after the first. There are two groups of consumers, namely *captive consumers* and *searchers*. Searchers have equal valuation for the good of both firms and, based on their valuation, optimally choose which firms to visit.

Directed consumer search strongly facilitates the transmission of data advantages into competitive advantages. In equilibrium, a large majority of searchers only visit the firm with data. The firm without data is just visited by searchers with very high valuations. As the share of searchers goes to 1, so does the market share of the firm with data, independent of the extent of the data advantage.

While I have considered a framework in which data is only used to price discriminate, the insights apply more generally. Consider, for instance, an insurance market: Consumers with low risk benefit if a firm has information about their traits, because this would translate into more favorable contract terms. Thus, these consumers would all prefer to visit a firm with a data advantage, which improves the overall risk profile of consumers who visit this firm. The generally better contract terms this firm can offer as a result would, in turn, attract even more consumers, mirroring the unraveling channels present in my model.

A Proofs - section 4

A.1 Proof of lemma 1

Part 1: In an equilibrium in which firms play pure strategies, $p^H < p^L$ cannot hold.

To see this, note that the conditional valuation distribution of consumers that generate $\tilde{v}^H$ at the firm with data strictly hazard rate dominates the conditional distribution for consumers that generate $\tilde{v}^L$ (both of which are Lipschitz continuous) at the firm with data.

Consumers visit the firm with data with probability $g(v) = \rho d(v) + 0.5(1+\rho)$, which is measurable. I first define the conditional cumulative density functions:

$$F^k(x) = Pr(v \leq x|\tilde{v}^k) = \frac{1}{Pr(\tilde{v}^k)}\int_0^1 Pr(v \leq x \wedge \tilde{v}^k|v)g(v)dv = \frac{1}{Pr(\tilde{v}^k)}\int_0^x Pr(\tilde{v}^k|v)g(v)dv$$

We can define a probability density function $f^k(x) = (1/Pr(\tilde{v}^k))Pr(\tilde{v}^k|x)g(x)$ corresponding to $F^k(x)$. The hazard rates for these distributions are $h^k(x) = \frac{f^k(x)}{1-F^k(x)}$. Thus, $h^H(x) < h^L(x)$ $\forall x \in (0,1)$, i.e. $F^H(x)$ strictly hazard ratio dominates $F^L(x)$, since:

$$h^H(x) < h^L(x) \iff \int_x^1 \underbrace{(Pr(\tilde{v}^L|v)/Pr(\tilde{v}^l|x))}_{<1} g(v)dv < \int_x^1 \underbrace{(Pr(\tilde{v}^H|v)/Pr(\tilde{v}^H|x))}_{>1} g(v)dv$$

Since p^L and p^H are available to set after any signal, there must be no profitable deviation from p^L to p^H after $\tilde{v}^L$ and no profitable deviation from p^H to p^L after $\tilde{v}^H$. But given the hazard ratio ordering established above, it will either be profitable to deviate from p^L to p^H when observing $\tilde{v}^L$ or vice versa, a contradiction.

Part 2: In an equilibrium in which firms play pure strategies, $p^L = p^H$ cannot hold.

If $p^L = p^H > p^{nd}$, or $p^L = p^H < p^{nd}$, all consumers with valuation above the lowest equilibrium price visit a given firm. This will imply that there is either a deviation from p^L or p^{nd}. If $p^{nd} < p^L = p^H$, for instance, all searchers with $v \geq p^{nd}$ visit the firm without data. Thus, $p^L = p^{L,M}$ and $p^H = p^{H,M}$ must hold, which contradicts $p^L < p^H$. Thus, suppose that $p^L = p^H = p^{nd}$. But then, all searchers randomize, which means that the distribution of valuations that visit either firm is the same. As a result, the prices $p^L < p^{nd} < p^H$ would be optimal, a contradiction.

Part 3: In an equilibrium in which firms play pure strategies, $p^L < p^{nd} < p^H$ must hold.

Previous arguments establish that $p^L < p^H$ must hold. Suppose, for a contradiction, that $p^{nd} \leq p^L$. Then, all searchers with $v > p^{nd}$ visit the firm without data, since $p^L < p^H$. For $p_j \geq p^{nd}$, the profit function of the firm without data thus becomes:

$$\Pi^{nd}(p_j) = p_j \left[\rho \int_{p_j}^{1} (1)dv + 0.5(1-\rho) \int_{p_j}^{1} (1)dv \right]$$

We know that the derivative of this is strictly positive when $p_j < 0.5$. Thus, the equilibrium price p^{nd} must satisfy $p^{nd} \geq 0.5$ to avoid the existence of a profitable upward deviation.

This implies that $p^L \geq p^{nd} \geq 0.5$. In equilibrium, no searcher with $v > p^L$ arrives at the firm with data - thus, equilibrium profits at the firm with data are only attained from captive consumers. Because $p^L \geq 0.5$, the low signal profits from captive consumers can be strictly increased by a downward deviation, a contradiction.

Suppose instead that $p^H \leq p^{nd}$. Since $p^L < p^H \leq p^{nd}$ thus holds, all searchers with $v > p^L$ surely visit the firm with data in equilibrium. For prices in an open ball around p^H, the high signal profits of the firm with data are:

$$\Pi^{H}(p_j) = p_j \left[\rho \int_{p_j}^{1} Pr^H(v)dv + 0.5(1-\rho) \int_{p_j}^{1} Pr^H(v)dv \right]$$

Thus, the price p^H must set a corresponding first-order condition equal to 0 and hence $p^H = p^{H,M} > 0.5$ must hold. Thus, $p^{nd} > 0.5$ also holds. In equilibrium, no searcher with $v > p^{nd}$ arrives at the firm with data - this firm only obtains profits from captive consumers. By deviating downward towards 0.5, this firm strictly raises the profits it makes from captive consumers, while weakly raising the profits it may obtain from searchers, a contradiction.

Part 4: Existence of the cutoff $\bar{v}$.

In equilibrium, we must have $p^L < p^{nd} < p^H$. We know that all consumers with $v \in (p^L, p^{nd}]$ visit the firm with data, since their utility at the firm without data is 0.

Consider consumers with $v \in (p^{nd}, p^H]$, for whom the preference for the firm with data is: $P^D(v) = \left[Pr^L(v)(v - p^L)\right] - (v - p^{nd})$. This is strictly falling in v.

Now consider consumers with $v \in [p^H, 1]$. For them, the preference for the firm with data is $P^D(v) = \left[Pr^L(v)(v - p^L) + Pr^H(v)(v - p^H)\right] - (v - p^{nd})$, which is continuous at p^H. For consumers with $v \in [p^H, 1]$, the preference for the firm with data is strictly falling in v.

Thus, there must be a unique $\bar{v}$, because all searchers with $v \in (p^L, p^{nd}]$ strictly prefer the firm with data and the preference for this firm is strictly decreasing in v thereafter.

A.2 Proof of proposition 1

In equilibrium, the inequality $p^{nd} < \bar{v}$ must hold. Suppose, for a contradiction, that $p^{nd} \geq \bar{v}$ holds. Because $p^L < p^{nd}$, consumers with $v \in (p^L, p^{nd}]$ strictly prefer to visit the firm with data. Since the expected utilities at the two firms are continuous in v, consumers with v just above p^{nd} will also prefer to visit the firm with data. Thus, we have a contradiction, since these consumers visit the firm without data in the supposed equilibrium.

Now suppose, for a contradiction, that we have an equilibrium in which $\bar{v} < 0.5(1 + \rho)$. In equilibrium, the objective function of the firm with data for prices $p_j \in (p^L, \bar{v})$ is:

$$\Pi^{nd}(p_j) = p_j \left[\rho \int_{\bar{v}}^{1} (1) d\nu + \frac{1-\rho}{2} \int_{p_j}^{1} d\nu\right]$$

We consider any $p^{nd} < \bar{v}$ and any $\bar{v} < 0.5(1 + \rho)$. Here, this derivative at p^{nd} satisfies:

$$\left.\frac{\partial \Pi^{nd}(p_j)}{\partial p_j}\right|_{p^{nd} \in (p^L, \bar{v})} = \rho(1 - \bar{v}) + 0.5(1 - \rho) - p^{nd}(1 - \rho) > \rho(1 - \bar{v}) + 0.5(1 - \rho) - \bar{v}(1 - \rho) =$$

$$\rho - \rho\bar{v} + 0.5(1 - \rho) - \bar{v} + \rho\bar{v} = 0.5(1 + \rho) - \bar{v} > 0$$

This is a contradiction - there would surely exist a profitable upward deviation.

A.3 Proof of proposition 2

Part 1: Preliminiaries - definition and properties of $\bar{v}^{HC}$.

I define a cutoff $\bar{v}^{HC}$ that solves: $\max_{p_j \leq \bar{v}^{HC}} \Pi^H(p_j; \bar{v}^{HC}) = 0.5(1 - \rho)\Pi^{H,M}(p^{H,M})$.

For any $\bar{v} > \bar{v}^{HC}$, $\max_{p_j \leq \bar{v}} \Pi^H(p_j; \bar{v}) > \max_{p_j \leq \bar{v}^{HC}} \Pi^H(p_j; \bar{v}^{HC})$, because the firm could set the old $arg \max_{p_j \leq \bar{v}} \Pi^H(p_j; \bar{v})$, but obtain strictly higher demand there via increases in $\bar{v}$.

(i) $\bar{v}^{HC} < p^{H,M}$.

Suppose $\bar{v}^{HC} = p^{H,M}$. Then, the left derivative of $\Pi^H(p_j; \bar{v}^{HC})$ at $p_j = \bar{v}^{HC} = p^{H,M}$ would be strictly negative - thus, profits could be strictly increased by a downward movement from $p_j = \bar{v}^{HC}$ and we could not have said equality.

Suppose $\bar{v}^{HC} > p^{H,M}$. Then, setting $p_j = p^{H,M} < \bar{v}^{HC}$ is available within $[0, \bar{v}^{HC}]$ - this would yield strictly higher profits than the scaled monopoly profits from captive consumers only, so said equality could not be satisfied.

(ii) If $\bar{v} \geq \bar{v}^{HC}$, i.e. $\max_{p_j \leq \bar{v}} \Pi^H(p_j; \bar{v}) \geq 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$, the locally optimal price $p_j \leq \bar{v}$ must be strictly below $\bar{v}$ and thus solve the FOC.

If it were equal to $\bar{v}$, the left derivative there would need to be weakly positive, i.e.

$$\left.\frac{\partial \Pi^H(p_j; \bar{v})}{\partial p_j}\right|_{p_j \uparrow \bar{v}} = 0.5(1-\rho)\int_{\bar{v}}^{1} Pr^k(v)dv - \bar{v}\big[0.5(1+\rho)Pr^k(\bar{v})\big] \geq 0$$

In order for this derivative to be weakly positive, we know that $\bar{v} < p^{H,M}$ must hold, since at $\bar{v} \geq p^{H,M}$, this derivative would be strictly negative. But thus, we have a contradiction, as the (strictly higher) and thus positive right derivative at $p_j = \bar{v}$ would imply that setting $p_j = p^{H,M}$ would yield strictly higher profits, a contradiction to the assumed case.

We distinguish two cases in the proof, namely (a) $\bar{v}^{HC} \leq 0.5(1+\rho)$ and (b) $0.5(1+\rho) < \bar{v}^{HC}$.

Part 2: Suppose (a) $\bar{v}^{HC} \leq 0.5(1+\rho)$. For any $\bar{v} \geq 0.5(1+\rho)$, $p^{L,*}(\bar{v}) < p^{H,*}(\bar{v}) < \bar{v}$.

The price $p^{H,*}(\bar{v})$ will lie below $\bar{v}$. This is because $\bar{v} \geq \bar{v}^{HC}$. Moreover, $p^{L,*}(\bar{v}) < \bar{v}$ holds for any $\bar{v} \geq 0.5(1+\rho)$ since $p^{L,M} < 0.5$. Thus, these prices must solve:

$$\rho \int_{p^{k,*}}^{\bar{v}} \big(Pr^k(v)/Pr^k(p^{k,*})\big)dv + 0.5(1-\rho)\int_{p^{k,*}}^{1} \big(Pr^k(v)/Pr^k(p^{k,*})\big)dv = p^{k,*}\big[0.5(1+\rho)\big] \quad (25)$$

Since $Pr^H(v)$ is strictly increasing, the following holds for all $\forall v > p^{k,*}$: $\big(Pr^H(v)/Pr^H(p^{k,*})\big) > 1 > \big(Pr^L(v)/Pr^L(p^{k,*})\big)$. This implies that $p^{L,*}(\bar{v}) < p^{H,*}(\bar{v})$ must hold, because otherwise the respective two first-order conditions could not be jointly satisfied.

Part 3: Suppose (a) $\bar{v}^{HC} \leq 0.5(1+\rho)$. There exists a $\bar{v} \in [0.5(1+\rho), 1]$ such that $\bar{v} = \hat{v}^B(\bar{v})$.

At $\bar{v} = 0.5(1+\rho)$, we have $\hat{v}^B(0.5(1+\rho)) = 1$. To see this, note that $p^{nd,*}(0.5(1+\rho)) = 0.5(1+\rho)$. In addition, we have $p^{H,*}(0.5(1+\rho)) < 0.5(1+\rho)$. Thus, both prices at the firm with data are strictly below p^{nd}, i.e. $\hat{v}^B(0.5(1+\rho)) = 1$.

I define $D(v, \bar{v}) := Pr^H(v)p^{H,*}(\bar{v}) + Pr^L(v)p^{L,*}(\bar{v}) - p^{nd,*}(\bar{v})$. Moreover, we can define $\bar{v}' \in (0.5(1+\rho), 1]$ at a cutoff such that:

$$D(1, \bar{v}') := Pr^H(1)p^{H,*}(\bar{v}') + Pr^L(1)p^{L,*}(\bar{v}') - p^{nd,*}(\bar{v}') = 0$$

At most one such such $\bar{v}'$ exists because $p^{H,*}(\bar{v})$ and $p^{L,*}(\bar{v})$ are strictly rising in $\bar{v}$ for $\bar{v} \in (0.5(1+\rho), 1)$ and $p^{nd,*}(\bar{v})$ is strictly falling in $\bar{v}$ on this interval. Thus, the function $D(1, \bar{v})$ is strictly increasing in $\bar{v}$. At $\bar{v} = 0.5(1+\rho)$, we know that $D(v, \bar{v}) < 0$ holds for any v.

Suppose that no such $\bar{v}'$ exists. Then, $D(1, \bar{v}) < 0$ must hold for any $\bar{v} \in [0.5(1+\rho), 1]$. If this inequality were reversed for any $\bar{v}$, we would have a contradiction by continuity of $D(.)$. Note that $D(v, \bar{v})$ is strictly increasing in v for any $\bar{v}$, because $p^{H,*}(\bar{v}) > p^{L,*}(\bar{v})$ holds for any $\bar{v}$ and $Pr^H(v)$ is strictly increasing in v. Thus, $D(v, \bar{v}) \leq D(1, \bar{v}) < 0$ holds in this scenario and hence $\hat{v}^B(\bar{v}) = 1$ for any $\bar{v} \in [0.5(1+\rho), 1]$, which means that $\hat{v}^B(\bar{v}) = \bar{v} \iff \bar{v} = 1$.

Suppose instead that a $\bar{v}' \in [0.5(1+\rho), 1]$ exists - this can be $\bar{v}' = 1$. At $\bar{v} \leq \bar{v}'$, $D(1, \bar{v}) \leq 0$, i.e. $D(v, \bar{v}) < 0$ holds for any $v < 1$, and hence $\hat{v}^B(\bar{v}) = 1$ holds for any $\bar{v} \in [0.5(1+\rho), \bar{v}']$.

At $\bar{v} \in [\bar{v}', 1]$, $\hat{v}^B(\bar{v})$ will be the unique solution to:

$$D(\hat{v}^B, \bar{v}) := Pr^H(\hat{v}^B)p^{H,*}(\bar{v}) + Pr^L(\hat{v}^B)p^{L,*}(\bar{v}) - p^{nd,*}(\bar{v}) = 0$$

Note that $\hat{v}^B(\bar{v})$ is continuous in $\bar{v}$ on $\bar{v} \in [\bar{v}', 1]$ by the implicit function theorem, since the above equation is continuous in both arguments. Because $\hat{v}^B(1) \leq 1$, there exists a solution to $\hat{v}^B(\bar{v}) = \bar{v}$ on $[\bar{v}', 1]$ when $\bar{v}' \leq 1$ exists (by the intermediate value theorem).

Part 4: Suppose (a) $\bar{v}^{HC} \leq 0.5(1+\rho)$. If $\hat{v}^B(\bar{v}) = \bar{v}$ holds at $\bar{v} > 0.5(1+\rho)$, the resulting combination $p^{L,*}(\bar{v}), p^{H,*}(\bar{v}), p^{nd,*}(\bar{v})$ and $\bar{v}$ constitute a perfect Bayesian equilibrium.

Recall that: $\hat{v}^B(\bar{v}) := \sup\big\{v \in [0,1] : \underbrace{Pr(\tilde{v}^L|v)p^{L,*}(\bar{v}) + Pr(\tilde{v}^H|v)p^{H,*}(\bar{v})}_{\text{exp. price at firm with data}} < p^{nd,*}(\bar{v})\big\}$.

By construction, all prices of the firm with data are set optimally if searchers visit the firm with data if and only if $v < \bar{v}$.

Since $\bar{v} = \hat{v}^B(\bar{v})$, it is optimal for searchers to visit the firm with data if and only if $v < \bar{v} = \hat{v}^B(\bar{v})$. We know $p^L < p^H < \bar{v} = \hat{v}^B(\bar{v})$ holds because $\bar{v} > 0.5(1+\rho)$. Moreover, we know that $p^L < p^{nd} < p^H$ holds for the best response prices listed above. If $\bar{v} = 1$, this holds by construction. If $\bar{v} < 1$, the expected prices must be exactly equal at $\bar{v}$, which requires that $p^L < p^{nd} < p^H$.

Thus, the expected price at the firm with data is strictly lower for consumers with $v = p^H$, i.e. all searchers with $v \in (p^L, p^H)$ strictly prefer to visit the firm with data. All consumers with $v \leq p^L$ visit the firm with data. The sign of the expected price relationship flips at $\bar{v}$ for consumers with $v > p^H$, which implies that it is optimal for them to sort according to $\bar{v}$.

Part 5: Suppose (b) $0.5(1+\rho) < \bar{v}^{HC}$. An equilibrium exists.

In general, the set of searchers who prefer to visit the firm with data, given the optimal pricing of firms, is given by:

$$\hat{v}^G(\bar{v}) = \sup\left\{v \in [0,1] : Pr^L(v)\max\{v - p^{L,*}(\bar{v}), 0\} + Pr^H(v)\max\{v - p^{H,*}(\bar{v}), 0\} \geq \max\{v - p^{nd,*}(\bar{v}), 0\}\right\}$$

Depending on the optimal high signal price, this function takes different forms:

$$\hat{v}^{HM}(\bar{v}) = \sup\left\{v \in [0,1] : Pr^L(v)\max\{v - p^{L,*}(\bar{v}), 0\} + Pr^H(v)\max\{v - p^{HM}, 0\} \geq \max\{v - p^{nd,*}(\bar{v}), 0\}\right\}$$

$$\hat{v}^{FC}(\bar{v}) = \sup\left\{v \in [0,1] : Pr^L(v)\max\{v - p^{L,*}(\bar{v}), 0\} + Pr^H(v)\max\{v - p^{H,FC}(\bar{v}), 0\} \geq \max\{v - p^{nd,*}(\bar{v}), 0\}\right\}$$

Note that these suprema are not defined over the expected prices as was the case for $\hat{v}^B(\bar{v})$, but over expected utilities. Note also that $p^{H,FC}(\bar{v})$ solves the FOC given in equation (25).

<u>Case 1:</u> $\hat{v}^{HM}(\bar{v}^{HC}) \geq \bar{v}^{HC}$.

At $\bar{v} = \bar{v}^{HC}$, we have that $p^{H,FC}(\bar{v}^{HC}) < \bar{v}^{HC}$ and p^{HM} will yield exactly the same high signal profits. Thus, setting the FOC high signal price would be optimal for the firm with data. Also, $p^{H,FC}(\bar{v}^{HC}) < \bar{v}^{HC} < p^{HM}$. Thus, we have that $\hat{v}^{HM}(\bar{v}^{HC}) \leq \hat{v}^{FC}(\bar{v}^{HC})$ (more people would visit the firm with data if offered the high signal FOC price there).

By implication, $\hat{v}^{FC}(\bar{v}^{HC}) \geq \hat{v}^{HM}(\bar{v}^{HC}) \geq \bar{v}^{HC} > p^{H,FC}(\bar{v}^{HC})$. Thus, the indifference valuation must be above $p^{H,FC}(\bar{v}^{HC})$ at this $\bar{v}$ and hence set the expected prices equal. This establishes that $\hat{v}^{B}(\bar{v}^{HC}) \geq \bar{v}^{HC}$, which suffices as a boundary condition for existence of a solution to $\hat{v}^{B}(\bar{v}) = \bar{v}$ on $\bar{v} \in [\bar{v}^{HC}, 1]$, because $\hat{v}^{B}(1) \leq 1$ and $\hat{v}^{B}(\bar{v})$ is continuous on $[\bar{v}^{HC}, 1]$.

Such a solution constitutes an equilibrium, because for $\bar{v} \in [\bar{v}^{HC}, 1]$, the price $p^{H,*}(\bar{v})$ is strictly below $\bar{v}$ - thus, the expected price at the firm with data is lower for $v = p^{H,FC}(\bar{v})$ at this solution $\bar{v}$ and thus, the implied search behavior is optimal, since $\hat{v}^{B}(\bar{v}) = \bar{v}$.

<u>Case 2:</u> $\hat{v}^{HM}(\bar{v}^{HC}) < \bar{v}^{HC}$ (we are still in the case $0.5(1+\rho) < \bar{v}^{HC}$)

On the interval $\bar{v} \in [0.5(1+\rho), \bar{v}^{HC}]$, the optimal prices set by the firm with data are continuous. Since $\bar{v} > 0.5(1+\rho)$, so is $p^{nd,*}(\bar{v})$. The function $\hat{v}^{HM}(\bar{v}^{HC})$ is thus continuous on this interval. This is because the supremum $\hat{v}^{HM}$ must lie weakly above $p^{nd,*}(\bar{v})$, and must thus solve the following (if the supremum is below 1):

$$Pr^{L}(\hat{v}^{HM})(\hat{v}^{HM} - p^{L,*}(\bar{v})) + Pr^{H}(\hat{v}^{HM}) \max\{\hat{v}^{HM} - p^{HM}, 0\} - \left(\hat{v}^{HM} - p^{nd,*}(\bar{v})\right) = 0$$

This function is continuous in both objects. Thus, the solution $\hat{v}^{HM}(\bar{v})$ must be continuous as well. At $\bar{v} = 0.5(1+\rho)$, we know $p^{nd}(0.5(1+\rho)) = 0.5(1+\rho)$, while $p^{L,*}(0.5(1+\rho)) < 0.5(1+\rho)$. Thus, $\hat{v}^{HM}(0.5(1+\rho)) \geq 0.5(1+\rho)$. Together with the fact $\hat{v}^{HM}(\bar{v}^{HC}) < \bar{v}^{HC}$, the intermediate value theorem thus guarantees the existence of a solution to $\hat{v}^{HM}(\bar{v}) = \bar{v}$ on $\bar{v} \in [0.5(1+\rho), \bar{v}^{HC}]$, which thus also solves $\hat{v}^{G}(\bar{v}) = \bar{v}$.

This constitutes an equilibrium because the prices $p^{L,*}(\bar{v})$, $p^{nd,*}(\bar{v})$ and $p^{H,M}$ are optimal. Moreover, $p^{L,*}(\bar{v}) < p^{nd,*}(\bar{v})$. Finally, $p^{nd,*}(\bar{v}) < p^{H,M}$. This is because $p^{nd,*}(\bar{v}) < \bar{v} < \bar{v}^{HC} < p^{H,M}$. By definition, searching according to this cutoff is then optimal, because it's argument is strictly falling in v for $v > p^{nd,*}(\bar{v})$, since $p^{L,*}(\bar{v}) < p^{H,M}$.

A.4 Proof of proposition 3

A proof of a more general statement may be found in the proof of proposition 6.

A.5 Proof of corollary 1

I work with the equilibrium $\bar{v}$ for a given signal distribution as a function of ρ, i.e. $\bar{v}^*(\rho)$. An equilibrium with $p^L < p^H$ always exists. First, note that $\lim_{\rho\to 1}\bar{v}^*(\rho) = 1$, by the squeeze theorem, because, for any $\rho \in (0,1)$, we have that $0.5(1+\rho) \leq \bar{v}^*(\rho) \leq 1$.

The total demand that the firm without data receives in equilibrium is:

$$D^{nd^*}(\rho) = \rho \int_{\bar{v}^*(\rho)}^{1} dv + 0.5(1-\rho)\int_{p^{nd}(\rho)}^{1} dv = \rho[1-\bar{v}^*(\rho)] + 0.5(1-\rho)\int_{p^{nd}(\rho)}^{1} dv$$

Now consider the limit of this as $\rho \to 1$, noting that all components of demand are continuous in ρ and that $\int_{p^{nd}}^{1} dv \in [0,1]$. Thus, we have $\lim_{\rho\to 1} D^{nd^*}(\rho) = (1)(0) + (0)\int_{p^{nd}}^{1} dv = 0$. Since the demand of the firm without data approaches 0 when $\rho \to 1$, the market share of the firm with data approaches 1 by any definition of the market share (sales or profit).

A.6 Statement and proof of lemma 5

Lemma 5 *Consider the sequential search framework. In any equilibrium in which firms play pure strategies:*

- *The ordering $p^L < p^{nd} < p^H$ must hold.*
- *There exists an $\epsilon > 0$ such that any searcher who visits the firm without data first in equilibrium will not search when offered a price $p_j \in [0, p^{nd} + \epsilon]$ at this firm.*
- *There exists a $\bar{v} > p^L$ such that all searchers with $v \in (p^L, \bar{v})$ visit the firm with data first and all searchers with $v \in (\bar{v}, 1]$ visit the firm without data first.*
- *The ordering $\bar{v} \geq 0.5(1+\rho)$ holds.*

Proof:

Part 1: In equilibrium, $p^L < p^H$ must hold.

Part 1a: There exists no equilibrium in which $p^H < p^L$.

Previous arguments imply the following: In any equilibrium with $p^H \neq p^L$, the ordering $p^{min} = \min\{p^L, p^H\} < p^{nd}$ and $p^{max} = \max\{p^L, p^H\} > p^{nd}$ must hold.

Suppose $p^L \leq p^{nd} + s$, i.e. no searcher will leave the firm with data to search at the equilibrium prices. Thus, for $p_j \in [p^H, p^L]$, all consumers who arrive at the firm with data buy there iff the price is below their v - then, the structure of equilibrium profits equal the one defined in the proof of lemma 1 and there is either a deviation from p^L to p^H or vice versa.

Suppose instead that $p^L > p^{nd} + s$. Then, the firm with data only sells to its captive consumers at p^L and thus $p^L = p^{L,M}$. All searchers who arrive at the firm without data buy in an open ball around p^{nd}. Hence, $p^{nd} \geq 0.5$ holds, and $p^{nd} \geq 0.5 > p^L$, a contradiction.

<u>Part 1b:</u> There exists no equilibrium in which $p^L = p^H$.

This follows from the same arguments made in the baseline model. The only possible equilibrium candidate is $p^L = p^H = p^{nd}$, which is ruled out under our tie-breaking rule.

Part 2: In equilibrium, $p^{nd} \in (p^L, p^H)$ must hold.

This follows from the arguments made in the proof of lemma 1.

Part 3: Any searcher who optimally visits the firm without data first must find it strictly optimal to not search when receiving p^{nd}.

To see this, define $U^{nd,s}(v)$ and $U^{nd,ns}(v)$ as the expected utilities of visiting the firm without data first and searching or not searching, respectively. Define $U^{d,s}(v)$ as the expected utility of visiting the firm with data and searching if and only if p^H is received there.

Consider a consumer that optimally visits the firm without data first, who must have $\nu > p^{nd}$. Suppose, for a contradiction, that $Pr^L(v)(p^{nd} - p^L) - s \geq 0 \iff U^{nd,s}(v) \geq U^{nd,ns}(v)$. Crucially, $U^{d,s}(v) > U^{nd,s}(v)$ will also hold, because:

$$Pr^L(v)(v - p^L) + Pr^H(v)(v - p^{nd} - s) > Pr^L(v)(v - p^L) + Pr^H(v)(v - p^{nd}) - s$$

The utility of visiting the firm without data is $U^{nd,s}(v)$, while the utility of visiting the firm with data first is at least $U^{d,s}(v)$. It would thus be strictly optimal for this consumer to visit the firm with data first, a contradiction. Hence, $Pr^L(v)(p^{nd} - p^L) - s < 0$ must hold for any consumer that visits the firm without data first in equilibrium, which implies that there exists an $\epsilon > 0$ such that these consumers would also not search for prices $p_j \leq p^{nd} + \epsilon$.

Part 4: Uniqueness of cutoff $\bar{v}$ for equilibria with $p^{nd} + s \geq p^H$

We consider an equilibrium candidate and define a $\tilde{v}^I$ that solves $Pr^L(\tilde{v}^I)(p^{nd} - p^L) - s = 0$.

All consumers with $v \in (p^L, \tilde{v}^I)$ will surely visit the firm with data first by previous arguments, because search would be optimal for them after visiting the firm without data (if $v > p^{nd}$). Now, we just need to make a case distinction w.r.t the ordering of p^H and $\tilde{v}^I$.

(i) $p^H < \tilde{v}^I$: For any searcher with $v > \tilde{v}^I$, the preference for the firm with data is strictly falling in v, because it is equal to:

$$P^D(v) = Pr^L(v)(v-p^L)+Pr^H(v)(v-p^H)-(v-p^{nd}) \implies \frac{\partial P^D(v)}{\partial v} = \frac{\partial Pr^L(v)}{\partial v}(p^H-p^L) < 0$$

(ii) $\tilde{v}^I \leq p^{nd} < p^H$: All searchers with $v \in (p^L, p^{nd})$ visit the firm with data. No searcher will search after receiving p^H since $p^{nd} + s \geq p^H$ by assumption. Thus, the preference for the firm with data is the following for all $v \in (p^{nd}, p^H]$:

$$P^D(v) = Pr^L(v)(v-p^L)+Pr^H(v)(0)-(v-p^{nd}) \implies \frac{\partial P^D(v)}{\partial v} = Pr^L(v)-1+\frac{\partial Pr^L(v)}{\partial v}(v-p^L) < 0$$

For $v > p^H$, we have $P^D(v) = Pr^L(v)(v - p^L) + Pr^H(v)(v - p^H) - (v - p^{nd})$. Thus, $P^D(v)$ is continuous at $v = p^H$ and falling globally. This implies the result.

(iii) $\tilde{v}^I \in (p^{nd}, p^H)$: All searchers with $v \leq \tilde{v}^I$ visit the firm with data first. Analogous arguments show that $P^D(v)$ is continuous at $v = p^H$ and strictly falling in v.

Part 5: Uniqueness of cutoff $\bar{v}$ in equilibria with $p^{nd} + s < p^H$

Searchers leave the firm with data to search when receiving p^H if and only if $v > p^{nd} + s$. As before, $\tilde{v}^I$ solves $Pr^L(\tilde{v}^I)(p^{nd} - p^L) - s = 0$. Calculating the relative preferences for the firm with data for two separate cases, namely (i) $p^{nd} + s < \tilde{v}^I$ and (ii) $\tilde{v}^I \leq p^{nd} + s$ yields the desired result based on steps that mirror those taken in the previous part.

Part 6: Establishing that $\bar{v} \geq 0.5(1 + \rho)$ holds true.

First, note that $p^{nd} < \bar{v}$ must hold. Suppose, for a contradiction, that we have an equilibrium with $p^L < p^H$ in which searchers apply a cutoff strategy with $\bar{v} \leq p^{nd}$.

This cannot be an equilibrium - consider any consumer with v just above p^{nd} who visits the firm without data first. This consumer would not search thereafter (else, she would optimally visit the firm with data). Thus, her utility at the firm without data is $v - p^{nd}$, which converges to 0 as $v \to p^{nd}$. By contrast, their utility at the firm with data is at least $Pr^L(v)(v - p^L)$, which remains strictly positive for any such v. Since expected utilities are continuous in v, these consumers would rather visit the firm with data, a contradiction.

Hence, $p^{nd} < \bar{v}$ must hold in an equilibrium (by the choices of searchers). However, if $\bar{v} < 0.5(1 + \rho)$, such prices are not optimal for the firm without data:

(i) Case 1: $p^H \leq p^{nd} + s$.

No arriving searcher will leave the firm without data for prices in an open ball around p^{nd}. A zero measure of searchers arrives at the firm without data after visiting the its rival, because $p^H \leq p^{nd} + s$ (if $p^H = p^{nd} + s$, there would else be undercutting by the firm with data). Thus, for prices in an open ball around p^{nd}, the profits of the firm without data are:

$$\Pi^{nd}(p_j; \bar{v}) = p_j \left[\rho \int_{\bar{v}}^{1} (1) dv + 0.5(1 - \rho) \int_{p_j}^{1} (1) dv \right]$$

To constitute an equilibrium, p^{nd} must lie strictly below $\bar{v}$. But for $\bar{v} < 0.5(1 + \rho)$, the derivative at any such price is strictly positive, a contradiction.

(ii) Case 2: $p^H > p^{nd} + s$.

If $\bar{v} \leq p^{nd} + s$, previous arguments directly imply the result. If $\bar{v} > p^{nd} + s$, searchers with $v \in (p^{nd} + s, \bar{v})$ visit the firm with data first and then search if and only if they generate $\tilde{v}^H$. Since $p^H > p^{nd} + s$ holds by assumption, these consumers buy in an open ball around p^{nd}. In an open ball around p^{nd}, the profits at the firm without data are hence:

$$\Pi^{nd}(p_j; \bar{v}) = p_j \rho \int_{p^{nd}+s}^{\bar{v}} Pr^H(v) dv + p_j \rho \int_{\bar{v}}^{1} (1) dv + p_j 0.5(1 - \rho) \int_{p_j}^{1} (1) dv$$

For any $\bar{v} < 0.5(1 + \rho)$, the derivative of the second component is strictly positive for any $p_j < \bar{v}$. The derivative of the first component is positive. Hence, $\bar{v} \geq 0.5(1 + \rho)$ must hold.

A.7 Proof of proposition 4

Any searcher who visits two firms with positive probability must either (i) visit the firm with data first with positive probability and search thereafter with positive probability or (ii) visit the firm without data first and search thereafter with positive probability. Lemma 5 implies that $p^L < p^H$.

Part 1: Any consumer who visits the firm without data first will find it strictly optimal to not search thereafter.

This follows from the arguments made in the proof of lemma 5, part 3. Thus, the set of consumers who visit the firm without data first (with positive probability) and search with positive probability thereafter must have measure zero.

Part 2: If $p^H < p^{nd} + s$, the set of consumers who visit the firm with data first (with positive probability) and search thereafter has measure zero.

Here, any searcher would find it strictly optimal to not search after any price the firm with data would offer to her in equilibrium. This implies the result.

Part 3: If $p^H = p^{nd} + s$, the set of consumers who visit the firm with data first (with positive probability) and search thereafter must have measure zero.

By lemma 5, it must hold that $p^L < p^{nd}$. Thus, p^L does not induce search. In a hypothetical equilibrium like this, the firm with data would prefer to undercut p^H, since this deters search by all consumers.

Part 4: Under assumption 2, there exists no equilibrium in which $p^{nd} + s < p^H$ and a strictly positive measure of searchers visit the firm with data first and search thereafter with positive probability.

Suppose, for a contradiction, that there exists such an equilibrium. By lemma 5, the search strategy is a cutoff rule with a cutoff $\bar{v}$ in such equilibria.

<u>Part 4a:</u> In order for a positive measure of searchers to search after visiting the firm with data, the ordering $\bar{v} > p^{nd} + s$ must hold.

Suppose that $\bar{v} \leq p^{nd} + s$. By lemma 5, searchers who visit the firm with data first must have $v \in [0, \bar{v}]$. Moreover, search is only optimal if $v \geq p^{nd} + s$. But since $\bar{v} \leq p^{nd} + s$, the set of searchers who visit the firm with data first & search thereafter with positive probability is a subset of $[0, \bar{v}] \cap [p^{nd} + s, 1]$, which has zero measure, a contradiction.

Part 4b: Optimal pricing of the firm without data.

We have proven that $p^{nd} + s < \bar{v}$ must hold. All searchers with $v \in [p^{nd} + s, \bar{v}]$ will visit the firm with data first and search when being offered p^H, which occurs with probability $Pr^H(v)$. Thus, the firm without data makes the sale to all these searchers at the price p^{nd} with probability $Pr^H(v)$. Since $p^{nd} + s < \bar{v}$, the firm without data will also make the sale to all searchers who initially visit it. For p_j in an open ball around p^{nd}, the profit function of the firm without data is hence:

$$\Pi^{nd}(p_j) = p_j \left[\rho \int_{p^{nd}+s}^{\bar{v}} Pr^H(v) dv + \rho \int_{\bar{v}}^{1} dv + 0.5(1-\rho) \int_{p_j}^{1} dv \right]$$

Thus, an equilibrium p^{nd} must equal $p^{nd,3}(\bar{v})$, which solves:

$$\left[\rho \int_{p^{nd,3}+s}^{\bar{v}} Pr^H(v) dv + \rho \int_{\bar{v}}^{1} dv + 0.5(1-\rho) \int_{p^{nd,3}}^{1} dv \right] - 0.5(1-\rho) p^{nd,3} = 0$$

Part 4c: If $p^{nd,3}(1) + s > p^{HM}$ (which holds by assumption 3), such an equi. does not exist.

In this equilibrium, $p^H = p^{H,M} > p^{nd} + s$ must be satisfied, where $p^{nd} = p^{nd,3}(\bar{v})$ must hold for the equilibrium level of $\bar{v}$, whatever this may be. Note that the function $p^{nd,3}(\bar{v})$ is falling in $\bar{v}$. Thus, we have $p^{nd,3}(1) + s \leq p^{nd,3}(\bar{v}) + s$ for any possible $\bar{v}$.

Suppose that $p^{nd,3}(1) + s > p^{HM}$, noting that $p^{nd,3}(1) = p^{nd,s}$ as defined in assumption 4. Since $p^H = p^{H,M} < p^{nd,3}(\bar{v}) + s = p^{nd} + s$, this equilibrium cannot exist, because there exists no $\bar{v}$ at which the necessary conditions for the existence of this equilibrium are satisfied.

A.8 Proof of proposition 5

Part 1: The first two bullet points hold by lemma 5.

Part 2: When $\Pi^H(0.5; 0.5(1+\rho)) > 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$ (assumption 5), the optimal

$p^{H,*}(\bar{v})$ lies strictly below $\bar{v}$ for any $\bar{v} \geq 0.5(1+\rho)$. Hence, $p^{H,*}(\bar{v})$ solves:

$$\rho \int_{p^{H,*}}^{\bar{v}} Pr^H(v)dv + 0.5(1-\rho) \int_{p^{H,*}}^{1} Pr^H(v)dv - p^{H,*}\big[0.5(1+\rho)Pr^H(p^{H,*})\big] = 0$$

By assumption, we have $\Pi^H(0.5;\bar{v}) \geq \Pi^H(0.5;0.5(1+\rho)) > 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$ for any $\bar{v} \geq 0.5(1+\rho)$. Thus, the optimal high signal price must lie strictly below $\bar{v}$. A $p^{H,*}(\bar{v})$ as defined above always exists by the intermediate value theorem.

Part 3: Consider an equilibrium candidate in which $p^L < p^{nd} < p^H$, $p^H \leq p^{nd} + s$, $p^H < \bar{v}$, and $\hat{v}(p^L, p^H, p^{nd}) = \bar{v}$. It is optimal for searchers to visit the firm with data if and only if $v > \bar{v}$ and never search thereafter.

Part 3a: In such an equilibrium candidate, the cutoff $\tilde{v}^I(p^L, p^H, p^{nd})$ will lie strictly below $\hat{v}^I(p^L, p^H, p^{nd})$, where these cutoffs are defined as follows:

$$Pr^L(\tilde{v}^I(.))\big(p^{nd} - p^L\big) - s = 0 \quad ; \quad Pr^L(\hat{v}^I(.))p^L + Pr^H(\hat{v}^I(.))p^H - p^{nd} = 0$$

Note first that $Pr^L(v)p^L + Pr^H(v)p^H = p^{nd} \iff Pr^L(v)\big(p^{nd} - p^L\big) + Pr^H(v)\big(p^{nd} - p^H\big) = 0$.

Now note that $p^{nd} + s \geq p^H$ by assumption, i.e. $p^{nd} - p^H \geq -s$. Thus:

$$0 = Pr^L(\tilde{v}^I)\big(p^{nd} - p^L\big) - s < Pr^L(\tilde{v}^I)\big(p^{nd} - p^L\big) - Pr^H(\tilde{v}^I)s \leq Pr^L(\tilde{v}^I)\big(p^{nd} - p^L\big) + Pr^H(\tilde{v}^I)(p^{nd} - p^H)$$

Since $\frac{\partial}{\partial v}\big[Pr^L(v)\big(p^{nd} - p^L\big) + Pr^H(v)\big(p^{nd} - p^H\big)\big] < 0$, we have $\tilde{v}^I(p^L, p^H, p^{nd}) < \hat{v}^I(p^L, p^H, p^{nd})$.

Part 3b: The postulated search behaviour is optimal:

Because $\hat{v}(p^L, p^H, p^{nd}) = \bar{v}$, $\bar{v}$ is either equal to $\hat{v}^I(p^L, p^H, p^{nd})$ or 1. Define $p = (p^L, p^H, p^{nd})$. It was established that $\hat{v}^I(p^1) > \tilde{v}^I(p^1)$.

Suppose $\tilde{v}^I(p) \geq 1$ in equilibrium, which then implies that $\hat{v}^I(p) > 1$, and thus $\hat{v}(p) = 1 = \bar{v}$. For all consumers with $v < 1 \leq \tilde{v}^I(p)$, it is strictly optimal to visit the firm with data in equilibrium, i.e. to visit according to the rule represented by $\bar{v} = 1$. No searcher will search after visiting the firm with data since $p^H \leq p^{nd} + s$. No searcher that arrives at the firm without data first will find it optimal to search afterwards (since no such consumer exists).

Suppose $\tilde{v}^I(p) < 1$. Because $\tilde{v}^I(p) < \hat{v}^I(p)$ will also hold, $\hat{v}(p)$ is either 1 when $\hat{v}^I(p) \geq 1$ or $\hat{v}(p) = \hat{v}^I(p)$. In either case, $\tilde{v}^I(p) < \bar{v}$. Thus, any consumer with $v \geq \bar{v}$ finds it strictly optimal to not search after visiting the firm without data first. Because $p^H \leq p^{nd} + s$ and $\hat{v}(p) = \bar{v}$, she will visit the firm without data first and not search thereafter.

Any searcher with $v < \tilde{v}^I(p)$ visits the firm with data first and does not search thereafter. Any searchers with $v \in [\tilde{v}^I(p), \bar{v}]$ would not search after visiting either firm. Because $\bar{v} = \hat{v}(p)$, they hence optimally visit the firm with data.

Part 4: Consider $\bar{v}^1$, where $\hat{v}^G(\bar{v}^1) = \bar{v}^1$, and $\bar{v}^1 \in [0.5(1+\rho), 1]$, which exists by proposition 2. If the accompanying $(p^{L,1}, p^{H,1}, p^{nd,1})$ satisfy $p^{H,1} \leq p^{nd,1} + s$, it is an equilibrium.

Search: It is optimal for searchers to visit the firm with data if and only if $v > \bar{v}$ and never search thereafter.

The ordering $p^{L,1} < p^{nd,1} < p^{H,1}$ holds by construction, since $\hat{v}^B(\bar{v}^1) = \bar{v}^1$. The latter holds because assumption 5 guarantees that a solution to $\hat{v}^G(\bar{v}) = \bar{v}$ on $\bar{v} \in [0.5(1+\rho), 1]$ also solves $\hat{v}^B(\bar{v}) = \bar{v}$. By assumption 5, we also have $p^{H,1} < \bar{v}^1$. By specification, $p^{H,1} \leq p^{nd,1} + s$. Thus, the insights of part 3 apply and the result follows.

Pricing: There are no profitable deviations from the equilibrium prices, given that searchers split according to $\bar{v}^1$ and do not search thereafter (for equilibrium prices).

Consider first the firm without data. True competitive profits are bounded from above $\Pi^{nd}(p_j; \bar{v}^1)$. This is because no consumers arrive after search. For prices $p_j \in [0, p^{nd} + \epsilon]$, true profits equal this function. For prices sufficiently high, searchers leave this firm to search, implying that true profits are below $\Pi^{nd}(p_j; \bar{v}^1)$. By construction, $p^{nd,1}$ maximizes $\Pi^{nd}(p_j; \bar{v}^1)$, and so there will be no profitable deviations.

Analogous arguments show that the firm with data has no profitable deviations, because competitive profits are bounded from above by $\Pi^k(p_j; \bar{v}^1)$, conditional on the signal $\tilde{v}^k$.

Part 5: These exists a candidate for an equilibrium of category 2 when $p^{H,1} > p^{nd,1} + s$.

Part 5a: Setup

In an equilibrium of this category, we have $p^L = p^{L,*}(\bar{v})$, $p^{nd} = p^{nd,*}(\bar{v})$, $p^H = p^{nd,*}(\bar{v}) + s$. Recall that $p^{L,1} = p^{L,*}(\bar{v}^1)$, $p^{nd,1} = p^{nd,*}(\bar{v}^1)$, and that $p^{H,1} = p^{H,*}(\bar{v}^1)$. We are looking for a $\bar{v}$ that solves:

$$\bar{v} = \hat{v}^S(\bar{v}) := \sup\big\{v \in [0,1] : \underbrace{Pr^L(v)p^{L,*}(\bar{v}) + Pr^H(v)(p^{nd,*}(\bar{v}) + s) - p^{nd,*}(\bar{v})}_{D^S(v;\bar{v}) := Pr^L(v)(p^{L,*}(\bar{v}) - p^{nd,*}(\bar{v})) + Pr^H(v)s} < 0\big\}$$

For any level of $\bar{v} \geq 0.5(1+\rho)$, we have $p^{L,*}(\bar{v}) < \bar{v}$ and $p^{nd,*}(\bar{v}) \leq \bar{v}$. Since $p^{L,*}(\bar{v}) \leq p^{LM} < 0.5$ and $p^{nd,*}(\bar{v}) \geq 0.5$, we have $p^{L,*}(\bar{v}) < p^{nd,*}(\bar{v})$. Since $p^{H,1} > p^{nd,1} + s$, we have:

$$\underbrace{\hat{v}(p^{L,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1) + s, p^{nd,*}(\bar{v}^1))}_{=\hat{v}^S(\bar{v}^1)} \geq \underbrace{\hat{v}(p^{L,*}(\bar{v}^1), p^{H,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1))}_{=\hat{v}^B(\bar{v}^1)}$$

<u>Part 5b:</u> If $p^{H,*}(\bar{v}^1) > p^{nd,*}(\bar{v}^1) + s$, there exists a $\bar{v}^2 \in [0.5(1+\rho), 1]$ s.t. $\hat{v}^S(\bar{v}^2) = \bar{v}^2$.

To show this, we work towards applying the intermediate value theorem. We know that (i) $\hat{v}^S(1) \leq 1$ and (ii) $\hat{v}^S(\bar{v}^1) \geq \bar{v}^1$. The second point follows from our previous results, i.e.:

$$\underbrace{\hat{v}(p^{L,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1) + s, p^{nd,*}(\bar{v}^1))}_{=\hat{v}^s(\bar{v}^1)} \geq \underbrace{\hat{v}(p^{L,*}(\bar{v}^1), p^{H,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1)))}_{=\hat{v}^B(\bar{v}^1)} = \bar{v}^1$$

As previously, we can show continuity of the function $\hat{v}^S(\bar{v})$ on the interval $\bar{v} \in [\bar{v}^1, 1]$, because all price functions will be strictly below $\bar{v}$ and hence continuous. Thus, the intermediate value theorem establishes existence of an appropriate solution.

(i) Subcase 1: $D^S(1;\bar{v}^1) > 0$ (interior search cutoff at $\bar{v}^1$).

Note that $D^S(1,\bar{v})$ is rising in $\bar{v}$ - in this subcase, we hence have that $D^S(1;\bar{v}) > 0$ holds for all $\bar{v} \in [\bar{v}^1, 1]$. Thus, the object $\hat{v}^S(\bar{v})$ must solve the following for all $\bar{v} \in [\bar{v}^1, 1]$:

$$D(\hat{v}^S, \bar{v}) = Pr^L(\hat{v}^S)\big(p^{L,*}(\bar{v}) - p^{nd,*}(\bar{v})\big) + Pr^H(\hat{v}^S)s = 0$$

Such a solution always exists in the relevant interval. This function is continuous in both arguments - thus, the implicit function theorem guarantees that the solution function $\hat{v}^S(\bar{v})$ will also be continuous (we can apply this theorem because the derivative of $D(\hat{v}^S, \bar{v})$ w.r.t $\hat{v}^S$ will never be zero, given that $p^{L,*}(\bar{v}) < p^{nd,*}(\bar{v})$ holds for the relevant $\bar{v}$).

(ii) Subcase 2: $D^S(1;\bar{v}^1) \leq 0$ (all searchers would visit firm with data at $\bar{v}^1$).

Find $\bar{v}' \geq \bar{v}^1$ that solves $D^S(1;\bar{v}') = 0$. If this does not exist, we know that $D^S(1;\bar{v}) < 0$ for all $\bar{v} \in [\bar{v}^1, 1]$, which implies that we will have a solution at 1, i.e. $\hat{v}^S(1) = 1$. If such a $\bar{v}'$ exists, we know that $\hat{v}^S(\bar{v})$ must solve the following equation for all $\bar{v} \in [\bar{v}', 1]$:

$$D(\hat{v}^S, \bar{v}) = Pr^L(\hat{v}^S)\big(p^{L,*}(\bar{v}) - p^{nd,*}(\bar{v})\big) + Pr^H(\hat{v}^S)s = 0$$

At $\bar{v} = \bar{v}'$, we know $\hat{v}^S(\bar{v}') = 1$ and at the opposite end, we know it is weakly below 1. The solution function is continuous - existence of a fixed point is hence guaranteed.

Part 6: Suppose $p^{H,1} > p^{nd,1} + s$. At $\bar{v}^2$, the following two conditions are satisfied: (i) $p^{H,2} < \bar{v}^2$ and (ii) $p^{H,2} < arg\max_{p_j} \Pi^H(p_j;\bar{v}^2) := p^{H,*}(\bar{v}^2)$.

<u>Part 6a:</u> If $p^{H,*}(\bar{v}^1) > p^{nd,*}(\bar{v}^1) + s$ holds, previous results imply that $\bar{v}^2 \geq \bar{v}^1$.

To see this, recall $\hat{v}^S(\bar{v}) := \sup\big\{v \in [0,1] : Pr^{L,*}(v)p^L(\bar{v}) + Pr^H(v)\big(p^{nd,*}(\bar{v}) + s\big) < p^{nd,*}(\bar{v})\big\}$.

Suppose $\bar{v}^1 = 1$. Then, $\bar{v}^2 = \bar{v}^1$ must be true, by previous arguments.

Suppose instead that $\bar{v}^1 < 1$. Previous arguments have established that:

$$\hat{v}(p^{L,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1) + s, p^{nd,*}(\bar{v}^1)) - \bar{v}^1 > \hat{v}(p^{L,*}(\bar{v}^1), p^{H,*}(\bar{v}^1), p^{nd,*}(\bar{v}^1)) - \bar{v}^1 = 0$$

Note that $\hat{v}(p^{L,*}(\bar{v}), p^{nd,*}(\bar{v}) + s, p^{nd,*}(\bar{v}))$ is weakly decreasing in $\bar{v}$ because $p^{L,*}(\bar{v})$ is rising in $\bar{v}$ and $p^{nd,*}(\bar{v})$ is falling in $\bar{v}$. Thus, any consumer that previously already had a higher expected price at the firm with data will continue to do so when $\bar{v}$ rises.

Thus, we have that the function $\hat{v}^S(p^{L,*}(\bar{v}), p^{nd,*}(\bar{v}) + s, p^{nd,*}(\bar{v})) - \bar{v}$ is strictly decreasing in $\bar{v}$ and that this function is strictly positive at $\bar{v}^1$. Hence, $\bar{v}^2 \geq \bar{v}^1$ must hold.

<u>Part 6b:</u> Since $\bar{v}^2 \geq \bar{v}^1$, $p^{H,2} < arg\max_{p_j} \Pi^H(p_j;\bar{v}^2) = p^{H,*}(\bar{v}^2)$ and $p^{H,2} < \bar{v}^2$ hold.

Note that $\bar{v}^1 \in [0.5(1+\rho), 1]$. Because $\bar{v}^2 \geq \bar{v}^1 \geq 0.5(1+\rho)$, $arg\max_{p_j} \Pi^H(p_j;\bar{v}^2) = p^{H,*}(\bar{v}^2)$ will be strictly below $\bar{v}^2$ and solve a FOC. Because $\bar{v}^2 \geq \bar{v}^1$, we know that the prices satisfy: (i) $p^{k,*}(\bar{v}^2) \geq p^{k,*}(\bar{v}^1)$ and (ii) $p^{nd,*}(\bar{v}^2) \leq p^{nd,*}(\bar{v}^1)$. Thus:

$$p^{H,2} = p^{nd,*}(\bar{v}^2) + s \leq p^{nd,*}(\bar{v}^1) + s < p^{H,*}(\bar{v}^1) \leq p^{H,*}(\bar{v}^2) < \bar{v}^2$$

Part 7: The solution $\bar{v}^2$, with accompanying prices, is an equilibrium if $p^{H,1} > p^{nd,1} + s$.

Part 7a: At $\bar{v}^2$, $\Pi^H(p^{2,H}; \bar{v}^2) \geq 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$ holds by assumption 5.

Note that $p^{H,2} = p^{nd,*}(\bar{v}^2) + s > 0.5$, since $p^{nd,*}(\bar{v}^2) > 0.5$. High signal profits for $p_j < \bar{v}^2$, which includes $p^{H,2}$ since $p^{H,2} < p^{H,*}(\bar{v}^2) < \bar{v}^2$ are:

$$\Pi^H(p_j; \bar{v}^2) = p_j \rho \int_{p_j}^{\bar{v}^2} Pr^H(v)dv + p_j 0.5(1-\rho) \int_{p^H}^{1} Pr^H(v)dv$$

We know that this function is strictly concave on $p_j \in [0, \bar{v}^2]$ and that $0.5 < p^{H,2} < p^{H,*}(\bar{v}^2) < \bar{v}^2$. Thus, profits from setting the price $p_j = 0.5$ will be below equilibrium profits. Moreover, we have $\bar{v}^2 > 0.5 + 0.5\rho$, which also implies that $\Pi^H(0.5; \bar{v}^2) \geq \Pi^H(0.5; 0.5 + 0.5\rho)$. By assumption, the final component is above $\Pi^{H,M}(p^{H,M})$.

Part 7b: Search

First, we note that the search behavior represented by the cutoff $\bar{v}^2$ will be optimal by the arguments in part 3. This is because $p^{L,2} < p^{nd,2} < p^{H,2}$ and $p^{H,2} = p^{nd,2} + s$ hold by construction, $p^{H,2} < \bar{v}^2$ by part 6, and $\hat{v}(p^{L,2}, p^{nd,2}, p^{H,2}) = \bar{v}^2$ holds by definition.

Part 7c: Pricing

Now consider optimal pricing. Since no consumer leaves to search on the equilibrium path, we know that $\Pi^{nd}(p_j; \bar{v})$ and $\Pi^L(p_j; \bar{v})$ are upper bounds for the true respective objective functions. Since the former are both globally maximized by our prices for the given $\bar{v}^2 > 0.5(1+\rho)$, we know there can be no profitable deviations from them p^{nd} or p^L.

Now consider the optimal pricing calculus of the firm with data when observing $\tilde{v}^H$. Since $p^{H,1} > p^{nd,1} + s$, part 6 established that $p^{H,2} < arg \max_{p_j} \Pi^H(p_j; \bar{v}^2)$ and $p^{H,2} < \bar{v}^2$. Because the high signal objective function is strictly concave on $p_j \in [0, \bar{v}]$, we thus know that there cannot be any downward deviations that are profitable. Any upward deviation would, at best, yield profits equal to $0.5(1-\rho)\Pi^{H,M}(p^{H,M})$. This deviation is not profitable by the result in part 7a. No other possible deviations remain.

A.9 Proof of proposition 6

Define the maximal and minimal prices that the firms offer as $\bar{p}^j$ and $\underline{p}^j$, respectively. We work with the search strategy $d(v)$ and with the cutoff prices $\hat{p}^j(v)$, where $j \in nd, d$. I show that there exists no equilibrium in which firms mix for three different possible cases:

(i) Suppose $\underline{p}^{nd} < \underline{p}^d$.

The price $\underline{p}^{nd}$ is played with probability 1. Suppose it is part of a mixed strategy. Then, there exists an interval of prices below $\underline{p}^d$ that are played by the firm without data.

All searchers with $v < \underline{p}^d$ will surely visit the firm without data first. For any $p_j < \underline{p}^d$, no searcher who arrives at the firm without data first will search. Any searcher arrives at the firm without data first with probability $1 - d(v)$.

Consider consumers who arrive at the firm without data second. Any such consumer has $v > \underline{p}^d$ and must have received a price strictly above $\underline{p}^d$ - thus, they entail fully inelastic demand. These consumers arrive at the firm without data with probability $d(v)Pr(p^d > \hat{p}^d(v))$.

Thus, the firm with data makes the following profits when setting any price $p_j \in [\underline{p}^{nd}, \underline{p}^d]$:

$$\Pi^{nd}(p_j) = p_j \left[\rho \int_{p_j}^{\underline{p}^d} (1)dv + \rho \int_{\underline{p}^d}^{1} \left[d(v)Pr(p^d > \hat{p}^d(v)) + (1 - d(v)) \right] dv + 0.5(1-\rho) \int_{p_j}^{1} (1)dv \right]$$

This function is strictly concave, which implies that there cannot be any other prices in the interval $p_j \in [\underline{p}^{nd}, \underline{p}^d]$ that are played by the firm with data. Thus, the price $\underline{p}^{nd}$ must be played with probability 1 by our restriction of connected support.

Now consider the prices of the firm with data. Because $\underline{p}^{nd}$ is played with probability 1 and $\underline{p}^{nd} < \underline{p}^d$, all searchers visit the firm without data and never arrive at the firm with data. Thus, the firm with data just makes the sale to its captive consumers for any price $p_j > \underline{p}^d$. This implies that the firm with data would not mix, since its monopoly profit functions are strictly concave. Summing up, firms would not mix in an equilibrium of category (i).

(ii) Suppose $\underline{p}^d < \underline{p}^{nd}$

As before, $\underline{p}^d$ must be played with probability 1 by the firm with data in the corresponding information set. This is because all searchers with valuation below $\underline{p}^{nd}$ visit the firm with data. All searchers with valuation above $\underline{p}^{nd}$ will generate inelastic demand for the firm with data around $\underline{p}^{d}$, because no such consumer would search when receiving a price below $\underline{p}^{nd}$.

Now consider the optimal pricing of the firm without data. A key step in this proof is to show that no consumer who visits the firm without data first would search after receiving $\underline{p}^{nd}$.

Suppose, for a contradiction, that a searcher with valuation v visits the firm without data first and finds it weakly optimal to search when receiving the price $\underline{p}^{nd}$, which means that she will search for any price she can receive at this firm. Define the signal which warrants the offering of $\underline{p}^{d}$ as $\tilde{v}^1$ and the converse signal as $\tilde{v}^2$, with corresponding probability $Pr^2(v)$. Because $\underline{p}^{d} < \underline{p}^{nd}$, the expected utility this consumer will attain by visiting the firm without data first is the following (because searching is always weakly optimal):

$$U^{nd}(v) = Pr^1(v)(v - \underline{p}^d) + Pr^2(v) \int_{\underline{p}^H}^{\bar{p}^H} \int_{\underline{p}^{nd}}^{\bar{p}^{nd}} \max\{v - p^{nd}, v - p^H, 0\} dF^{nd}(p^{nd}) dF^2(p^H) - s$$

Alternatively, this consumer (who must have $v > \underline{p}^d$) can visit the firm with data first and search if and only if $\tilde{v}^2$ is generated. This yields strictly higher expected utility, namely:

$$U^{d}(v) = Pr^1(v)(v - \underline{p}^d) + Pr^2(v) \int_{\underline{p}^H}^{\bar{p}^H} \int_{\underline{p}^{nd}}^{\bar{p}^{nd}} \left[\max\{v - p^{nd}, v - p^H, 0\} - s\right] dF^{nd}(p^{nd}) dF^2(p^H)$$

Thus, any such consumer would not visit the firm without data first, a contradiction. Thus, any searcher who visits the firm without data first must find it strictly optimal to not search at $\underline{p}^{nd}$. Because search preferences are continuous in the initial price, searchers will also not search for prices just above it (by the dominated convergence theorem).

There exist $\epsilon > 0$ and $\delta > 0$ such that:

- Searchers with $v \in [\underline{p}^d, \underline{p}^{nd} + \epsilon]$ visit the firm with data first. Setting ϵ small enough also implies that these consumers would never search thereafter.
- Searchers who visit the firm without data first will not search if offered a price $p_j \in [\underline{p}^{nd}, \underline{p}^{nd} + \delta]$.
- Searchers who arrive at the firm without data second buy if offered $p_j \in [\underline{p}^{nd}, \underline{p}^{nd} + s]$.

Set $\tau = \min\{\epsilon, \delta, s\}$. For all $p_j \in [\underline{p}^{nd}, \underline{p}^{nd} + \tau]$, the profits of the firm without data are:

$$\Pi^{nd}(p_j) = p_j \left[\rho \int_{p^{nd}+\epsilon}^{1} \left[d(v) Pr(\hat{p}^d(v) > p^d) + (1 - d(v))(1) \right] dv + 0.5(1-\rho) \int_{p_j}^{1} (1) dv \right]$$

The demand implied by searchers is fully inelastic for these prices. This means that profits are strictly concave, which implies that $\underline{p}^{nd}$ must be played with probability 1 by the restriction of connected support.

We have established that $\underline{p}^d$ and $\underline{p}^{nd}$ both have to be played with probability 1. Thus, the only possibility of mixing is that the firm with data mixes after one of the two signals. Define that the firm with data mixes after receiving $\tilde{v}^m$, with prices in the support $[\underline{p}^m, \bar{p}^m]$. No consumer who visits the firm without data first will search thereafter (by an option value logic, as above). Moreover, any searchers who visit the firm with data first will leave to search if $v > \underline{p}^{nd} + s$ and the price they receive is above this cutoff.

It cannot hold that $\underline{p}^{nd} + s < \bar{p}^m$. Then, all searchers will surely not consume at the firm with data for $p_j \in [\underline{p}^{nd} + s, \bar{p}^m]$, which means profits only accrue from captive consumers. Since these are strictly concave, there is a contradiction.

Finally, suppose that $\bar{p}^m \leq \underline{p}^{nd} + s$, i.e. that none of the prices played after $\tilde{v}^m$ trigger search. Then, we can show that $\underline{p}^d$, which is strictly lower than $\bar{p}^m$ by the assumption that the firm with data is mixing, must be played after the low signal. If $\underline{p}^d$ were played after $\tilde{v}^H$, there would be a contradiction by hazard ratio ordering arguments (since no price triggers search).

Since $\underline{p}^d < \bar{p}^m$ and $\underline{p}^d$ is played after $\tilde{v}^L$, the strategy of searchers ($d(v)$) will be a cutoff rule, because the price distribution at the firm with data becomes strictly less favorable as a consumer's valuation increases. Thus, searchers will visit the firm with data only if their valuation is below $\bar{v}$. Because $\bar{p}^H \leq \bar{v}$ must hold (else the firm only sells to captive consumers for a subinterval of prices), demand from any price $p_j \in [\underline{p}^H, \bar{p}^H]$ is:

$$\Pi^H(p_j) = p_j \left[\rho \int_{p_j}^{\bar{v}} (1) Pr^H(v) dv + 0.5(1-\rho) \int_{p_j}^{1} (1) Pr^H(v) dv \right]$$

But this is strictly concave, so the firm with data would also never mix.

(iii) Suppose $\underline{p}^d = \underline{p}^{nd}$.

For prices in an open ball above the lowest price, no consumer that arrives at any firm will leave to search. Consider this open interval of prices, and call it $[\underline{p}^d, \underline{p}^d + \epsilon]$. If ϵ is set small enough, consumers with valuation in this open ball will generally never leave to search.

Even if some individual prices in this interval are played with positive probability, the preferences that consumers with $v \in [\underline{p}^d, \underline{p}^d + \epsilon]$ have over which firm to visit will be continuous in v. This can be shown by applying the dominated convergence theorem.

Suppose that there exists a $v' \in [\underline{p}^d, \underline{p}^d + \epsilon]$ such that a searcher with valuation v' strictly prefers to visit the firm with data. Then, consumers with valuation in an open ball with radius δ around v' will also strictly prefer to visit the firm with data first. As a result, setting any price $p_j \in [v', \underline{p}^d + \epsilon]$ will yield the following profits for the firm with data:

$$p_j\left[\rho\int_{p_j}^{v'+\delta} Pr^k(v)dv+\rho\int_{v'+\delta}^{1}\left[d(v)+(1-d(v))Pr(\hat{p}^{nd}(v) < p^{nd})\right]Pr^k(v)dv+0.5(1-\rho)\int_{p_j}^{1} Pr^k(v)dv\right]$$

This profit function is strictly concave in this domain, a contradiction to mixing indifference.

Similar arguments rule out that some searchers with $v \in [\underline{p}^d, \underline{p}^d+\epsilon]$ strictly prefer to visit the firm without data. Thus, they must all be indifferent and randomize by our tie-breaking rule. But then, the firm without data would make the following profits for any price $p_j \in [\underline{p}^d, \underline{p}^d+\epsilon]$:

$$p_j\left[\rho\int_{p_j}^{\underline{p}^d+\epsilon}(0.5)dv + \rho\int_{\underline{p}^d+\epsilon}^{1}\left[(1-d(v)) + d(v)Pr(\hat{p}^d(v) < p^d)\right]dv + 0.5(1-\rho)\int_{p_j}^{1}(1)dv\right]$$

But this profit function is strictly concave once more. Thus, the firm without data would have to set this lowest price with probability 1. If the firm with data sets this price with probability 1 as well, we have no MSE. Alternatively, it sets it with probability below 1. Then, all searchers with $v > \underline{p}^{nd}$ visit the firm without data and don't search. Thus, the firm with data would not mix, because it sells only to captive consumers for any of its prices.

A.10 Proof of corollary 2

Part 1: As $\rho \to 1$, assumptions 4 and 5 both hold.

Consider any $s > 0$. Recall that assumption 4 required that $p^{nd,s} + s > p^{H,M}$, where

$p^{nd,s}(\rho)$ solves the following:

$$\rho \int_{p^{nd,s}(\rho)+s}^{1} Pr(\tilde{v}^H|v)dv + 0.5(1-\rho)\left[\int_{p^{nd,s}(\rho)}^{1} dv - p^{nd,s}(\rho)\right] = 0$$

One can show that $p^{nd,s}(\rho) \to \max\{0.5, 1-s\}$ as $\rho \to 1$. This means that $\lim_{\rho\to 1}[p^{nd,s}(\rho) + s] \geq 1 > p^{H,M}$, i.e. assumption 4 is satisfied.

Now consider assumption 5, namely $\Pi^H(0.5; 0.5+0.5\rho) > 0.5(1-\rho)\Pi^{H,M}(p^{H,M})$. As $\rho \to 1$, the LHS goes to something strictly positive, while the LHS goes to 0. Thus, the assumption is satisfied as well.

Part 2: As $\rho \to 1$, there is no search on the equilibrium path and $\bar{v} \geq 0.5(1-\rho)$ holds in equilibrium by the previous results. Thus, $\bar{v} \to 1$ as $\rho \to 1$, which implies the result.

A.11 Proof of corollary 3

We consider $s < p^{H,1} - p^{nd,1}$. Then, the equilibrium $(p^{L,2}, p^{H,2}, p^{nd,2}, \bar{v}^2)$ is played, in which:

$$\hat{v}(p^{L,*}(\bar{v}^2), p^{nd,*}(\bar{v}^2) + s, p^{nd,*}(\bar{v}^2)) - \bar{v}^2 = 0$$

Consider two values of s for which $s < p^{H,1} - p^{nd,1}$ and call them s' and s'', with $s'' > s'$. Define the resulting equilibrium levels of $\bar{v}$ as $\bar{v}^2(s'') := \bar{v}^{2,''}$ and $\bar{v}^2(s') := \bar{v}^{2,'}$. I show that $\bar{v}^2(s'') \leq \bar{v}^2(s')$. If $\bar{v}^2(s') = 1$, the result is immediate.

Thus, suppose that $\bar{v}^{2,'} < 1$. Then, $\bar{v}^{2,'}$ must set the expected prices exactly equal. For $s'' > s'$, we thus have: $Pr^L(\bar{v}^{2,'})(p^{L,*}(\bar{v}^{2,'}) - p^{nd,*}(\bar{v}^{2,'})) + Pr^H(\bar{v}^{2,'})s'' > 0$. Because $p^{L,*}(\bar{v}^{2,'}) - p^{nd,*}(\bar{v}^{2,'}) < 0$ and $p^{nd,*}(\bar{v})$ $(p^{L,*}(\bar{v}))$ is falling (rising) in $\bar{v}$, this expression is rising in $\bar{v}$. As a result, $\bar{v}^{2,''} < \bar{v}^{2,'}$, which completes the proof.

B Proofs - section 5

B.1 Proof of proposition 7

If $e > 0$, it is optimal to exercise this right only if $v \geq p^a + e$. Suppose the right to anonymity is exercised by a positive measure of consumers. Thus, the corresponding information set for the firm with data is on-path and this firm understands that a consumer who has anonymized is a searcher and has $v > p^a + e$. Thus, there would be a profitable upward deviation from p^a.

Now consider $e = 0$ and suppose that a strictly positive measure of consumers exercises the right to anonymity.

Suppose $p^{nd} < p^a$. Then, any searcher with $v > p^{nd}$ would not visit the firm with data and utilize their right to data portability. If a consumer exercises this right, she must have $v < p^{nd}$. But then, setting the price p^a would be suboptimal, a contradiction.

Suppose $p^a \leq p^{nd}$. Then, all searchers weakly prefer to visit the firm with data. Suppose $p^L \neq p^H$. Then, $p^{min} < p^a < p^{max}$ must hold. But then, consumers with $v \leq p^a$ won't exercise the right to anonymity. Thus, there is a profitable upward deviation from p^a, as the firm with data knows that any consumer who anonymizes has a valuation strictly above p^a.

Thus, $p^L = p^H$ must hold. If p^a is not exactly equal, either no consumer will anonymize (contradiction) or all searchers anonymize (then contradiction to ordering, since valuations are uniform). The final case is hence $p^L = p^H = p^a$. By assumption, consumers then randomize between anonymizing and not anonymizing - and then, valuations would be uniform once more and we obtain several contradictions. For instance, $p^L = p^H$ would not be optimal.

B.2 Proof of proposition 8

There are five equilibrium prices - the prices of the firm with data (p^L, p^H), the uniform price of the firm without data (p^{nd}) and the signal prices at this firm $(p^{nd,L}, p^{nd,H})$. Consider an equilibrium in which (i) searchers with $v \in [0, v^t)$ visit the firm without data and port their data, and (ii) searchers with $v \in (v^t, 1]$ visit the firm without data but do not port their data.

Pricing & candidate existence:

We are searching for an equilibrium with the property $p^{nd,L} < p^{nd} < p^{nd,H}$ - such that the cutoff v^t can be interior. The prices $p^{L,nd}$ and $p^{H,nd}$ must, given v^t, solve:

$$p^{nd,k}(v^t) = arg \max_{p_j} \left[p_j \int_{p_j}^{v^t} \rho Pr^k(v) dv \right]$$

This optimal price will always be strictly below v^t. The price p^{nd} must maximize:

$$\Pi^{nd}(p_j; v^t) = p_j \left[\rho \int_{v^t}^{1} \mathbb{1}[p_j \leq v] dv + 0.5(1-\rho) \int_{0}^{1} \mathbb{1}[p_j \leq v] dv \right]$$

In order for the search behavior we posited to be optimal, we need to have $p^{nd} < v^t$. This, in turn, implies that $v^t \geq 0.5(1+\rho)$ must hold. For any $v^t \geq 0.5(1+\rho)$, the price p^{nd} will equal $p^{nd,*}(v^t)$. Because all prices must be below v^t in equilibrium, v^t must solve:

$$v^t = \underbrace{\sup\left\{v \in [0,1] : Pr^H(v)p^{nd,H}(v^t) + Pr^L(v)p^{nd,L}(v^t) - p^{nd}(v^t) < 0\right\}}_{:=\hat{v}^T(v^t)}$$

Previous arguments show that the function in this supremum is rising in v, which means we have a well-defined supremum. Moreover, $p^{nd,L}(v^t)$ and $p^{nd,H}(v^t)$ are both rising in v^t, while the uniform price $p^{nd}(v^t)$ is falling in this object. This proves that $\hat{v}^T(v^t)$ is falling in v^t.

This function is also continuous because all price functions are continuous in v^t. Together with two border conditions on $v^T(.)$, the intermediate value theorem guarantees the existence of a solution v^t to $\hat{v}^T(v^t) - v^t = 0$. These border conditions are: At $v^t = 0.5(1+\rho)$, we have $\hat{v}^T(0.5(1+\rho)) = 1 > 0.5(1+\rho)$. At $v^t = 1$, we have $\hat{v}^T(1) \leq 1$ by definition.

Thus, we know that $v^t = \hat{v}^T(v^t)$ holds at some $v^t \geq 0.5(1+\rho)$ and the prices at the firm with no data are set optimally, given this search behaviour. We also know that all implied optimal prices lie below v^t. In the supposed equilibrium, the firm with data is only visited by captive consumers and thus sets the monopoly prices.

Search: If $v^t = \hat{v}^T(v^t)$, it is optimal for consumers to search in the posited way.

For all searchers with $v < v^t$, we know that porting the data will be strictly better than remaining anonymous at the firm without data. This is because the expected price when porting the data lies below p^{nd} for a consumer with $v = p^{nd,H}$. For any consumer with $v > p^{nd,H}$, the preferences for porting are strictly falling in v and switch sign at v^t.

For consumers with $v > v^t$, it is better to remain anonymous at the firm without data than to port the data. Now I argue that, for any searcher, it is better to port the data to the firm without data than to visit the firm with data. In equilibrium, the firm with data set the prices $p^L = p^{L,M}$ and $p^H = p^{H,M}$.

Since $v^t \leq 1$, we know that $p^{L,nd} \leq p^L = p^{L,M}$ and $p^{H,nd} \leq p^H = p^{H,M}$. Thus, any searcher prefers porting the data to visiting the firm with data since:

$$Pr^H(v)\max\{v-p^{H,nd},0\}+Pr^L(v)\max\{v-p^{L,nd},0\} \geq Pr^H(v)\max\{v-p^H,0\}+Pr^L(v)\max\{v-p^L,0\}$$

Thus, all searchers with $v < v^t$ port the data. All searchers with $v > v^t$ prefer to visit the firm without data anonymously over porting the data, which they in turn prefer to visiting the firm with data. Hence, the postulated search behavior is optimal.

C Proofs - section 6

C.1 Proof of lemma 2

For a given valuation v, the random variable $\tilde{v}$ is uniformly distributed with mean v and support $[v - \bar{\epsilon}, v + \bar{\epsilon}]$. Hence, the conditional density is $f_{\tilde{v}|v} = 1/2\bar{\epsilon}$ for $\tilde{v} \in [v - \bar{\epsilon}, v + \bar{\epsilon}]$ and 0 otherwise.

Part 1: The price the firm with data will set is weakly rising in $\tilde{v}$.

I consider the cumulative density function of v, conditional on $\tilde{v}$. The density of arriving consumers valuations is $g(v)$, which is strictly positive and bounded throughout. The probability that the consumer's v is below x, conditional on this consumer generating $\tilde{v}^r$, is:

$$Pr(v \leq x|\tilde{v}^r) = \int_0^x f_{v|\tilde{v}}(v|\tilde{v}^r)dv = \int_0^x \frac{f_{\tilde{v}|v}(\tilde{v}^r|v)g(v)}{f_{\tilde{v}}(\tilde{v}^r)}dv = \int_0^x \frac{\mathbb{1}[v \in [\tilde{v}^r - \bar{\epsilon}, \tilde{v}^r + \bar{\epsilon}]](1/2\bar{\epsilon})g(v)}{f_{\tilde{v}}(\tilde{v}^r)}dv$$

Call the corresponding cdf $B(x|\tilde{v})$. The corresponding density is: $b(x|\tilde{v}) = (1/f_{\tilde{v}}(\tilde{v}))\mathbb{1}[x \in [\tilde{v} - \bar{\epsilon}, \tilde{v} + \bar{\epsilon}]](1/2\bar{\epsilon})g(x)$. We can define the hazard ratio:

$$h(x|\tilde{v}) = \frac{\mathbb{1}[x \in [\tilde{v} - \bar{\epsilon}, \tilde{v} + \bar{\epsilon}]](1/2\bar{\epsilon})g(x)}{\int_x^1 \mathbb{1}[v \in [\tilde{v} - \bar{\epsilon}, \tilde{v} + \bar{\epsilon}]](1/2\bar{\epsilon})g(v)dv}$$

Consider two signal realizations $\tilde{v}^1, \tilde{v}^2$ with $\tilde{v}^1 < \tilde{v}^2$. Suppose that there is an overlap between the supports. If there is no overlap, i.e. $\tilde{v}^1 + \bar{\epsilon} \leq \tilde{v}^2 - \bar{\epsilon}$, then $p(\tilde{v}^1) < p(\tilde{v}^2)$ must hold, because each price must be in the support of valuations corresponding to a signal. The interval of valuations where the two signals overlap is $[\tilde{v}^2 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon}]$.

Suppose, for a contradiction, that $p^d(\tilde{v}^2) := p^2 < p^d(\tilde{v}^1) := p^1$. The price p^2 must satisfy $p^2 \geq \tilde{v}^2 - \bar{\epsilon}$ - else, there is a profitable upward deviation. Similarly, $p^1 < \tilde{v}^1 + \bar{\epsilon}$ must hold. Thus, both prices must lie in $[\tilde{v}^2 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon}]$. For any $x \in [\tilde{v}^2 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon}]$, the hazard ratios satisfy $h(x|\tilde{v}^1) > h(x|\tilde{v}^2)$, because $\tilde{v}^2 > \tilde{v}^1$. But then, $p^d(\tilde{v})$ must be weakly rising in the signal $\tilde{v}$ - else, there is a profitable deviation from one of these prices.

Part 2: Consider any $v > 0$. There exists a $\tilde{v}' > \tilde{v}^{lb}(v) := v - \bar{\epsilon}$ such that $p^d(\tilde{v}') < v$.

Suppose there exists no such $\tilde{v}' > \tilde{v}^{lb}(v)$. By implication, $p^d(\tilde{v}') \geq v$ must hold for any $\tilde{v}' > \tilde{v}^{lb}(v)$. Consider the profits the firm with data would make after any $\tilde{v}'$ for some $p_j \in [\max\{\tilde{v}' - \bar{\epsilon}, 0\}, \tilde{v}' + \bar{\epsilon}]$, which are:

$$\Pi(p_j; \tilde{v}') = p_j \frac{1}{f_{\tilde{v}}(\tilde{v}')} \int_{p_j}^{\tilde{v}'+\bar{\epsilon}} (1/2\bar{\epsilon}) g(v) dv$$

We can take the following limit of the postulated equilibrium profits:

$$\lim_{\tilde{v}' \to \tilde{v}^{lb}(v)} \Pi(p^d(\tilde{v}'); \tilde{v}') = \big(\lim_{\tilde{v}' \to \tilde{v}^{lb}(v)} p^d(\tilde{v}') \big) \frac{1}{f_{\tilde{v}}(\tilde{v}')} \int_{\lim_{\tilde{v}' \to \tilde{v}^{lb}(v)} p^d(\tilde{v}')}^{\tilde{v}^{lb}(v)+\bar{\epsilon}} (1/2\bar{\epsilon}) g(v) dv = 0$$

The latter condition holds because $\tilde{v}^{lb}(v) + \bar{\epsilon} = v$ and $\lim_{\tilde{v}' \to \tilde{v}^{lb}(v)} p^d(\tilde{v}') \geq v$ by assumption. Thus, profits converge to zero in this case as $\tilde{v} \to \tilde{v}^{lb}(v)$.

Alternatively, the firm with data could set the price $0.5\big(\tilde{v}' + \bar{\epsilon}\big)$, at which it would be guaranteed to make positive profits. For any $\tilde{v}$, the profits from this approach are:

$$\Pi(0.5\big(\tilde{v}' + \bar{\epsilon}\big); \tilde{v}') = 0.5\big(\tilde{v}' + \bar{\epsilon}\big) \left[\frac{1}{f_{\tilde{v}}(\tilde{v}')} \int_{\max\{\tilde{v}'-\bar{\epsilon}, 0\}}^{\tilde{v}'+\bar{\epsilon}} (1/2\bar{\epsilon}) g(v) \mathbb{1}[v > 0.5(\tilde{v}' + \bar{\epsilon})] dv \right]$$

This remains strictly positive even in the limit at $\tilde{v}' \to \tilde{v}^{lb}(v)$ because $v > 0$. Thus, we would have a profitable deviation for some signal close enough to $\tilde{v}^{lb}(v)$, and thus a contradiction.

Part 3: For any $v > 0$, the expected utility of visiting the firm with data is strictly positive.

On the interval $\tilde{v} \in [\tilde{v}^{lb}(v), \tilde{v}']$, we will have $p^d(\tilde{v}) < v$ (by monotonicity of $p^d(\tilde{v})$). Thus, the expected utility of the consumer, namely $\int_{v-\bar{\epsilon}}^{v+\bar{\epsilon}} \max\{v - p^d(v), 0\}(1/2\bar{\epsilon}) dv$, is positive.

Part 4: The expected utility of visiting the firm with data is Lipschitz continuous.

The utility of visiting the firm with data is $U^d(v) = \int_{v-\bar{\epsilon}}^{v+\bar{\epsilon}} \max\{v - p^d(\tilde{v}), 0\}(1/2\bar{\epsilon}) d\tilde{v}$. There exists $K = \big(\frac{2}{2\bar{\epsilon}} + 1\big) \in \mathbb{R}^+$ such that $|U^d(v^1) - U^d(v^2)| \leq K|v^1 - v^2|$ for any v^1, v^2.

Part 5: By the previous results, the search rule of consumers is a cutoff rule.

Consider consumers with $v \leq p^{nd}$ receive 0 utility at the firm with data, but strictly positive utility at the firm with data (by previous arguments), so they all prefer the firm without data. If $v > p^{nd}$, the expected utility of visiting the firm without data is $U^{nd}(v) = v - p^{nd}$. It was shown that $U^d(v)$ is Lipschitz continuous, hence differentiable almost everywhere. The derivative of $U^d(v)$ is strictly below 1, since the price distribution at the firm with data changes in v. Thus, the preference for the firm with data, namely $U^d(v) - U^{nd}(v)$, is strictly falling in v. This establishes the result.

C.2 Proof of proposition 9

Part 1: If searchers search according to a cutoff rule, the optimal price function at the firm with data is:

$$p^d(\tilde{v}) = \begin{cases} 0.5(\tilde{v} + \bar{\epsilon}) & \tilde{v} \in [-\bar{\epsilon}, 3\bar{\epsilon}] \\ (\tilde{v} - \bar{\epsilon}) & \tilde{v} \in [3\bar{\epsilon}, 1 + \bar{\epsilon}] \end{cases}$$

Part 1a: Monopoly pricing

The maximization problem of a monopolist with data upon observing $\tilde{v}^1 \in [\bar{\epsilon}, 1 - \bar{\epsilon}]$ is to maximize the following through choice of $p_j \in [\tilde{v}^1 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon}]$:

$$\Pi(p_j; \tilde{v}^1) = p_j \int_{p_j}^{\tilde{v}^1 + \bar{\epsilon}} (1/2\bar{\epsilon}) dv = p_j \left[\tilde{v}^1 + \bar{\epsilon} - p_j\right]$$

The first-order condition of this expression is equal to 0 at $p(\tilde{v}^1) = 0.5(\tilde{v}^1 + \bar{\epsilon})$. Whether this constitutes an interior price depends on $\bar{\epsilon}$. The lower bound is $\tilde{v}^1 - \bar{\epsilon}$. At $\tilde{v}^1 = 3\bar{\epsilon}$, we have that $p(3\bar{\epsilon}) = 2\bar{\epsilon}$, which is exactly equal to the lower bound. Thus, we can compute the following pricing schedule on $[\bar{\epsilon}, 1 - \bar{\epsilon}]$ because $3\bar{\epsilon} < 1 - \bar{\epsilon}$ by the fact that $\bar{\epsilon} < 0.25$:

$$p^*(\tilde{v}^1) = \begin{cases} 0.5(\tilde{v}^1 + \bar{\epsilon}) & \tilde{v}^1 \in [\bar{\epsilon}, 3\bar{\epsilon}] \\ (\tilde{v}^1 - \bar{\epsilon}) & \tilde{v}^1 \in [3\bar{\epsilon}, 1 - \bar{\epsilon}] \end{cases}$$

Analogous arguments establish the optimal prices for $\tilde{v}^1 < \bar{\epsilon}$ and $\tilde{v}^1 > 1 - \bar{\epsilon}$. Summing up, the optimal prices of a monopolist are given by the schedule listed at the beginning.

Part 1b: In the competitive equilibrium, the optimal price function remains unchanged.

In equilibrium, searchers will visit the firm with data if $v < \bar{v}$ and vice versa. Because they can all obtain strictly positive utility at the firm with data, $p^{nd} < \bar{v}$ must hold. Thus, searchers will push up p^{nd}, and thus $\bar{v} > p^{nd} \geq 0.5$ must hold in an equilibrium.

Consider a signal $\tilde{v}^1 < 3\bar{\epsilon}$, where $\tilde{v}^1 + \bar{\epsilon} < 4\bar{\epsilon} < 0.5 < \bar{v}$. Thus, all searchers with $v \in [\tilde{v}^1 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon}]$ arrive at the firm with data. The optimal price thus remains $0.5(\tilde{v}^1 + \bar{\epsilon})$.

Thus, consider a signal $\tilde{v}^1 > 3\bar{\epsilon}$, for which the upper bound of valuations may lie above $\bar{v}$. If $\tilde{v}^1 + \bar{\epsilon} \leq \bar{v}$, as before, then nothing changes - the optimal price is $\tilde{v}^1 - \bar{\epsilon}$.

Suppose instead that $\bar{v} < \tilde{v}^1 + \bar{\epsilon}$. Then, there are two further possibilities.

(i) $\bar{v} \leq \tilde{v}^1 - \bar{\epsilon}$. Then, no searchers would be arriving at this firm and generate $\tilde{v}^1$ - which means the optimal price is $\tilde{v}^1 - \bar{\epsilon}$ (the sale is only made to captive consumers).

(ii) $\bar{v} \in (\tilde{v}^1 - \bar{\epsilon}, \tilde{v}^1 + \bar{\epsilon})$: Some arriving searchers will generate $\tilde{v}^1$.

Suppose we have a price at which the sale is only made to captive consumers, which implies that $p_j \geq \bar{v} > \tilde{v}^1 - \bar{\epsilon}$. But because $\tilde{v}^1 > 3\bar{\epsilon}$, there is a downward deviation (since this raises the profits from captive consumers). Thus, consider a price $p_j < \bar{v}$, at which:

$$\Pi(p_j; \tilde{v}^1) = p_j \left[\rho \int_{p_j}^{\bar{v}} (1/2\bar{\epsilon}) dv + 0.5(1-\rho) \int_{p_j}^{\min\{\tilde{v}^1 + \bar{\epsilon}, 1\}} (1/2\bar{\epsilon}) dv \right]$$

Thus, the derivative at any p_j will be weakly below the monopoly case, which means that the optimal price must also be directly at the lower bound here.

Part 2: In equilibrium, $\bar{v} \geq 0.5(1+\rho)$ must hold.

For $\bar{v} < 0.5(1+\rho)$, we have $p^{nd,*}(\bar{v}) \geq \bar{v}$. However, this is not consistent with optimal search behaviour. Any searcher with $v > 0$ receives strictly positive utility at the firm with data. Thus, $p^{nd} < \bar{v}$ must hold in an equilibrium. But such a price would never be optimally set if $\bar{v} < 0.5(1+\rho)$.

Part 3: Establishing equilibrium existence.

For any $\bar{v} \geq 0.5(1+\rho)$, the firm with data will price according $p^*(\tilde{v})$. The strategy $d(v)$ is a cutoff rule. The price p^{nd} depends on $\bar{v}$. I work with the following object:

$$\hat{v}^C(\bar{v}) = \sup\left\{v \in [0,1] : \int_{v-\bar{\epsilon}}^{v+\bar{\epsilon}} \max\{v - p^d(\tilde{v}), 0\}(1/2\bar{\epsilon})d\tilde{v} - \max\{v - p^{nd}(\bar{v}), 0\} \geq 0\right\}$$

Since the function in this supremum is strictly falling in v, the supremum separates the groups of searchers who visit the firm with data from those who visit the firm without data.

As before, we work with boundary conditions and the intermediate value theorem. At $\bar{v} = 0.5(1+\rho)$, it holds that $p^{nd} = 0.5(1+\rho)$, which means that $\hat{v}^C(\bar{v}) > \bar{v}$ will hold at $\bar{v} = 0.5(1+\rho)$. At $\bar{v} = 1$, $\hat{v}^C(1) \leq 1$ holds by construction. Finally, the utility of visiting the firm with data will not be affected by changes in $\bar{v}$, but only p^{nd} responds to changes in $\bar{v}$. This establishes continuity of $\hat{v}^C(\bar{v})$, which confirms the existence of an equilibrium.

C.3 Proof of lemma 3

Because there is no search after visiting the first firm by assumption, previous arguments show that any equilibrium must satisfy the ordering $p^{LW} \leq p^{HW}$ and $p^{LB} \leq p^{HB}$.

Moreover, there exists no simple equilibrium in which all consumers randomize under assumption 4. If all consumers randomize, $p^{LB} = p^{LB,M} < p^{LW,M} = p^{LW}$ would be optimally set - but then, searchers with $v \in (p^{LB}, p^{LW})$ would not randomize.

Part 1: In a simple equilibrium, $p^{LB} \leq p^{LW}$ must hold.

Suppose, for a contradiction, that $p^{LW} < p^{LB}$. We know that all consumers with $v < p^{LB}$ will surely visit the firm with worse data.

Suppose that $p^{HW} \leq p^{LB}$ holds as well. Given that $p^{LW} \leq p^{HW}$ and $p^{LB} \leq p^{HB}$ must hold as well, $p^{LW} \leq p^{HW} \leq p^{LB} \leq p^{HB}$ holds. Then, all searchers will visit the firm with worse data (since $p^{LW} < p^{LB}$), which would then imply a contradiction, since $p^{LB} < p^{LW}$ holds true when the valuations of consumers are uniformly distributed by assumption 6.

Thus, the only possible equilibria with $p^{LW} < p^{LB}$ must satisfy $p^{LB} < p^{HW}$. Then, there are two possibilities now (i) $p^{LW} < p^{LB} < p^{HW} \leq p^{HB}$, and (ii) $p^{LW} < p^{LB} \leq p^{HB} < p^{HW}$. Neither of these can constitute an equilibrium, as I will show now.

(i) Ruling out $p^{LW} < p^{LB} < p^{HW} \leq p^{HB}$.

Note first that all consumers with $v \leq p^{LB}$ will surely visit the firm with worse data. The expected utilities of consumers with $v \in [p^{HB}, 1]$ are:

$$U^B(v) = v - \left(Pr^{HB}(v)p^{HB} + Pr^{LB}(v)p^{LB}\right) \quad ; \quad U^W(v) = v - \left(Pr^{HW}(v)p^{HW} + Pr^{LW}(v)p^{LW}\right)$$

For a consumer with $v > 0.5$, the expected price at the firm with better data is higher:

$$Pr^{HB}(v)p^{HB} + Pr^{LB}(v)p^{LB} > Pr^{HW}(v)p^{HB} + Pr^{LW}(v)p^{LB} > Pr^{HW}(v)p^{HW} + Pr^{LW}(v)p^{LW}$$

But this yields a contradiction. Consumers with $v > \max\{p^{HB}, 0.5\}$ will prefer visiting the firm with worse data, because the expected price there is lower for them. In any equilibrium, $p^{HB} < 1$ must hold. Thus, both consumers with valuations $v \in (\max\{p^{HB}, 0.5\}, 1]$ and low-valuation consumers prefer the firm with worse data. Either the definition of the simple equilibrium fails or all consumers prefer the firm with worse data - in which case our results on the resulting optimal prices under uniform valuations imply that $p^{LB} < p^{LW}$, a contradiction.

(ii) Ruling out $p^{LW} < p^{LB} \leq p^{HB} < p^{HW}$.

Once again, all searchers with valuations $v \leq p^{LB}$ surely visit the firm with worse data. Thus, the cutoff $\bar{v}$ in a simple equilibrium must be such that all consumers with $v > \bar{v}$ visit the firm with better data.

Suppose, for a contradiction, that $\bar{v} < p^{HB}$. Then, no searchers will buy at the firm with worse data (since searchers only visit the firm with worse data if $v \leq \bar{v}$), i.e. $p^{HW} = p^{HW,M}$. The price of the firm with better data is strictly above $\bar{v}$, i.e. has to satisfy $p^{HB} = p^{HB,M}$. However, we know that $p^{HW,M} < p^{HB,M}$, which implies that $p^{HW} < p^{HB}$, a contradiction.

Suppose, instead, that $\bar{v} \in [p^{HB}, p^{HW})$. Then, searchers put upward pressure on p^{HB} and vice versa. Thus, the two prices will optimally satisfy $p^{HW} < p^{HB}$, a contradiction.

Thus, we have ruled out any candidate for an equilibrium in which $p^{LW} < p^{LB}$.

Part 2: In a simple equilibrium (where $p^{LB} \leq p^{LW}$ by part 1), the cutoff $\bar{v}$ must be such that all consumers with $v > \bar{v}$ visit the firm with worse data.

In such an equilibrium, all consumers with $v \leq p^{LW}$ surely visit the firm with better data. If all consumers weakly prefer the firm with better data, then their strategy can be described by cutoff rule $\bar{v} = 1$, where they visit the firm with better data if and only if $v \leq \bar{v}$.

Suppose there exists some consumer who strictly prefers the firm with worse data. Then, any cutoff $\bar{v}$ must be interior, i.e. $\bar{v} \in (0, 1]$. Because all consumers with $v \leq p^{LW}$ surely visit the firm with better data, there exists no $\bar{v}$ such that all consumers with $v < \bar{v}$ visit the firm with worse data and vice versa.

Part 3: If $\bar{v} < \bar{v}^{LW}$, $p^{LW}(\bar{v}) \geq \bar{v}$ will be optimally set by the firm with worse data.

Consider the function $H(\bar{v}) := \rho \int_{\bar{v}}^{1} Pr^{LW}(v)dv + 0.5(1-\rho)\Bigg[\int_{\bar{v}}^{1} Pr^{LW}(v)dv - \bar{v}Pr^{LW}(\bar{v})\Bigg]$.

Our assumption on concavity of the low signal monopoly profit function implies that this function is strictly decreasing in $\bar{v}$. An appropriate $\bar{v}^{LW}$ that sets the above equation to zero exists by the intermediate value theorem.

Now consider a $\bar{v} < \bar{v}^{LW}$, at which $H(\bar{v}) > 0$. Consider the optimal pricing problem of the firm with worse data when observing the low signal. For prices $p_j < \bar{v}$, the profits are:

$$\Pi^{LW}(p_j) = \rho p_j \int_{\bar{v}}^{1} Pr^{LW}(v)dv + 0.5(1-\rho)p_j \int_{p_j}^{1} Pr^{LW}(v)dv$$

The derivative $\frac{\partial \Pi^{LW}(p_j)}{\partial p_j}$ is strictly decreasing in p_j by previous arguments. At $p_j = \bar{v}$, we know that the (left) derivative is strictly positive because $\bar{v} < \bar{v}^{LW}$. Thus, this derivative must be strictly positive for all $p_j < \bar{v}$. Thus, $p^{LW}(\bar{v}) \geq \bar{v}$ must hold.

Part 4: In a simple equilibrium, $\bar{v}^{LW} \leq \bar{v}$ must hold.

Suppose, for a contradiction, that $\bar{v} < \bar{v}^{LW}$. Then, we have established that $p^{LW}(\bar{v}) \geq \bar{v}$ will hold. Note that $p^{HW}(\bar{v}) \geq p^{LW}(\bar{v})$ will generally hold. Also recall that $p^{LB} \leq p^{LW}$ must hold in a simple equilibrium.

If $p^{LB} < p^{LW}$, a consumer with $v = p^{LW}$ will strictly prefer to visit the firm with better data - and by continuity arguments, so will consumers with valuation v just above p^{LW}. This represents a contradiction to the properties of $\bar{v}$. This is because $\bar{v} \leq p^{LW}$, but consumers with a valuation in an open ball above p^{LW}, i.e. with $v > \bar{v}$, would strictly prefer to visit the firm with better data but visit the firm with worse data in equilibrium.

Suppose that $p^{LB} = p^{LW}$. There exists no simple equilibrium in which a firm sets a uniform price. Thus, $p^{LB} < p^{HB}$ and $p^{LW} < p^{HW}$ must hold in such an equilibrium.

Suppose that $p^{LB} = p^{LW} < 0.5$. Then, all consumers with valuation in an open ball above p^{LW} will strictly prefer to visit the firm with better data, because they receive the low signal price there with a higher probability, a contradiction because $\bar{v} \leq p^{LW}$.

Suppose alternatively that $p^{LB} = p^{LW} \geq 0.5$. Then, consumers with $v < 0.5$ will visit the firm with better data (since they offer lowest the equilibrium price with higher probability to them) and consumers with v just above 0.5 will visit the firm with worse data, since they receive the low signal price with higher probability at the firm with worse data. Thus, this cutoff must then be exactly equal to $\bar{v} = 0.5$ in a simple equilibrium. Searchers visit the firm with better data if and only if $v < \bar{v} = 0.5$. But then, the optimal price of the firm with better data is strictly below 0.5, a contradiction.

C.4 Proof of proposition 10

Part 1: For any $\bar{v} \geq \max\{p^{HB}, \bar{v}^{LW}\}$, the function $\hat{v}^X(\bar{v})$ will be continuous.

The prices of the firm with better data are both strictly below $\bar{v}$ because $p^{HB} \leq \bar{v}$, which implies that $p^{L,*}(\bar{v})$ and $p^{H,*}(\bar{v})$ must solve the FOCs and be continuous.

We can also generally prove continuity of the prices of the firm with worse data. The low signal price has to be weakly below $\bar{v}$ and always solve a FOC that is continuous for any $\bar{v} \geq \bar{v}^{LW}$ - this establishes this part of the result. The high signal price is also continuous. It solves a FOC for all $\bar{v} \geq \bar{v}^{HW}$. For any $\bar{v} \in [p^{HW,M}, \bar{v}^{HW})$, the optimal price is $\bar{v}$. For any $\bar{v} < p^{HW,M}$, the optimal price is $p^{HW,M}$.

Part 2: Suppose $p^{HB} \leq \bar{v}^{LW}$. A solution $\bar{v} \in [\bar{v}^{LW}, 1]$ to $\hat{v}^X(\bar{v}) = \bar{v}$ exists.

Then, our assumption tells us that the expected price functions at $\bar{v} = \bar{v}^{LW}$ are such that $\hat{v}^X(\bar{v}^{LW}) > \bar{v}^{LW}$. For the prices at $\bar{v} = \bar{v}^{LW}$, we know that $EP^B(\bar{v}^{LW}; \bar{v}^{LW}) < EP^W(\bar{v}^{LW}; \bar{v}^{LW})$ by assumption, which implies that consumers with v in an open ball around $v = \bar{v}^{LW}$ would have a strictly lower expected price at the firm with better data, which establishes that the supremum of the corresponding set must lie above $\bar{v}^{LW}$. At $\bar{v} = 1$, we know that $\hat{v}^X(1) \leq 1$. Because $p^{HB} \leq \bar{v}^{LW}$, we are guaranteed continuity of this function in the interval $[\bar{v}^{LW}, 1]$ and hence existence of a solution.

Part 3: Suppose $\bar{v}^{LW} < p^{HB,M}$. A solution $\bar{v} \in [p^{HB,M}, 1]$ to $\hat{v}^X(\bar{v}) = \bar{v}$ exists.

Our assumption guarantees that $\hat{v}^X(p^{HB,M}) > p^{HB,M}$ by previous arguments. Thus, the supremum of the corresponding set must lie above $p^{HB,M}$. Continuity of the function proves existence of an appropriate solution, together with the fact that$\hat{v}^X(1) > 1$.

Part 4: Consider the $\bar{v} \in [\bar{v}^{LW}, 1]$ with $\hat{v}^X(\bar{v}) = \bar{v}$. By construction, the prices at this $\bar{v}$ are optimal. By assumption, the search behavior is optimal, i.e. we have an equilibrium.

C.5 Proof of lemma 4

Part 1: Density of valuations at the firms

Note that the type $\theta \sim U[0,1]$ of any consumer is a random variable with density $f(\theta) = 1$. Moreover, the event "ds" denotes that a consumer is a searcher, while the event "lc" denotes the probability that a consumer is captive.

Consider any interval I. We can write the probability that $\theta^{nd} \in I$ as follows:

$$Pr(\theta^{nd} \in I) = Pr(\theta \in I | I^{nd}) = \frac{Pr(\theta \in I \wedge I^{nd})}{Pr(I^{nd})} = \frac{1}{Pr(I^{nd})} \int_0^1 Pr(\theta \in I \wedge I^{nd} | \theta) d\theta$$

Note that:

$$Pr(\theta \in I \wedge I^{nd} | \theta \wedge ds) Pr(ds|\theta) = \rho \mathbb{1}\big[\theta \in I\big] \big[\mathbb{1}[\theta > \bar{\theta}](1-g^H) + \mathbb{1}[\theta = \bar{\theta}](1-\bar{g}) + \mathbb{1}[\theta < \bar{\theta}](1-g^L)\big]$$

Similarly, we have that $Pr(\theta \in I \wedge I^{nd} | \theta \wedge lc) Pr(l|\theta) = 0.5(1-\rho)\mathbb{1}\big[\theta \in I\big]$. Thus:

$$Pr(\theta^{nd} \in I) = \int_I \Bigg(\underbrace{\frac{\rho\big[\mathbb{1}[\theta > \bar{\theta}](1-g^H) + \mathbb{1}[\theta = \bar{\theta}](1-\bar{g}) + \mathbb{1}[\theta < \bar{\theta}](1-g^L)\big] + 0.5(1-\rho)}{Pr(I^{nd})}}_{:= f^{nd}(\theta)} \Bigg) d\theta$$

Now consider the firm with data. The valuations of consumers that visit the firm with data and generate the signal $\tilde{v}^k$ is a random variable - call this θ^k. This is also an absolutely continuous random variable - the following holds for any interval I:

$$Pr(\theta^k \in I) = Pr(\theta \in I | I^k) = \frac{Pr(\theta \in I \wedge I^k)}{Pr(I^k)} = \frac{1}{Pr(I^k)} \int_0^1 Pr(\theta \in I \wedge I^k | \theta) d\theta$$

Note that:

$$Pr(\theta \in I \wedge I^k | \theta \wedge ds) Pr(ds|\theta) = \rho \mathbb{1}\big[\theta \in I\big] Pr^k(\theta) \big[\mathbb{1}[\theta > \bar{\theta}] g^H + \mathbb{1}[\theta = \bar{\theta}] \bar{g} + \mathbb{1}[\theta < \bar{\theta}] g^L\big]$$

Note further that $Pr(\theta \in I \wedge I^k | \theta \wedge lc) Pr(lc|\theta) = \mathbb{1}[\theta \in I] Pr^k(\theta) 0.5(1-\rho)$. Thus:

$$Pr(\theta^k \in I) = \int_I \underbrace{\frac{1}{Pr(I^k)} \Big(\rho Pr^k(\theta) \big[\mathbb{1}[\theta > \bar{\theta}] g^H + \mathbb{1}[\theta = \bar{\theta}] \bar{g} + \mathbb{1}[\theta < \bar{\theta}] g^L\big] + Pr^k(\theta) 0.5(1-\rho)\Big)}_{:= f^k(\theta)} d\theta$$

All densities are bounded from above and measurable, i.e. integrable, and thus well-defined.

Part 2: Calculating the virtual valuation functions

I omit the conditioning on $(\bar{\theta}, g)$ in the following arguments for ease of exposition.

Consider first the firm without data. For any $\theta < \bar{\theta}$, we can note that:

$$1 - F^{nd}(\theta) = \frac{1}{Pr(I^{nd})} \int_\theta^{\bar{\theta}} \big(\rho(1-g^L) + 0.5(1-\rho)\big) dx + \frac{1}{Pr(I^{nd})} \int_{\bar{\theta}}^1 \big(\rho(1-g^H) + 0.5(1-\rho)\big) dx$$

$$\implies J^{nd}(\theta) = \theta - \frac{\int_\theta^{\bar{\theta}} \big(\rho(1-g^L) + 0.5(1-\rho)\big) dx + \int_{\bar{\theta}}^1 \big(\rho(1-g^H) + 0.5(1-\rho)\big) dx}{\rho(1-g^L) + 0.5(1-\rho)}$$

For $\theta > \bar{\theta}$, the virtual valuation at the firm without data can be calculated as follows:

$$1 - F^{nd}(\theta) = \frac{1}{Pr(I^{nd})} \int_{\bar{\theta}}^1 \big(\rho(1-g^H) + 0.5(1-\rho)\big) dx \implies J^{nd}(\theta) = \theta - \frac{\int_\theta^1 \big(\rho(1-g^H) + 0.5(1-\rho)\big) dx}{\rho(1-g^H) + 0.5(1-\rho)}$$

One can show that $J^{nd}(\theta)$ is always piecewise strictly increasing.

Now let's calculate the virtual valuations at the firm with data. The virtual valuation takes the following form when $\theta < \bar{\theta}$:

$$J^k(\theta) = \theta - \frac{1 - F^k(\theta)}{f^k(\theta)} = \theta - \frac{\int_\theta^{\bar{\theta}} \big(\rho g^L + 0.5(1-\rho)\big) Pr^k(x) dx + \int_{\bar{\theta}}^1 \big(\rho g^H + 0.5(1-\rho)\big) Pr^k(x) dx}{\big(\rho g^L + 0.5(1-\rho)\big) Pr^k(\theta)}$$

The virtual valuation takes the following form when $\theta > \bar{\theta}$:

$$J^k(\theta) = \theta - \frac{1 - F^k(\theta)}{f^k(\theta)} = \theta - \frac{\int_\theta^1 \big(\rho g^H + 0.5(1-\rho)\big) Pr^k(x)dx}{\big(\rho g^H + 0.5(1-\rho)\big) Pr^k(\theta)}$$

One can show that $J^H(\theta)$ is always piecewise strictly increasing.

Part 3: Setting up the expected revenues of the firms

Any mechanism is incentive compatible if and only of it satisfies both the integrability and the monotonicity condition. This follows because one can apply the envelope theorem from Milgrom & Segal (2002). Using standard arguments, the expected revenue becomes:

$$-\mathbb{E}[t(\theta)] = -U(0) + \int_0^1 q(\theta)\left(\theta - \frac{1 - F^{nd}(\theta)}{f^{nd}(\theta)}\right) f^{nd}(\theta) d\theta$$

Similar arguments prove that the firm with data, when observing a given signal, would also face an expected revenue function that is equal to:

$$-\mathbb{E}[t^k(\theta)] = -U^k(0) + \int_0^1 q^k(\theta)\left(\theta - \frac{1 - F^k(\theta)}{f^k(\theta)}\right) f^k(\theta) d\theta$$

Part 4: Ordering of cutoffs when consumers visit randomly:

The virtual valuations at the firm with data and without data are:

$$J^k(\theta) = \theta - \int_\theta^1 \big(Pr^k(x)/Pr^k(\theta)\big)dx \quad : \quad J^{nd}(\theta) = \theta - \int_\theta^1 (1)dx$$

Because $Pr^H(\theta)$ is strictly increasing, $Pr^L(x)/Pr^L(\theta) < 1 < Pr^H(x)/Pr^H(\theta)$ holds $\forall x > \theta$: This implies that, for any $\theta \in (0,1)$, the ordering $J^H(\theta) < J^{nd}(\theta) < J^L(\theta)$ holds.

To see this, note that the functions $J^{nd}(\theta)$ and $J^H(\theta)$ are monotonically increasing and continuous. Because $J^H(1) > 0$, we know that there must exist a $\hat{\theta}^H < 1$ such that $J^H(\hat{\theta}^H) = 0$ and that the virtual valuation is strictly positive for all $\theta \geq \hat{\theta}^H$. At $\theta > \hat{\theta}^H$, both other virtual valuations will be strictly positive, which implies that the associated cutoffs must both lie strictly below $\hat{\theta}^H$. That $\hat{\theta}^L < \hat{\theta}^{nd}$ must hold follows by analogous arguments.

Part 5: In any simple equilibrium, there must exist a $\bar{\theta}$ such that all consumers with $\theta < \bar{\theta}$ visit the firm with data and vice versa.

Part 5a: There exists no simple equilibrium in which all searchers randomize between firms.

In that setting, $\hat{\theta}^L < \hat{\theta}^{nd}$. The firm without data will offer quality 1 to all consumers with $\theta > \hat{\theta}^{nd}$ by monotonicity of $J^{nd}(\theta)$. Then, the firm with data must offer 0 quality to all consumers with $\theta \leq \hat{\theta}^{nd}$ - else, these consumers would not randomize. But then, we obtain a contradiction, as the utility of consumers with $\theta \geq \hat{\theta}^{nd}$ at the firm without data is $\theta - \hat{\theta}^{nd}$, but strictly below $(\theta - \hat{\theta}^{nd})$ at the firm with data since $\hat{\theta}^{nd} < \hat{\theta}^H$. Thus, they would not randomize, a contradiction.

Part 5b: Initial steps:

Suppose we are in a simple equilibrium where there exists no $\bar{\theta}$ such that all consumers with type above it visit the firm without data and vice versa. Thus, there must exist a $\bar{\theta}$ such that searchers with $\theta < \bar{\theta}$ visit the firm without data and all consumers with $\theta > \bar{\theta}$ visit the firm with data, i.e. $g^L = 0, g^H = 1$. In that case, we know that the virtual valuations of the firm with data jump up at $\bar{\theta}$, because:

$$\lim_{\theta \uparrow \bar{\theta}} J^k(\theta) < \lim_{\theta \downarrow \bar{\theta}} J^k(\theta) \iff 0.5(1-\rho) < \rho + 0.5(1-\rho)$$

Part 5c: $\hat{\theta}^L < \hat{\theta}^{nd}$ must hold in such a simple equilibrium with $g^L = 0,\ g^H = 1$.

Assume, for a contradiction, that $\hat{\theta}^L \geq \hat{\theta}^{nd}$ in the supposed equilibrium with $g^L = 0,\ g^H = 1$.

First, note that $\hat{\theta}^L \leq \hat{\theta}^H$ holds once more. Suppose, for a contradiction, that $\hat{\theta}^H < \hat{\theta}^L$ holds in such an equilbrium. When $\hat{\theta}^H \neq \bar{\theta}$, we know that continuity of the virtual valuation function implies that $\hat{\theta}^L < \hat{\theta}^H$ must hold, a contradiction.

Thus, suppose that $\hat{\theta}^H = \bar{\theta}$. Suppose that $\hat{\theta}^H < \hat{\theta}^L$ holds. This is impossible, as any $\theta > \hat{\theta}^H = \bar{\theta}$ must satisfy $J^H(\theta) \geq 0$ by monotonicity of $J^H(\theta)$ - else, we would have a contradiction. But this implies that $J^L(\theta) > 0$ holds for any $\theta > \bar{\theta}$, which implies that $\hat{\theta}^L \leq \bar{\theta} = \hat{\theta}^H$ must hold in this case, a contradiction.

Thus, we know that $\hat{\theta}^L \leq \hat{\theta}^H$ must hold in any equilibrium of this type. Moreover, recall that we have assumed (for a contradiction) that $\hat{\theta}^{nd} \leq \hat{\theta}^L$ holds.

Note that $\hat{\theta}^{nd} \neq \bar{\theta}$ must hold, since the virtual valuation at the firm with data jumps down in this subcase. There are thus two subcases two distinguish: (i) $\hat{\theta}^{nd} > \bar{\theta}$ and (ii) $\hat{\theta}^{nd} < \bar{\theta}$.

(i) Subcase 1: $\bar{\theta} < \hat{\theta}^{nd}$

Suppose that $\hat{\theta}^{nd} < \hat{\theta}^{L}$ and that $\bar{\theta} < \hat{\theta}^{nd}$ holds. Because $\bar{\theta} < \hat{\theta}^{nd}$, the firm without data assigns quality $q^{nd}(\theta) = 1$ to all $\theta > \hat{\theta}^{nd}$ and 0 to all other types - because the virtual valuation at the firm without data is always monotonic. Then, all consumers with $\theta \in (\hat{\theta}^{nd}, \hat{\theta}^{L})$ would get strictly positive utility only at the firm without data and would prefer this firm but visit the firm with data in equilibrium, a contradiction.

Suppose that $\hat{\theta}^{nd} = \hat{\theta}^{L}$ and that $\bar{\theta} < \hat{\theta}^{nd}$ holds. Suppose further that $\hat{\theta}^{L} = \hat{\theta}^{H}$. But then, we have $\bar{\theta} < \hat{\theta}^{nd} = \hat{\theta}^{L} = \hat{\theta}^{H}$, which cannot be true. Hence, $\hat{\theta}^{L} < \hat{\theta}^{H}$ must hold true.

Thus, we have the ordering $\bar{\theta} < \hat{\theta}^{nd} = \hat{\theta}^{L} < \hat{\theta}^{H}$. Consumers with $\theta \in (\hat{\theta}^{nd}, \hat{\theta}^{H})$ get the utility $\theta - \hat{\theta}^{nd}$ at the firm without data and at most the utility $Pr^{L}(\theta)(\theta - \hat{\theta}^{L})$ at the firm with data. Because $\hat{\theta}^{nd} = \hat{\theta}^{L}$, they all strictly prefer the firm without data, but visit the firm with data in equilibrium (since $\hat{\theta}^{nd} > \bar{\theta}$), a contradiction.

(ii) Subcase 2: $\hat{\theta}^{nd} < \bar{\theta}$ and there exist **no** $\theta > \bar{\theta}$ for which $J^{nd}(\theta) < 0$.

For all types $\theta > \bar{\theta}$, the virtual value $J^{nd}(\theta)$ must then be strictly positive (the converse creates a contradiction by monotonicity of $J^{nd}(\theta)$). In that case, the optimal mechanism of the firm without data sets $q^{nd}(\theta) = 1$ for all $\theta > \hat{\theta}^{nd}$ - once again, because the virtual valuation function at the firm without data is always monotonic.

Suppose $\hat{\theta}^{nd} < \hat{\theta}^{L}$. Then, consumers with $\theta > \bar{\theta}$ get the utility $\theta - \hat{\theta}^{nd} > 0$ at the firm without data. The utility they get at the firm with data is weakly smaller than $\theta - \hat{\theta}^{L} < \theta - \hat{\theta}^{nd}$. Thus, all consumers with $\theta > \bar{\theta}$ strictly prefer the firm without data, a contradiction.

Suppose $\hat{\theta}^{nd} = \hat{\theta}^{L}$. Then, we must have $\hat{\theta}^{nd} = \hat{\theta}^{L} < \bar{\theta}$ and hence $\hat{\theta}^{L} < \hat{\theta}^{H}$ must hold. Once again, consumers with $\theta > \bar{\theta}$ get the utility $\theta - \hat{\theta}^{nd} > 0$ at the firm without data. The utility they get at the firm with data is strictly smaller than $\theta - \hat{\theta}^{L}$, because $\hat{\theta}^{L} < \hat{\theta}^{H}$. Thus, all consumers with $\theta > \bar{\theta}$ prefer the firm without data, a contradiction.

(iii) Subcase 3: $\hat{\theta}^{nd} < \bar{\theta}$ and there exist **some** $\theta > \bar{\theta}$ for which $J^{nd}(\theta) < 0$.

In order for such $\theta > \bar{\theta}$ for which $J^{nd}(\theta) < 0$ to exist, the right limit of the virtual valuation at the firm without data must be weakly negative - by monotonicity of $J^{nd}(\theta)$. For any $\bar{\theta} > 0.5$, this right limit will be strictly positive, because it is $\lim_{\theta \downarrow \bar{\theta}} J^{nd}(\theta) = 2\bar{\theta} - 1 > 0$.

Thus, $\bar{\theta} \leq 0.5$ must hold in such an equilibrium. We are still assuming, for a contradiction, that $\hat{\theta}^L \geq \hat{\theta}^{nd}$, and that $g^H = 1$ (high valuation searchers visit firm with data).

Assume, within this subcase, that $\bar{\theta} < \hat{\theta}^L$, which also implies that $\hat{\theta}^L < \hat{\theta}^H$. Then, the provided quality schedule of the firm without data must satisfy $q^{nd}(\theta) = 0$ for any $\theta \in [0, \hat{\theta}^L)$. If this (expected) quality is strictly positive for any such type, then the cutoff $\bar{\theta}$ would not be below $\hat{\theta}^L$ - this follows from the search behaviour of consumers, since a consumer with $\theta = \hat{\theta}^L$ would attain strictly positive utility only at the firm without data.

Thus, the lowest type that gets quality in equilibrium is $\hat{\theta}^L$. By our equilibrium refinement, $\bar{\theta} < \hat{\theta}^L$ thus cannot hold - all consumers with $\theta < \hat{\theta}^L$ visit the same firm, a contradiction.

Suppose instead that $\bar{\theta} = \hat{\theta}^L$, which implies the ordering $\hat{\theta}^{nd} < \bar{\theta} = \hat{\theta}^L$. Once again, any type $\theta \in [0, \hat{\theta}^L) = [0, \bar{\theta})$ must receive $q^{nd}(\theta) = 0$ - else, the cutoff must lie strictly above $\hat{\theta}^L = \bar{\theta}$ by monotonicity of $q^{nd}(\theta)$ in an equilibrium and the optimal search behavior of consumers.

Thus, all types $\theta \geq \bar{\theta} = \hat{\theta}^L$ for which $J^{nd}(\theta) < 0$ must also get zero quality - if they get positive quality, there would be a profitable deviation by monotonicity of $J^{nd}(\theta)$.

Thus, $\underline{\theta}^{nd} > \bar{\theta}$ would hold. Then, either $\underline{\theta}^d \leq \bar{\theta}$ or $\underline{\theta}^d > \bar{\theta}$ must hold. If $\underline{\theta}^d \leq \bar{\theta}$, we have $\underline{\theta}^d \leq \bar{\theta} < \underline{\theta}^{nd}$, and thus all consumers with $\theta < \bar{\theta}$ visit the firm with data by our first tie-breaking rule, a contradiction. If $\underline{\theta}^d > \bar{\theta}$, our tie-breaking rules imply a contradiction as well.

Thus, we must have $\hat{\theta}^L < \bar{\theta}$ in this sort of problematic equilibrium. It was previously also established that the equilibrium under consideration must satisfy $\bar{\theta} \leq 0.5$. For any such $\bar{\theta}$, our assumption implies that $\hat{\theta}^L \geq \bar{\theta}$, which means that no such equilibrium can exist.

Thus, we are done. We have shown that $\hat{\theta}^L < \hat{\theta}^{nd}$ must hold.

<u>Part 5d:</u> There exists no simple equilibrium in which $g^L = 0$ and $g^H = 1$ (all types above $\bar{\theta}$ go to the firm with data).

We have established that $\hat{\theta}^L < \hat{\theta}^{nd}$ holds in a simple equilibrium in which there exists no $\bar{\theta}$ such that consumers with type above it go to the firm without data and vice versa. In this sort of equilibrium, it must also hold that $\hat{\theta}^{nd} \neq \bar{\theta}$. If all consumers visit the firm with data in equilibrium, we can write $\bar{\theta} = 1$ and all consumers with type below this visit the firm with data, a contradiction.

Suppose that $\bar{\theta} < \hat{\theta}^{nd}$, which means that $\underline{\theta}^{nd} = \hat{\theta}^{nd}$. Recall that all consumers with $\theta < \bar{\theta}$ supposedly visit the firm without data. Then, $\underline{\theta}^d$ must be weakly above $\hat{\theta}^{nd}$ - else, all consumers with $\theta < \bar{\theta}$ must visit the firm with data by our refinement. But this implies that all consumers with $\theta < \hat{\theta}^{nd}$ either all visit the firm without data (if $\underline{\theta}^{nd} < \underline{\theta}^d$) or $\underline{\theta}^d = \underline{\theta}^{nd}$ - in either case, they all visit the same firm by our tie-breaking rule, a contradiction.

Finally, suppose that $\hat{\theta}^{nd} < \bar{\theta}$ and hence $\hat{\theta}^L < \bar{\theta}$.

No type $\theta \in (0, \hat{\theta}^{nd})$ can receive positive quality at the firm with data. Else, all types $\theta < \hat{\theta}^{nd}$ would visit the firm with data by our refinement but visit the firm without data in the supposed equilibrium, a contradiction. Thus, we have $\hat{\theta}^{nd} \leq \underline{\theta}^d$.

Moreover, the fact that $\hat{\theta}^L < \bar{\theta}$ implies that $\bar{\theta} \geq 0.5$ must hold. Previous results imply that the virtual valuation of the firm without data thus stays weakly positive for all $\theta > \hat{\theta}^{nd}$. Thus, the firm without data will offer quality 1 to all $\theta > \hat{\theta}^{nd}$. Because $\hat{\theta}^{nd} = \underline{\theta}^{nd} \leq \underline{\theta}^d$all consumers weakly prefer the firm without data.

Suppose (within the subcase $\hat{\theta}^{nd} < \bar{\theta}$) that $\hat{\theta}^{nd} < \underline{\theta}^d$. Then, all consumers visit the firm without data and we could express the search behavior by $\bar{\theta} = 0$, where all consumers with type above $\bar{\theta}$ visit the firm without data, a contradiction.

Thus, suppose (within the subcase $\hat{\theta}^{nd} < \bar{\theta}$) that $\hat{\theta}^{nd} = \underline{\theta}^d$. Unless both firms offer exactly the same menu, all consumers with $\theta > \hat{\theta}^{nd}$ will strictly prefer the firm without data. If the preference is strict, all consumers will visit the firm without data, since $\bar{\theta} > 0$ holds in the supposed equilibrium and all consumers with $\theta > \hat{\theta}^{nd}$ strictly prefer the firm without data. But then, we could have expressed the strategy with a cutoff $\bar{\theta} = 0$ s.t. all consumers with $\theta < \bar{\theta}$ visit the firm with data and vice versa, a contradiction.

If the menus are the same, all searchers randomize, which cannot be a simple equilibrium.

C.6 Proof of proposition 11

Part 1: In a simple equilibrium, $\hat{\theta}^L < \hat{\theta}^{nd}$ holds

Suppose we are in a simple equilibrium, in which searchers with $\theta < \bar{\theta}$ visit the firm with data. Suppose, for the contradiction we seek, that $\hat{\theta}^{nd} \leq \hat{\theta}^L$. In the equilibrium we study, the virtual valuation functions at the firm with data jump down at $\bar{\theta}$, i.e. $\lim_{\theta \uparrow \bar{\theta}} J^k(\theta) > \lim_{\theta \downarrow \bar{\theta}} J^k(\theta)$.

Note first that $\hat{\theta}^L < \hat{\theta}^H$ must hold. This is because neither cutoff $\hat{\theta}^k$ can be exactly at $\bar{\theta}$ - then, $\lim_{\theta \downarrow \bar{\theta}} J^k(\theta) \geq 0$ would have to hold and hence the virtual corresponding valuation would also be strictly positive for values just below $\bar{\theta}$. Thus, $\hat{\theta}^L$ and $\hat{\theta}^H$ must be at points where the virtual valuation is continuous, i.e. they must set the corresponding virtual valuations to zero. A $\hat{\theta}^H$, we would thus have $J^H(\hat{\theta}^H) = 0$ and hence $J^L(\hat{\theta}^H) > 0$. Because $J^L(\hat{\theta}^H) > 0$ would hold there, thus would also hold for values just below $\hat{\theta}^H$. Hence, we know that $\hat{\theta}^L < \hat{\theta}^H$ must hold in an equilibrium where consumers seperate in this way.

The virtual valuation $J^{nd}(\theta)$ jumps up at $\bar{\theta}$, i.e. the utility any consumer attains at the firm with data is $\max\{\theta - \hat{\theta}^{nd}, 0\}$. Thus, all searchers with $\theta > \hat{\theta}^{nd}$ prefer the firm without data, since $\hat{\theta}^{nd} \leq \hat{\theta}^L < \hat{\theta}^H$. Thus, $\bar{\theta} \leq \hat{\theta}^{nd}$ must hold, Because $\hat{\theta}^{nd} \leq \hat{\theta}^L$ by assumption, we obtain a contradiction, since $J^{nd}(\theta; \bar{\theta}) < J^L(\theta; \bar{\theta})$ for all $\theta \geq \bar{\theta}$ and thus, $\hat{\theta}^L < \hat{\theta}^{nd}$ would hold.

Part 2: There exists no simple equilibrium in which $\bar{\theta} \in [0.5, 0.5(1+\rho)]$.

Consider first $\bar{\theta} \in [0.5, 0.5(1+\rho)]$. We know that $\hat{\theta}^{nd} \geq \bar{\theta}$ holds for these values of $\bar{\theta}$, since $\lim_{\theta \uparrow \bar{\theta}} J^{nd}(\theta; \bar{\theta}) < 0$ for $\bar{\theta} < 0.5(1+\rho)$.

The virtual valuation $J^L(\theta)$ will be strictly positive for any $\theta > \hat{\theta}^L$. If $\hat{\theta}^L > \bar{\theta}$, this is true by monotonicity. The case $\hat{\theta}^L = \bar{\theta}$ cannot be true, because, in a simple equilibrium, the virtual valuations at the firm with data jump down.

Thus, consider the third case where $\hat{\theta}^L < \bar{\theta}$. While the virtual valuation will be strictly positive for any $\theta \in (\hat{\theta}^L, \bar{\theta})$, it may drop into the negative at $\bar{\theta}$. To discuss this, consider the right limit of $J^L(\theta)$ at $\bar{\theta}$. Recall that $J^L(\theta) = \theta - \int_\theta^1 \left(Pr^L(x)/Pr^L(\theta)\right)dx$ for any $\theta > \bar{\theta}$: We know that $Pr^L(x)/Pr^L(\theta) < 1$ holds for all $x > \theta$ at any θ. Thus, we have:

$$\lim_{\theta \downarrow \bar{\theta}} J^L(\theta) = \bar{\theta} - \int_{\bar{\theta}}^1 \left(Pr^L(x)/Pr^L(\bar{\theta})\right)dx > \bar{\theta} - \int_{\bar{\theta}}^1 \left(1\right)dx = \bar{\theta} - (1 - \bar{\theta}) = 2\bar{\theta} - 1$$

Because $\bar{\theta} \geq 0.5$, it follows that $\lim_{\theta \downarrow \bar{\theta}} J^L(\theta) > 0$. Thus, the low signal optimal mechanism is uniquely pinned down - all types above $\hat{\theta}^L$ will be assigned the quality level $q^L(\theta) = 1$.

Recall that $\hat{\theta}^L < \hat{\theta}^{nd}$ must hold in such an equilibrium - but this contradicts the statement $\hat{\theta}^{nd} \geq \bar{\theta}$. All consumers with type below $\hat{\theta}^{nd}$ surely prefer the firm with data. Moreover, consumers with $\theta = \hat{\theta}^{nd}$ attain utility 0 at the firm with data, but strictly positive utility at the firm without data - so they visit the firm with data, and so will consumers with a type just above $\hat{\theta}^{nd}$. In equilibrium, they visit the firm without data, a contradiction.

Part 3: There exists no simple equilibrium in which $\bar{\theta} \in [0, 0.5)$.

First, note that $\hat{\theta}^{nd} = 0.5$ holds for any such $\bar{\theta} < 0.5$. The virtual valuation at the firm without data jumps up at $\bar{\theta}$. Thus, we consider the right limit of $J^{nd}(\theta)$ at $\bar{\theta}$. Recall that this virtual valuation equals $J^{nd}(\theta) = \theta - \int_{\theta}^{1}(1)dx$ for any $\theta > \bar{\theta}$. Thus, we have $\lim_{\theta \downarrow \bar{\theta}} J^{nd}(\theta) = \bar{\theta} - (1 - \bar{\theta}) = 2\bar{\theta} - 1 < 0$. Moreover, we have that:

$$\lim_{\theta \downarrow \bar{\theta}} J^{nd}(\theta) = \bar{\theta} - \int_{\bar{\theta}}^{1}(1)dx > \bar{\theta} - \int_{\bar{\theta}}^{1}\left(\frac{\rho + 0.5(1-\rho)}{0.5(1-\rho)}\right)dx = \lim_{\theta \uparrow \bar{\theta}} J^{nd}(\theta)$$

This proves that $\hat{\theta}^{nd} > \bar{\theta}$ holds for such $\bar{\theta} < 0.5$. Moreover, $\hat{\theta}^{nd} = 0.5$ will hold exactly.

For any $\theta > \bar{\theta}$, the virtual valuation $J^L(\theta)$ is strictly greater than $J^{nd}(\theta)$. Thus, we have some $\tilde{\theta}^L > \bar{\theta}$ with: (i) $\tilde{\theta}^L < \hat{\theta}^{nd}$ and (ii) $J^L(\theta) > 0$ for any $\theta > \tilde{\theta}^L$.

Suppose we have a simple equilibrium in which $\bar{\theta} < 0.5$. In equilibrium, the firm with data must optimally assign strictly positive quality to a strictly positive measure of consumers with $\theta \in (\tilde{\theta}^L, \hat{\theta}^{nd})$.

Suppose, for a contradiction, that all consumers with $\theta \in (\tilde{\theta}^L, \hat{\theta}^{nd})$ receive the quality level 0 according to the mechanism $q^L(\theta)$. By monotonicity, all consumers with $\theta \leq \tilde{\theta}^L$ must also receive the quality level 0. But this is a contradiction - all consumers with $\theta > \tilde{\theta}^L$ have strictly positive virtual valuations. Thus, the firm is guaranteed to make more higher revenue by setting $q^L(\theta) = 1$ to all $\theta \in (\tilde{\theta}^L, 1)$. This also satisfies monotonicity, so it is feasible, and we cannot have an equilibrium.

Thus, there must exist a type $\theta \in (\tilde{\theta}^L, \hat{\theta}^{nd})$ that receives a strictly positive quality level $q^L(\theta)$ at the firm with data. By monotonicity, all types above this must also receive a strictly positive $q^L(\theta)$. Thus, the utility of a consumer with $\theta = \hat{\theta}^{nd} > \bar{\theta}$ at the firm with data will be strictly positive. The utility this consumer would receive at the firm without data would be zero - this is a contradiction, as consumers with θ just above $\bar{\theta}$ would similarly prefer the firm with data, but visit the firm without data in equilibrium.

Part 4: Under said assumptions, there exists a simple equilibrium with $\bar{\theta} \geq [0.5(1+\rho), 1]$.

Properties of the virtual valuations under the assumptions:

The virtual valuation of the firm without data always satisfies monotonicity. Moreover, this virtual valuation $J^{nd}(\theta)$ jumps up around $\bar{\theta}$ in the equilibrium we study.

When $\theta < \bar{\theta}$, we have:

$$\frac{\partial J^k}{\partial \theta} = 2 + \left(\frac{\int_\theta^{\bar{\theta}} \big(\rho g^L + 0.5(1-\rho)\big) Pr^k(x)dx + \int_{\bar{\theta}}^1 \big(\rho g^H + 0.5(1-\rho)\big) Pr^k(x)dx}{\big(\rho g^L + 0.5(1-\rho)\big) Pr^k(\theta)} \right) \left[\big(Pr^k(\theta)\big)^{-1} \frac{\partial Pr^k(\theta)}{\partial \theta} \right]$$

For $\theta > \bar{\theta}$, this derivative is:

$$\frac{\partial J^k}{\partial \theta} = 2 - \int_\theta^1 \big(Pr^k(x)/Pr^k(\theta)\big) dx \left[- \big(Pr^k(\theta)\big)^{-1} \frac{\partial Pr^k(\theta)}{\partial \theta} \right]$$

Thus, the high signal virtual valuation is generally piecewise increasing, while the low signal virtual valuation is rising under our assumptions for $\bar{\theta} \in [0.5(1+\rho), 1]$ and $(g^L = 1, g^H = 0)$.

It was further assumed that both virtual valuations do not jump into the negative region at $\bar{\theta}$ for $\bar{\theta} \geq 0.5(1+\rho)$. Formally, it was stated that $\lim_{\theta \downarrow \bar{\theta}} J^k(\theta; \bar{\theta}) > 0$. Recall that $\lim_{\theta \uparrow \bar{\theta}} J^k(\theta; \bar{\theta}) > \lim_{\theta \downarrow \bar{\theta}} J^k(\theta; \bar{\theta})$ holds, which implies that $\hat{\theta}^k < \bar{\theta}$ must hold.

Summing it all up, these assumptions imply that the virtual valuation functions will be strictly positive for all types above the respective cutoffs $\hat{\theta}^{nd}$ and $\hat{\theta}^k$, which implies that the optimal mechanism just assigns quality 1 to all consumers with a positive virtual valuation.

General notions:

I describe the consumer's search behavior by:

$$\theta^*(\bar{\theta}) = \sup \left\{ \theta \in [0,1] : Pr^L(\theta)\hat{\theta}^L(\bar{\theta}) + Pr^H(\theta)\hat{\theta}^H(\bar{\theta}) < \hat{\theta}^{nd}(\bar{\theta}) \right\}$$

We have defined $\hat{\theta}^k(\bar{\theta}) = \inf\{\theta : J^k(\theta, \bar{\theta}) > 0\}$ and $\hat{\theta}^{nd}(\bar{\theta}) = \inf\{\theta : J^{nd}(\theta, \bar{\theta}) > 0\}$. Suppose we have found an $\bar{\theta} \geq 0.5(1+\rho)$.

To establish that this is an equilibrium, we make use of some ancillary results. By assumption, we have that $\hat{\theta}^H < \bar{\theta}$ holds for any $\bar{\theta} \geq 0.5(1+\rho)$. Previous arguments imply that $\hat{\theta}^{nd} \leq \bar{\theta}$ for any such $\bar{\theta} \geq 0.5(1+\rho)$. $\hat{\theta}^{nd}$ has to solve the following:

$$J^{nd}(\hat{\theta}, \bar{\theta}) = \hat{\theta} - \frac{\int_{\hat{\theta}}^{\bar{\theta}} 0.5(1-\rho)dx + \int_{\bar{\theta}}^{1} 0.5(1+\rho)dx}{0.5(1-\rho)} = 0$$

Thus, $\hat{\theta}^{nd}$ is falling in $\bar{\theta}$. Now consider $\hat{\theta}^k$, which has to solve the following since $\hat{\theta}^k < \bar{\theta}$:

$$J^k(\hat{\theta}^k, \bar{\theta}) = \hat{\theta}^k - \int_{\hat{\theta}^k}^{\bar{\theta}} \big(Pr^k(x)/Pr^k(\hat{\theta}^k)\big)dx - \int_{\bar{\theta}}^{1} \big((1-\rho)/(1+\rho)\big)\big(Pr^k(x)/Pr^k(\hat{\theta}^k)\big)dx = 0$$

Thus, we have $\frac{\partial \hat{\theta}^k}{\partial \bar{\theta}} = -\frac{\partial J/\partial \bar{\theta}}{\partial J/\partial \hat{\theta}^k} > 0$. Summing up, our assumptions imply that $\frac{\partial \hat{\theta}^k}{\partial \bar{\theta}} > 0$ and that $\frac{\partial \hat{\theta}^{nd}}{\partial \bar{\theta}} < 0$. Because $\hat{\theta}^L(1) < \hat{\theta}^{nd}(1)$, the following holds for any $\bar{\theta} \in [0.5(1+\rho), 1]$:

$$\hat{\theta}^L(\bar{\theta}) \leq \hat{\theta}^L(1) < \hat{\theta}^{nd}(1) \leq \hat{\theta}^{nd}(\bar{\theta})$$

Thus, we must have $\hat{\theta}^L < \hat{\theta}^{nd}$ in the equilibrium candidate we have found.

Optimal menus:

Consider first the firm without data, for which $J^{nd}(\theta)$ is monotonic and jumps upward at $\bar{\theta}$. The virtual valuation $J^{nd}(\theta)$ is thus strictly positive if and only if $\theta > \hat{\theta}^{nd}$. Thus, their optimal menu is to offer $q^{nd}(\theta) = 1$ to all $\theta > \hat{\theta}^{nd}$ and quality 0 to all other types.

Now consider the firm with data. The virtual valuations are piecewise monotonic by assumption and don't jump into the negative region at $\bar{\theta}$, again by assumption. Thus, the virtual valuation $J^k(\theta)$ will be strictly positive iff $\theta > \hat{\theta}^k$, which means their optimal mechanism will also assign $q^k(\theta) = 1$ to all $\theta > \hat{\theta}^k$ and quality 0 to all other types.

Search:

Because $\hat{\theta}^L < \hat{\theta}^{nd}$, all searchers with $\theta < \hat{\theta}^{nd}$ visit the firm with data. The preference for the firm with data of consumers with $\theta \in [\hat{\theta}^{nd}, \hat{\theta}^H]$ is $P^d(\theta) = Pr^L(\theta)(\theta - \hat{\theta}^L) - (\theta - \hat{\theta}^{nd})$.

This is falling in θ. Because $\hat{\theta}^H$ lies strictly below $\theta^*(\bar{\theta}) = \bar{\theta}$ by assumption and the fact that the LHS in this sup-expression is strictly rising in θ, we have:

$$Pr^L(\hat{\theta}^H)\hat{\theta}^L + Pr^H(\hat{\theta}^H)\hat{\theta}^H < \hat{\theta}^{nd} \iff Pr^L(\theta)\big(\hat{\theta}^H - \hat{\theta}^L\big) - \big(\hat{\theta}^H - \hat{\theta}^{nd}\big) > 0$$

Thus, the consumer with $\theta = \hat{\theta}^H$ prefers the firm with data, and so will all $\theta \in [\hat{\theta}^{nd}, \hat{\theta}^H]$.

Now consider $\theta \in [\hat{\theta}^H, \bar{\theta})$. We know that $q^H(\theta) = 1$ for any $\theta \in (\hat{\theta}^H, \bar{\theta})$ and $q^L(\theta) = 1$ for any $\theta \in (\hat{\theta}^L, \bar{\theta})$. Thus, the utility that this consumer attains at the firm with data is $U^d(\theta) = Pr^L(\theta)(\theta - \hat{\theta}^L) + Pr^H(\theta)(\theta - \hat{\theta}^H)$. Similarly, $q^{nd}(\theta) = 1$ for any $\theta \in (\hat{\theta}^{nd}, \bar{\theta})$, i.e. their utility at the firm without data is $\theta - \hat{\theta}^{nd}$. Thus, any such searchers prefer the firm with data because $\hat{\theta}^{nd} > Pr^L(\theta)\hat{\theta}^L + Pr^H(\theta)\hat{\theta}^H$. Consumers with $\theta > \bar{\theta}$ prefer to visit the firm without data by analogous arguments.

Thus: When $\bar{\theta} = \theta^*(\hat{\theta})$ and $\bar{\theta} > 0.5(1+\rho)$, we have an equilibrium.

Existence of a solution to $\bar{\theta} = \theta^*(\hat{\theta})$ on $\bar{\theta} > 0.5(1+\rho)$

It remains to show that such a value exists. To see this, recall first that $\hat{\theta}^k$ are both strictly rising in $\bar{\theta}$, while $\hat{\theta}^{nd}$ is strictly falling in $\bar{\theta}$. We work with the object $\bar{\theta}'$. At $\bar{\theta} = \bar{\theta}'$, we have:

$$Pr^L(1)\hat{\theta}^L(\bar{\theta}') + Pr^H(1)\hat{\theta}^H(\bar{\theta}') - \hat{\theta}^{nd}(\bar{\theta}') = 0$$

Note that this function is strictly rising in θ, which implies that $\theta^*(\bar{\theta}) = 1$ for any such $\bar{\theta} \in [0.5(1+\rho), \bar{\theta}']$. Further note that $\bar{\theta}'$ must be strictly above $\bar{\theta} = 0.5(1+\rho)$, since:

$$\hat{\theta}^L(0.5(1+\rho)) < \hat{\theta}^H(0.5(1+\rho)) < 0.5(1+\rho) = \hat{\theta}^{nd}(0.5(1+\rho))$$

For any $\bar{\theta} \in [\bar{\theta}', 1]$, we have $Pr^L(1)\hat{\theta}^L(\bar{\theta}') + Pr^H(1)\hat{\theta}^H(\bar{\theta}') - \hat{\theta}^{nd}(\bar{\theta}') \geq 0$. This implies that $\theta^*(\bar{\theta})$ must solve:

$$Pr^L(\theta^*)\hat{\theta}^L(\bar{\theta}') + Pr^H(\theta^*)\hat{\theta}^H(\bar{\theta}') - \hat{\theta}^{nd}(\bar{\theta}') = 0$$

Thus: (i) At $\bar{\theta} = \bar{\theta}'$, we have $\theta^*(\bar{\theta}) \geq \bar{\theta}$. (ii) At $\bar{\theta} = 1$, $\theta^*(\bar{\theta}) \leq \bar{\theta}$. By assumption and our result, all cutoffs are strictly below $\bar{\theta}$ for $\bar{\theta} \geq 0.5(1+\rho)$, Thus, their solutions are continuous in $\bar{\theta}$, and so is $\theta^*(\bar{\theta})$. Application of the intermediate value theorem to the equation $\theta^*(\bar{\theta}) - \bar{\theta} = 0$ just laid out, together with the border conditions, guarantee existence of such a $\bar{\theta}$.

D Omitted results

D.1 Baseline model - monopoly solution

The high signal profit function of a monopolist is: $\Pi^M(p_j|\tilde{v}^H) = p_j \int_{p_j}^1 Pr^H(v)dv$. Our assumptions guarantee that this is strictly concave and differentiable. Thus, the optimal high signal monopoly price (p^{HM}) has to satisfy:

$$\int_{p^{HM}}^1 Pr^H(v)dv - p^{HM} Pr^H(p^{HM}) = 0$$

Assume, for a contradiction, that $p^{HM} \leq 0.5$, i.e. $1 - p^{HM} \geq 0.5 \geq p^{HM}$. Since $Pr^H(v)$ is strictly increasing and $p^{HM} < 1$ must hold, we have: $\int_{p^{HM}}^1 Pr^H(v)dv > p^{H,M} Pr^H(p^{HM})$. This is a contradiction. Analogous arguments show that $p^{L,M} < 0.5$ holds.

D.2 Baseline model - equilibria without data advantages

Suppose no firm receives an informative signal in the baseline model, i.e. $Pr^H(v) = 0.5 \ \forall v$.

(i) Ruling out equilibria in which $p^L = p^H \neq p^{nd}$.

If $p^L < p^{nd}$, all searchers with $v > p^L$ visit the firm with data. Thus, $p^{nd} = 0.5$, since the firm without data is only visited by captive consumers. But then, there exists a profitable upward deviation from p^L. If $p^{nd} < p^L$, all searchers with $v > p^{nd}$ visit the firm without data. Thus, $p^L = 0.5$, and there is a profitable upward deviation from p^{nd}.

(ii) Ruling out equilibria with $p^L < p^H$ (analogous arguments rule out equilibria with $p^H < p^L$, since the signal is not informative):

Previous arguments establish that $p^L < p^{nd} < p^H$ must hold. As a result, there exists an $\epsilon > 0$ such that any searcher with $v \in (p^L, p^{nd} + \epsilon]$ will visit the firm with data. Thus, searchers put upward pressure on p^{nd}, and hence $p^{nd} \geq 0.5$ must hold. One can show that the average price at the two firms must be equal. Then, all searchers with $v \in (p^L, p^H)$ strictly prefer the firm with data, while those with $v \geq p^H$ are indifferent. Searchers put upward pressure on p^{nd}, which implies that $p^{nd} \geq 0.5$ and hence $p^H > 0.5$. But this is a contradiction, as searchers put downward pressure on this price, i.e. $p^H \leq 0.5$.

D.3 Dispersed data framework - no data advantages

First, one can show that $p^{LB} \leq p^{LW}$ must hold and that the reverse, namely $p^{LB} \geq p^{LW}$, must also hold true. This is based on arguments analogous to those made in the proof of lemma 3, part 1. Thus, $p^{LW} = p^{LB}$. Based on this, one can show that $p^{HB} = p^{HW}$ must hold.

Suppose that $p^{HB} < p^{HW}$, noting that both low signal prices must be the same. Then, all consumers with $v \geq p^{HB}$ will strictly prefer to visit the firm with better data. In order to constitute a simple equilibrium, the cutoff must be set in such a way that all searchers with $v > \bar{v}$ (where $\bar{v} \leq p^{HB}$) visit the firm with better data.

But by this logic, the firm with worse data will only make the sale to captive consumers in an open ball around p^{HW}, which implies that $p^{HW} = p^{HW,M}$. Moreover, because all searchers with $v \geq p^{HB}$ visit the firm with better data, we have $p^{HB} \geq p^{HB,M}$, as there would be an upward deviation otherwise. Hence, we have $p^{HB} \geq p^{HB,M} = p^{HW}$, a contradiction. Similar logic rules out the other case - hence, all prices have to be equal and all searchers randomize.

E Numerical illustrations

E.1 Existence results for the dispersed data framework

Consider the equilibrium of proposition 10. Every graph corresponds to a fixed level of ρ. Different levels of α_b are plotted on the x-axis and different levels of α_w on the y-axis. For a given parameter combination, a green dot indicates that the price - search combination from equation (20) constitutes a perfect Bayesian equilibrium.

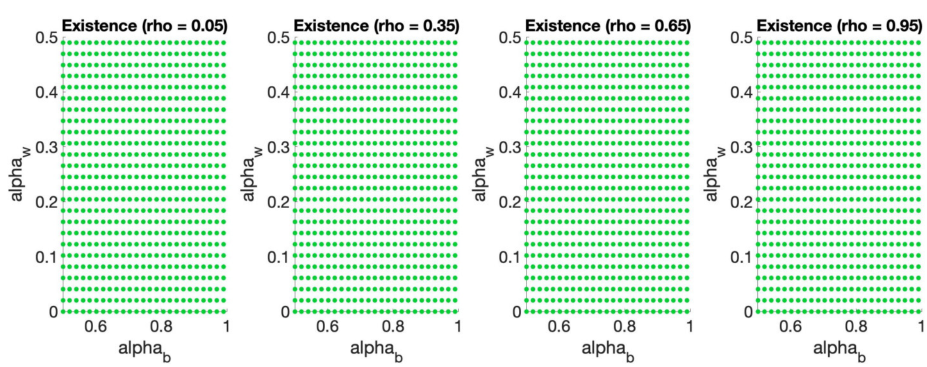

Figure 7: Visualization - existence of a simple equilibrium

References

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of economic Literature*, *54*(2), 442–92.

Armstrong, M., & Zhou, J. (2016). Search deterrence. *The Review of Economic Studies*, *83*(1), 26–57.

Atayev, A. (2021). Search cost based price discrimination. *Working paper*.

Belleflamme, P., Lam, W. M. W., & Vergote, W. (2020). Competitive imperfect price discrimination and market power. *Marketing Science*, *39*(5), 996–1015.

Belleflamme, P., & Vergote, W. (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters*, *149*, 141–144.

Bergemann, D., & Bonatti, A. (2022). Data, competition, and digital platforms.

Bergemann, D., Brooks, B., & Morris, S. (2021). Search, information, and prices. *Journal of Political Economy*, *129*(8), 2275–2319.

Bounie, D., Dubus, A., & Waelbroeck, P. (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*(2), 283–313.

Braghieri, L. (2019). Targeted advertising and price discrimination in intermediated online markets. *Working paper*.

Burdett, K., & Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, 955–969.

Clavorà Braulin, F. (2021). The effects of personal information on competition: Consumer privacy and partial price discrimination. *ZEW-Centre for European Economic Research Discussion Paper*(21-007).

Delbono, F., Reggiani, C., & Sandrini, L. (2022). Strategic data sales to competing firms. *Available at SSRN*.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory*, *3*(2), 156–168.

Eeckhout, J., & Veldkamp, L. (2022). *Data and market power* (Tech. Rep.). National Bureau of Economic Research.

Escobari, D., Rupp, N. G., & Meskey, J. (2019). An analysis of dynamic price discrimination in airlines. *Southern Economic Journal*, *85*(3), 639–662.

Esteves, R.-B. (2014). Price discrimination with private and imperfect information. *The Scandinavian Journal of Economics*, *116*(3), 766–796.

European Commission. (2019). Directive (eu) 2019/2161 of the european parliament and of the council. *Official Journal of the European Union*.

European Commission. (2020). A proposal for a regulation of the european parliament and of the council on contestable and fair markets in the digital sector (digital markets act).

Fabra, N., & Reguant, M. (2020). A model of search with price discrimination. *European Economic Review*, *129*, 103571.

Freedman, C., & Sagredo, J. (2022). The invisible faang. *Available at SSRN 3948210*.

Garcia, D. (2022). Harmonic price targeting. *Information Economics and Policy*, *60*, 100984.

Garrett, D. F., Gomes, R., & Maestri, L. (2019). Competitive screening under heterogeneous information. *The Review of Economic Studies*, *86*(4), 1590–1630.

Groh, C.-C. (2022). Competitive price discrimination, imperfect information, and consumer search. *Working paper*.

Gu, Y., Madio, L., & Reggiani, C. (2019). Exclusive data, price manipulation and market leadership.

Guembel, A., & Hege, U. (2021). *Data, product targeting and competition* (Tech. Rep.). Working Paper.

Hannak, A., Soeller, G., Lazer, D., Mislove, A., & Wilson, C. (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference* (pp. 305–318).

Ke, T. T., Lin, S., & Lu, M. Y. (2022). Information design of online platforms. *Available at SSRN*.

Kirpalani, R., & Philippon, T. (2021). *Data sharing and market power with two-sided platforms* (Tech. Rep.). National Bureau of Economic Research.

Larson, J., Mattu, S., & Angwin, J. (2015). Unintended consequences of geographic targeting. *Technology Science*.

Li, X., Lo, A. W., & Thakor, R. T. (2021). *Paying off the competition: Market power and innovation incentives* (Tech. Rep.). National Bureau of Economic Research.

Marshall, G. (2020). Search and wholesale price discrimination. *The RAND Journal of Economics*, *51*(2), 346–374.

Mauring, E. (2021). Search and price discrimination online. *Working paper*.

Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, *70*(2), 583–601.

Mussa, M., & Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, *18*(2), 301–317.

OECD Secretariat. (2016). Executive summary of the roundtable on price discrimination.

Osório, A. (2022). Data and competitive markets: Some notes on competition, concentration and welfare. *The BE Journal of Theoretical Economics*.

Peiseler, F., Rasch, A., & Shekhar, S. (2021). Imperfect information, algorithmic price discrimination, and collusion. *The Scandinavian Journal of Economics*.

Preuss, M. (2021). Search, learning and tracking. *Working paper*.

Rhodes, A., & Zhou, J. (2022). Personalized pricing and competition. *Available at SSRN 4103763*.

Stahl, D. O. (1989). Oligopolistic pricing with sequential consumer search. *American Economic Review*, 700–712.

Statista. (2021). *Leading e-commerce websites in the united states as of june 2021, based on number of monthly visits.* Retrieved 2022-11-09, from `https://www.statista.com/statistics/271450/monthly-unique-visitors-to-us-retail-websites/`

Statista. (2022). *Worldwide desktop market share of leading search engines from january 2010 to july 2022.* Retrieved 2022-09-16, from `https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/`