NBER WORKING PAPER SERIES

MARKET POWER IN ARTIFICIAL INTELLIGENCE

Joshua S. Gans

Working Paper 32270
http://www.nber.org/papers/w32270

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2024

Joshua Gans has drawn on the findings of his research for both compensated speaking engagements and consulting engagements. He has written the books Prediction Machines, Power & Prediction, and Innovation + Equality on the economics of AI for which he receives royalties. He is also chief economist of the Creative Destruction Lab, a University of Toronto-based program that helps seed stage companies, from which he receives compensation. He conducts consulting on anti-trust and intellectual property matters with an association with Charles River Associates and his ownership of Core Economic Research Ltd. He also has equity and advisory relationships with a number of startup firms. Thanks to Andrei Haigu, Chad Jones and the participants at the Rochester Antitrust Workshop for their helpful comments. All errors remain my own. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Market Power in Artificial Intelligence
Joshua S. Gans
NBER Working Paper No. 32270
March 2024
JEL No. L15,L40,O34

ABSTRACT

This paper surveys the relevant existing literature that can help researchers and policy makers understand the drivers of competition in markets that constitute the provision of artificial intelligence products. The focus is on three broad markets: training data, input data, and AI predictions. It is shown that a key factor in determining the emergence and persistence of market power will be the operation of markets for data that would allow for trading data across firm boundaries.

Joshua S. Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
and NBER
joshua.gans@rotman.utoronto.ca

1 Introduction

New technologies often give rise to concerns that the new technological opportunities will be exploited by those with existing market power and, hence, diminish the rate of adoption and innovation in those technologies. For instance, concerns about the market power of telecommunications infrastructure providers were the focus of much regulatory attention with respect to the Internet although, ultimately, it was newer firms that ended up being the later focus of antitrust inquiries. The same discussions are taking place with respect to artificial intelligence (AI). There are simultaneously concerns that incumbents who have large repositories of data will end up leveraging their market power into the provision of artificial intelligence technologies that rely on such data and also concerns that new entrants who have built foundational models are signs of a highly concentrated market for AI provision.

The purpose of this survey is to examine these issues of market power in the provision of AI. The focus here is on the latest generation of AI technologies that have received a large amount of investment attention since the improvements in deep learning were demonstrated around 2012 and, more recently, with the emergence of generative AI and, in particular, large language models. Those technologies are all properly characterised as advances in computational statistics; in particular, the statistics of prediction (Agrawal et al., 2018). This means that the resources that are used to generate AI predictions (the output of AI technologies) are data (both input and training), talent (data scientists and AI developers) and compute (the hardware to train AI and run inference tasks). There is potential for concentration amongst each of these inputs and also in the provision of AI output.

The focus of this survey is on market power issues that uniquely arise in the AI context. Thus, we set aside considerations of talent and compute because these are factors that apply more generally in industrial organisation and antitrust. Instead, the focus is on data and predictions. To that end, the analysis here is divided into three areas along broad market lines. First, the determination of competition in markets for training data – the data that is used to develop AI predictive algorithms – is examined. Here, the central focus is on data feedback loops that can allow market power to beget market power. The message here is that these effects can be subtle (as anticipated by Varian (2019) and Agrawal et al. (2022)). However, it is demonstrated that the presence of data sharing or data markets can significantly impact the potential for concentration in the provision of training data. Second, market power in input data markets is examined. This data is used to 'feed' existing AI algorithms (already trained and developed) to generate AI predictions. The key issue identified again is how such input data traverses firm boundaries – especially amongst competitors in a market. It is here that the potential for data-driven mergers is analysed. Finally, market power in the market for predictions themselves – the output of AI algorithms – is analysed. This area, while relatively nascent, has received attention in matching markets where match quality is important.

2 Market Power in Training Data

AI prediction algorithms generate predictions by being fed input data. The prediction algorithms themselves are created using training data. The training data and input data are often distinct, with the training data being generated in the past, whereas the input data is often to be generated in the future. Sometimes, data might be both input and training data, for instance, where input data is used to update prediction algorithms.

From a competition perspective, a useful although admittedly not perfect attribution of the distinct roles of training and input data is that training data, by enabling the creation of AI prediction algorithms, is a key component that drives entry into a market, whereas input data, by allowing prediction algorithms to run effectively, is a key component that drives cost and/or quality of offerings by firms in a market.

In this section, market power in the provision of training data is examined by considering situations where incumbent firms have greater access to training data and that this creates a potential barrier to entry. There are other AI-related resources, such as compute, that can also create barriers to entry. Here, however, the focus is on training data, not just because it can be an essential input but also because its non-rival nature means that it can be made freely available to incumbents and entrants alike. Thus, it is the incentives of incumbents not to make that data available, even at a price, that lead to higher entry barriers and market power by those who generate or hold training data.

2.1 The Data Feedback Loop

At the heart of many considerations regarding the existence of market power is whether the adoption of a technology involves a production function with increasing returns to scale. With respect to AI, the issue of returns to scale has been predominantly explored with respect to data. Agrawal et al. (2018) described how Amazon was able to generate a virtuous cycle by launching its recommendation engines sooner, attracting more customers that generate feedback data that improves the predicted recommendations, attracts even more customers and so on (see Figure 1). This implies that, at least over an important range, the use of data in production may exhibit increasing returns to scale.[1]

[1] For a pricing treatment see, Gurkan and de Véricourt (2022).

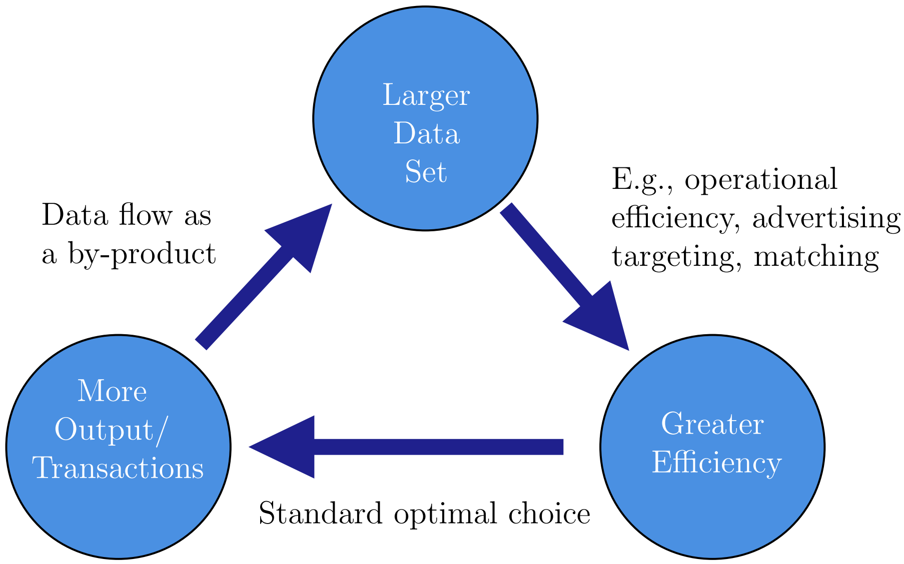

Figure 1: The Data Feedback Loop

Farboodi and Veldkamp (2023) provide a simple model that generates this outcome. They assume that output at time t, Y_t, is a function of productivity (A_t) and capital (K_t) at that time.

$$Y_t = A_t K_t^{\alpha}$$

where $\alpha < 1$. Productivity is dependent on data available, D_t:

$$A_t = A(D_t)$$

Finally, data is accumulated through the sales of products – that is, there is what Rosenberg (1982) termed "learning by using."

$$D_{t+1} = (1 - \delta)D_t + zY_t$$

z is a scaling parameter to track how output translates into a measure of data. This model originates in the examples of Wilson (1975) and is enriched in Farboodi and Veldkamp (2021) and Jones and Tonetti (2020).

While this data feedback loop is at the core of how AI may lead to market power accrued by some firms in some markets, empirically, it appears that there may be limits to the operation of the data feedback loop. In a study of Amazon's retail forecasting, Bajari et al. (2019). They note that statistical learning theory implies that there are diminishing returns to the size of the dataset in terms of prediction performance, even though observations across a greater period of time should improve forecast accuracy. However, as Amazon's data covers millions of products, as the number of products considered expands, this should improve prediction performance. They found, however, that the number of products had little impact on prediction performance, even though as the number of observations per product increased, prediction performance did improve. Over time, however, there was a trend towards better performance, suggesting a 'learning by doing' type effect that comes from improvements in organisational practices and other investments.[2] This analysis, of course, does not take into account interactions with the competitive environment between firms. Calvano et al. (2023) examine this theoretically using a commonly used approach underlying consumer recommendation systems and show that, while more data may lead to more concentration, the end result may be lower prices and higher consumer surplus.[3]

2.2 Competition with Data Feedback

A natural way of modelling the data feedback loop is to assume that when a firm is able to attract customers, it can use the data generated by those customers to improve product quality for any future customer it might supply. For instance, if N_i is a measure of the data gathered from past customers – perhaps an index of the number of periods that it has attracted a large number of customers in a market – then one component of a consumer's willingness to pay for their product is $f_i(N_i)$ which is a non-decreasing function. There is another component of a consumer's willingness to pay for i's product is a stand-alone value, $s_i(\geq 0)$, which does not depend on data gathered and used by i but is something intrinsic to the product that i offers. Suppose there is a continuum of $[0, 1]$ identical consumers in the market in any period. It is assumed that, in a period, the total willingness to pay for i's product, $s_i + f_i(N_i)$ is the same for all consumers and consumers care about $s_i + f_i(N_i) - p_i$, where p_i's price in that period when selecting a product to purchase. Consumers purchase a product from only one firm in the market at a time. It will, however, be assumed that there are no constraints on consumers from choosing to buy from one firm today and selecting another in the future; i.e., there are no direct switching costs.

This is the set-up of Hagiu and Wright (2023) for what they term 'across consumer' learning. For our purposes here, the idea is that past experience with consumers generates training data to provide improved product quality for consumers in the future. Thus, it has the ingredients of a data feedback loop, as the greater the number of customers a firm has in the past will determine the product quality offerings to all consumers in the future. This,

[2] In other studies, there is more evidence of a data feedback loop. See, for example, Schaefer and Sapi (2023).

[3] Similarly, Zhou (2022) shows that while improved information for search may create pressures for firms to increase prices, changes in the way improved information causes consumers to search may create downward pressure on prices.

in turn, potentially leads them to attract more future customers, completing the loop.

The simplest setup involves $i \in \{I, E\}$ with an incumbent (I) and an entrant (E). Each has a marginal cost of production of c and no fixed costs, meaning they will both always be active in the market even if they do not produce anything. Suppose each has data at the outset equivalent to (N_I, N_E) with $N_I > N_E$ (as would be expected from an incumbent). They compete over an infinite number of discrete time periods with discount factor, $\delta \in (0,1)$. Because consumers are symmetric in their preferences, when these firms compete on the basis of price, all consumers are choosing to buy from either I or E in a period. Specifically, they purchase from I in the first period if:

$$s_I + f_I(N_I) - p_I \geq s_E + f_E(N_E) - p_E$$

There is, however, more at stake for each firm when setting its price than just supplying consumers in the present period. By winning now, they secure data that they can use to improve product quality and make them more competitive in the future. Therefore, each firm may want to discount their price today in order to acquire training data to increase future profits.

Given this, what is the total value to a firm i of winning the market in the current period? One way to examine this is to suppose there were just two periods. In this case, working backwards, if, say, firm I had been the market leader in the first period, then the second-period profits of each would be, for E, $\max\{s_E + f_E(N_E) - s_I - f_I(N_I+1), 0\}$ and, for I, $\max\{s_I + f_I(N_I+1) - s_E - f_E(N_E), 0\}$; with the profits if E was the first-period winner being analogous.[4] There are, therefore, three possible outcomes:

1. E always wins in period 2, if $s_E + f_E(N_E) - s_I - f_I(N_I+1) \geq 0$;

2. I always wins in period 2, if $s_I + f_I(N_I) - s_E - f_E(N_E+1) \geq 0$; or

3. the firm that wins period 1, wins period 2, if $f_I(N_I) - f_E(N_E+1) \leq s_E - s_I \leq f_I(N_I+1) - f_E(N_E)$.

Turning to period 1, the firm that wins that period will be able to extract the difference in surplus it offers in period 2. Thus, for instance, E will win in period 1 if:

$$\begin{aligned} & s_E + f_E(N_E) + \delta(s_E + f_E(N_E+1) - s_I - f_I(N_I)) \\ & \geq s_I + f_I(N_I) + \delta(s_I + f_I(N_I+1) - s_E - f_E(N_E)) \end{aligned}$$

[4] That is, if E had won period 1, second-period profits would be for E, $\max\{s_E + f_E(N_E+1) - s_I - f_I(N_I), 0\}$ and, for I, $\max\{s_I + f_I(N_I) - s_E - f_E(N_E+1), 0\}$.

or

$$s_E - s_I \geq \Delta(N_I, N_E)$$

where

$$\Delta(N_I, N_E) \equiv \frac{(1+\delta)(f_I(N_I) - f_E(N_E))}{1+2\delta} + \frac{\delta(f_I(N_I+1) - f_E(N_E+1))}{1+2\delta}$$

Here $\Delta(N_I, N_E)$ represents I's data advantage. If this is positive, the standalone value of E must exceed that of I for E to win. Notice that while I's data advantage is higher if $N_I > N_E$, it also depends on the relative marginal benefits from data each gets from winning in period 1 and also on the discount factor. Importantly, $N_I > N_E$ does not imply that $\Delta(N_I, N_E) \geq 0$.

What is going on here is that firms are willing to subsidise consumers in period 1 in order to improve their prospects for competition in period 2. Note, however, that this does not guarantee a socially optimal outcome in either period. The condition for socially optimality in the two period model is that E wins if:

$$\begin{aligned} & s_E + f_E(N_E) + \delta(s_E + f_E(N_E+1)) \\ & \geq s_I + f_I(N_I) + \delta(s_I + f_I(N_I+1)) \end{aligned}$$

which does not align with the competitive outcome as $s_I + f_I(N_I) \neq s_E + f_E(N_E)$. The condition for E to win places excessive weight on first-period surpluses because that surplus also is extracted as profit by the winning firm in the second period. However, Hagiu and Wright (2023) show that in the infinite horizon game, this distortion is spread over infinitely many periods and so disappears, thereby, implying the competitive outcome is also socially optimal.

In the infinite horizon game, the data advantage takes the following form:

$$\Delta(N_I, N_E) = (1-\delta)\left(\sum_{j=0}^{\infty} \delta^j f_I(N_I + j) - \sum_{j=0}^{\infty} \delta^j f_E(N_E + j)\right)$$

If $s_E - s_I > \Delta(N_I, N_E)$, then E wins in all periods otherwise I wins in all periods. Importantly, because $s_E - s_I > \Delta(N_I, N_E)$ even if $N_I > N_E$, I's initial data advantage may not even allow it to preserve market leadership from the start. In other words, this model exhibits a data feedback loop, but the shape of the data-enabled learning functions, $f_E(.)$ and $f_I(.)$, might be such that the cycle favours the firm that is initially lagging. Intuitively, Hagiu and Wright (2023) show via an induction argument that if a firm wins in the current period, it wins in all periods and, therefore, the current period competition is effectively

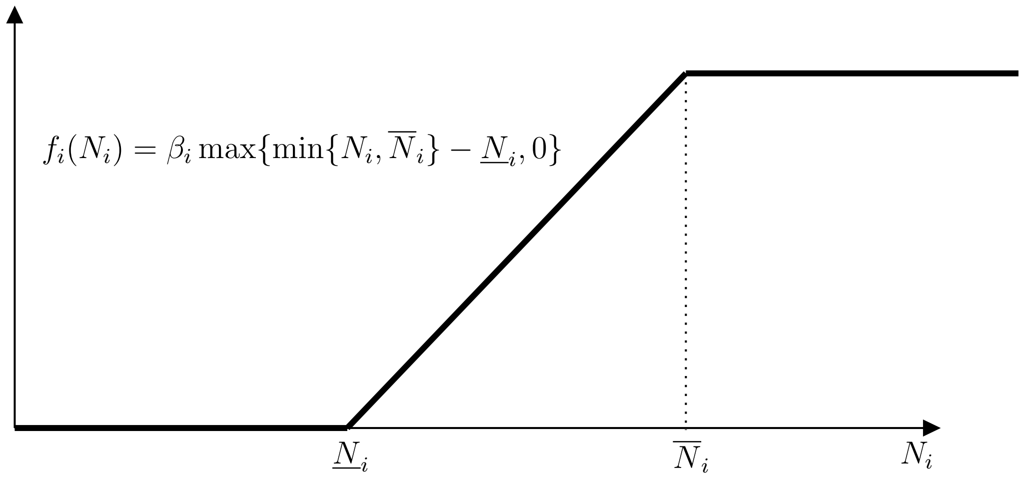

Figure 2: S-Curve Data-Enabled Learning Function

'for the market.' This involves the accumulated, discounted difference in surplus in future periods with the expectation that as the advantage of the market leader grows over time, the discounts offered to consumers to maintain that leadership fall.

There are two broad cases where the shape of the data-enabled learning functions might be decisive. First, suppose that $f_i(N_i) = \beta_i N_i$. In this case,

$$\Delta(N_I, N_E) = \beta_I \left(N_I + \frac{\delta}{1-\delta} \right) - \beta_E \left(N_E + \frac{\delta}{1-\delta} \right)$$

In this case, if $\beta_E > \beta_I$, even if $\beta_I N_I > \beta_E N_E$, E may prevail if $s_E - s_I > \Delta(N_I, N_E)$. Second, suppose that for some N_i a limit is reached whereby $f_i(N_i) = f_i(\bar{N}_i)$ for all $N_i \geq \bar{N}_i$. For instance, the data-enabled learning function may take the form as in Figure 2. In this case, notice that if I has already reached its limit, then $\Delta(\bar{N}_I, N_E) = f_I(\bar{N}_I) - (1 - \delta) \sum_{j=0}^{\infty} \delta^j f_E(N_E + j)$. Thus, any discount that I offers to continue to lead the market at this point is purely defensive; that is, it prevents E from obtaining a data advantage. However, this also implies that with a high enough discount factor and with $\bar{N}_E >> \bar{N}_I$, I's blocking ability will be limited.

The pressure of intertemporal competition determines what share of surplus accrues to consumers. Consumers benefit when firms set prices lower in order to maintain or obtain a future data advantage. Any increase in the data advantage of I (weakly) decreases consumer surplus. This effect is stronger if I and E are evenly matched so that even the firm that is behind finds it optimal to offer a subsidy in competition. In this situation, increasing N_I leads to a reduction in subsidization by E, lowering consumer surplus.

Thus, if policies are directed at improving consumer welfare, those policies need to maintain that competitive pressure. For instance, Hagiu and Wright (2023) demonstrates that a policy that required firms to share data would have ambiguous effects. When one firm is at a clear disadvantage (for instance, if $s_I - s_E$ is extremely high or extremely low), by closing that gap, competitive pressure is increased, and consumers benefit. However, when there is already close competition, data sharing reduces the incentives of the losing firm to discount and apply pressure to the winning firm as the consequences of losing are diminished. In this situation, consumers are worse off when data is shared.[5]

2.3 Trade in Data

This section examines the power firms might have resulting from control of their own data that may impact the extent of competition they face. A firm generates data that is used to train AI algorithms by observing aspects of their production as well as interactions with their own consumers. Jones and Tonetti (2020) argue that such data is not only useful to the firm gathering it but to other firms that may not be in direct competition with the firm. For instance, consider self-driving cars. If Tesla (firm i) observes the outcomes from total distance travelled, q_i, that data may be useful to others developing self-driving cars, such as Uber or Waymo. They formalise this as the total data input used for training by i, D_i is equal to $\alpha x_i q_i + (1-\alpha)B$ where x_i is the fraction of data generated, and B is a bundle of data that is purchased by i. B is larger the more data that is shared/sold which for firm i is $\tilde{x}_i$.

Absent any other considerations (such as consumer privacy), the firm will capture as much data as possible; i.e., $x_i = 1$. The amount of data shared, $\tilde{x}_i$, is also a choice of the firm. If the purchasers of B are incumbent competitors to i, this may mute the amount of data that i will be willing to release. However, as B is an aggregate of data collected from all firms, it is arguable that, in many situations, the marginal increment to B from i's data is relatively small. Thus, if $p_{s,i}$ is the price per unit of data sold by firm i, then the firm will set that price to maximise $p_{s,i}D_{s,i}$ where $D_{s,i} = \tilde{x}_i q_i$.

What is the demand curve facing the firm as a data seller? Jones and Tonetti (2020) assume that data is exchanged through an intermediary. That intermediary purchases data from firms and then assembles it into the bundle, B, before selling the bundle to firms. The aggregator function for B takes a CES form:

$$B = \left(\frac{1}{N^{\frac{1}{\epsilon}}} \int_0^N D_{s,i}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}$$

[5] Interestingly, when one firm has reached its limit of data-enabled learning while the other has not, that firm has a greater incentive to take actions that limit learning or data acquisition by their rival than their rival does to take counter actions. This is explored further with respect to input data mergers below.

where $\epsilon < 1$ is the elasticity of substitution between data supplied by firms across N sectors. The data intermediary chooses each $D_{s,i}$ given $p_{s,i}$ to maximise:

$$p_b NB - \int_0^N p_{s,i} D_{s,i} di$$

Taking the derivative of this with respect to $D_{s,i}$, setting it equal to 0 and re-arranging gives:

$$p_{s,i} = p_b N \left(\frac{B}{ND_{s,i}} \right)^{\frac{1}{\epsilon}}$$

Thus, firms face a constant elasticity demand function for their data and will exercise some market power over its provision.

2.3.1 Innovation Game

It is now time to more carefully specify firm choices. It is assumed that there are N sectors in the economy over which shared data is potentially valuable. Each sector has at most two firms present at any given time – an incumbent (I) and a potential entrant (E). In this regard, the individual sector market structure follows the assumptions of Segal and Whinston (2007). Each firm has a common discount factor, $\delta \in (0, 1)$ and time is discrete with an infinite horizon. At any given point in time, one of these firms holds a patent for a product that offers superior quality to previous products in that consumers in that sector value the product Δ more than the previous generation; that is, if there have been j generations of a product, the value of the current generation to consumers is $j\Delta$. Moreover, consumer value can be enhanced further using an AI trained on available data so that total consumer value is $(j-1)\Delta + \gamma D_i \Delta$ for sector i.[6] Here, $\gamma > 1$ is a parameter that controls the importance of data in generating value from the latest product. There is a continuum of identical consumers in each sector of measure 1. Thus, if the firm charges a price to consumers of $\gamma D_i \Delta$,[7] that becomes their total revenue. In that regard, effectively $q_i = 1$ for all i and, therefore, $D_i = \alpha x_i + (1-\alpha)B$. Thus, the *flow* profits for the patent holder of the highest-value product are:

$$\pi_m = \gamma D_i \Delta - p_b B + p_{s,i} \tilde{x}_i$$

The other firm, E, engages in research and development each period towards the next generation of product for the sector. They choose an innovation rate, $\phi_i \in [0, 1]$, at a

[6]The assumption that AI only enhances consumer value on the increment between the jth and $(j-1)$th products is made to simplify calculations in what follows.

[7]This assumes that the previous generation of the product, $j-1$, is freely available.

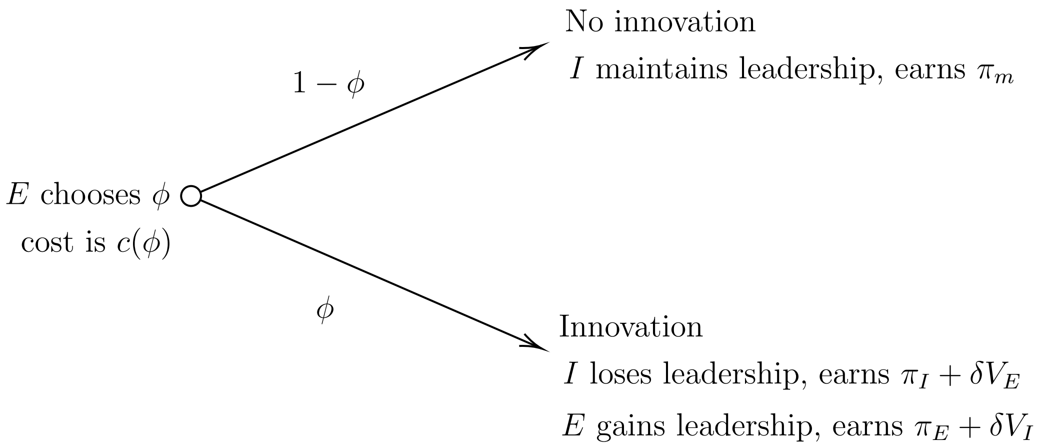

Figure 3: Per Period Outcomes

diffentiable, increasing and (strictly) convex cost of $c(\phi_i)$ where $c(0) = 0$ and $c'(1) = \infty$. In each period, t, with probability $1 - \phi_i$ they fail to innovate and with probability ϕ_i they innovate, secure a patent and can enter into competition with the current incumbent. That initiates one period of competition where both firms are present in the product market. Let D_E denote E's data during that period when they have not been able to generate training data of their own. E can, however, purchase data to assist in enhancing product value and, therefore, $D_E = (1-\alpha)B$ at a cost to them of $p_b B$. That said, if E manages to supply consumers upon entry, then at the end of that period, they are able to sell that data when they are assumed to become the monopolist incumbent. Moreover, it will allow them to collect training data for the next period. This allows them to earn $p_{s,i}\tilde{x}_i + \delta\alpha\Delta x_i$ more than they would if they did not succeed in selling to any customers during the competitive period.

Note that similar considerations apply to I. If it succeeds in selling to customers during the competitive period, it will generate data that it can sell as it exits at the end of that period, providing revenue of $p_{s,i}\tilde{x}_I$. However, as it no longer has a productive role following the competitive period and has effectively lost market leadership, the incumbent has no incentive to withhold data from sale. Therefore, $\tilde{x}_I = 1$.

Given this, I and E engage in Bertrand price competition. There are two outcomes:

1. (I continues to lead in competition with E) Suppose that $\gamma D_I \Delta + p_{s,i} > \Delta + \gamma D_E \Delta + p_{s,i}\tilde{x}_i + \delta\alpha\Delta$.[8] Thus, the maximum surplus E could offer customers is $\Delta + \gamma(1-\alpha)B\Delta + p_{s,i}\tilde{x}_i + \delta\alpha\Delta - p_b B$. Given that I can offer more surplus, this means that $\pi_I = \gamma D_I \Delta - p_b B + p_{s,i} - ((\gamma D_E + 1)\Delta - p_{s,i}\tilde{x}_E + p_{s,i} + \delta\alpha\Delta - p_b B) = ((\gamma-\delta)\alpha - 1)\Delta + p_{s,i}(1-\tilde{x}_i)$

[8] That is, the entrant's product is worth $j\Delta + \gamma D_E \Delta$ and the incumbent's product is worth $(j-1)\Delta + \gamma D_I \Delta$ to consumers plus the additional revenue I can receive by selling their data and generating training data for the future, so that the inequality involves the incumbent's product still being superior to the entrants.

while $\pi_E = 0$. Here the fact that $x_I = 1$ and the fact that $D_I - D_E = \gamma\alpha$ is used to calculate I's profit.

2. (E leads in competition with I) Suppose that $\gamma D_I \Delta + p_{s,i} < \Delta + \gamma D_E \Delta + p_{s,i}\tilde{x}_i + \delta\alpha\Delta$. Thus, the maximum surplus I could offer customers is $\gamma D_I \Delta + p_{s,i} - p_b B$. Given that E can offer more surplus, this means that $\pi_E = \Delta + \gamma D_E \Delta - p_b B + p_{s,i}\tilde{x}_i - (\gamma D_I \Delta + p_{s,i} - p_b B) = (1 - \gamma\alpha)\Delta - p_{s,i}(1 - \tilde{x}_i)$ while $\pi_I = 0$.

It is useful to compare these outcomes to a benchmark case where all data was collected and made freely available. This is a benchmark because data is nonrival in nature. In this case, E would be able to generate a product of value $j\Delta + \gamma D_E \Delta$ where $D_E = \alpha + (1-\alpha)N$ and would compete against I's product that has value of $(j-1)\Delta + \gamma D_I \Delta$ where D_I also equals $\alpha + (1-\alpha)N$. In this case, $\pi_E^* = \Delta$ and $\pi_I^* = 0$. Note that when data is not freely available or available at all, then entrant profit is lower and incumbent profit is higher than the benchmark case. Moreover, observe that π_E^* is less than π_I (where I remains the leader) and π_E (where E is the leader); that is, the sum of profits, $\pi_E + \pi_I$ is less than $\pi_E^* + \pi_I^*$.

Following Segal and Whinston (2007), the innovation choice of E is examined as a Markov perfect equilibrium of the dynamic game. Let V_E be E's expected present discounted profit and let V_I be I's expected present discounted profit (evaluated at the beginning of a period). If the probability that E innovates is ϕ, each of these can be calculated from the following:

$$V_E = \delta V_E + \phi\left(\pi_E \delta(V_I - V_E)\right) - c(\phi) \tag{VE}$$

$$V_I = \pi_m + \delta V_I + \phi\left(\pi_I - \pi_m + \delta(V_E - V_I)\right) \tag{VI}$$

In each period, E chooses ϕ to maximise its expected present discounted profits. If $w = \pi_E + \delta(V_I - V_E)$ is the benefit from successfully innovating (that Segal and Whinston (2007) term the *innovation prize*), the optimal innovation level is:

$$\Phi(w) =_{\phi \in [0,1]} \{\phi w - c(\phi)\}$$

$\Phi(w)$ is an innovation supply (IS) function whose inverse is graphed in Figure 4. Note that it is upward-sloping due to the assumptions on $c(.)$.

The innovation prize, w, is determined by subtracting (VE) from (VI) and solving for $(V_I - V_E)$ and then substituting that into $w = \pi_E + \delta(V_I - V_E)$ to give $W(\phi)$:

$$W(\phi) = \pi_E + \delta \frac{\phi\pi_I + (1-\phi)\pi_m - \phi\pi_E + c(\phi)}{1 - \delta + 2\delta\phi} \tag{IB}$$

This is depicted in Figure 4 and is downward sloping. The equilibrium level of innovation,

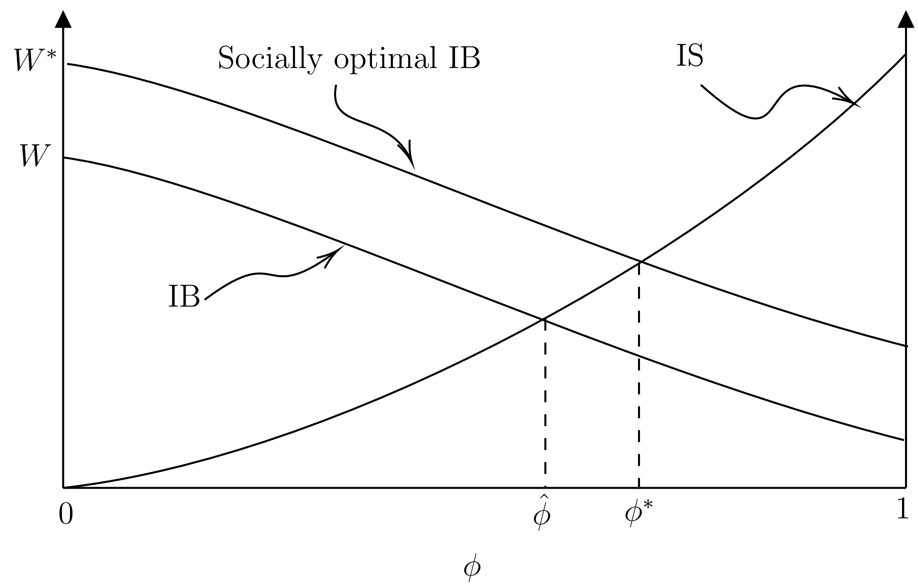

Figure 4: Socially Optimal and Equilibrium Innovation Rate

$\hat{\phi}$, arises where $\Phi(w)$ and $W(\phi)$ intersect. Importantly, note that changes that impact flow profits will impact on the level of $W(\phi)$ and, thus, the impact of policy changes on innovation can be assessed by looking at their impact on the (IB) curve only. Note also that the entrant is not only motivated by the profits they earn when in direct competition with I but also the value of becoming the incumbent themselves. Thus, policies that increase π_E may not end up encouraging a higher rate of innovation.

2.3.2 Is data an entry barrier?

We are now in a position to examine I's choice of $\tilde{x}_i$ when it is the monopoly leader in a sector. Recall that the inverse demand curve I faces as an information seller is $p_{s,i} = p_b N \left(\frac{B}{ND_{s,i}}\right)^{\frac{1}{\epsilon}}$ and substituting this into π_m gives:

$$\pi_m = \gamma D_i \Delta - p_b B + p_b B^{\frac{1}{\epsilon}} (N\tilde{x}_i)^{\frac{\epsilon-1}{\epsilon}}$$

Note that because sectors are infinitesimal, I's choice of $\tilde{x}_i$ does not impact on B and, therefore, does not impact on either π_E or π_I in the next period should that period consist of competition nor the next period's π_m if there is no new innovation. Thus, for an interior solution, the chosen $\tilde{x}_i$ will satisfy the first order condition:

$$p_b B^{\frac{1}{\epsilon}} \left(\frac{\epsilon-1}{\epsilon}\right) (N\tilde{x}_i)^{-\frac{1}{\epsilon}} \leq 0$$

However, the LHS equals $p_{s,i}N > 0$. Therefore, $\tilde{x}_i = 1$. Applied to all sectors $B = N$. Therefore,

$$\pi_m = \gamma D_i \Delta - p_b N + p_b N = \gamma D_i \Delta$$

Note, for completeness, that Jones and Tonetti (2020) assume the data intermediary is a perfectly contestable monopolist with profit, $p_b N^2 - p_{s,i} N = 0$, which implies that $p_b = \frac{p_{s,i}}{N}$.

As $x_i = \tilde{x}_i = 1$, note that $\gamma(\alpha + N(1-\alpha))\Delta < \Delta(1+\gamma N(1-\alpha)) + \delta\alpha\Delta$ or $\Delta(1-(\gamma-\delta)\alpha) > 0$ so that E becomes the leader in competition with I with $\pi_E = (1-\gamma\alpha)\Delta$ ad $\pi_I = 0$. Therefore:

$$W(\phi) = \frac{(1-\delta(1-\phi))(1-\gamma\alpha) + (1-\phi)\gamma(\alpha + N(1-\alpha))}{1-\delta+2\delta\phi}\Delta$$

By contrast, the socially optimal IB curve is found by allowing E to utilise I's data fully, resulting in:

$$W^*(\phi) = \frac{(1-\delta(1-\phi)) + (1-\phi)\gamma(\alpha + N(1-\alpha))}{1-\delta+2\delta\phi}\Delta$$

It is clear that $W^*(\phi) > W(\phi)$ for all ϕ (Figure 4). The reason is that if E does not have I's data, then I is relatively more competitive and reduces E's competitive profit by $\gamma\alpha\Delta$. This implies that the socially optimal innovation rate ϕ^* exceeds the equilibrium rate, $\hat{\phi}$. In effect, there is an entry barrier facing E because it does not have access to I's data, creating an asymmetry when competing for consumers.[9]

The above analysis assumes that when I sells its data to the intermediary, the data is only disclosed to others through the bundle, B. However, what if such data could be inferred or possibly leaked as a result of I engaging in the data market in the first place? Suppose that with probability, ρ, a firm's individualised data, $\tilde{x}_i$ could become available to E prior to the competitive period should it arise. In this case, $E[D_E] = \rho\alpha\tilde{x}_i + (1-\alpha)B$. In this case, $\mathbb{E}[\hat{\phi}(\rho\alpha\tilde{x}_i)]$ is increasing in $\tilde{x}_i$. Therefore, in choosing how much data to sell, I would take this possibility of leakage into account. Thus, I's first order condition in choosing $\tilde{x}_i$ becomes:

$$p_b B^{\frac{1}{\epsilon}} \left(\frac{\epsilon - 1}{\epsilon}\right) (N\tilde{x}_i)^{-\frac{1}{\epsilon}} + \frac{\partial \mathbb{E}[\hat{\phi}]}{\partial \tilde{x}_i} (\pi_I - \pi_m + \delta(V_E - V_I)) \leq 0$$

If this is an interior solution, then $\tilde{x}_i < 1$.

[9] Jones and Tonetti (2020) note that a regulatory environment that prevented any data sharing would be a poor outcome as $\tilde{x}_i = 0$ for all i. They also examine situations where consumers own data and have privacy concerns. This also leads to less data sharing, so in the context of the model, this would reduce the rate of innovation in each sector.

2.3.3 Data Neutrality

Given that the level of data barriers to entry are higher in equilibrium when incumbents are concerned that data sharing should increase the rate of creative destruction specific to their sector/market, are there policy interventions that could reduce these barriers? Note that, in equilibrium, the data gathered and used by incumbents internally is greater than the data they choose to share with others, including, in this model, potential entrants. One policy intervention that could eliminate this data advantage would be a policy of neutrality where $\tilde{x}_i$ is constrained to be no less than x_i. In equilibrium, $x_i = 1$ in the absence of this policy and, therefore, superficially, a policy of neutrality might raise the innovation rate.

This intuition presumes, however, that I does not change x_i in response to the policy. As $x_i = \tilde{x}_i$, the first order condition for I choosing x_i (substituting x_i for $\tilde{x}_i$) becomes:

$$\gamma\alpha\Delta + p_b B^{\frac{1}{\epsilon}} \left(\frac{\epsilon - 1}{\epsilon}\right) (Nx_i)^{-\frac{1}{\epsilon}} + \frac{\partial \mathbb{E}[\hat{\phi}]}{\partial x_i} (\pi_I - \pi_m + \delta(V_E - V_I)) \leq 0$$

If this is an interior solution, then $\hat{x}_i < 1$ although $\tilde{x}_i$ and hence B will be higher under this policy than in its absence. In equilibrium, π_m will equal $\gamma(\alpha\hat{x}_i + (1-\alpha)B)\Delta$. This will increase the rate of innovation if it increases W; that is, if

$$\frac{\partial \pi_E}{\partial x_i} + \frac{\delta}{1 - \delta(1-\phi)} \left((1-\phi)\frac{\partial \pi_m}{\partial x_i} + \phi\frac{\partial \pi_I}{\partial x_i}\right) > 0$$

Note, however, that $\frac{\partial \pi_m}{\partial x_i} = -\gamma\alpha\Delta$ and when it leads in competition, $\frac{\partial \pi_I}{\partial x_i} = 0$, as I will already be replaced next period and so is unconcerned about sharing. Meanwhile, $\frac{\partial \pi_E}{\partial x_i} = 0$ as a higher B provides E with no advantage in competition with I and, in competition, I is unconcerned about sharing. Thus, this condition becomes:

$$-\frac{\delta}{1 - \delta(1-\phi)}(1-\phi)\gamma\alpha\Delta < 0$$

In this model, a policy of neutrality reduces the level of data sharing without any impact on the payoffs of I and E during the competitive stage and thus also reduces the equilibrium innovation rate.

3 Market Power in Input Data

Input data is the key input to generating a prediction from a given AI algorithm. As mentioned above, relative to training data, input data can be thought of as a driver of competitiveness within a market rather than as something that determines market structure. Specifically, input data is often generated as a result of ongoing observations in the market, such as being able to predict changes in demand conditions over time. But it is also the case that such input data can be shared – or traded in an arms-length market – allowing any differences in each firm's predictions to be eliminated. As we will see, whether this occurs will have important implications for policies designed to take into account any market power a firm might have with respect to input data.

3.1 Input Data Advantages

Earlier in this paper, we considered the model of Hagiu and Wright (2023) with respect to firms gathering data over time in order to train superior AI prediction algorithms. Their model also considers another form of data-enabled learning whereby firms can gather data regarding individual customers and create AI algorithms that serve those particular customers with higher-quality products.

Hagiu and Wright (2023) show that a similar condition determines the winner and loser in within consumer learning as it did for across consumer learning. Specifically, E will win a consumer (although not necessarily the whole market) if, for that consumer, $s_E - s_I \geq \Delta(N_I, N_E)$ where N_I and N_I now represent data quantities held by each firm about that consumer. The nature of competition is different, however. When a consumer stays with a firm, they are granting that firm the ability to improve their product offerings relative to the competitor and, in the process, charge a higher price for it. Thus, to acquire the consumer requires a subsidy from the start. Hagiu and Wright (2023) show that this type of competition can be more beneficial to individual consumers than competition 'for the market' created by across-consumer data. In effect, competition in this situation creates consumer switching costs that create nuances in our assessment of the impact of firm market power.[10]

3.2 Sharing of Input Data

Above, the sharing/trading of training data was considered with respect to its impact on entry barriers and innovation. Here, we consider the sharing/trading of input data whose primary impact is on the short-term pricing and output outcomes arising from competition between firms when the market structure is fixed. Information sharing (of which sharing of input data is an important component) has extensive literature in industrial organisation beginning with work such as Vives (1984) and ultimately synthesised by Raith (1996).

[10]See, for example, Klemperer (1987).

Here, the model of Jansen (2008) is explored as it nests easily within the two-state setup pursued throughout this book. Specifically, demand for two firms is symmetric with an intercept, θ, that is a random variable which is $\underline{\theta}$ with prior probability $1-\rho$ and $\bar{\theta}$ with probability ρ. Each firm i chooses e_i which is the probability that it learns θ at a cost of $c(e_i) = \eta e_i$ ($\eta > 0$). With probability $1 - e_i$, the firm learns nothing. In this respect, η is a parameter that captures the cost of predicting the level of demand.

After the first stage, where the firms invest and potentially learn θ, they have an opportunity to share that information with their rival firm. If the firm reveals its input data, that information is verifiable, but what is not verifiable is whether the firm has information or not. Thus, if they do have a prediction of demand based on input data gathered, with probability $\sigma_i(\theta) \in [0,1]$, they disclose that data and, with probability $1-\sigma_i(\theta)$, they send an uninformative message to their rival (which is denoted '0'). If a firm does not have a prediction of demand based on their input data, that uninformative message is the only message that firm can send. Firm i's disclosure rule is $D_i = \{\sigma_i(\underline{\theta}), \sigma_i(\bar{\theta})\}$. Each firm chooses its disclosure rule simultaneously.

Following the information-sharing stage, the firms compete. It is assumed that marginal costs are 0. Jansen (2008) focuses on Cournot (quantity) competition where each firm i chooses q_i to maximise its profits holding the quantity of the other firm constant. If there was perfect information regarding θ, firm i would choose q_i to maximise:

$$\pi(q_i, q_j; \theta) = (\theta - q_i - \gamma q_j)q_i$$

where $j \neq i$ and $\gamma \in (0,1]$ represents the degree of product differentiation. This perfect information outcome arises whenever one firm learns θ and discloses that information to the other firm. In this case, $\hat{q}^*(\theta) = \frac{\theta}{2+\gamma}$. Note that $\pi_f = \hat{q}^*(\theta)^2$.

There are three possible disclosure strategies for firms: commit to disclose nothing ($D_i = \{0,0\}$), commit to disclose everything ($D_i = \{1,1\}$), or disclose selectively (i.e., disclose when $\theta = \underline{\theta}$ but not when $\theta = \bar{\theta}$ or $D_i = \{1,0\}$). The perhaps surprising thing is that competing firms might share data with their rival at all. However, as will be shown, in some cases, this changes a rival's competitive behaviour in a way that benefits a firm.

Here, we examine each strategy in turn and determine outcomes under an assumption of symmetry with $D_i = D_j$ (symmetric disclosure rules) and $e_1 = e_2 = e$ (symmetric data investments).

Disclose everything: If at least one firm is informed, then the competitive outcome will be the same as the complete information case; that is, the quantity chosen by each firm is $\hat{q}^f(\theta) = \hat{q}^*(\theta)$. This happens with probability $1-(1-e)^2$, resulting in expected profits are $\mathbb{E}[\pi^f|\theta] = \frac{\rho\bar{\theta}^2+(1-\rho)\underline{\theta}^2}{(2+\gamma)^2}$. If neither is informed, which occurs with probability $(1-e)^2$, then each firm's posterior equals their prior on θ, and they base their choices on expected demand $\rho\bar{\theta}+(1-\rho)\underline{\theta}-p_i-\gamma p_j$. Expected profits in this eventuality are $\mathbb{E}[\hat{q}^f] = \frac{\mathbb{E}[\theta]}{1+\gamma}$ and $\mathbb{E}[\pi^f] = \frac{\mathbb{E}[\theta]^2}{(2+\gamma)^2}$. Thus, $\mathbb{E}[\pi^f|e] = (1-(1-e)^2)\frac{\rho\bar{\theta}^2+(1-\rho)\underline{\theta}^2}{(2+\gamma)^2} + (1-e)^2\frac{(\rho\bar{\theta}+(1-\rho)\underline{\theta})^2}{(2+\gamma)^2}$.

Disclose nothing: When neither firm discloses anything, the belief a firm has is that with probability, e, their rival knows θ and does not know it otherwise. If a firm learns θ, it bases its quantity choice on that knowledge, while if it does not learn θ, it bases its choice on $\mathbb{E}[\theta] = \rho\bar{\theta}+(1-\rho)\underline{\theta}$. In equilibrium, the output of a firm that knows θ is $\hat{q}^0(\theta) = \hat{q}^f(\theta) + \frac{\gamma(1-e)}{(2+\gamma)(2+\gamma e)}(\theta-\mathbb{E}[\theta])$ while the output of a firm that does not know θ is $\hat{q}^0(0) = \hat{q}^f(\theta) - \frac{2+\gamma e}{(2+\gamma)(2+\gamma e)}(\theta-\mathbb{E}[\theta])$. Thus, $\mathbb{E}[\pi^0|e] = e(\rho\hat{q}^0(\bar{\theta})^2 + (1-\rho)\hat{q}^0(\underline{\theta})^2) + (1-e)\mathbb{E}[\hat{q}^0(0)^2]$.

Disclose only when demand is low In this case, when one firm discloses, the probability a firm assigns to whether their rival knows demand is 0 when demand is low. By contrast, no disclosure may mean that the firm has observed high demand or has obtained no prediction. Thus, in the absence of disclosure, a firm assesses that its rival knows with probability e that demand is high. Given this, the posterior probability that demand is higher for a firm that does not observe its own prediction is:

$$\tilde{\rho} = \frac{\rho}{\rho+(1-\rho)(1-e)}$$

Therefore, the equilibrium quantities for a firm who knows $\underline{\theta}$, $\bar{\theta}$ and does not observe or receive a prediction are:

$$\hat{q}^s(\underline{\theta}) = \hat{q}^f(\underline{\theta})$$

$$\hat{q}^s(\bar{\theta}) = \hat{q}^f(\underline{\theta}) + \frac{2+\gamma(1-\tilde{\rho})(\bar{\theta}-\underline{\theta})}{(2+\gamma)(2+\gamma(1-\tilde{\rho})e)}$$

$$\hat{q}^s(0) = \hat{q}^f(\underline{\theta}) + \frac{2\tilde{\rho}(\bar{\theta}-\underline{\theta})}{(2+\gamma)(2+\gamma(1-\tilde{\rho})e)}$$

Thus, $\mathbb{E}[\pi^s|e] = e(\rho\hat{q}^s(\bar{\theta})^2 + (1-\rho)\hat{q}^f(\underline{\theta})^2) + (1-e)\mathbb{E}[\hat{q}^s(0)^2]$.

Jansen (2008) finds that the rankings of these various equilibrium quantities (and with them expected profits) are as follows:

$$\hat{q}^0(\underline{\theta}) < \hat{q}^s(\underline{\theta}) = \hat{q}^f(\underline{\theta}) < \hat{q}^f(0) = \hat{q}^0(0) < \hat{q}^s(0) < \hat{q}^f(\bar{\theta}) < \hat{q}^s(\bar{\theta}) < \hat{q}^0(\bar{\theta})$$

Put simply, conditional on receiving a prediction, the more disclosure there is, the lower the variance associated with a firm's observation of a prediction. Given this, we can see that the effect of an increase in e is on equilibrium outputs. When there is no disclosure, if e is higher, each firm believes it is more likely their rival is informed. Informed firms with low demand expect a less aggressive response from their rivals, and so expand output, whereas the reverse is true when there is high demand. No similar effects occur, of course, when everything is disclosed, but the effects are asymmetric under strategic disclosure of low demand only.[11]

This analysis allows Jansen (2008) to characterise the ex ante expected profits for the firm under each regime, taking into account a given investment, e, in input data.

Proposition 1 *Ex ante, expected equilibrium profits are lowest under a regime whereby disclosure is only of low demand predictions. (i) If $\gamma > 2\sqrt{2} - 2$, ex ante expected profits are greatest under full disclosure. (ii) If If $\gamma < 2\sqrt{2} - 2$ then, for sufficiently low (high) e, ex ante expected profits are greatest under no disclosure (full disclosure).*

Intuitively, if a firm receives no information from its rival, its profits are higher under no disclosure than full disclosure. However, given levels of input data generation, disclosure of input data from a competitor will generate some additional profits. Depending on the degree of differentiation (a proxy for market power), one factor outweighs the other. Specifically, the lower the degree of differentiation, the higher the returns from full disclosure of input data, especially if it is expensive to generate that input data.

These potential equilibrium quantity and profit outcomes influence the choice by each firm in their input data to generate a demand prediction (e_i). Jansen (2008) shows that expected equilibrium profits for firm i when the disclosure regime is $l \in \{f, 0, s\}$ and i's investment is e_i and expected rival investment is e is:

$$\Pi^l(e_i, e) = \mathbb{E}[\pi^l(\theta)] - \psi^l(e) + e_i(\psi^l(e) - \eta)$$

where

$$\psi^l(e) \equiv \mathbb{E}[\pi^l(\theta)] - \pi^l(0) - e\mathbb{E}[\sigma^l(\theta)(\pi^l(\theta) - \pi^l(0)]$$

Here, $\mathbb{E}[\pi^l(\theta)] - \phi^l(e)$ is expected profit to i in the absence of any input data generated by i; e.g., this is $e\mathbb{E}[\pi^f(\theta)] + (1-e)\pi^f(0)$ under full disclosure and $\pi^0(0)$ under no disclosure. The remaining part of the expected profit is $e_i(\psi^l(e) - \eta)$, which is the contribution to profit arising from a firm's own input data generation. Thus, in equilibrium, $\psi^l(e)$ is the marginal benefit from input data generation that must be equated to the marginal cost of η. That marginal benefit, $\psi^l(e)$, itself has two components. The first is the marginal value of using a prediction versus not using one. The second is a substitution (or 'free rider') effect from their rival's input data generation and disclosure ($\mathbb{E}[\pi^l(\theta)] - \pi^l(0)$). Figure 5 shows the

[11] Jansen (2008) shows that when firms cannot commit to a disclosure regime and can choose to disclosure after observing their own demand prediction, the only equilibrium outcome is disclosure when demand is low. This critically relies on $e < 1$. When $e = 1$, other equilibrium outcomes are possible and consistent with other papers in the literature full information outcomes result.

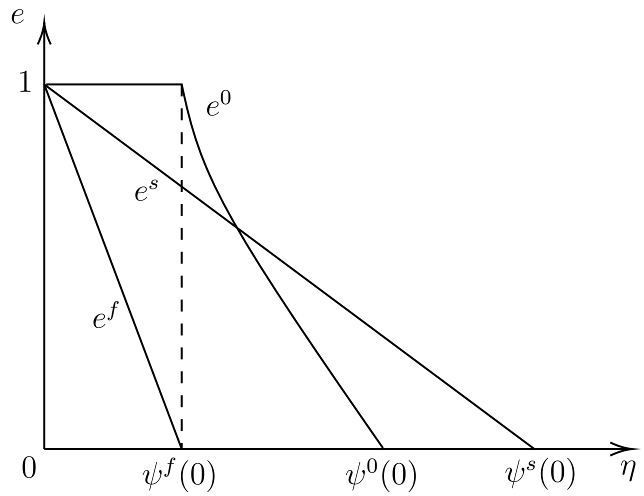

Figure 5: Input Data Investments

equilibrium investments in input data as a function of η. Notice that those investments are lower when there is a full disclosure regime.

This leads, however, to a rather remarkable result: the ex ante expected equilibrium profit outcomes for a firm under full disclosure and no disclosure are exactly the same. Note that, from 5, if $\psi^f(0) > \eta$, then $e_i^0 = 1$ for both firms under no disclosure. The equilibrium profit of the firms is, therefore, $\Pi^0[\hat{e}^0] = \mathbb{E}[\pi^f(\theta)] - \eta$. By contrast, under full disclosure, there is lower input data investment, $\hat{e}^f < 1$ (where $\hat{e}^f$ satisfied $\psi^f(\hat{e}^f) = \eta$). The lower input data generation lowers profits losing firms $(1 - \hat{e}^f)\psi^f(\hat{e}^f)$ but there is also reduced investment cost with savings of $(1 - \hat{e}^f)\eta$. These are, of course, equal in equilibrium, and so $\Pi^f[\hat{e}^f] = \Pi^0[\hat{e}^0]$. A similar argument holds for $\eta > \psi^f(0)$.[12]

Typically, the ex ante expected profits are higher for 'strategic' disclosure, whereby only low-demand outcomes are shared. This reverses the profit ranking compared to the case where e is exogenous. However, when the level of differentiation is sufficiently low (i.e., γ is high), these are out-ranked by the profits from full or no disclosure when input data generation costs are sufficiently high.

3.3 Input Data Mergers

The above analysis considers sharing data amongst competitors within a market. In some situations, there has been concern about sharing input data about consumers who make purchases in different markets. The idea is that data gathered in one market can be valuable to sellers in another market who serve the same consumers. This has also been the rationale

[12]This result relies on the linearity of input data generation costs. If these were convex, Jansen (2008) shows that full disclosure profits are higher than profits under no disclosure.

behind some cross-market merger investigations by antitrust authorities.

de Corniere and Taylor (2023) provide a simple model to analyse such cases.[13] They assume there is one market, A, that is served by a monopolist and another market, B, that is served by duopolists, B_1 and B_2. If q_A customers are served in market A, it is assumed that this creates q_A is data that potentially can be used by firms in market B.[14] If neither B firm has access to A-data, their profits are $\pi(0,0)$. If one has access but the other doesn't, the profits are $\pi(q_A,0)$ for the firm with access and $\pi(0,q_A)$ for the firm without access. And if both have access, their profits are $\pi(q_A,q_A)$. It is assumed that data increases profits – i.e., $\pi(q_A,0) > \pi(0,0)$ – and rival data reduces profits – i.e., $\pi(0,q_A) < \pi(0,0)$ and $\pi(q_A,q_A) < \pi(q_A,0)$. It is assumed that these data effects do not impact consumer behaviour in market A (perhaps because they are myopic to its effect).

The model involves, first, A collecting data, then data trade (if any) occurring, followed by competition and the realisation of profits in market B. Of key importance is whether the data trade that occurs prior to competition involves A selling data exclusively to one firm or non-exclusively to both. If A chooses to offer data exclusively, the payment they can receive is $\pi(q_A,0) - \pi(0,q_A)$ whereas if they offer the data non-exclusivity, the payment from any one firm is $\pi(q_A,q_A) - \pi(0,q_A)$. Thus, A will prefer non-exclusivity if:

$$\pi(q_A,0) - \pi(0,q_A) \leq 2(\pi(q_A,q_A) - \pi(0,q_A)) \Leftrightarrow \pi(q_A,0) + \pi(0,q_A) \leq 2\pi(q_A,q_A)$$

That is, if non-exclusivity results in higher total B-market profits.

Interestingly, this same condition determines whether exclusivity is optimal even if A and, say, B_1 have merged, implying that B_1 always has access to A's data. The maximum A can receive by also selling the data to B_2 is $\pi(q_A,q_A) - \pi(0,q_A)$. However, for that trade to be worthwhile, it must be the case that it exceeds $\pi(q_A,q_A) - \pi(q_A,0)$. In other words, that $\pi(q_A,0)+\pi(0,q_A) \leq 2\pi(q_A,q_A)$. Otherwise, A will not trade following the merger. This means that in comparing the merger to the non-merger outcome, there is no effect change in the amount of data traded.

A merger does potentially have an effect on A's incentives; specifically, its choice of q_A will not just be driven by what it can earn in market A but also what it might earn from selling its data. To understand that, suppose that the demand for A's product is $q_A = \max\{\alpha - p_A, 0\}$, and it faces no production costs. In setting q_A, the monopolist looks to maximise $q_A(\alpha - q_A) + R(q_A)$ where $R(q_A)$ are A's expected data revenues. Now, if, for some reason, A cannot sell data, then, by acquiring B_1, it allows B_1 to utilise that data.

[13] Chen et al. (2022) have a related model regarding the impact of data-driven mergers that focuses on foreclosure issues and does not take into account pre-merger data trading that might occur.

[14] The fact that the data units are the same as the customer units is a notational convenience.

Thus, the merged firm's profits will be $q_A(\alpha - q_A) + \pi(q_A, 0)$. As $\pi(q_A, 0)$ is increasing in q_A, this means that after the merger, A will set q_A higher. This will benefit both consumers in the A market, who face a lower price, and consumers in the B market, who now get the benefits of data being used to, say, increase product quality for B_1 and reduce prices, at least in quality-adjusted terms.

What happens if data trade is possible? Suppose that, prior to the merger, exclusive data selling was optimal so that $R(q_A) = \pi(q_A, 0) - \pi(0, q_A)$. Then pre-merger, A chooses q_A to maximise $q_A(\alpha - q_A) + \pi(q_A, 0) - \pi(0, q_A)$. By contrast, post-merger, A will not trade data outside of the merged entity and so will choose q_A to maximise $q_A(\alpha - q_A) + \pi(q_A, 0)$. In this case, because $\pi(0, q_A)$ is decreasing in q_A, A will have an incentive to set q_A lower following the merger. This will result in higher prices and lower consumer surplus in both markets.[15]

What if data trade was non-exclusive prior to the merger? Then $R(q_A) = 2(\pi(q_A, q_A) - \pi(0, q_A))$ and A's objective is to maximise $q_A(\alpha - q_A) + 2(\pi(q_A, q_A) - \pi(0, q_A))$. Post-merger, this objective will be $q_A(\alpha - q_A) + 2\pi(q_A, q_A) - \pi(0, q_A)$. So, once again, the merger will lead to higher prices and lower consumer surplus in both markets.

The conclusion of this analysis suggests that assessing the impact of a merger on market power will critically depend on whether input data can be traded pre-merger or not. Even if A is acquiring the one firm it has an exclusive data arrangement with; this will still have the effect of changing A's incentives and potentially reducing consumer surplus in both markets.

de Corniere and Taylor (2023) show that these conclusions depend critically on how data impacts the B market. For instance, if data allows price discrimination that also harms consumers, then the merger, by reducing the amount of input data used in the B market, may have some benefits for consumers.

In summary, the implication of this section is that assessments of how input data may impact market power in other markets depend critically on the type of and level of incentives to share or trade that input data. The conditions for that inform regulators of the likely changes that may occur to that trade or to the generation of input data following, say, a merger or perhaps the impact of other policies such as privacy laws.

[15] Hagiu and Wright (2023) whose model was exposited earlier have a related implication in their environment. There, if one of the competing firms acquires a firm in another market for the purpose of gathering additional consumer data, then this can lead to socially inefficient outcomes as the rents accrue to the firm being acquired rather than consumers. Moreover, it may be the case that a firm that has already reached the threshold of data usage may have an incentive to acquire the outside firm in order to block its rival from the acquisition. They term this a 'killer data acquisition.'

4 Market Power in Predictions

As AI develops, there may emerge firms whose main product is to sell AI predictions. At the moment, the closest to this that has developed in markets is where firms predict the quality of matches between consumers and firms. This has had an impact on advertising markets. In this section of the paper, models where predictions are bought and sold are examined. To be sure, some of the models with training or input data involve similar considerations, but these particular models rest on the quality of the prediction output and, hence, arguably are more about those outputs than the inputs driving them.

4.1 Predictions of Match Quality

In modern digital platforms, advertisers engage in various targeting strategies that involve limiting the reach of advertising to those consumers to which advertising will be more effective; that is, where they have a higher predicted match quality. This has advantages to those selling advertising space as scarce advertising inventory (itself limited by the amount of consumer attention on a platform) can be optimised with respect to revenue per unit of space.[16] However, due to differences in data, different platforms have different capabilities to provide predictions of match quality. If relevant data were more freely available, this would impact competition for the provision of such predictions and potentially competition between platforms. Here, how such market power in predictions of match quality manifests itself is examined.

Here, the environment of Bergemann and Bonatti (2015) is developed. Their motivation began with "cookies," which were information gathered by websites on the identity and other information of users who visited those websites. That information may be gathered directly by platforms or sold through intermediate markets. However, in terms of their implications, the cookies represent not so much data (or, specifically, input data) but training data that allows a richer prediction of the match quality of any given consumer. So long as that consumer can be tracked (also made possible by cookies), when they visit a website, they are accompanied by a prediction of match quality for advertisers on that site.

There is a unit continuum of consumers ($i \in [0, 1]$), a unit continuum of firms (or "advertisers") ($j \in [0, 1]$), a single publisher (or website) and initially, a monopolist provider of match predictions. Each consumer and advertiser generates a potential match value, v, uniformly distributed on $[0, 1]$. Each advertiser chooses $q_{ij} \geq 0$ directed at consumer i, which is the probability that consumer i is aware of j's product or *match intensity*. Given this, the

[16] For discussions see Athey and Gans (2010) and Bergemann and Bonatti (2011).

expected profit of an advertiser choosing q is:

$$\pi(v, q) = vq - cm(q)$$

where c is the cost of advertising space and $m(q)$ is the amount of advertising space purchased to generate q. While Bergemann and Bonatti (2015) consider more general environments, here the focus will be on a linear cost environment where $m(q) = q$.

Advertisers purchase predictions of match values, v, can a given set of consumers, $A \subset V$. That is, if advertisers purchase a prediction that includes match value v that allows it to identify all consumers with that match value. So, while predictions are for the match value of a given identified consumer, i, the thing that advertisers are interested in is the type of consumers predicted. Thus, their demand for predictions is with respect to the predicted match value rather than the match value of a specific consumer. Put simply, advertisers only care about match value and so their willingness to pay for predictions is with respect to v rather than the v_{ij} for specific i's. One thing this framing allows is that if the advertiser purchases predictions, A, i.e., the *targeted* set, it also receives a signal regarding whether a consumer visiting a website is in its complement, A^C, the *residual* set. This will play an important role in what follows.

If an advertiser knew v, then it would choose to advertise if $v > c$ and not otherwise. Thus, an advertiser's complete information profits are $\pi(v) = \max\{v - c, 0\}$. If the advertiser purchases a prediction for a set A, it will be able to achieve this level of profits on that set. On the residual set, however, it may also choose to set $q^*(A^C) = 1$ if and only if $\mathbb{E}[v|v\ inA^C] \geq c$. While it is often thought that an advertiser will want information on those consumers with the highest v as these are the consumers it wants to target with ads, that is, *positive targeting*, this specification immediately exposes a distinct strategy, *negative targeting*, where advertisers obtain information on those consumers with the lowest v and then explicitly do not advertise to them. Thus, in purchasing predictions, an advertiser using positive targeting will choose a threshold, v^*, thereby setting $A = [v^*, 1]$ while under negative targeting, a threshold is chosen, thereby setting $A = [0, v^*]$. The choice of strategy and threshold will depend on c and also on the price of predictions, p.

The threshold, v^*, depends on whether positive or negative targeting is adopted. Under positive targeting, the surplus from an ad is $v - c$, and so the advertiser will purchase predictions up to the point of indifference; that is, $v^* - c = p$ or $v^* = c + p$. Under negative targeting, the benefit of a prediction is to avoid advertising to a consumer and saving $c - v$, which implies the point of indifference will be $c - v^* = p$ or $v^* = p - c$. It is the cost of advertising space, c, that will determine which targeting strategy is optimal. If c is higher, the advertiser wants to limit ads to a select group of the highest-value consumers, so positive targeting is optimal. If c is relatively low, the advertiser finds it optimal to advertise to a wide set of higher-value consumers, and so negative targeting is optimal. Thus, it can be shown that the optimal set of predictions, $A(c,p)$ is:

$$A(c,p) = \begin{cases} [0, \max\{c-p, 0\}] & \text{if } c < \frac{1}{2}; \\ [\min\{c+p, 1\}, 1] & \text{if } c \geq \frac{1}{2}. \end{cases}$$

One feature of this is that the advertiser always chooses a different q on the targeted set and the residual set – with advertisements shown to the targeted set and not the residual set under positive targeting and advertisements shown to the residual set and not the targeted set under negative targeting.[17]

Given this specification of demand, we are now in a position to examine the price of predictions (p), assuming first a monopoly seller of predictions before considering the behaviour of many sellers. Interestingly, while it is natural to assume that predictions are a straightforward input into advertising, it turns out that the pricing strategy is more subtle than simply pricing an input. Predictions allow advertisers to concentrate their advertising on a smaller set of consumers and thus, predictions are an option to concentrate whenever advertising space becomes more expensive. In other words, while intuition would suggest that prediction purchases complement advertising decisions because, when advertising space is relatively cheap, negative targeting is optimal, prediction purchases may substitute for advertising decisions. Thus, the monopoly price of predictions is given by:

$$p^*(c) = \begin{cases} \arg\max_p [p(c-p)] & \text{if } c < \frac{1}{2}, \\ \arg\max_p [p(1-c-p)] & \text{if } c \geq \frac{1}{2}, \end{cases}$$

and therefore

$$p^*(c) = \tfrac{1}{2}\min\{c, 1-c\}$$

It follows from the earlier equation of prediction demand that the equilibrium level of prediction sales is:

$$A(c, p^*(c)) = \begin{cases} [0, \frac{c}{2}] & \text{if } c < \frac{1}{2}; \\ [\frac{1+c}{2}, 1] & \text{if } c \geq \frac{1}{2}. \end{cases}$$

Finally, profits rise with c until $c = \frac{1}{2}$ and then fall thereafter. This implies that predictions are most valuable when advertising space is neither very expensive nor very cheap. When

[17]The fact that these two strategies are demarcated by $c = \frac{1}{2}$ is due to the assumption that v is uniformly distributed on $[0, 1]$.

advertising space is very expensive or very cheap, prediction only alters the optimal advertising action (i.e., advertising to none or advertising to all) for a small set of consumers and, hence, has little value.

What happens when there are more sellers of predictions? Clearly, if more than one seller had access to the complete set of predictions, V, then competition would be in Bertrand and fall to the unit price of a prediction. In this case, advertisers would gain, relative to the monopoly provider case, a greater amount in surplus where advertising space is of intermediate cost. However, Bergemann and Bonatti (2015) analyses an interesting form of multiple seller environment where each seller has exclusive predictions for a specific consumer set i. Therefore, there is a continuum of sellers, each setting a price of p for that consumer. This might arise when consumers sell their data to a platform or when data is exchanged, and providers purchase data on individuals.

To keep the analysis simple, suppose that c is sufficiently high that positive targeting is optimal. Advertisers then are interested in purchasing a prediction set $A = [v_2, 1]$. The question is what determines the threshold, v_2. If the advertiser does not purchase predictions for $v < v_2$, then, under positive targeting, it will set $q = 0$ for those consumers. Thus, their willingness to pay, $p(v, v_2)$ for a prediction with valuation $v < v_2$ is:

$$p(v, v_2) \equiv (\pi(v, q^*(v)) - \pi(v, q^*([0, v_2]))$$

Given that match values are identically distributed, a symmetric pricing equilibrium can be reformulated as a prediction seller choosing a threshold, v_2 to maximise:

$$\arg\max_{v} p(v, v_2)(1 - F(v))$$

This differs from the monopoly problem in that the residual advertising intensity, $q^*([v, v_2])$, cannot be influenced by any one seller. Interestingly, the monopolist has a greater incentive to lower prices and expand supply. By expanding supply, the gap between the complete and incomplete profits of the marginal advertiser is increased (reducing v_2) and requiring the monopolist to lower price. This also changes the composition of the residual set – decreasing the average value in that set. This means that the marginal advertiser just outside the targeted set has a higher value of information, creating a further incentive to lower prices and expand supply. By contrast, this positive externality coming from the change in composition is not internalised by competing sellers. They, therefore, have a lower incentive to reduce prices. In equilibrium, prices, therefore, will be *higher* when prediction sellers are fragmented.

Selling predictions can result in subtle external effects that impact incentives to purchase predictions when purchasers are competing against one another. Here, this is not the case but the sales of one prediction impact on the marginal value of other predictions (positively) and multiple sellers do not internalise these effects. The end result can be reduced prediction sales in the market.

4.2 Bundling Attention and Prediction

Online platforms attract consumer attention and sell that attention to advertisers. Suppose any given consumer values one and only one of k products, and there are Nk consumers. If all consumers are observationally identical to the platform, then an advertiser of one of the k products has a probability of $\frac{1}{k}$ of being a match with a consumer they advertise to. Consequently, an advertiser's willingness to pay to place an ad in front of one consumer on a platform is $\frac{v}{k}$ where v is the value of a correct match (i.e., the amount the advertiser expects to earn from the consumer). If, on the other hand, the platform is able to predict matches with perfect accuracy, an advertiser's willingness to pay for an ad placed in front of the correct match is v. Thus, an advertiser who wants to ensure that the correct consumer sees their advertisement will either purchase N ads (if the match can be predicted perfectly) or Nk ads if there is no prediction. In each case, the expected payoff to the advertiser is Nv, while the cost depends on the price charged by the platform to advertise to each consumer. In either case, the maximum the platform can charge for a targeted (predicted) or non-targeted campaign is Nv (Athey and Gans, 2010).

Suppose that advertising value, v, is distributed uniformly on $[0, 1]$; that is, there is a continuum of advertisers for each product with differing values from advertising. Then, if ads are non-targeted, if there is advertising space of a ads per consumer, then the platform can supply at most Na campaigns and so its profits are $(1 - a)Na$. When ads are targeted, the platform can supply Nka campaigns and so its profits are $(1 - ka)Nka$. Note, however, that the platform can also choose a and for the non-targeted case, it would choose $a_n^* = \frac{1}{2}$ and for the targeted case it would choose $a_t^* = \frac{1}{2k} < a_n^*$. In the end, profits in each case would be $\frac{N}{4}$ and the campaign price would be $\frac{1}{2}$ with each ad price being k times lower in the non-targeted than the targeted case. In other words, when advertising space can be freely chosen, there is no payoff to the platform from being able to predict matches. However, if a is limited, then in the non-targeted case, profits will be lower with targeting, allowing the platform to better allocate that scarce ad space efficiently. In this case, prediction will increase a platform's profits.

This demonstrates that what makes prediction valuable is whether there is scarcity, in this case, in attention. What limits the ability to show ads is the amount of consumer attention that can be 'sold.' This may come from simple cognitive constraints but also from other factors, such as advertisements being annoying for consumers, creating incentives for them to avoid advertisements if there are too many of them (Anderson and Gans, 2011).

Of course, advertising does not exist in a vacuum, and it is a potentially essential tool that firms can use in competition with one another. Prat and Valletti (2022) show that this creates more potential for platforms to bundle attention and prediction together with consequent impacts on competition, not just in prediction but also in advertisers' own markets. They assume that each of the k products is currently served by a monopolist incumbent firm that does not, in fact, need to advertise because consumers are already aware of their product. This would seem to be bad news for platforms, except that for each market, there is a potential entrant. Entrants can enter successfully if consumers are made aware of their product. Suppose that the monopolist in each market earns π_m per consumer while if there is competition, its per consumer profit falls to π_c, which is the same as the entrant.

Prat and Valletti (2022) demonstrate that this potential competition now creates an incentive for both the incumbent and entrant to advertise. The entrant needs to advertise to a consumer just once so that the consumer is aware of its product. Thus, their willingness to pay for a single ad is π_c. On the other hand, the incumbent can, by purchasing ads to every consumer interested in their product, that is, a ads, preserve their monopoly profits. Therefore, the incumbent's willingness to pay for a ads is $\pi_m - \pi_c$. Importantly, this means that a platform will want to sell a campaign to the incumbent or the entrant and not both. It will be profitable to sell to the incumbent if $\pi_m - \pi_c > \pi_c$ or $\pi_m > 2\pi_c$. Otherwise, it is profitable to sell to the entrant.[18]

This analysis assumes that there is only a single platform capturing a consumer's attention. If consumers multi-home across more than one platform, then it remains the case that the entrant only needs to place an ad on a single platform to give it an entry opportunity. However, for the incumbent, being able to block the entrant requires advertising on all platforms. As an entrant is willing to pay up to π_c on each platform, if there are J platforms upon which a consumer's attention is distributed, the incumbent must pay at least $J\pi_c$ to block the entrant. Thus, the incumbent will only find this worthwhile if $\pi_m - \pi_c > J\pi_c$, which is a condition less likely to hold as J increases. Thus, a platform's ability to broker attention and sell targeted ads is critically related to the exclusivity of the access it has to any given consumer.

What does this analysis imply for the value of prediction itself? When prediction does not allow for ad-targeting, both the incumbent and entrant have a $\frac{1}{k}$ chance that the ad

[18] Prat and Valletti (2022) show this in a framework where each ad slot is auctioned off by a second-price auction, and this same condition drives whether the incumbent or entrant purchases an ad campaign.

shown to a given consumer is relevant to serve their intention of blocking or information, respectively. However, interestingly, incumbents benefit from advertising by other incumbents because each of their ads has a $\frac{1}{k}$ probability of being a blocking ad for their own consumer. This suggests that incumbent blocking is easier to achieve under non-targeting than with targeting. Indeed, if incumbents purchase J ads as before, then in total, they will purchase kJ ads, and all entrants will be blocked. However, any given incumbent could choose not to purchase any ads. In this case, the chance that their entrant is able to 'breakthrough' is still very small. While working out the precise strategy is complex, the fact that incumbent advertising under non-targeting creates positive externalities for other incumbents implies that there is more chance that entrants in some markets will actually enter. By contrast, with targeting, that possibility can be reduced to zero.

5 Conclusion

Precisely how market power might impact the various markets that constitute the provision of AI is an open question. This paper has examined some of the theories that outline the main trade-offs in assessing market power in markets for training data, input data and predictions themselves. Of course, by separating out these analyses, some of the linkages between those markets have not been considered. For instance, what happens if firms are integrated into two or more of these markets and have market power in one or more of them? On the one hand, this may enable those firms to take advantage of dynamic effects and feedback loops that might be present to take advantage of continuous learning opportunities. At the same time, this may ultimately lead to a situation where competitive entry is difficult as it would require operations in all of those markets. Finally, the focus here has been on markets that are responsible for generating AI. However, AI is itself an input into many other industries and markets and thus, the potential for the leverage of market power in either direction is something that could prove relevant for policy responses to mitigate market power in AI.

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Prediction machines: The Simple Economics of Artificial Intelligence*, Harvard Business Press, 2018.

_ , Joshua S Gans, and Avi Goldfarb, *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Harvard Business Review Press, 2022.

Anderson, Simon P and Joshua S Gans, "Platform siphoning: Ad-avoidance and media content," *American Economic Journal: Microeconomics*, 2011, *3* (4), 1–34.

Athey, Susan and Joshua S Gans, "The impact of targeting technology on advertising markets and media competition," *American Economic Review*, 2010, *100* (2), 608–613.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki, "The impact of big data on firm performance: An empirical investigation," in "AEA papers and proceedings," Vol. 109 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2019, pp. 33–37.

Bergemann, Dirk and Alessandro Bonatti, "Targeting in advertising markets: implications for offline versus online media," *The RAND Journal of Economics*, 2011, *42* (3), 417–443.

_ and _ , "Selling cookies," *American Economic Journal: Microeconomics*, 2015, *7* (3), 259–294.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello, "Artificial Intelligence, Algorithmic Recommendations and Competition," *Algorithmic Recommendations and Competition (May 14, 2023)*, 2023.

Chen, Zhijun, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima, "Data-driven mergers and personalization," *The RAND Journal of Economics*, 2022, *53* (1), 3–31.

de Corniere, Alexandre and Greg Taylor, "Data-Driven Mergers," Technical Report 2023.

Farboodi, Maryam and Laura Veldkamp, "A model of the data economy," Technical Report, National Bureau of Economic Research 2021.

_ and _ , "Data and markets," *Annual Review of Economics*, 2023, *15*, 23–40.

Gurkan, Huseyin and Francis de Véricourt, "Contracting, pricing, and data collection under the AI flywheel effect," *Management Science*, 2022, *68* (12), 8791–8808.

Hagiu, Andrei and Julian Wright, "Data-enabled learning, network effects and competitive advantage," *RAND Journal of Economics*, 2023.

Jansen, Jos, "Information Acquisition and Strategic Disclosure in Oligopoly," *Journal of Economics & Management Strategy*, 2008, *17* (1), 113–148.

Jones, Charles I and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Klemperer, Paul, "The competitiveness of markets with switching costs," *The RAND Journal of Economics*, 1987, pp. 138–150.

Prat, Andrea and Tommaso Valletti, "Attention Oligopoly," *American Economic Journal: Microeconomics*, August 2022, *14* (3), 530–57.

Raith, Michael, "A general model of information sharing in oligopoly," *Journal of economic theory*, 1996, *71* (1), 260–288.

Rosenberg, Nathan, *Inside the black box: technology and economics*, cambridge university press, 1982.

Schaefer, Maximilian and Geza Sapi, "Complementarities in learning from data: Insights from general search," *Information Economics and Policy*, 2023, *65*, 101063.

Segal, Ilya and Michael D Whinston, "Antitrust in Innovative Industries," *American Economic Review*, 2007, *97* (5), 1703–1730.

Varian, Hal, "Artificial Intelligence, Economics, and Industrial Organization," *The Economics of Artificial Intelligence: An Agenda*, 2019, p. 399.

Vives, Xavier, "Duopoly information equilibrium: Cournot and Bertrand," *Journal of economic theory*, 1984, *34* (1), 71–94.

Wilson, Robert, "Informational economies of scale," *The Bell Journal of Economics*, 1975, pp. 184–195.

Zhou, Jidong, "Improved information in search markets," Technical Report 2022.