

Sen, Anindya

Article

Are data markets a solution to big tech market power? A competitive analysis

Journal of Government and Economics (JGE)

Provided in Cooperation with:

Elsevier

Suggested Citation: Sen, Anindya (2022) : Are data markets a solution to big tech market power? A competitive analysis, Journal of Government and Economics (JGE), ISSN 2667-3193, Elsevier, Amsterdam, Vol. 7, pp. 1-12,
<https://doi.org/10.1016/j.jge.2022.100052>

This Version is available at:

<https://hdl.handle.net/10419/340698>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Are data markets a solution to big tech market power? A competitive analysis[☆]

Anindya Sen

Department of Economics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

JEL classification:

L1
L2
L4

Keywords:

Big tech
Data markets
Data portability
Data governance
Targeted advertising

ABSTRACT

This paper explores whether the establishment of data markets based on individual data portability can result in better societal outcomes. The results suggest that markets where individuals can sell data generated through their online engagement to third parties, could result in pareto improving outcomes for subscribers to digital platforms and purchasers of targeted advertising services. Data markets would enable third parties to combine their own proprietary data with other individual level data and produce information for targeted advertising, reducing the market power of Big Tech firms. However, successful data markets require strong regulatory measures by governments that ensure privacy and that data collected by Big Tech firms are considered the property of individuals. Such policies have the potential to guarantee that the benefits of Big Data are not confined to a few large firms.

Introduction

The past decade has witnessed the exponential growth of Big Tech firms that are Multi-Sided Platforms (MSPs), which are platforms or technologies that create value by facilitating direct interactions between two or more participant groups.¹ Google, Facebook, and Amazon are the most successful of these firms in terms of harnessing data generated through consumer use of services available from their platforms and employing Artificial Intelligence (AI) based methods to convert them into information used for targeted advertising. The objective of this

paper is to understand whether the creation of data markets that allow third party firms to purchase data from MSPs, have the potential to result in more competition for individual data and in the market for targeted advertising services.

Recent calls for government regulation of MSPs have been prompted by a variety of public policy concerns relating to the market power of such firms. First, the massive amount of data collected by MSPs have become a significant barrier to entry, as any incumbent would need comparable information to become a viable competitor. While there might be some debate on whether Facebook, Google, and Amazon are

[☆] I acknowledge comments from a very helpful referee, Andy Bazilauskas, Amit Bubna, Phil Curry, Steffen Ziss, Daniel Goetz, and Juan Morales. I thank Rishav Raj Agarwal and John Baker for excellent research assistance. The usual disclaimer applies.

E-mail address: asen@uwaterloo.ca.

¹ This definition is taken from Hagi (2013). Aside from Facebook and Google, other examples of MSPs noted by Hagi (2013) include: Alibaba.com and eBay which connect buyers and sellers; Airbnb which allows renters to find vacancies from dwelling owners; Uber and Lyft that allow passengers and professional drivers to connect in real time and passengers; Apple's iOS which provides pathways between application developers and users; and American Express, PayPal and Square which provide online payment platforms between merchants and consumers.

monopolies, there can be no doubt that the data collected by these firms have given them considerable market power, and they have been charged for a variety of anti-competitive practices.² A recent example of the adverse impacts of this market power was Facebook's 2021 decision to temporarily block news sharing on its platform in Australia. There is evidence that this action negatively impacted local organizations that relied on Facebook to disseminate information of public importance.³ Second, there is growing demand for some redistribution of MSP profits back to platform users on the principles of fairness and equity, as advertising revenues for these firms are driven by data generated by individual subscribers to their platforms.⁴⁵ A third reason behind the demand for regulation relates to public accountability for the data collected by MSPs and concerns that individual privacy may be compromised, with such data being used in ways that are harmful to society.⁶⁷

On the other hand, it cannot be denied that subscribers do receive benefits from platform membership. Facebook and Google offer substantial free services. Further, targeted advertising not only benefits MSPs but also platform users, as they are offered an extensive menu of possible purchasing opportunities based on preferences gleaned from their own individual interaction with platforms. Arguably, this leads to a reduction in search costs and cheaper prices as consumers are efficiently made aware of competing options. However, researchers have suggested

² Since 2017 Google has faced fines worth more than US\$ 9 billion for abuse of its dominant position in online search advertising, giving prominence to its shopping platforms after internet searches, and pushing its own apps on smartphone users at the expense of products from other competitors. [Riley and Ivana Kottasová \(2019\)](#) contain a concise summary of these charges. These charges are consistent with [Zuboff \(2019\)](#) who points out that data accumulation by MSPs have also enabled them to construct accurate predictions of consumer behaviour and develop AI algorithms that nudge individuals to certain purchasing decisions, rather than presenting unbiased information that can be used to make independent decisions. Facebook has also faced competition issues. In February 2019, The German competition authority (Bundeskartellamt) arrived at the conclusion that Facebook abused its market dominance in collecting, merging and using user data across different social media platforms and forbade Facebook from combining users' data from separate apps like WhatsApp and Instagram without explicit voluntary user consent. This order was subsequently dismissed by a German high court. For further details please see [Colangelo \(2019\)](#). Finally, it is also important to note the recent actions by The House Judiciary Subcommittee on Antitrust in calling CEOs to address concerns on anti-competitive practices by Big Tech companies. Please see [Foer \(2020\)](#) for further details.

³ For further details please see <https://www.bbc.com/news/world-australia-56165015>. Facebook's decision was prompted by its dispute with the Australian Government on payment to news publishers for content. Google was also impacted by this proposed policy and threatened to shut down its search engine in Australia if the Australian government implemented such regulation.

⁴ This is in fact, the main source of revenue for both Facebook and Google, with Amazon experiencing massive growth in its own advertising business. Google and Amazon currently have over \$1 trillion mark in market capitalization and despite declining share prices over the past few years, Facebook is still at an impressive \$450 billion. For further details on Amazon's growth in advertising revenues please see [Liu \(2020\)](#). [Graham \(2020\)](#) has a discussion of Amazon's advertising growth.

⁵ Please see [Lanier and Weyl \(2018\)](#) for an insightful discussion on how a more equitable redistribution of profits from MSPs to individual subscribers can be based on the concept of equity and 'data dignity'.

⁶ The best example, of course, is the Cambridge Analytica scandal in which the personal information of over 50 million Facebook users was hacked and harvested to create fake news and false advertising and influence the decisions of voters, which impacted the outcome of the 2016 U.S. Presidential election.

⁷ [Savona \(2019a, 2019b\)](#) provides an excellent discussion of these issues.

that the value of such services are lower than payments individuals should be receiving for generating data that are the source of significant profits enjoyed by these firms.⁸

This paper relies on simple partial equilibrium models that are traditionally used in competition policy/antitrust analysis to analyze the surplus creation and distributional effects that may occur with the introduction of public accessible data markets. To ensure that the modelling and main results can be followed by a general audience, most of the analysis is done through simple graphs and we do not provide mathematical proofs on the existence and sustainability of different equilibria. The key assumption of this study is that individual subscribers have the property right to data that are created from online engagement with MSPs. They also have the freedom to port their data from one platform to another, and therefore sell data generated through online engagement with MSPs to third parties.⁹ The concept of data portability is embedded in the idea that data are created by individuals and are their intellectual property. This thinking then implies that MSPs are publishers of data created by individuals and owe them payment for creating content.¹⁰ Arguably, efficient pricing for data created by individuals can be accomplished through competitive markets.

The results of this study suggest that the creation of data markets facilitating data portability and hence entry into the market for targeted advertising may result in efficient outcomes for platform members and purchasers of such advertising. Platform subscribers willing to sell their data are better off through payments for their data, while purchasers of targeted advertising enjoy the benefits of more choice and lower prices for advertising. What is critical is the type of entrant and whether they possess unique datasets of their own. Entrants who exclusively rely on data offered by platform subscribers will not be able to offer sustainable and effective competition to the incumbent. This is because they will not have the economies of scale in predictive analytics possessed by the incumbent. Coupled with additional costs from having to purchase data, the entrant experiences higher costs than the incumbent and will therefore exit the market in the long run.

On the other hand, entry by firms possessing their own unique data should be more successful. If such firms have comparable economies of scale from analyzing their own data, they will be able to offer competitive predictive analytics to the incumbent. Further, an attempt by the incumbent to indirectly raise the entrant's costs by charging platform users a subscription fee, has the potential to adversely impact it, through reduced demand for its own targeted advertising services. This is a consequence of a loss in data from reduced platform membership, which ultimately impacts the quality of predictive analytics conducted by the incumbent and hence, information used for targeted advertising. Entry may actually make the incumbent no worse off, as it could potentially recoup lost profits by charging a fee to entrants for the ability to advertise to platform subscribers. In summary, if structured properly and with entry by experienced firms, data markets could lead to a distribution of profits from targeted advertising among more firms and result in better outcomes for platform members and purchasers for targeted advertising services. The results of this research should be of interest to policymakers given the difficulties of identifying appropriate

⁸ For example, [Arrietta-Ibarra et al. \(2018\)](#) estimate that a family of four could receive as much as \$20,000 in additional annual income through information generated from different platforms, conditional on the existence of markets for data.

⁹ This assumption is certainly consistent with Article 20 of the General Data Protection Regulation (GDPR) implemented by the European Union (EU), which permits data portability across platforms, when technically feasible. The General Data Protection Regulation ("GDPR") was introduced by the European Union (EU) as a means to require businesses to protect the personal data and privacy of European Union (EU) citizens for transactions that occur within member states. It requires compliance from all firms that conduct business transactions with EU citizens, irrespective of country of residence.

¹⁰ For further discussion please refer to [Savona \(2020\)](#).

regulations to deal with the market power of Big Tech firms and the need for workable models of data governance. However, governments must enact appropriate regulation to ensure that data markets do not fail.

There is increasing evidence of individuals' willingness to sell information, which indicates that data markets should be successful. For example, apps such as Premise (www.premise.com) enable individuals to be paid conditional on the execution of different data collection tasks. Premise then aggregates this information and sells it to clients that include government agencies and non-profit organizations. Survey participants can make between US\$5-US\$40 a month, conditional on the number of tasks.¹¹ Along with the collection of specific data on different projects, individuals also share information on their personal characteristics such as gender, employment status, household income, family characteristics, and living conditions.

The remainder of the paper is structured as follows. The next section discusses the data economy, regulations imposed in some countries to protect individual privacy, and literature on data markets. Section III establishes the baseline model and associated assumptions. Section IV compares the effects of different types of entry. Section V discusses necessary conditions for the existence of data markets and the role for government regulation. Section VI concludes with a summary of the key findings.

Related literature and data regulation

The data economy and privacy regulation

The implications of data markets go far beyond than just as a mechanism to curb the market power of MSPs, as data based economic activities are becoming more prominent contributors to the GDP of several countries. A 2017 European Union (E.U.) study finds that the overall value of the data economy in EU nations in 2016 was roughly US \$300 billion.¹² A recent estimate by the China Academy of Information and Communications Technology (CAICT) suggests that the 2020-21 value of China's digital economy was even larger.¹³ The size of China's digital economy should not be surprising given that in 2017, there were 731 million internet users and 695 million mobile internet users, generating vast amounts of data on a daily basis (Aho and Duffield, 2020). Further, recent data suggest that Chinese consumers are responsible for 40 per cent of the value of all global e-commerce transactions (CNNIC, 2017). The growth of China's digital economy has been fueled by its tech giants such as Tencent and Alibaba and in part, the willingness of these firms to enter the world of financial technology through their widely used WeChat Pay and Alipay digital payment services.¹⁴ Using measures such as the volume of broadband consumed per capita, Chakravorti et al. (2019) rank the size of data economies of different countries/regions. and find the U.S., China and the E.U. as the top data producers.

The growth of data economies has implications on societal welfare beyond GDP growth, which is a result of different levels of government

intervention across countries/regions. A key issue is the loss of individual privacy. Let us take the example of two very different privacy regimes in China and the European Union. China's social credit system (SCS) minimizes the importance of individual privacy, as it relies on extensive data collected through different forms of online engagement and the world's largest camera surveillance network (170 million CCTV cameras installed by 2017, BBC, 2017). As detailed by Aho and Duffield (2020), firms are assigned scores through algorithms managed by a central government authority, which allow the government to encourage certain social and economic behaviours, while discouraging others. While such data collection by a central authority can also be used to suppress social dissent, there is evidence of high public support for the SCS and its surveillance imperatives to reduce corruption and encourage honesty and other outcomes beneficial to society (Kostka, 2018). Despite the system being used to socially engineer certain behavioural outcomes through rewards and punishments, Aho and Duffield (2020) state (on page 195):

"Rather, the system should be seen primarily as an innovative means to compensate for weak institutions and to increase statecraft efficiency. Given the SCS's ability to be monitored and adjusted in real-time thanks to digital data collection, the system will provide the Chinese state with the ability to roll out new policies and programmes at speeds unparalleled in any other country. Government agencies can thus quickly observe the effects of their policies and interventions and adjust accordingly to maximize effect, enabling a process of rapid regulatory learning."

In contrast, privacy protection in most European countries is quite different from China, as governments are not involved in comparable levels of data collection. The GDPR is intended to give individuals more control over their data and restrict the data collection abilities of firms. Personal consent and protection of individual privacy are the most important pillars of the GDPR, in contrast to societal benefit in China (as interpreted through its social credit system). Under the GDPR, individuals have rights on knowing how their data are used, view information being collected on them, be able to download the information in a machine-readable format, and not be subject to automated decision making. This precludes the large-scale use of social engineering through behavioral analytics, even in the presence of positive outcomes for society. However, this is an acceptable opportunity cost given research that firmly establishes the existence of discrimination and other unintended consequences that can be perpetuated through Machine Learning (ML) algorithms (Veale & Binns, 2017). On the other hand, Aho and Duffield (2020) point out that such regulation incentivizes the movement of AI development to the U.S., which does not have substantial federal level data privacy laws. This is unsurprising, given the lobbying power of the U.S. tech industry.

Literature on data markets

The idea of data markets is not a novel suggestion. Lanier and Weyl (2018), Arrieta-Ibarra et al. (2018), and Posner and Weyl (2018) recommend the development of markets based on easy portability of data across platforms, and the ability of individuals to sell their information directly or indirectly to different technological platforms.¹⁵ The consensus view of these studies is that such data markets have the potential to lead to a better distribution of economic surplus between large MSPs and consumers, and also to more surplus by incentivizing

¹¹ A review of this app is available at <https://www.swiftsalary.com/premise-app-review/>, last accessed August 7, 2022.

¹² Refer to 'Measuring the size and trends of the EU data economy', available at <https://digital-strategy.ec.europa.eu/en/library/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy>, last accessed August 4th, 2022.

¹³ Please see 'China's digital economy to help drive world growth' by Cheng Yu, published November 6th 2021 and available at <https://www.chinadaily.com.cn/a/202111/06/WS6185c943a310cdd39bc73b28.html>.

¹⁴ For further details refer to 'As information increasingly drives economies, China is set to overtake the US in race for data', by Saheli Roy Choudhury, published February 13th 2019 and available at <https://www.cnn.com/2019/02/14/china-will-create-more-data-than-the-us-by-2025-idx-report.html#:~:text=The%20report%20said%20in%202018,ZB%20of%20data%20last%20year>, last accessed August 4th 2022.

¹⁵ One of the first papers with a specific focus on the exchange of data between consumers and firms through online engagement, is Bergemann and Bonatti (2015) who study the sale of information obtained from cookies that can record a consumer's browsing history.

efficiencies and innovation.¹⁶ Some other papers on data portability arrive at the opposite conclusion. Acemoglu et al. (2021) and Choi et al. (2019) note that data portability may not be efficiency enhancing for all, as individuals willing to share or sell their personal information impose a negative externality on others who have similar preferences, but do not want firms to have the ability to extrapolate their preferences.¹⁷ Krämer and Stüdle (2019) and Lam and Liu (2020) also find that data portability may not be welfare enhancing for consumers, as entrants paradoxically become less competitive.

In contrast, theoretical modeling by Jones and Tonetti (2020) implies that data portability could result in improved societal welfare. Specifically, if data are nonrival, then benefits from economic activities by a large number of users might be considerable. However, firms have an incentive to hoard their data from fear of becoming obsolete, if other competitors are more productive in exploiting the data. Similarly, governments are excessively cautious on data sharing because of privacy costs to individuals and will therefore limit access to data. On the other hand, allowing consumers to choose how much data to sell results in an efficient outcome, as individuals are able to rationally weigh benefits against privacy loss.¹⁸ Further, enabling consumers to take decisions on access to their data also enables society to reach an efficient scale in the number of users. The key difference between this study and Jones and Tonetti (2020), is that they do not distinguish between different types of firms with varying levels of market power. We study the effects of data markets on individuals, the incumbent, and different types of entrants.

In summary, while some of the above research are based on models that look at the effects of entry facilitated by data portability, there is a gap on understanding the effects that different types of entry can have on the distribution and amount of total surplus generated. This study analyzes and compares the impacts of: (1) data markets with entry by third party firms reselling information generated through data obtained from platform subscribers without other independent data; and (2) a scenario where third-party firms already possess large and unique datasets of their own, that can be combined with data from platform subscribers. Traditional models of societal surplus that are commonly used by competition authorities are employed to analyze the impacts of data portability on different economic agents. Specifically, the surplus calculations are based on; changes to a platform's profitability in both the upstream market that collects information from subscribers and downstream market that sells targeted advertising to third parties; the effects on profits of third-party advertisers; and surplus of platform subscribers.

Basic model assumptions

Assume the existence of an upstream market where a single monopoly firm offers consumers free access to a digital platform in return for a variety of utility increasing services.¹⁹ The demand curve for the platform's services is downward sloping. Although the monopolist does not pay users for their data, obtaining the data is not costless.

¹⁶ Lanier and Weyl (2018), in particular, champion the idea of "mediators of individual data", which are organizations that essentially keep the data of members and bargain on their behalf with interested purchasers.

¹⁷ In contrast, Bergemann et al. (2019) demonstrate that to the extent that data externalities generated by consumers are predictive of the behaviour of others, they reduce the cost of the data intermediary in acquiring information, which can then be sold to the highest bidding firms that use the data to better tailor their products and prices. Information externalities may, therefore, be positive.

¹⁸ For further discussion on how consumers weigh privacy loss against benefits of information disclosure to sellers please see Campbell et al. (2015) and Ichihashi (2020).

¹⁹ We do not consider a case where an individual subscriber is able to choose different types of access, such as more platform services for a higher fee. The model is motivated by what is currently the case with many digital platforms.

Establishing and maintaining the infrastructure for the digital platform and its associated services entails significant initial fixed costs. As is the case with most network infrastructure, average costs decline with the number of users as there are significant economies of scale. The marginal cost of providing platform access lies below the average cost curve, with the gap between the two receding, as the number of users increase. The two curves almost converge at the point where demand is at its maximum point (where price is equal to zero). The monopolist can charge a subscription fee to platform users.

Data collected by the monopolist through its digital platform, are used as inputs to conduct predictive analytics and construct AI algorithms in the downstream market and produce information for targeted advertising. As with other research (e.g., Koutroumpis et al., 2019), we consider data as intermediate goods that are combined that are transformed by advanced data analytics to create other information goods. Targeted advertising opportunities can be sold to other firms that are interested in reaching the network of users available from the monopolist's digital platform. Therefore, the monopolist has an incumbency advantage of not only obtaining data and conducting analysis on how to approach different customer segments, but also using the same platform to deliver targeted advertisements. The demand curve for targeted advertising is downward sloping. The supply curve is the monopolist's marginal cost of predictive analytics and AI algorithms that convert the data into information, and infrastructure costs of placing advertisements on its platform. Consistent with Agarwal et al. (2018), it is assumed that the marginal cost of data to information conversion (AI and predictive analytics) is downward sloping. The marginal cost of infrastructure costs should also be downward sloping. For simplicity it is assumed that marginal costs are equal to average costs. Further, only the monopolist can place targeted advertisements on its platform.

In summary, while individuals consume the services offered by the digital platform, their consumption results in the production of data and ultimately, information that are monetized by the monopolist. Subscribers may download data they generate from online engagement through the platform but cannot access data generated by others. Individuals experience positive marginal costs in producing data as they are giving up time that could be spent on other activities. From the monopolist's perspective, data generated through its platform possesses both public and private good characteristics. The marginal cost of sharing data with other firms is probably close to zero. On the other hand, data sharing reduces the monopolist's profits, as other firms would be able to replicate its predictive analytics and possibly reduce profits earned in the downstream advertising market.

Equilibrium in data markets

Baseline model without entry & regulation

The monopolist cannot choose subscription prices for its platform without taking into account the impacts on the downstream advertising market. Accessing more data through a lower platform subscription price results in richer data analytics and more precise information for targeted advertising services. In other words, $\partial Q^a / \partial P^d < 0$ where Q^a is the demand for targeted advertising and P^d is the subscription fee charged to digital platform users. A higher platform subscription fee results in reduced data for the monopolist and predictions with lower accuracy from AI models. This leads to less useful information, and therefore a decrease in demand for the monopolist's targeted advertising. As assumed earlier, there are economies of scale in marginal cost of data to information conversion, which is denoted by $\partial c(Q^a) / \partial Q^a < 0$. To reduce notation, we assume that this also captures the monopolist's marginal cost of advertisements on its platform.

Hence, the monopolist maximizes a joint profit function consisting of profits from both upstream and downstream markets that takes into account the effects that platform prices have on advertising profits. The externality that pricing in the data market affects sales in targeted

advertising can be seen through the following functions;

$$\pi^d = p^d q^d(p^d) - c^d(q^d(p^d)) \quad (1)$$

$$\pi^a = p^a Q^a(p^a, p^d) - c^a(Q^a(p^a, p^d)) \quad (2)$$

q^d is the number of users on the monopolist's digital platform, which represents the amount of data produced. Each user produces a fixed amount of data. Therefore, the monopolist benefits from having more users but with diminishing returns ($\partial^2 Q^a / \partial q^{d2} < 0$). p^d is the digital platform subscription price paid by a user. Suppose the monopolist maximizes (1) with respect to p^d and (2) with respect to p^a . The First Order Conditions for (1) and (2) are

$$p^d \frac{\partial q^d}{\partial p^d} + q^d = \frac{\partial c^d}{\partial q^d} \frac{\partial q^d}{\partial p^d} \quad (3)$$

$$p^a \frac{\partial Q^a}{\partial p^a} + Q^a = \frac{\partial c^a}{\partial Q^a} \frac{\partial Q^a}{\partial p^a} \quad (4)$$

Maximizing these profit functions with respect to prices in both markets leads to the traditional finding that a monopolist's choice of prices in both markets is determined at the point where marginal revenues are equal to marginal costs of production in each market. However, the price in platform access is inefficiently high as it does not take into account the effects of platform prices on advertising sales. In comparison the First Order Condition (FOC) if the monopolist jointly maximizes (1) and (2) with respect to p^d is;

$$q^d + p^d \frac{\partial q^d}{\partial p^d} + p^a \frac{\partial Q^a}{\partial p^d} = \frac{\partial c^d}{\partial q^d} \frac{\partial q^d}{\partial p^d} + \frac{\partial c^a}{\partial Q^a} \frac{\partial Q^a}{\partial p^d} \quad (5)$$

In other words, overall profit maximization requires the monopolist to consider the effects of prices for platform subscription on targeted advertising sales. $q^d + p^d \frac{\partial q^d}{\partial p^d}$ is the monopoly marginal revenue from the platform market while $p^a \frac{\partial Q^a}{\partial p^d}$ represents the loss in targeted advertising sales that would occur from an increase in platform access price p^d that would result less data and hence, reduced quality in information. The above equation can be further manipulated to obtain a clearer idea of the trade-off between p^a and p^d . Dividing (5) by q^d and with some algebraic manipulation we obtain;

$$1 - \epsilon^d + \frac{p^a}{q^d} \frac{\partial Q^a}{\partial p^d} = \frac{MC^d}{q^d} \frac{\partial q^d}{\partial p^d} + \frac{MC^a}{q^d} \frac{\partial Q^a}{\partial p^d} \quad (6)$$

where $\epsilon^d = -\frac{\partial q^d}{\partial p^d} \frac{p^d}{q^d}$, $MC^d = \frac{\partial c^d}{\partial q^d}$, $MC^a = \frac{\partial c^a}{\partial Q^a}$. Manipulating this further,

$$1 - \epsilon^d + \epsilon^d \frac{MC^d}{p^d} = -p^a \frac{R^a}{R^d} \frac{\partial q^d}{\partial p^d} \frac{p^d}{Q^a} + \frac{R^a}{R^d} MC^a \frac{\partial q^d}{\partial p^d} \frac{p^d}{Q^a} \quad (7)$$

and

$$1 - \epsilon^d \left(\frac{p^d - MC^d}{p^d} \right) = \epsilon^d \frac{R^a}{R^d} \left(\frac{p^a - MC^a}{p^a} \right) \quad (8)$$

Where $\epsilon_d^a = -\frac{\partial q^a}{\partial p^d} \frac{p^d}{q^a}$, $R^a = p^a Q^a$, $R^d = p^d q^d$

Eq. (8) yields the trade-off that the monopolist faces in increasing platform subscription price p^d . A higher margin in the platform market (weighted by the price elasticity of demand of platform prices) is offset by lower margin in targeted advertising, weighted by the elasticity of targeted advertising sales in response to platform prices and targeted advertising revenues as a proportion of platform revenues.

Examining the monopolist's incentives through conventional demand and supply diagrams adds further clarity especially with respect to impacts on consumer surplus. Consistent with what it is observed in many digital platforms, it is assumed that joint profit maximization entails the setting of an equilibrium price of zero for access to the digital platform. Fig. 1 gives the loss (rectangle ABCO) that the monopolist

incurs at a zero price and maximum demand for platform services given by point C on the horizontal axis.²⁰ In contrast, $P_1^d E_1 DF$ gives the profit that would have been earned at the monopoly price P_1^d and q_1^d .

Fig. 2 captures the dynamics in the targeted advertising market. If the monopolist charges the separate profit maximizing price for each market - P_1^d and P_1^a - then profits in targeted advertising are $P_1^a E_1 GH$ at q_1^d . However, with lower prices in the digital platform the demand for targeted advertising shifts from D_1 to D_2 , as purchasers are willing to buy more with the improvement in predictions stemming from the availability of more data. Prices for targeted advertising increase to P_2^a and profits from advertising are at $P_2^a E_2 IJ$. If $P_2^a E_2 IJ - ABCO > P_1^a E_1 GH + P_1^d E_1 DF$, then making losses on the digital platform ultimately leads to higher overall profits as opposed to charging the profit maximizing price for each individual market.

In terms of consumer surplus, platform subscribers are unambiguously better off with zero prices. It is also possible that free access to the platform results in higher consumer surplus in the downstream targeted advertising market because of the improvement in predictive analytics, which leads to higher marginal valuation and more purchases. While this is not necessary and is conditional on the improvement facilitated by the collection of more data, it does suggest that the monopolist's strategy of allowing free access to its digital platform may be efficiency-improving for itself as well as consumers in both upstream and downstream markets, relative to the charging of monopoly prices in both markets.

It is also important to note that the above analysis does not incorporate the positive spillover that targeted advertising generates for consumers, independent of total surplus in the downstream market. Specifically, targeted advertising has the potential to improve awareness of the availability of different consumption options, reduce search costs for consumers, and lead to purchases of goods and services at lower prices than what would have happened in the absence of the monopolist's ability to analyze platform generated data and conduct predictive analytics.²¹ However, these are the general effects of any type of advertising that is factual, and it is important to assess whether welfare impacts to parties are robust to the exclusion of such impacts.

Entry and a market for data

From a social welfare perspective, competition is desirable. However, it would be extremely difficult for another firm to enter the market for targeted advertising. Successful entry would incur only if: (1) the entrant possesses large enough datasets that would result in comparable predictive analytics to the monopolist's, leading to accurate targeted advertising; and (2) also offer other businesses a means to connect with potential customers, as enabled by the social media firm through its platform. Hence, the existence of data portability is assumed, which enables consumers to transfer the information they generate through the incumbent's platform to the entrant. The entrant can then compete with the incumbent in the upstream targeted advertising market without having to establish a separate digital platform.

This is a plausible assumption given that it is possible for users to port the data they generate from online activity (such as Facebook, Whatsapp, Instagram, Google) in the form of csv or comparable data files. Further, as noted above, data portability is an established right in Europe, mandated through the General Data Protection Regulation

²⁰ For simplicity, marginal revenue curves are omitted in monopoly diagrams.

²¹ It can be argued that if targeted advertising results in consumers switching purchases from a high-cost firm to a low-cost alternative, then the high-cost firm is worse off by the amount of lost sales. However, this is a transfer of sales to another firm and not a loss in economic activity.

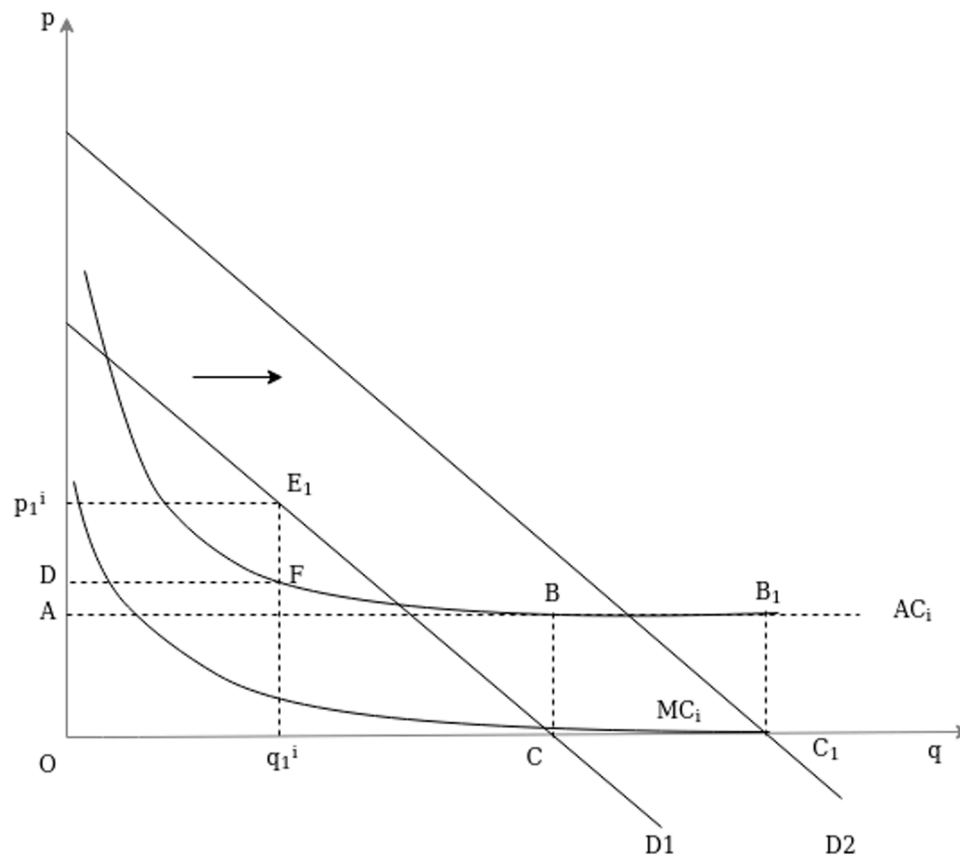


Fig. 1. Upstream digital platform.

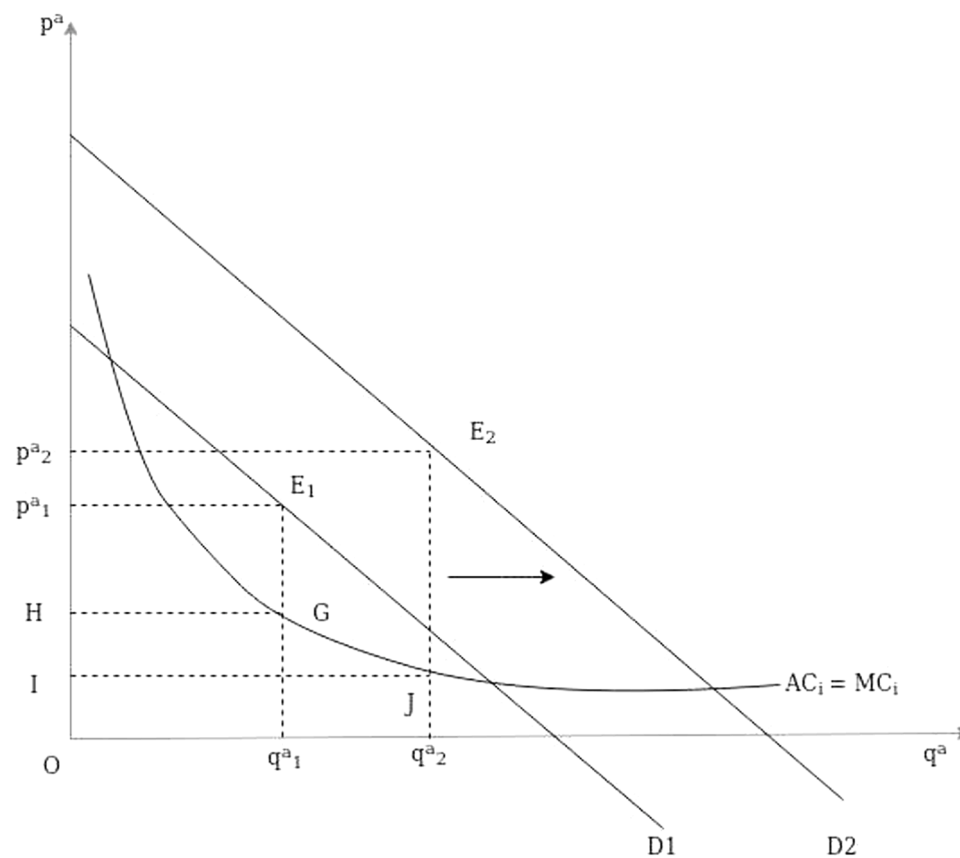


Fig. 2. Downstream targeted advertising.

(GDPR) and has also been recently enacted in California.²² However, an entrant would likely have to pay individuals for access to their platform data, as there is no other incentive for individuals to share their online activity with a third party. This is, of course, possible through the establishment of a data market.²³ The question is, whether entry on the basis of payment for platform data, would lead to improved societal welfare.

For the moment, it is assumed that establishing a data market is possible and is mandated and protected by the same laws and regulations governing most markets. Government regulations guarantee data portability to individuals and allows them to sell information they generate on the monopolist's platform to third parties. Further, the entrant can costlessly receive data from individuals through an app or data portal. Also, if given permission by the individual subscriber, the entrant can place advertisements on the subscriber's platform space.²⁴ The entrant is a price taker and pays platform subscribers a price p_e^d for their data, where $p_e^d = p_e^d(p^d)$ and $\frac{\partial p_e^d}{\partial p^d} > 0$. A simple pricing scheme is where the entrant pays a fixed fee above any platform price charged to it by the monopolist, or $p_e^d = f + p^d$, where f is the fixed fee.²⁵ Hence, if $f = 0$ the minimum price paid by an entrant to an individual is the platform subscription price. The incumbent sees the entrant's price for targeted advertising and decides its best response through its own prices for platform access and targeted advertising.

Profit maximizing decisions by the incumbent and entrant take place as follows. In the first stage the incumbent simultaneously chooses optimal prices for its platform and advertising services. The entrant chooses its optimal price for targeted advertising after the incumbent's decisions. Hence, the entrant maximizes the following profit maximizing function with respect to p_e^a ;

$$\pi^e = p_e^a q_e^a(p_e^a, p^a, p_e^d(p^d)) - g(q_e^a(p_e^a, p^a, p_e^d(p^d))) - r(q_e^a(p_e^a, p^a, p_e^d(p^d))) \quad (9)$$

Where p_e^a is the price charged by the entrant for its targeted advertising services, q_e^a is the entrant's sales, p^a is the incumbent's price for targeted advertising, and p_e^d is the price paid by the entrant to platform subscribers for data. $\frac{\partial q_e^a}{\partial p_e^a} < 0$, $\frac{\partial q_e^a}{\partial p^a} > 0$, and $\frac{\partial g}{\partial p_e^d} > 0$. As is the case with the incumbent, $\frac{\partial q_e^a}{\partial p_e^d} < 0$. In other words, as the entrant pays a higher price to subscribers, it obtains less data leading to lower quality predictive analytics and lower targeted advertising sales. The entrant's cost function is in two parts. $g(\cdot)$ captures costs associated with the production of predictive analytics, which has economies of scale. On the other hand, a significant portion of entrant costs are also driven by data purchases. $r(\cdot)$ represents total costs of data purchases, which is increasing and convex with respect to quantity of targeting advertising services produced. Hence, $\frac{\partial g}{\partial p_e^d} < 0$, $\frac{\partial^2 g}{\partial p_e^d^2} > 0$ and $\frac{\partial r}{\partial p_e^d} > 0$, $\frac{\partial^2 r}{\partial p_e^d^2} > 0$. Maximizing this profit function with respect to p_e^a yields the following FOCs;

$$q_e^a + p_e^a \frac{\partial q_e^a}{\partial p_e^a} = \frac{\partial g}{\partial q_e^a} \frac{\partial q_e^a}{\partial p_e^a} + \frac{\partial r}{\partial q_e^a} \frac{\partial q_e^a}{\partial p_e^a} \quad (10)$$

(10) is the standard profit maximizing condition marginal revenue equal

to marginal cost condition with respect to an oligopoly. Part of the marginal cost is driven by the costs of predictive analytics converting data to relevant information for advertising ($\frac{\partial g}{\partial q_e^a} \frac{\partial q_e^a}{\partial p_e^a}$), while the other portion is from expenses associated with data purchases ($\frac{\partial r}{\partial q_e^a} \frac{\partial q_e^a}{\partial p_e^a}$). (10) can be further manipulated to obtain

$$p_e^a = f(p^a, p_e^d(p^d)) \quad (11)$$

A functional form of the relationship between p_e^a and p^a can be derived from (11). The sign of this relationship depends on the relative magnitudes of $g(\cdot)$ and $r(\cdot)$. A dominant $g(\cdot)$ implies $\frac{\partial p_e^a}{\partial p^a} < 0$, while a dominant $r(\cdot)$ yields $\frac{\partial p_e^a}{\partial p^a} > 0$.

The profit maximizing choice of upstream and downstream prices by the incumbent can be obtained by maximizing the following function;

$$\pi^i = p^d q^d(p^d) - c^d(q^d(p^d)) + p^a q^a(p^a, p_e^a(p^a, p_e^d(p^d)), p^d) - c^a(q^a(p^a, p_e^a(p^a, p_e^d(p^d)), p^d)) \quad (12)$$

The FOC with respect to p^a and p^d are;

$$p^a \left[\frac{\partial q^a}{\partial p^a} + \frac{\partial q^a}{\partial p_e^a} \frac{\partial p_e^a}{\partial p^a} \right] + q^a = \frac{\partial c}{\partial q^a} \frac{\partial q^a}{\partial p^a} + \frac{\partial c}{\partial q^a} \frac{\partial q^a}{\partial p_e^a} \frac{\partial p_e^a}{\partial p^a} \quad (13)$$

$$p^d \frac{\partial q^d}{\partial p^d} + q^d + p^a \frac{\partial q^a}{\partial p^d} + p^a \frac{\partial q^a}{\partial p_e^a} \frac{\partial p_e^a}{\partial p^d} \frac{\partial p_e^d}{\partial p^d} = \frac{\partial c^d}{\partial q^d} \frac{\partial q^d}{\partial p^d} + \frac{\partial c^a}{\partial q^a} \frac{\partial q^a}{\partial p^d} + \frac{\partial c^a}{\partial q^a} \frac{\partial q^a}{\partial p_e^a} \frac{\partial p_e^a}{\partial p^d} \frac{\partial p_e^d}{\partial p^d} \quad (14)$$

Total industry output is $Q^a = q^a + q_e^a$. Eq. (13) is the standard marginal revenue equal to marginal cost relationship where a firm has to consider not only the effects of changes in its own price on sales ($\frac{\partial q^a}{\partial p^a}$) but also the corresponding impacts on its competitor's price ($\frac{\partial q_e^a}{\partial p^a} \frac{\partial p_e^a}{\partial p^a}$). This effect is conditional on the sign of $\frac{\partial p_e^a}{\partial p^a}$. As noted above, with strong economies of scale $\frac{\partial p_e^a}{\partial p^a} < 0$, which would imply that a higher price for targeted advertising sales charged by the incumbent leads to a rightward shift of the demand curve for the entrant's product and a lower p_e^a and reduces targeted advertising sales for the incumbent. The opposite would be true if the entrant's costs are dominated by data purchase expenses.

Eq. (14) gives the equilibrium relationship that determines how the incumbent sets p^d . As assumed, lowering p^d on platform membership results in more data and consequently, an increase in the incumbent's targeted advertising sales. However, with competition, a lower p^d reduces the entrant's data purchase price ($\frac{\partial p_e^d}{\partial p^d} > 0$) as well, which leads to a lower price charged by the entrant for targeted advertising ($\frac{\partial p_e^a}{\partial p_e^d} > 0$) and therefore, diminished targeted advertising sales for the incumbent ($\frac{\partial q^a}{\partial p_e^d} > 0$). Hence, while a lower platform price benefits the incumbent it also implies a stronger competitor.

These relationships are further explored through the following figures. First, consider the upstream platform market. The presence of an entrant willing to pay for an individual's data will shift the demand curve for the monopolist's platform services to the right from D_1 to D_2 , as more people seek to derive monetary benefit from the generation of online content. This increase in demand will result in further financial losses for the monopolist, given by the area BB_1CC_1 in Fig. 1. However, from a total surplus perspective, these losses are probably offset by the increase in consumer surplus, which should be quantified by payments received from the entrant.

The entrant's sustainability is a function of the strength of its predictive analytics. For the entrant to be successful, purchasers of its targeted advertising services must find at the very least, similar increases in marginal revenues relative to what it would have obtained from purchasing the same services from the incumbent. Further, the entrant must be able to make profits after payments it needs to make to platform

²² <https://www.nytimes.com/2020/01/03/us/ccpa-california-privacy-law.html>

²³ It should be noted that an extremely strong and large market does exist where data brokers facilitate the sale of information culled from online activity by individuals. What does not exist are markets where individuals can directly sell their information to interested parties.

²⁴ This is similar to the concept of smaller mobility providers being able to access the network infrastructure of larger telecommunications firms.

²⁵ Of course, other schemes are possible, such as ones where individuals are compensated based on the amount of information they produce.

subscribers for their data. Assume that even with a limited dataset, the entrant is able to produce comparable targeted advertising services relative to the incumbent, as there are significant correlations in personal characteristics and behaviour across individuals. Hence, a complete dataset consisting of all platform users is not required.

If the services of the incumbent and entrant are relatively homogeneous, then the market for targeted advertising will be split between the two firms. Assuming symmetric marginal costs between firms, aggregate duopoly profits are lower in comparison to the monopoly case.²⁶ However, it is unlikely that marginal and average costs will be symmetric between firms. Specifically, the incumbent's experience in predictive analytics and the fact that it is obtaining data for free from its platform implies that the entrant's average and marginal cost curves will be higher than the incumbent's, which reflect the costs of predictive analytics and data purchases. Further, marginal costs from data purchases implies that the decline in the incumbent's overall average and marginal costs driven by economies of scale from experience in predictive analytics, will stop at a certain point beyond which the incumbent will continue to reap the benefits of declining costs.

These features are captured in Fig. 3, where the average cost curves of the incumbent and entrant are given by AC_i and AC_e . The figure also depicts the residual demand curve of the incumbent, which shifts inwards from D_1 to D_2 , because of entry. The new equilibrium price charged by the incumbent is P_2^a . Given the economies of scale in terms of predictive analytics, the decline in demand for the incumbent's targeted advertising services will unambiguously result in lower profits because of the drop in price and quantity of sales, along with the fact that the incumbent will now be operating at a higher point on its average and marginal cost curves. For the reasons outlined above, the entrant will be making even lower profits, as its marginal and average costs will be higher than the incumbent's. This is because the incumbent has to pay subscribers for their data and will also not have the same economies of scale in predictive analytics enjoyed by the incumbent. In the short run, purchases of targeted advertising services are better off from the reduction in prices.

What is important to note is that this equilibrium is not sustainable. As is evident from Fig. 3, the incumbent could capture the market for targeted advertising through a modest reduction in price that the entrant will not be able to match if the differences between the entrant's and its cost structures are significant. Further, it is possible that the incumbent monopolist might charge a fee for platform access to subscribers, to offset lost profits in the downstream targeted advertising market. The result would be further deadweight loss with some subscribers being priced out of the market. With the assumption of correlation in personal characteristics across individuals, the incumbent may not experience a significant loss in predictive analytics. On the other hand, this would further push up the entrant's marginal costs of business as it would do the same for platform subscribers. Specifically, having a fee for platform access then becomes the marginal cost to the platform subscriber, in addition to the opportunity costs of time in producing information through online engagement. The entrant would need to pay a higher fee to individual subscribers for their data, increasing the likelihood of its exit. However, raising the entrant's costs through a platform subscription fee does have the possibility of hurting the incumbent's predictive analytics through loss in data.

Hence, entry with the incumbent's data - purchased through subscribers - will result in short term gains only to platform subscribers, through payments for information. Purchasers of targeted advertising are also better off in the short run as prices are lower because of entry. However, total industry profits are correspondingly reduced. Further, benefits to platform subscribers will be offset if the incumbent decides to

charge a platform subscription fee, which would have the effect of increasing costs to the entrant and therefore the potential for exit. A possible means to prevent this is through regulation capping the platform subscription fee at zero. On the other hand, this is inconsequential if the incumbent decides to reduce prices for targeted advertising in order to retain its monopoly status. A final point is whether allowing the incumbent to charge a fee to the entrant for posting advertisements would make a difference to total surplus. However, this is a less attractive strategy for the incumbent, than reverting to a monopoly position as the entrant is only cannibalizing the sales of the incumbent, and not increasing market size through new customer segments. As the entrant has higher marginal and average costs, the incumbent would be making less profit relative to being a monopolist, even if it could charge the entrant a fee or two-part tariff through which it could extract most of the incumbent's profits.

Entry by a differentiated firm

Now consider the case where a firm enters with data that are not based exclusively on information from the incumbent's platform. The entrant possesses data, which contains extensive details on individual preferences and choices related to goods and services that are not related to the information possessed by the incumbent. In this scenario, it is possible that predictions generated by the entrant's data along with any other information that can be purchased by it from subscribers, are superior for certain sectors relative to predictions produced by the incumbent. For example, credit companies and banks possess significant amounts of individual level data, and it is possible that they would be able to produce better targeted advertising for financial related products. It is also assumed that the incumbent is unable to purchase data from the entrant and match it to corresponding individual level information that it possesses. In other words, the entrant is participating in the data market as a purchaser but not as a seller. As previously assumed, the entrant can implement targeted advertising through the incumbent's platform. The analysis in this section will be restricted to basic graphs in order to ensure clarity for policy makers and also because the implicit modelling is quite similar to the previous section. In terms of assumptions, the entrant's marginal and average costs of prediction analytics are similar to those of the incumbent's. However, the entrant does incur additional costs, which are a function of the amount of data that it purchases from platform subscribers.

Entry implies that purchasers of targeted advertising are better off through lower prices. This is a consequence of the demand curve for the incumbent shifting inwards to the left, as some of its customers gravitate towards the entrant, which offers predictions that are more suitable to its customer base. From an industry perspective, the loss in incumbent revenue will be offset by the entrant's revenue. It is possible that the increase in entrant's revenue is larger than the loss in incumbent revenue, if the entrant's targeted advertising analytics are superior to the incumbent's for some new customer segments. For this to happen the aggregate market demand curve for targeted advertising services must necessarily increase. Arguably, this is possible if there is a higher willingness to pay by customers for the targeted advertising generated by the entrant. From a total surplus perspective, the question is whether industry profits will also increase. If the marginal costs of the entrant are as low as the incumbent's then the net results of entry may be potentially distributional as the entrant gains profits previously earned by the incumbent, and as well, the creation of new surplus. This would occur if the incumbent charges a fee to the entrant for advertising space on its platform.

These dynamics are illustrated through Fig. 4. The incumbent's monopoly profits are given by $P_1^a E_1 GF$ while post-entry profits are $P_2^a E_2 HI$. Suppose that the incumbent is able to charge a fee, which is equivalent to a percentage (λ) of the entrant's profits. If $P_1^a E_1 GF - P_2^a E_2 HI = \lambda$, then the incumbent is no worse off with entry, as any loss in

²⁶ In the monopoly case a firm only has to take into account the effects of its pricing on overall demand. In a duopoly a firm also needs to consider the effects of pricing decisions on its rival's output.

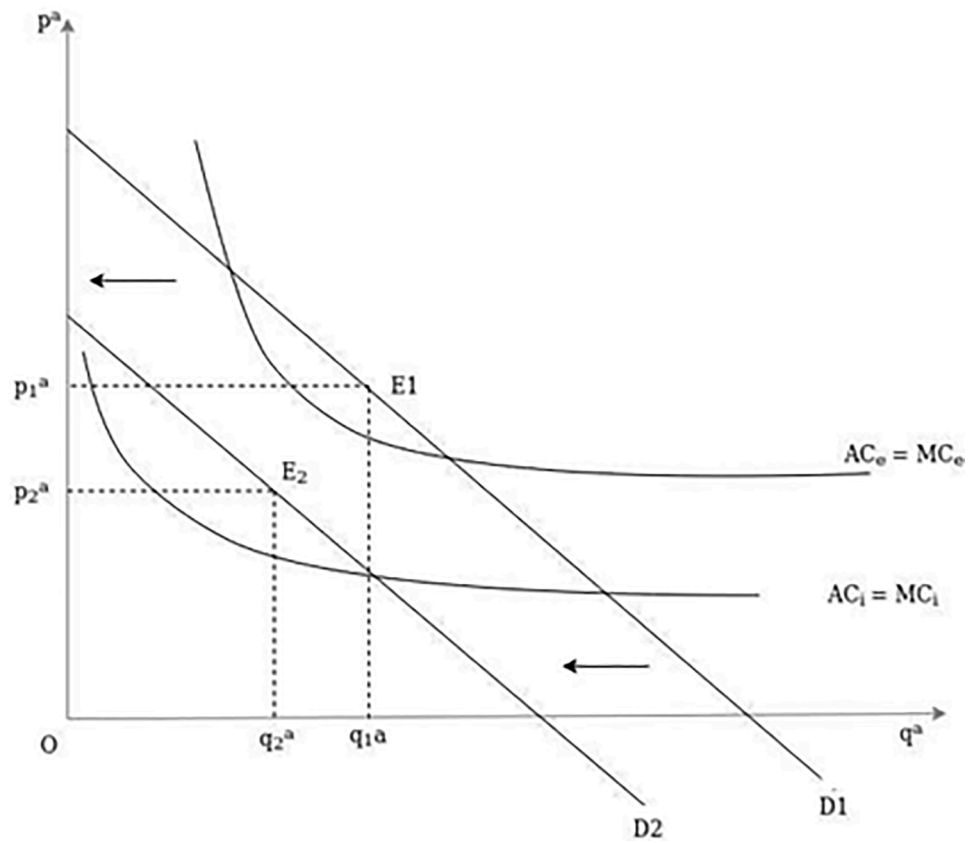


Fig. 3. Effects of entry without incumbent's data.

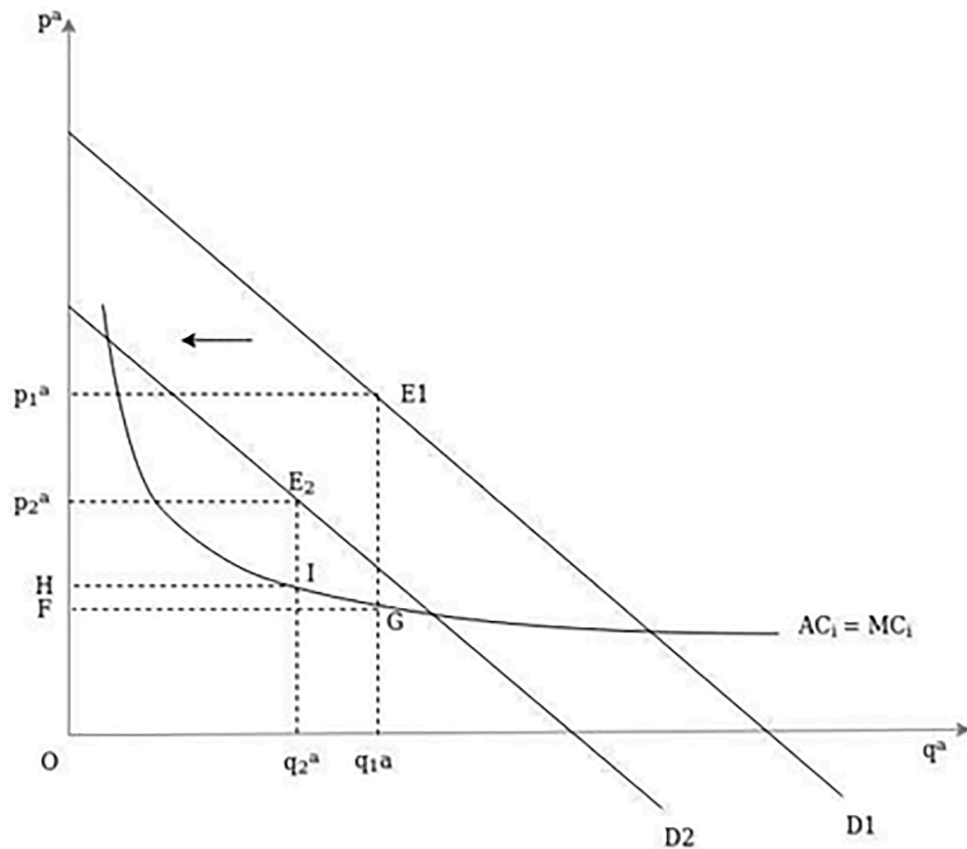


Fig. 4. Effects of entry with proprietary data.

targeted advertising profits is offset through revenue earned by the incumbent by allowing the entrant to place advertising directed to platform subscribers. If these revenues are substantial, the incumbent will have no incentive to recuperate lost monopoly profits by charging platform fees to subscribers, in an effort to increase costs to the entrant. By doing that, the incumbent will be reducing the amount of data collected for itself and the entrant, assuming that a platform membership fee will reduce the number of subscribers. Less data implies lower demand for its targeted advertising services. Further, since the entrant does not exclusively rely on the incumbent's data, a platform subscription fee has the potential to negatively impact the incumbent more than the entrant.

Role of government

This above analysis suggests a pathway to curbing the economic power of MSPs and ensuring competition. Successful entry could occur if a firm has large datasets of its own that could be used in tandem with data purchased from platform subscribers. MSPs do not possess data on every choice or purchase by individuals. The best example of this is Google's purchase of Fitbit, which industry watchers acknowledge, allows Google to access health data that could be used for its advertising services.²⁷ This section identifies some key conditions that must be met for successful data markets from the perspectives of purchasers and sellers, as well as the role of governments in implementing appropriate regulations.²⁸

The first condition is that a significant number of individuals must be willing to supply their data, and this will occur if they feel that they are adequately compensated for the value of their information as well as the time it might take to collect and aggregate their data from digital platforms. Government regulation has a role here in requiring MSPs to enable individuals to download their information in machine readable files, facilitating data portability. The GDPR does require this, and such features are available to Facebook and Google subscribers. However, in most cases, data availability to subscribers are only possible as 'data dumps' through machine readable formats. MSPs should be required to build and offer technological features, which enable individuals to easily sift through their collected data and choose the information they would like to specifically download. Such regulation would significantly reduce opportunity costs associated with participation in data markets.

Second, if individuals are the ultimate producers of data, they must be assured that participating in data markets does not result in their privacy being compromised. Studies in computer science have demonstrated that it is possible to identify individuals from what might be considered to be anonymous data (Ohm, 2010). The supply of data might further decrease if individuals feel that their data are being used for purposes, which meet their approval. Data purchasers must be able to provide evidence of their ability to securely store data storage and preserve individual privacy. Governments have a role in identifying and implementing mandatory standards that all firms must adhere to, with respecting to data protection.

Third, firms must be able to provide secure application program interfaces (APIs) that enable consumers to share their data with purchasers. This is another area, which requires the identification of minimum standards by governments. There are steps being taken in this direction through the development of Open Banking that is occurring in Australia, Canada, the United Kingdom, the European Union and Japan. In its consultation document, Canada's Department of Finance defined open banking as "a framework where consumers and businesses can authorize third party financial service providers to access their financial

transaction data, using secure online channels". Open Banking enables consumers to port and share their financial data across different financial institutions through secure application program interfaces (APIs). This is the foundation of how data markets may work in practice. Given this, it is certainly plausible for consumers to share other non-financial information with banks, if such data can be combined with historical financial transactions to offer better financial management services. It is possible to think of firms requiring 'data licenses' from governments to be able to purchase information from individuals. Such a license would be granted if a firm can offer evidence of compliance with minimum standards of data storage, steps for ensuring individual privacy, and the use of secure APIs to obtain data directly from individuals. Penalties should be imposed on firms that do not take acceptable measures to protect the data of individuals.

Fourth, firms must feel that they have a reasonable likelihood of obtaining high quality data. Koutroumpis et al. (2020) suggest that conventional market mechanisms can only facilitate large markets with little control or small markets with greater (but not full) control. While there are multiple reasons behind their conclusion, an important point is the existence of informational asymmetries, if data purchasers have limited information on data quality and their ability to produce valuable insights. The potential for data to produce useful information can only be verified through a study of their statistical properties and performance in statistical modelling. This informational asymmetry between buyers and sellers will result in limited demand from adverse selection, and therefore, 'thin' markets that are unsustainable.

Independent of actual statistical performance, buyers need to be assured that the data they are purchasing are from desired sources, and these sources must be verifiable. Without this property, buyers will naturally be hesitant in purchasing data, irrespective of how well the data might perform in generating insights. Of course, the inability of potential buyers to ascertain data sources and collection method are not the only reason for possible market failure. Further, governments should also make concerted efforts to stop illegal data collection and sales by private data brokers who scrape such information without proper permission or authorization.²⁹

Firms participating in data markets may be interested in aggregating data and selling them to other participants in the market for data analytics. In this respect, firms must be reassured that their products will not be appropriated by third parties who repackaging and sell their data. Koutroumpis et al. (2020) note that intellectual property protection for data is quite weak in most jurisdictions. Specifically, while copyright laws might be considered the best means for protecting sellers against appropriation of data products, such regulation usually only protects the structure and organization of the data conditional on the existence of an original contribution through its construction. On the other hand, copyright regulation does not protect the individual observations. Strong governance, embodied by targeted legislation with significant penalties are required to address concerns by buyers and sellers. Intellectual property rights should be protected through significant penalties on firms that appropriate data collected by other competitors and attempt to sell the data as part of their own offerings.

Koutroumpis et al. (2020) point to novel technological solutions such as distributed ledger technologies (DLTs) (Evans, 2014; Catalini and Gans, 2016) that can address some of the above concerns, as they automatically track transactions from source to sale. This technology is similar to blockchains that are used in cryptocurrency transactions. Recording through DLTs can address concerns regarding data source and also prevent unauthorized appropriation. A DLT should also have the ability to confirm the use of an individual's data, which would facilitate payments to subscribers based on the number of times their data are

²⁷ For further details please see Austin (2019).

²⁸ Some of the points in this section are based on recommendations from Koutroumpis et al. (2020) and general principles of market design from Roth (2008).

²⁹ For further details on illegal activities by data brokers and their lack of transparency please refer to Office of the Privacy Commissioner of Canada (2014).

used for targeted advertising. There is also a need for government regulation in defining steps that should be taken by data aggregators to ensure that individuals cannot be 'deanonymized'.

Conclusion

The vast amount of data and economic power accumulated by MSPs have attracted policy and public concern. Economists have proposed the creation of data markets as a possible solution. This paper contributes to the literature by offering some specific guidance on how such markets should be structured. European Union nations have moved forward and implemented regulation, which guarantees data portability as a right. Data portability is the foundation through which individuals can sell information through regulated markets. This study relies on simple models to determine some necessary features for the successful functioning of such markets. The key to the economic power of most successful MSPs is their ability to collect data from free online platforms and use this information to construct predictive analytics aimed at targeted advertising. Hence, a data market, which enables other firms to directly purchase information from individuals to use as inputs for predictive analytics, should encourage competition to MSPs and ensure a more equitable distribution of societal surplus.

This study assesses the societal impacts of data markets on consumers and producers in both upstream data and downstream targeted advertising markets. The results demonstrate that the type of entrant is critical in ensuring competitive markets. Entrants who can only offer predictive analytics through data purchased from platform subscribers, will not be able to offer sustainable competition. Relative to the incumbent, they will not possess comparable economies of scale in predictive analytics and have higher marginal and average costs in terms of payments to individuals for their data. This is consistent with the arguments made by Arrieta-Ibarra et al. (2018). On the other hand, large firms with their own data – that cannot be accessed by incumbent MSPs – possess the potential to be viable competitors. This would occur if such firms: (1) have similar economies of scale to MSPs; and (2) can conduct comparable predictive analytics based on either their own data or a combination of it, with information purchased from subscribers. Incumbent MSPs will then have no incentive to charge platform subscription fees as that would result in reduced data for themselves and lower quality predictive analytics. This market structure would result in lower prices to targeted advertisers and a redistribution of existing profits away from the incumbent to the entrant. However, in the long run, the incumbent may only be marginally worse off, if it charges a fee to the entrant for advertisements on its platform. Hence, competition will result in redistribution as well as creating new societal surplus.

Governments have a strong role to play in regulating data markets through the setting of appropriate standards for data secure storage by firms and protecting the privacy of individuals and ensuring that they can easily download their data from digital platforms and share information through secure APIs. Of course, the most important policy innovation should be to ensure that individuals clearly have property rights to the data they generate through online engagement with MSPs. Effective intellectual property protection must be implemented to protect firms that sell data purchased from individuals and to prevent data appropriation.

There are important societal consequences from data markets that should be noted. Savona (2020) points out that individuals willing to sell their data might primarily be low-income individuals. Their data would be in less demand for targeted advertising purposes, which would not result in sustainable data markets. A related concern is that data markets will be regressive as only individuals with middle or high incomes would be able to sell their data. While this is a possibility, data market transactions can be taxed, resulting in government revenue that can be used to implement either cash or in-kind transfers to low-income households. Further, the existence of apps such as Premise suggest that the low-income individuals would benefit from formal data markets that

could generate part-time employment opportunities that are no different from the 'gig' activities such as being an Instacart shopper or Uber/Uber Eats driver.

The existence of data markets can benefit government and public policy in other ways. There is no reason why the government cannot become a purchaser of information that it would otherwise be unable to collect through voluntary surveys. These data could be released to the academic community for research purposes. This potential has emerged clearly during the current pandemic. For example, there has been a clear deficit in locational data that can be used to map out movements of individuals and therefore analyze factors behind the spread of COVID-19. On the other hand, the collection and use of such data should be balanced by the need for individual privacy and security. However, the establishment of data markets could accelerate the pace of debate and ensure that relevant concerns are resolved before the occurrence of similar future crises, and the existence of market-based mechanisms that could help policymakers in obtaining real time information that have the potential to make critical impacts (Eqs. (3), (4), (6), (9), (12)–(14)).

Funding

No specific funding was received for this study.

Declarations of Competing Interest

The authors declare that they have no conflict of interest.

References

- Acemoglu, D., Makhdomi, A., Malekian, A., Ozdaglar, A., 2021. Too much data: prices and inefficiencies in data markets. *Forthcoming Am. Econ. J. Microecon.*
- Agarwal, A., Gans, J., Goldfarb, A., 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.
- Aho, B.A., Duffield, R., 2020. Beyond surveillance capitalism: privacy, regulation and big data in Europe and China. *Econ. Soc.* 49 (2), 187–212. <https://doi.org/10.1080/03085147.2019.1690275>.
- Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., Glen Weyl, E., 2018. Should we treat data as labor? Moving beyond "Free". In: *Proceedings of the AEA Papers*, 108, pp. 38–42.
- Austin, P.L., (2019). "The real reason Google is buying Fitbit". Available at <https://time.com/5717726/google-fitbit/>.
- Bergemann, D., Bonatti, A., 2015. Selling cookies. *Am. Econ. J. Microecon.* 7 (3), 259–294.
- BBC News. (Producer), 2017. "China: the world's biggest camera surveillance network. BBC News Video available at <https://www.youtube.com/watch?v=pNf4-d6fDoY>.
- Campbell, J., Goldfarb, A., Tucker, C., 2015. Privacy regulation and market structure. *J. Econ. Manag. Strat.* 24 (1), 47–73.
- Bergemann, D., A. Bonatti, and T. Gan, 2019. "The economics of social data". Working Paper. URL: <http://www.mit.edu/~bonatti/social.pdf>.
- Catalini, C., Gans, J.S., 2016. Some simple economics of the blockchain. In: *NBER Working Paper Series*. National Bureau of Economic Research, Cambridge, MA, p. 30.
- Chakravorti, B., Bhalla, A., Chaturvedi, R.S., 2019. Which countries are leading the data economy? *Harv. Bus. Rev.* Available at <https://hbr.org/2019/01/which-countries-are-leading-the-data-economy>.
- Choi, J.P., Jeon, D.S., Kim, B.C., 2019. Privacy and personal data collection with information externalities. *J. Public Econ.* 173, 113–124.
- China Internet Network Information Center (CNNIC), 2017. Statistical report on internet development in China. Retrieved from <https://cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf>.
- Colangelo, G., 2019. Facebook and the bundeskartellamt's winter of discontent. Available at: <https://www.competitionpolicyinternational.com/facebook-and-bundeskartellamts-winter-of-discontent/>.
- Evans, D.S., 2014. Economic aspects of bitcoin and other decentralized public-ledger currency platforms. Coase-Sandor Institute for Law & Economics Working Paper, Coase-Sandor Institute for Law and Economics, Chicago, IL.
- Foer, F., 2020. The tech giants are dangerous, and congress knows it. Available at <https://www.theatlantic.com/ideas/archive/2020/07/tech-giants-are-dangerous-and-congress-knows-it/614740/>.
- Graham, M., 2020. Amazon's Ad business is booming and there's still room to grow. Available at <https://www.cnbc.com/2020/01/31/amazons-ad-business-is-booming-and-theres-still-room-to-grow.html>.
- Hagiu, A., 2013. Strategic decisions for multisided platforms. *MIT Sloan Manag. Rev.* Available at <https://sloanreview.mit.edu/article/strategic-decisions-for-multisided-platforms/>.
- Ichihashi, S., 2020. Online privacy and information disclosure by consumers. *Am. Econ. Rev.* 110 (2), 569–595. <https://doi.org/10.1257/aer.20181052>.

- Jones, C.I., Tonetti, C., 2020. Nonrivalry and the economics of data. *Am. Econ. Rev.* 110 (9), 2819–2858.
- Kostka, G., 2018. China's social credit systems and public opinion: explaining high levels of approval. SSRN. Retrieved from <https://ssrn.com/abstract=3215138>.
- Koutroumpis, P., Leiponen, A., Thomas, L.D.W., 2020. Markets for data. *Ind. Corp. Change* 29 (3), 645–660. <https://doi.org/10.1093/icc/dtaa002>.
- Office of the Privacy Commissioner of Canada (2014). "Data brokers: a look at the Canadian and American landscape". Report prepared by the Research Group of the Office of the Privacy Commissioner of Canada. Available at https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2014/db_201409/.
- Ohm, P., 2010. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Rev.* 57, 1701.
- Posner, E., Weyl, G.E., 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
- Riley, C. and I. Kottasová, 2019. Europe hits google with a third, \$1.7 Billion Antitrust fine. Available at <https://www.cnn.com/2019/03/20/tech/google-eu-antitrust/index.html>.
- Roth, A.E., 2008. What have we learned from market design? *Econ. J.* 118 (527), 285–310.
- Savona, M. (2019), "The value of data: towards a framework to redistribute it", SPRU working paper 2019-21.
- Savona, M., 2019b. If Data are the 'New Oil', How Can Their Value be Shared Fairly? Policy@Sussex.
- Savona, M., 2020. Governance Models for Redistribution of Data Value. VOX. CEPR Policy Portal. Available at: <https://voxeu.org/article/governance-models-redistribution-data-value>.
- Veale, M., Binns, Reuben, 2017. Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. *Big Data Soc.* <https://doi.org/10.1177/2053951717743530>.
- Zuboff, S., 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power Hardcover*. Published by Public Affairs.