Privacy in information markets

Andreas Bech

September 2023

1 Introduction

Markets for information are an increasingly important aspect of economic activity. The ability of market participants to collect and analyze data can have large implications for the profitability of firms and their advantage in competition. Data in this context is information that is useful for inferring the private types of consumers. Consumer's types can be private information but might also be unknown to consumers themselves, where consumers only observe a noisy signal of their own type. An example is customer reviews on the Amazon marketplace, which can be seen as a signal about the consumers preferences. The market for information is thus potentially useful for both sides of the market. For the firm, knowledge of consumer types can be used for price discrimination and recommending products. For a platform, consumer preferences can be used to sell advertising spots that match firms and consumers.

The revelation of data, however, raises the issue of privacy. There is no unifying model of privacy in the economics literature. Privacy might be modelled as direct consumer disutility from the disclosure of private information. A more economic approach, however, would be to model the economic implications in the market from the disclosure of private information – for example through third-degree price discrimination. An important issue that arises in this case is an externality from individual consumers sharing personal information. If consumer's types are related, the firm will use all its available data to infer each individual's type, regardless of how much information was shared by the consumer. This idea is quite natural and clearly arises in practice. Companies will estimate statistical models that take an individual's coviariates as input and output a prediction/recommendation. Such models clearly benefit from more data from all consumers. The externality, of course, arises as individuals only account for their personal loss from sharing their own data.

This kind of loss reminds of the classic argument by the Chicago School scholars (e.g., Posner (1978, 1981) and Stigler (1980)) that one main reason for demanding privacy is to avoid exploitation by potential trading partners who might take advantage of the released information against the revealing individuals.

What is a good model of the market for information? First, it should include a way for firms to exploit information, hence creating an endogenous demand in data markets as is clearly seen in practice. This could be through price discrimination with market segmentation, or through creating surplus through product recommendation and loosening search frictions. Second, the model should predict an equilibrium price of information and how it compares to a socially optimal price. Another aspect that would be desirable to include is the ratchet effect: The expected use of information influences a consumers willingness to reveal information. This likely requires a general equilibrium type model with more moving parts. Lastly, a model with clear predictions about welfare would make policy design and regulation more tractable.

In terms of regulation, GDPR in the European Union is the most comprehensive privacy regulation to date. GDPR, however, emphasizes the individuals control of their own data, and does not handle externalities at all.

Ideally, regulation should contain measures to increases overall social welfare in information markets. An obvious idea in regulation would be to put explicit privacy constraints on the full market/mechanism that incorporates all agent's decisions. Another alternative suggested by Posner and Weyl (2018) [3] would be to construct a platform where groups of agents can make a collective decision about how much information to reveal, thus internalizing the externatility. The major issue with regulation in this area, clearly, is feasibility.

In this paper, I explore a potential mechanism for improving welfare outcomes in information markets. Assuming that the key factor driving inefficiency in information markets is a negative externality, arising from the fact that consumers types are dependent on each other, a way to reduce the externality would be to "decorrelate" consumers signals such that a sharing decision by one consumer does not affect the welfare of other consumers. Using the framework of Acemoglu et al. (2022) [8] I analyze a game where consumers can choose to sell their data to a mediator that commits to transform (decorrelate) the data before sharing it with the platform.

The model will also try to answer the question of competition. Since the mediator will be competing directly with the platform, it is necessary to think how such competition works in games where players in the second stage have externalities in their action choices.

In section ....

As mentioned, the GDPR regulation might miss the mark in terms of data externalities and prices of information. Mechanisms for how to combat information revelation across consumers might be very interesting but perhaps very complicated. Taxes on the exchange of information is another option that needs to be explored more. In any case, a good modelling framework is necessary.

2 Literature

This paper builds on the model introduced by Acemoglu et al. (2022) [8]. This model uses a reduced-form disutility from privacy loss. A platform wants to minimize an expected squared loss of estimates of consumers true types,

$$\sum_{i \in \mathcal{V}} \left\{ E\left[(\hat{x}_i(\mathbf{S}) - X_i)^2 \right] - \sigma_i^2 + p_i \right\},$$

S is a vector of data that platform acquires from consumers at prices p_i and $\hat{x}_i$ is an estimate of consumers type. Consumers want to maximize this quantity. The platform compensates consumers for sharing their signal with individual prices. The idea here is that the data is dependent across consumers such that data from consumers $j \neq i$ is useful for estimating consumer i's type, i.e. data sharing by one user reveals data about others, i.e. an externality. Because the value of information in the model is sub-modular (each consumer sharing adds a little less information than the previous one), the data externality depresses the price, p_i, of the data. One consequence of this is that you can get non-monotonicity in payments to users, i.e. the platform might have a decrease in total payment by increasing the number of users sharing data. This will lead to "over-sharing" of data in equilibrium and there can be negative total surplus from the data sharing market.

In contrast Bergemann et al. (2022) [11] consider dependent types that are unknown to both the consumer and a monopolist firm. In this model a data intermediary buys data from the consumers and sells it back to both consumers and the firm. The firm wants to acquire information about consumers for price discrimination (firm charges personalized prices p_i). In this model the data is valuable both to the consumers and the firm, but as in Acemoglu et al. (2022) it hurts consumer's surplus when the firm gets a better signal, as the firm can then better extract surplus from the consumers. This is more of a general equilibrium model where the data is useful in a product market and is useful both to consumers (consumers also only observe a noisy signal) and the firm. This is opposed to Acemoglu et al. (2022) which is reduced form where the objective is simply to predict each type from the aggregate data. In this sense the paper endogenizes the notion of privacy which is very attractive. The model is very flexible and does not have any definite predictions about social surplus from the presence of the data intermediary - it depends on the data structure. One prediction of the model, however, is that the intermediary profitable only acquires anonymous (aggregated) data, as this data is cheaper to get from consumers.

Choi, Jeon, and Kim (2019) [4] also consider a model with information externalities. Here each consumer has a two-dimensional type, willingness to pay and a privacy preference parameter, and information shared by one user is useful to infer information about another. The data collector is a monopoly firm. The optimal pricing policy results in too much data sharing and loss of privacy.

In Ichihashi (2021) [7] they model the consumer's data as part of an information structure. As opposed to other papers where each consumer has a type that needs to be estimated, here there is an unknown state of the world X and each consumer has an experiment (as in Bayesian Persuasion literature) available for sale that is informative about the state. The consumer's experiments may not be independent. The firm (data collector) sets a price vector and then consumers play the game of whether to share their data/experiment. An allocation of data is an experiment available to each consumer. The question is, which allocations of data maximize or minimize consumers surplus and firms profit. If it is assumed that consumers are worse of if the firm learns more about the state, then consumer welfare is minimized and firm profit maximized with "substitutable" data. This is in line with Acemoglu et al. (2022) [8], where leaked information is submodular. Consumer surplus is maximized with complementary data, where $n-1$ sharing is uninformative about the state without the final consumer.

Ichihashi (2021) [6] presents an alternative explanation for low the compensation consumers get for sharing data. There are multiple data intermediaries and the consumer's data is nonrivalrous, thus the consumer can sell the same data to multiple intermediaries. In the downstream market (where intermediaries sell data to firms) this decreases the value of data. Thus in equilibrium intermediaries will not pay much to consumers to acquire data, as nonrivalry of data relaxes competition between intermediaries. This is interesting because a common argument for why consumers are not compensated well for their data is that there is a lack of competition. But with the nonrivalry of data, and Bertrand competition in downstream markets, more competition does not lead to better outcomes for consumers. It is related to vertical contracting too. With a rivalrous good, all surplus accrues to the upstream players, where in this case competition drives the price of data to 0 and the consumer is only compensated for their potential loss of data sharing.

There are also alternative notions of privacy in information markets. Another example of a reduced-form privacy model is to consider a mechanism where each agents is asked to send a message to a principal. The principal will then update their belief about each agents type using a prior and all of the messages. Eilat et al. (2021) [5] have a version of this model with one agent. Loss of privacy for the agent is then measured as the KL-divergence between the prior and posterior. In this case one can think of an external regulator adding a constraint that this loss of privacy be less than a given constant. This model is very general in measuring the loss of privacy but fails to capture the equilibrium effects among consumers of sharing information, and hence the potential externality.

Likewise, there are notions of so-called differential privacy, stemming from the computer science literature. Pai and Roth (2013) [2] use a definition where the message of single player has only a small effect on the outcome of the mechanism. This is a characterization of a mechanism where each individuals message does not matter too much. Obviously this becomes easier when the number of participants increase. In terms of privacy, however, it does not capture how much is learned about each consumer's type in the mechanism.

In Armstrong and Zhou (2022) [9] explore competition when information has to be acquired by firms. There is one consumer and two firms, and the consumer is looking to buy a single unit. The difference between the consumer's valuations for each firms product, x, is symmetrically distributed around 0 with CDF F. The consumer does not observe x, but observes a private signal $\sigma(s|x)$.

The posterior expected x for the consumer is called G, which is observed by the firm (since they know σ). Firms set prices simultaneously and solution concept is Bertrand-Nash. So in this paper, there are no data externalities on the consumer side. Instead, two firms compete and it is explored which signal structures give high consumer welfare and firm profits, respectively. For the signal structures, there is a trade-off b/w match quality and competition. Full information disclosure is optimal for neither consumer nor firms.

Finally, Bergemann and Bonatti (2022) [10] model an extension of the Mussa and Rosen (1978), where a platform sells access to the on-platform market which has better information to match consumers to one of J firms. Each firm offers two mechanisms $(q_j(\theta), p_j(\theta))$, one on and one off the platform. In this case platforms take on the role of intermediaries that are better informed than both consumers and firms and can thus better facilitate transactions. The platform then charges a fee for any firm that wants to use the platform. It is not so related to data externalities as consumers do not have private information to sell.

3 Model

The model is borrowed from Acemoglu et al. (2022) [8]. There are n users and one platform. Users have types $X = (X_1, \dots X_n)$ which are jointly normally distributed $\mathcal{N}(0, \Sigma)$. The diagonal of the correlation matrix will be denoted by σ_i. Each user has a signal, which can be interpreted as personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$. The platform offers prices, $p = (p_1, \dots, p_n)$ to users to acquire their data. Given prices, users make sharing decisions $\mathbf{a} \in \{0, 1\}^n$, where $a_i = 1$ corresponds to sharing data.

Let $\mathbf{S_a}$ index the users that choose to share their data. Then the *leaked information* is defined as

$$\mathcal{I}_i(a) = \sigma_i - \min_{\hat{x}_i} E[(X_i - \hat{x}_i(S_a))^2]$$

where σ_i is a normalization such that without any signals the leaked information is 0.

The platform wants to maximize

$$U(a, p) = \sum_i \mathcal{I}_i(a) - \sum_{i: a_i = 1} p_i$$

And users want to max $u_i(a, p) = p_i - v_i \mathcal{I}_i(a)$, where v_i is their preference for privacy.

4 Platform Monopolist

When platform offers price vector $\mathbf{p} = (p_1, \dots, p_n)$, a sharing profile $\mathbf{a}$ forms a user equilibrium if

$$a_i \in \underset{a\in\{0,1\}}{\arg\max} u_i\left(a_i = a, \mathbf{a}_{-i}, \mathbf{p}\right).$$

Let the set of equilibria formed by price, $\mathbf{p}$, be denoted by $\mathcal{A}(\mathbf{p}$. A pair $(\mathbf{a}^E, \mathbf{p}^E)$ if $\mathbf{a}^E \in \mathcal{A}(\mathbf{p}^E$ forms a pure strategy Stackelberg equilibrium of the game if

$$U\left(\mathbf{a}^{\mathrm{E}}, \mathbf{p}^{\mathrm{E}}\right) \geq U(\mathbf{a}, \mathbf{p}), \quad \text{for all } \mathbf{p} \text{ and for all } \mathbf{a} \in \mathcal{A}(\mathbf{p})$$

Acemoglu et al. (2022) [8] show that such an equilibrium exists and is given by prices $p_i^{\mathrm{E}} = v_i\left[\mathcal{I}_i\left(a_i = 1, \mathbf{a}_{-i}^{\mathrm{E}}\right) - \mathcal{I}_i\left(a_i = 0, \mathbf{a}_{-i}^{\mathrm{E}}\right)\right]$ such that users are indifferent between sharing and not sharing. The allocation preferred by the platform is then characterized by sharing decisions

$$a^E = \underset{a\in\{0,1\}^n}{\arg\max} \sum_i (1-v_i)\mathcal{I}_i(a) + v_i\mathcal{I}_i(a_{-1}, a_i = 0)$$

4.1 Inefficiency and Regulation

It can be shown that if a high value user ($v_i > 1$) is correlated with any other user then the equilibrium is inefficient for some v. Equilibrium surplus can even be negative in some cases, in particular when user with a high-value for privacy have a high correlation with low-value users, so that low-value users leak a lot of information about them.

When social surplus is negative, shutting down data markets altogether is clearly beneficial. Taxes can even restore the first best if regulation is allowed to use personalized taxes. It is straightforward to construct a Pigovian tax where users who should share in equilibrium are not taxed and the rest are taxed prohibitively. However, such a tax is practically infeasible as it requires knowledge of users private information which the regulator would have to extract.

5 Mediator

From a regulatory viewpoint, an alternative approach would be to construct a mechanism that can improve efficiency in the data market. Given that inefficiency is caused by the negative externality from the dependence among user's types, an immediate idea is to find a way to remove the dependence among user's signals.

Suppose that a (trusted) mediator could collect the data and transform it before giving it to the platform. The idea is that the mediator can "decorrelate" user's signals using the transformation

$$\tilde{\mathbf{S}} = \Sigma^{-1}\mathbf{S}$$

where $\mathbf{S}$ is the full vector of user signals. Removing correlation is meaningful is this case because the user types are jointly normally distributed. With this transformation the leaked information about user types given by

$$\tilde{\mathcal{I}}_i(\mathbf{a}) = \sigma_i^2 - \min_{\hat{x}_i} E\left[\left(X_i - \hat{x}_i\left(\tilde{\mathbf{S}}_{\mathbf{a}}\right)\right)^2\right] = \begin{cases} 0, & a_i = 0; \\ \mathcal{I}_i\left(a_i, \mathbf{a}_{-i}\right), & a_i = 1. \end{cases}$$

Thus, for users not sharing there is no leaked information, regardless of who else shares information. Furthermore, the leaked information about any user who shares is the same as without decorellation given $\mathbf{a_{-i}}$. The equilibrium sharing profile that is preferred by the platform with this scheme is given by

$$\tilde{a}^E = \arg\max_{a \in \{0,1\}^n} \sum_i (1 - v_i)\tilde{\mathcal{I}}_i(a)$$

and prices are given by $\tilde{p}_i^E = v_i \tilde{\mathcal{I}}_i(a)$ for all i such that $\tilde{a}_i^E = 1$. It can be shown that

$$\textit{Social surplus}(\tilde{a}_i^E) \geq \max\left\{\textit{Social surplus}(a_i^E), 0\right\}$$

which follows from the fact that high value users, who contribute negatively to social surplus, now get a non-negative surplus.

A question that remains is how such a mechanism could be implemented. In the above, the mediator was assumed to be trusted and have no objective of it's own. A more implementable alternative would be a "market" solution where a mediator is allowed to participate in the data market and users are allowed to choose freely who to sell their data to. In such a situation the platform and the mediator would compete for user data. In the following I will try to analyse such a game where it is assumed that the mediator can commit to a data transformation policy.

6 Competing

The mediator will buy data from users at prices $\mathbf{p}^M$ and then sell the transformed data, $\tilde{\mathbf{S}}$, to the platform. The mediator thus gets a payoff $U^M(q, \mathbf{p^M}) = q - \sum_{i:a_i=M} p^M$, where q is the price paid to mediator by the platform. The platform gets payoff as before $U^P = \sum_i \mathcal{I}_i(a) - \sum_{i:a_i=P} p_i - q$. If the mediator does not acquire any data then $q = 0$. The leaked information in the game with mediator and platform is given by

$$\mathcal{I}_i(a) = \sigma_i - \min_{\hat{x}_i} E[(X_i - \hat{x}_i(S_a, \tilde{S_a}))^2]$$

such that the platform updated the expected type using both the data it acquires directly and the data from the mediator, $\tilde{S_a}$

The equilibrium concept here is SPE between the Mediator and the Platform. The timing of the game is as follows.

1. The mediator commits to a data transformation

2. Platform and mediator simultaneously set price vectors $\mathbf{p}^M \in \mathbb{R}^n$ and $\mathbf{p} \in \mathbb{R}^n$
3. Users decide who to share data with, if any, i.e. $\mathbf{a} \in \{0, P, M\}^n$, where 0 denotes not sharing.
4. Finally, Mediator can share acquired and transformed data with platform at price $q \in \mathbb{R}$

The following observations simplify the game.

Observation 1. *All user who choose not to share their data through mediator will still prefer to give data to mediator so that mediator can remove correlation.*

Given this observation, the data that the mediator is selling to the platform is completely decorrelated, i.e. the users who are not sharing data through mediator will not have further leaked information when the platform buys $\tilde{S}_a$.

Observation 2. *In the final stage, the mediator will charge the platform's willingness to pay for the mediator's data, $\tilde{S}$. In stage 2, the mediator will not charge prices such that they pay more in equilibrium than the platforms willingness to pay in stage 4.*

The platform's willingness to pay for the mediators data is complicated. If the platform already has acquired data from a subset $T \subset \{1, \ldots, n\}$ and the mediator has data from a subset S with $T \cap S = \emptyset$. The platform will update posterior means using its own data T, and then update again will the decorrelated data from the mediator, $\tilde{S}$. I do not analyze how this works in the n player case. It is feasible with 2 users. It is thus unclear whether a closed form solution is available for the platforms willingness to pay for $\tilde{S}$ in the general case. The platforms willingness to pay is the difference $\sum_i \mathcal{I}_i(\mathbf{T}_a)$ and $\sum_i \mathcal{I}_i(\mathbf{T}_a + \tilde{\mathbf{S}}_a)$[1].

7 Existence

The game between the mediator and the platform is discontinuous as small changes in prices can change user equilibria and thus cause large changes in the payoff of the mediator and platform. Thus, the Nash-Glicksberg theorem does not guarantee any existence of a mixed strategy equilibrium. Existence can, however, be established using the Dasgupta-Maskin theorem [1].

8 Example

Consider game with two users, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $v_1 = v_2 = v$. This is the simplest, non-trivial version of the game possible.

[1] I am using the notation $\mathbf{T}_a$ to denote the data available for a player. In this case $\mathbf{T}_a$ is the set of signals shared with the platform so $\mathbf{T_a} = (S_j : j \in \{1, \ldots, n\}$ such that $a_j = P)$. The addition operator in $\mathbf{T}_a + \tilde{\mathbf{S}}_a$ then denotes the union.

8.1 Monopolists case

First I analyze the game where the platform is playing the game as a monopolist, with or without the mediator transforming the data.

Without the mediator, the platform sets prices and users choose whether or not to sell data. If there are multiple user equilibria, the platform gets to choose their preferred one. With $v \leq \frac{4}{(2-\rho^2)^2}$ the equilibrium is $a_1 = a_2 = 1$ with prices $p_1 = p_2 = v\frac{\left(2-r^2\right)^2}{2(4-r^2)}$. Otherwise, the equilibrium is $a_0 = a_2 = 0$. As such, with only two players the platform prefers to sell acquire data from either both or neither, depending on the value of privacy, v.

When the (trusted) mediator transforms the data before giving it to the platform, the equilibrium outcome changes. With $v \leq 1$ the mediator equilibrium is $\tilde{a}_1 = \tilde{a}_2 = 1$ with prices $\tilde{p_1} = \tilde{p_2} = v\frac{2}{4-\rho^2}$. Otherwise, the equilibrium is $\tilde{a}_0 = \tilde{a}_2 = 0$.

When there is sharing it holds that $\tilde{p} > p$. Since no information about a user is leaked when he does not share, the outside option is better for the user. This gives higher prices for the users. Social surplus is bigger with mediator, as previously pointed out. Also note that without the mediator, the sharing equilibrium $a_1 = a_2 = 1$ is more likely in terms of v. That is, without the mediator the platform chooses a sharing equilibrium even when $v > 1$, which is due to the submodularity of the leaked information – the sharing choice of one user affects the reservation value of the other user.

8.2 Competing

Consider the case where $v \leq \frac{4}{(2-\rho^2)^2}$. Then platform gets positive surplus $\frac{4}{4-\rho^2} - v\frac{(2-\rho^2)^2}{4-\rho^2}$. Thus the mediator can offer higher prices to buy both user's data and offer them to the platform. When both users share, the platform gets the same amount of leaked information from data from the mediator or from acquiring the data directly. The mediator and platform are thus in direct price competition. Notice, however, that the mediator has to offer the users higher prices since the outside option when the data is shared through the mediator is better (no leaked information). In particular, the mediator has to offer prices $p_i^M \geq v\frac{2}{4-\rho^2}$. The platforms break-even price is $p_i = \frac{2}{4-\rho^2}$. Thus for $v \in [1, \frac{4}{(2-\rho^2)^2}$ the mediator cannot profitably offer a higher price to outbid the platform. Thus, there is an equilibrium in pure strategies, where the platform offers prices $p_i = \frac{2}{4-\rho^2} - \epsilon$ (where ϵ depends on v), the mediator offers prices below that and users share data with the platform, $a_1 = a_2 = M$. In this equilibrium, as compared to the equilibrium with the platform as monopolist, the users get a higher surplus, through higher prices. The platform still gets a positive surplus, though it is reduced to ϵ.

When $v > \frac{4}{(2-\rho^2)^2}$ then no sharing, $a_1 = a_2 = 0$, is always an equilibrium. When $v < 1$ then platform and mediator outbid each other until price $p_i = \frac{2}{4-\rho^2}$ and both get payoff 0. Both users get payoff $(1-v)\frac{2}{4-\rho^2}$. It can be checked that these are the only pure strategy equilibria of the game, for any $v > 0$.

9 Discussion

The addition of the mediator in competition for user data is unclear. When the data is shared through the mediator, it increases the outside options of users and hence higher prices have to be offered. When values for privacy are in a medium range – as was seen in the two user example – these prices are prohibitively large, and the platform gets to make a positive profit. Even though, through the introduction of the mediator, users have the option to submit decorrelated signal, and thus get higher prices, in one equilibrium the mediator is completely bypassed. Even so, prices, and consumer surplus, are higher than without the mediator.

It is unclear what will happen in the n player case, more restrictions on the problem are required in order to make the game tractable and draw general conclusions.

References

[1] Partha Dasgupta and Eric Maskin. "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory". In: *The Review of Economic Studies* 53.1 (Jan. 1986), p. 1. ISSN: 00346527. DOI: 10.2307/2297588. URL: https://academic.oup.com/restud/article-lookup/doi/10.2307/2297588 (visited on 09/19/2023).

[2] Mallesh Pai and Aaron Roth. *Privacy and Mechanism Design*. 2013. arXiv: 1306.2083 [cs.GT].

[3] Eric A. Posner and E. Glen Weyl. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press, 2018. ISBN: 9780691177502. URL: http://www.jstor.org/stable/j.ctvc77c4f (visited on 04/25/2023).

[4] Jay Pil Choi, Doh-Shin Jeon, and Byung-Cheol Kim. "Privacy and personal data collection with information externalities". In: *Journal of Public Economics* 173 (May 2019), pp. 113–124. ISSN: 00472727. DOI: 10.1016/j.jpubeco.2019.02.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047272719300131 (visited on 03/05/2023).

[5] Ran Eilat, Kfir Eliaz, and Xiaosheng Mu. "Bayesian privacy". In: *Theoretical Economics* 16.4 (2021), pp. 1557–1603.

[6] Shota Ichihashi. "Competing data intermediaries". In: *The RAND Journal of Economics* 52.3 (Sept. 2021), pp. 515–537. ISSN: 0741-6261, 1756-2171. DOI: 10.1111/1756-2171.12382. URL: https://onlinelibrary.wiley.com/doi/10.1111/1756-2171.12382 (visited on 07/05/2023).

[7] Shota Ichihashi. "The economics of data externalities". In: *Journal of Economic Theory* 196 (Sept. 2021), p. 105316. ISSN: 00220531. DOI: 10.1016/j.jet.2021.105316. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022053121001332 (visited on 06/18/2023).

[8] Daron Acemoglu et al. "Too Much Data: Prices and Inefficiencies in Data Markets". In: *American Economic Journal: Microeconomics* 14.4 (Nov. 2022), pp. 218–56. DOI: 10.1257/mic.20200200. URL: https://www.aeaweb.org/articles?id=10.1257/mic.20200200.

[9] Mark Armstrong and Jidong Zhou. "Consumer Information and the Limits to Competition". In: *American Economic Review* 112.2 (Feb. 1, 2022), pp. 534–577. ISSN: 0002-8282. DOI: 10.1257/aer.20210083. URL: https://pubs.aeaweb.org/doi/10.1257/aer.20210083 (visited on 06/18/2023).

[10] Dirk Bergemann and Alessandro Bonatti. "Data, Competition, and Digital Platforms". In: *SSRN Electronic Journal* (2022). ISSN: 1556-5068. DOI: 10.2139/ssrn.4236337. URL: https://www.ssrn.com/abstract=4236337 (visited on 01/27/2023).

[11] Dirk Bergemann, Alessandro Bonatti, and Tan Gan. "The economics of social data". In: *The RAND Journal of Economics* 53.2 (2022), pp. 263–296. DOI: https://doi.org/10.1111/1756-2171.12407. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12407. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12407.