Consumer data effects on competition and market outcomes

The thesis focuses on digital markets, with a particular attention toward consumer data and their effects on market outcomes. Consumer data can be obtained through various means. Some, usually with low accuracy, can be obtained through public repositories, such as the average wages in a city's zone. Other data are instead generated by the interaction between consumers and firms, and the latter can then use that information when it interacts with that consumer again (or consumers with similar tastes).

In the first chapter, I provide a survey of the existing literature regarding consumer data and their effects. As this literature includes various strands, such as artificial intelligence and machine learning, I limit to the scope of the analysis to those papers that model data as an input to a decision problem. This choice thus excludes works regarding data-enabled technologies that do not explicitly model how the quantity or quality of data affect decisions. With the aim of extracting broader insights, I organize the existing literature depending on the data acquisition stage. Indeed, I find that the modelization choices of this stage often guide many of the market outcomes presented.

First, I focus on papers where data are exogenously available to firms, or where firms can obtain them without strategic interactions (such as buying them at an exogenous cost). This modelling choice does not usually allow firm to fully internalize data externalities: as such, they present an overuse of data with respect to the social optimum, which leads to an increase in competition between firms. While this effect should benefit consumers, data overcollection can raise privacy concerns that should be taken into account when examining the effects of data.

Second, I focus on models where firms can obtain data by interacting with consumers one or multiple times. In this class of models, I find that the pro or anti-competitive effect of data is strongly linked to firm symmetry. Indeed, the data use over repeated periods can exacerbate a firm's starting advantage, increasing concentration in the market or even leading to tipping. Moreover, firms can also trade or sell data among them, to limit consumer interaction and thus reducing the compensation they should pay to consumers. This feature is especially relevant when consumer data are correlated, enabling firms to gain information on consumers they do not directly interact with.

Finally, I analyze models where data are obtained from intermediaries, that can either directly interact or not with consumers to obtain such data. Intermediaries are able to internalize the effect that selling data to a firm has on its rivals, and choose their data selling strategy accordingly. Moreover, competing data intermediaries can strategically coordinate their actions to temper competition between them.

For each of these classes of models, I briefly describe the evolution of the theoretical literature, and the features that drive the main findings. Moreover, whenever possible, I integrate the analysis with empirical works to better frame the inisghts of the theoretical literature.

In the second chapter, I build on the existing literature by focusing on the effects that a DB has on a downstream market where entry is endogenous. In particular, I focus on the case where consumer data sold by the DB allows firms to price discriminate consumers. Most of the literature on DBs has focused on downstream markets with a fixed number of firms, highlighting a pro-competitive effect of data. Indeed, price discrimination induces firms to lower their prices, benefitting consumers as a whole.

I study a monopolistic DB that has data regarding all consumers in the downstream market, which is modelled as a circular city where firms can enter by paying a fixed cost. Firms can buy data from the DB, and then use them to price discriminate consumers, offering them tailored prices.

The analysis shows that the DB's equilibrium strategy depends on the selling mechanism he adopts. When the DB sells data through Take-It-Or-Leave-It offers, he sells data to all entering firms. In particular, he chooses to sell non overlapping partitions if the downstream market horizontal differentiation is high, and opts to sell the whole dataset to all entering firms otherwise. Instead, if the DB can change the offer he makes to a firm conditional on another firm's response such as through auction mechanisms, the DB opts to only serve a subset of the entering firms to further temper downstream competition. Regardless of the selling mechanism, the data sale results in a reduction of the number of entering firms, which I refer to as *entry barrier effect.* The entry reduction results in an increase in market concentration, which in turn harms consumers with respect to the benchmark model with no data.

In the third chapter, I expand the previous model by adding competition in the DB market and by analyzing the effect of information accuracy on market outcomes. To do so, I model a vertically differentiated duopoly in the DB market, where two DBS have datasets with different levels of accuracy. In equilibrium, only the most accurate DB sells partitions to downstream firms. However, the less accurate DB exerts competitive pressure on him, limiting the rents that he can extract from downstream firms. While the data sale always decreases downstream entry, the effects on consumer surplus with respect to the benchmark are more nuanced. Indeed, I find that the degree of information accuracy determines the magnitude of the effect that the data sale has consumer surplus. Instead, the degree of competition between DBs determines whether this effect will be positive or negative. In particular, consumer surplus is maximized (and higher than the benchmark model with no data) when both information and DB competition are perfect, and conversely it is minimized when information is perfect and the DB market is monopolistic.

Finally, in the fourth chapter I focus on a different type of data intermediary, that is a digital platform. In particular, the chapter aims to analyze the effects that mandating data sharing can have on hybrid marketplaces such as Amazon. The term hybrid marketplace indicates a platform that hosts sellers, but can also vertically integrate to compete with them. As data sharing is one of the main pillars of the upcoming Digital Markets Act, I focus on the intended and unintended effects this policy measure can have.

To do so, I model a continuum of downstream markets, each one containing a firm. We assume that all these markets belong to the same product group (such as "baby products" or "clothing and accessories") Firms pay a per-transaction fee to the platform, consistently with Amazon's referral fees, which is unique for the given product group. Moreover, we only allow the platform to enter a market where a firm is already present, as empirical evidence has shown that platforms like Amazon usually enter markets after having observed their profitability.

To broaden the scope of the analysis, I study the effects of mandated data sharing when the downstream markets sell homogenous or heterogenous goods. Moreover, I analyze two separate effects of data: whether they allow firms to price discriminate consumers, or whether they allow firms to decrease their marginal costs of. The results show that mandating data sharing can have unintended consequences on consumers, especially when data allows firms to price discriminate consumers. While total welfare increases, the platform can strategically use the per-transaction fee as a tool to temper competition with sellers. In particular, in the case of heterogeneous goods, the platform can set a fee so high that downstream sellers opt to set monopolistic prices, even if the market is fully covered in equilibrium. Instead, the cost reduction effect of data is welfare increasing and benefits consumers. However, cost efficient sellers may be particularly damaged by data sharing, as their benefit from cost reduction is limited and they suffer from the increase of the per-transaction fees.