Data, Targeted Advertising and Consumer Welfare*

Alejandro Rábano†

October 16, 2025

Updated regularly, please click here for the last version.

Abstract

Which firms and consumers benefit from the increasing use of data for targeted advertising? This paper develops a model of search frictions in product markets where heterogeneous firms use data to target different consumers. Firms endogenously adopt Big Data technologies that lower the cost of targeting. Using firm-level balance sheet data combined with a technology survey for France, I show that firms analyzing Big Data from social media (i) use it for advertising, (ii) are larger, (iii) are more productive, and (iv) face higher fixed costs. I calibrate the model to France in the late 2010s and show that it replicates the positive correlation between Big Data use, firm size, and fixed costs. The model implies that a technological improvement increasing the availability of consumer data benefits only large firms, which allocate more resources to fixed targeting costs, and shifts consumption toward high-valuation consumers. The equilibrium is inefficient and expenditure in targeting is excessive because firms do not internalize that targeting high-valuation consumers crowds out competitors from reaching them. Welfare in the competitive equilibrium rises with a digital advertising tax, which reduces the intensive use of targeted advertising by large firms. Finally, I use the model to evaluate the aggregate effects of lifting the GDPR. Relaxing consumer data regulation would encourage Big Data adoption but increase firm concentration.

Keywords: Big Data, Firm Behavior, Technological Adoption, Advertising
JEL Classifications: D21, D22, D83, M37, E22, O33

*I am grateful to my advisors Isaac Baley and Edouard Schaal for their constant guidance and support for this project, and to Elia Benveniste, Ayah Bohsali, Andrea Caggese, Pia Ennuschat, Santiago Iglesias, Marko Irisarri, Raquel Lorenzo, Gerard Martin-Escofet, Agustina Martínez, Marta Morazzoni, David Nagy, Pau Roldán-Blanco, José Nicolás Rosas, Andrea Sy and Jaume Ventura, for many helpful discussions. I am grateful to Maarten De Ridder for his comments and for the hospitality during my stay at LSE. This paper was previously circulated as "Firms and Big Data: Adoption, Use and Impacts", and has also benefited from discussions with participants at the UCL Macro Reading Group, LSE Macro Workshop, Universitat de València, BSE Jamboree, 22nd ZEW Conference on the Economics of ICT, XXVII Vigo Workshop on Dynamic Macro, 49th SAEe, VIII Bellaterra Macro Workshop and UPF. I would like to thank Andrea Caggese, the rest of the project participants and the EIB for granting me access to the INSEE data through the EIBURS grant (ECONPSRINDF-2023-05) on "Intangibles, Technology Diffusion, and Public Policies: Implications for Firm Investment, Market Structure, and Aggregate Productivity". I gratefully acknowledge support from the FPI grant (PRE2020-093287).

†Universitat Pompeu Fabra. Email: alejandro.rabano@upf.edu. Website

1 Introduction

New technologies are driving data collection at an unprecedented pace: the volume of available information increased sixtyfold between 2010 and 2023.[1] A major source of this information is social media, where firms can learn about consumer demographics such as age, gender, or location. These data allow firms to target ads toward consumer groups that place higher value on specific goods. The rise of targeted advertising has been a key driver of the threefold increase in digital advertising expenditure since 2017.[2] The growing reliance on consumer data has prompted regulatory responses, particularly in Europe: the General Data Protection Regulation (GDPR) imposed additional compliance costs on the use of consumer information, and there is an active policy debate about the regulation of digital advertising on social media.[3]

A strand of the literature has examined the aggregate effects of data (Farboodi and Veldkamp (2024); Jones and Tonetti (2020), among others), while recent research has focused on the macroeconomics of targeted advertising (Cavenaile, Celik, Perla, and Roldan-Blanco (2023); Greenwood, Ma, and Yorukoglu (2025)). However, existing frameworks of targeted advertising abstract from the interaction with the analysis of consumer data at the firm level. In addition, little attention has been paid to how targeted advertising interacts with firm heterogeneity. How does greater availability of consumer data affect firms across the size and productivity distributions? This paper bridges these two literatures by developing a novel framework of heterogeneous firms that engage in targeted advertising based on consumer data. The results suggest that technological advances increasing the availability of consumer data disproportionately benefit large firms, which use targeted advertising more intensively.

I present a model of search in product markets where heterogeneous firms adopt Big Data technologies that reduce the cost of processing consumer information. Firms choose how to target different consumers. They seek to target high-value consumers who have a greater willingness to pay, but targeting is costly because it requires processing consumer data. In order to discipline the model, I combine a technology survey with balance sheet data from France to show that firms using Big Data from social media—a proxy for consumer data—are larger, more productive, and face higher overhead costs. I use this empirical evidence to calibrate the model to France in the late 2010s and show that it replicates the positive correlation between Big Data use, firm size, and overhead costs. To illustrate the economic forces at work, I show that a technological improvement that lowers information-processing costs benefits only large firms, which allocate more resources to overhead targeting costs. With the purpose of understanding the need for policy intervention, I show that the competitive equilibrium is inefficient because firms do not internalize that targeting high valuation consumers crowds out competitors from reaching them. Welfare in the competitive equilibrium rises with a digital advertising tax, which reduces the intensive use of targeted advertising for large firms. Finally, I use the model to evaluate the aggregate effects of lifting the GDPR. Relaxing consumer data regulation would encourage Big Data adoption but increase firm concentration.

The theoretical framework combines standard elements from models of firm heterogeneity and technology adoption with recent insights from the literature on targeted search (Cheremukhin, Restrepo-Echavarria, and Tutino (2020); Wu (2024)). In traditional models, search is either random, when agents cannot choose where to search, or directed, when they can choose the specific market in which to search. The targeted search framework lies between these two extremes: agents can partially direct their search, but greater precision in targeting entails higher costs. I apply this framework to model targeted adver-

[1] Appendix Figure B.1b shows that worldwide Big Data grew from 2 zettabytes to 123 zettabytes over this period. Big Data is typically defined as information that is large in volume, diverse in format, and rapidly updated.

[2] Expenditure on digital advertising in Europe reached \$148 billion in 2025, nearly three times its 2017 value (Statista (2024)).

[3] The GDPR entered into force in the EU in 2018, increasing the regulatory burden on firms regarding the use of consumer data. In a recent policy brief, Acemoglu and Johnson (2024) proposed a 50% tax on digital advertising through social media.

tising in product markets, where firms search for heterogeneous consumers and Big Data technologies reduce the cost of targeting.

The model is static and there is a household that consumes a homogeneous good, supplies labor to firms and chooses how much search effort to exert. Within the household, there are different consumer types, with each type representing a demographic group sharing a similar valuation for the good. High valuation types represent demographic groups with a high taste for the good: increasing their consumption rises household welfare. Product markets are subject to search frictions, and a separate market exists for each consumer type. Consumers are matched with firms and the number of matches increases in the household search effort and the number of ads posted by firms. Prices are determined through bargaining between firms and consumers: higher-valuation types have greater willingness to pay and therefore face higher prices. For instance, in the tourism sector, consumers with a strong preference for traveling are willing to pay more for a hotel room, whereas stay-at-home consumers tend to spend less.

Firms differ in productivity and data processing costs and choose a targeted advertising strategy over consumer types. Targeting high valuation consumers yields higher average prices but requires processing large amounts of information, capturing the cost of learning about consumer demographics through social media. Firms can adopt either Big Data or traditional advertising technologies. Big Data lowers data-processing costs and enables more precise targeting across consumer types but requires a fixed adoption investment, reflecting the need to establish a data analytics team. In equilibrium, firms trade-off the higher prices of advertising to high valuation consumers with the targeting costs of processing consumer data. Advertising expenditures enter firms' overhead costs and raise prices for all goods. As a result, large firms have more incentives to spend more resources in targeting and adopt Big Data technologies.

To analyze the welfare implications of targeted advertising, I characterize the social planner's problem. While targeted advertising can raise welfare by increasing consumption by high valuation consumers, it can have adverse aggregate effects. When many firms target the same consumer type, this generates congestion on competitors and decreases the probability to sell. I find that the optimal targeted advertising policy is characterized by a variation of a Hosios rule, where the bargaining power should be equal to the matching elasticity. Because firms take the aggregate level of advertising as given, they fail to internalize the crowding-out effects that arise when all firms target the same consumers. The competitive equilibrium is inefficient. The resulting wedge with the social planner's allocation gives rise to two inefficiencies—a targeting distortion and a production distortion. The direction and magnitude of these distortions depend on model parameters, a point I return to after the calibration.

With the purpose of quantifying the relevance of targeted advertising, I present empirical evidence on Big Data adoption using more than 20,000 firm-year observations from the French Statistical Institute's Enterprise ICT Survey merged with balance-sheet data for the years 2015, 2017 and 2019. The survey asks whether firms use Big Data, defined as large-volume, high-velocity, and diverse information. I focus on Big Data in the context of social media to capture consumer-related information. Around 5% of French firms use Big Data from social media, with adoption rates higher in services. First, firms in 2015 are also asked about the use of information, which shows that Big Data from social media tends to be used for advertising. Second, adoption is positively associated with firm size: only 4% of firms in the bottom quintiles adopt, compared to nearly 8% in the top quintile. Third, I estimate revenue-based total factor productivity (TFPR) following the methodology by Ackerberg, Caves, and Frazer (2015). I find that TFPR is 4.5% higher for firms that use Big Data from social media than comparable firms in the same sector-year, after controlling for capital, age and other technologies. Finally, adopters also exhibit a ratio of overhead to total costs 2.8 percentage point higher than non-adopters.

The empirical evidence informs the model calibration for France in the late 2010s. The framework rationalizes the key empirical facts. It explains why Big Data from social media is used for advertising: lower costs of processing information facilitate targeting high-valuation consumers. A feedback loop between targeting and scale endogenously generates the positive correlation between adoption and firm size. Large firms have stronger incentives to adopt because advertising and technology expenditures are part of overhead costs. In addition, a precise targeting strategy increases firm price and the incentives to operate at a larger scale. While TFPR differences are targeted moments, model-based regressions show that higher prices, rather than higher productivity, drive the revenue premium for adopters. Finally, the model replicates untargeted patterns in costs: overhead shares are 2.9 percentage points higher for Big Data firms, close to the observed value in France.

To explore the key mechanisms, I use the calibrated model to analyze the aggregate effects of technological change. A counterfactual with higher Big Data processing costs shows heterogeneous impact across the firm size distribution. Only large firms increase output after technological change, since they engage in higher targeting expenditures and are more likely to use Big Data technologies. Through general equilibrium effects, traditional firms and small adopters decrease production as the competition to reach high-valuation consumers increases. Since Big Data users are at the top of the size distribution, technological change leads to an increase in the output share by firms at the top quintile. Aggregate welfare rises modestly (0.1%), but gains are concentrated among high-valuation consumers, who receive more ads and increase consumption, while lower-valuation consumers fall behind.

In the baseline calibration, the competitive equilibrium features excessive targeted advertising, particularly by large firms using Big Data. This suggests the need for digital advertising regulation, captured in the model by taxing the targeting expenditures by Big Data firms. A digital advertising tax curbs targeting and reduces production by large Big Data adopters, while increasing output among traditional firms and small adopters through general equilibrium effects. As a result, the production share of top firms declines, bringing the competitive equilibrium closer to the social planner allocation and raising welfare. However, aggregate welfare declines because high valuation consumers are targeted less intensively and thus reduce consumption. Introducing a 30% digital advertising tax in the competitive equilibrium maximizes welfare by balancing these opposing effects. A tax on general advertising yields qualitatively similar results, although the quantitative implications differ.

Finally, I use the model to explore the impact of consumer data regulation. I analyze the effects of lifting the GDPR, introduced in the model as a decrease in Big Data adoption costs in line with empirical estimates from the literature. Lifting the GDPR encourages the adoption of Big Data technologies that improve targeting, increasing consumption by high valuation types and raising welfare. However, it also increases firm concentration: relaxing consumer data regulations amplifies production dispersion since Big Data adopters are disproportionately large. Overall, these results highlight that alternative forms of consumer data regulation have heterogeneous effects on the firm size distribution and on consumer welfare.

1.1 Related Literature.

Evidence of Firms and Big Data. The literature has developed two main approaches to measure firms' use of Big Data. The first relies on inputs related to the use of data. Recent research estimates aggregate investments in data through data-related occupations and document a sharp increase in adoption over time.[4] Similarly, Wu, Hitt, and Lou (2020), Fedyk and Hodson (2023), and Abis and Veldkamp (2024) use job postings and CVs to trace how data capabilities influence firm-level innovation, financial val-

[4]See Goodridge, Haskel, and Edquist (2022), Corrado et al. (2025), Calderón and Rassier (2025), and Schmidt, Pilgrim, and Mourougane (2023).

uation, and labor share, respectively. Demirer, Hernández, Li, and Peng (2024) study the adoption of complementary cloud services to assess the impact of privacy regulation. In contrast, a second strand of research uses firm-level information about adoption. Brynjolfsson and McElheran (2016a) and Maurin, Kolev, and Revoltella (2023) exploit technology surveys to show that Big Data adoption is associated with higher productivity, while Galdon-Sanchez, Gil, and Uriz-Uharte (2025) reach similar conclusions for firms using Big Data provided by a bank. Building on this literature, this paper uses a firm-level technology survey from France to provide new evidence on the impact of social media data on firm size, productivity, and costs.

Models of Firms and Big Data. From the theoretical perspective, a strand of the literature models the aggregate implications of Big Data use: Farboodi and Veldkamp (2024) analyze firm dynamics and the implications for growth, Begenau, Farboodi, and Veldkamp (2018) consider the link to financing costs, Asriyan and Kohlhas (2025) analyze how data decreases firm uncertainty about supply and demand, while Eeckhout and Veldkamp (2023) study the relation to markups. Jones and Tonetti (2020), Cong, Xie, and Zhang (2021) and Freeman, Yang, and Zhang (2023) model Big Data as knowledge, exploring the connection between information and innovation. Baley and Veldkamp (2025) provide a comprehensive review of the theoretical frameworks and empirical applications that analyze the use of data.[5] In this paper, I capture the specific use of information for advertising, endogeneize the adoption of Big Data technologies for targeting and analyze the welfare consequences of consumer data regulation.

Macroeconomic Impact of ICT. This paper relates to a strand of the literature that combines empirical and theoretical insights to understand the macroeconomic consequences of firm-level use of Information and Communication Technologies (ICT). Weiss (2020) analyzes firm-level decisions of intangible investment, while De Ridder (2024) explores the consequences of software on innovation and market power. Brand, Demirer, Finucane, and Kreps (2024) discuss the efficiency in the use of cloud computing. The link between IT usage and returns to scale is explored by Lashkari, Bauer, and Boussard (2024). Other research focuses on leisure technologies (Rachel (2024)) or the decrease in costs of expansion, either geographically (Jiang (2024)), or across multiple markets (Aghion et al. (2023)). In relation to AI, Babina, Fedyk, He, and Hodson (2024) and Acemoglu, Autor, Hazell, and Restrepo (2022) explore its consequences on product expansion and employment, respectively. This paper characterizes the aggregate impact of a technological change that reduces the cost of processing information about consumers for targeted advertising.

Targeted Advertising. A series of papers considers the aggregate impacts of targeted advertising. Greenwood, Ma, and Yorukoglu (2025) consider the interaction between online advertising and the provision of media goods, while Baslandze, Greenwood, Marto, and Moreira (2023) relate the increase in digital advertising to a rise in the number of varieties produced.[6] Tadelis et al. (2023) show that advertisers that engage in data collection update their marketing strategies. The closest paper in this literature is Cavenaile, Celik, Perla, and Roldan-Blanco (2023), who provide a rich dynamic model where consumers become aware of their existence when products age, and targeted advertising can influence this awareness process. While they focus on an equilibrium where targeted advertising is symmetric across firms, I characterize the use of targeted advertising across the firm distribution, and analyze its interaction with the use of consumer data using firm-level empirical evidence.

Targeted Search. To model the role of Big Data, I build on recent targeted search frameworks, where

[5] I focus here on the macroeconomic effects of the use of data by firms. Other research analyzes data intermediation (Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022), Bergemann, Bonatti, and Gan (2022)), network effects (Hagiu and Wright (2023) or Prüfer and Schottmüller (2021)), competition effects (Hagiu and Wright (2023)) or price discrimination in oligopolies (Fudenberg and Villas-Boas (2012), Kehoe, Larsen, and Pastorino (2020) or Ichihashi (2020)).

[6] While I focus here on targeted advertising, the macroeconomic impacts of product market frictions and advertising expenditure have been widely studied (see Gourio and Rudanko (2014) Cavenaile and Roldan-Blanco (2021), Cavenaile, Celik, Roldan-Blanco, and Tian (2025), Chiavari (2024), Shen (2025) or Afrouzi, Drenik, and Kim (2023)).

agents choose search probabilities in frictional markets subject to an entropy cost of information processing.[7] Cheremukhin, Restrepo-Echavarria, and Tutino (2020) analyze a marriage market in which both sides search, while Wu (2024) studies a labor market where firms post wages and workers process information about vacancies. While I abstract from some of the features in these papers, I adapt this framework to product markets in which firms search for consumers, incorporating an intensive production margin and introducing an endogenous Big Data technology choice that reduces the cost of processing information.

Outline. Section 2 discusses the theoretical framework and Section 3 presents the empirical evidence, which is used in Section 4 to quantify the model. The implications of technological change and policy counterfactuals are presented in Section 5 and Section 6 concludes.

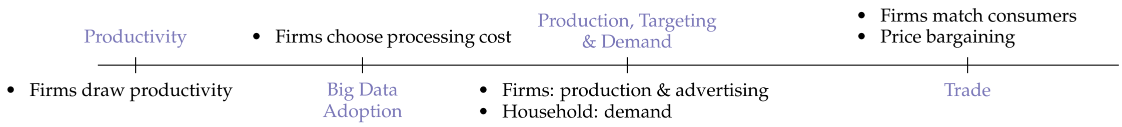
2 A Model of Data and Targeted Advertising

Environment. The model is static, with events occurring in different stages. There is one homogeneous good and there is a household composed of different types of consumers. Consumers with the same type share the same valuation for the homogeneous good: they can be considered as people with the same demographic characteristics (i.e. age, gender or municipality). The household supplies labor and exerts some search effort to buy goods.

There is a distribution of firms that choose how much to produce and decide a targeted advertising strategy, which characterizes how firms search for consumer types. Consumer types with high valuation are willing to pay more, but targeting them requires processing consumer data. Every firm is characterized by an exogenous productivity and an endogenous cost of processing data. Firms choose whether they would like to pay a fixed cost to adopt Big Data or stay with a traditional technology. Big Data provides a lower cost of processing information to target ads to specific consumer types.

Timing. Figure 1 characterizes the different stages of the model:

Figure 1: Model Timing



After productivity is realized, firms choose whether they would like to adopt Big Data technologies. Then, firms decide how much to produce and how to target their ads to different consumer types, and the Household supplies labor and decides her search effort. After ads are sent, firms match consumers and negotiate a price. Section 2.1 characterizes the decisions by the Household and describes how firms match and trade with different consumer types. Section 2.2 characterizes the production and advertising decisions of firms and Section 2.3 describes the technology choice of Big Data. Section 2.4 describes the competitive equilibrium, Section 2.5 presents the solution to this competitive equilibrium and Section 2.6 discusses the social planner problem.

[7] Entropy costs were first introduced to model rational inattention by Sims (2003). Building on this, Matějka and McKay (2015) establish the mapping between probabilistic choice and entropy costs, while Matějka and McKay (2012) apply the framework to product markets where consumers are rationally inattentive. Maćkowiak, Matějka, and Wiederholt (2023) provide a comprehensive review of rational inattention and its applications across fields. Lentz, Maibom, and Moen (2022) offer empirical evidence on the extent to which workers can target their search.

2.1 Household

Consumers. There is a household composed by a unit mass of consumers, which differ in the valuation for an homogeneous good denoted by ε. This valuation can take a continuum of values, and every consumer draws her realization from the distribution $g(\varepsilon)$. Since there is a unit mass of consumers, there is a measure $g(\varepsilon)$ with valuation ε. From now on, I characterize a consumer only by her realized valuation, and define a *consumer type* as the set of consumers with the same valuation. A consumer type can be interpreted as representing shoppers with a specific set of characteristics such as age, gender or location. As an example, we can consider the tourism sector offering a hotel room. Consumers with a strong taste for traveling would belong to a consumer type with a high valuation, while consumers that usually do not travel would belong to a low valuation type.

Product Markets. There are search frictions in product markets and consumers need to see advertisements to be able to buy. There is a separate market for each consumer type. I assume for the moment that a quantity of goods $a(\varepsilon)$ are posted to consumers of valuation ε, which will be characterized later as a result of firm decisions. The demand for goods from a consumer type has an extensive and an intensive margin. The extensive margin is given by the exogenous quantity of consumers $g(\varepsilon)$, while the intensive margin is characterized by the quantity of goods demanded per consumer, given by s. This is an endogenous choice of the household, as it will be described below.

Matching. Goods advertised to a consumer type will be matched one to one to the different shopping orders. The quantity of trades is given by the following matching function, which depends on the supply $a(\varepsilon)$ and the demand $sg(\varepsilon)$:

$$m\left(sg\left(\varepsilon\right), a\left(\varepsilon\right)\right) = \frac{sg\left(\varepsilon\right) a\left(\varepsilon\right)}{\left[\left(sg\left(\varepsilon\right)\right)^{\alpha} + a\left(\varepsilon\right)^{\alpha}\right]^{\frac{1}{\alpha}}} \tag{1}$$

Tightness, Purchase and Sale Probabilities. We can define the market tightness for consumers with valuation ε, denoted by $\theta(\varepsilon)$. It is defined from the perspective of buyers: it is high when there is a high amount of consumers' demand relative to the quantity of goods supplied:

$$\theta\left(\varepsilon\right) \equiv \frac{sg\left(\varepsilon\right)}{a\left(\varepsilon\right)} \tag{2}$$

As it is standard in the literature, I define $q(\varepsilon)$ as the probability that a specific shopping order from a consumer is matched with a good (or purchase probability), and $f(\varepsilon)$ as the probability that an advertised good is matched with a demand from a consumer (or sale probability). Since the matching function is homogeneous of degree 1, both probabilities can be written as follows:

$$q\left(\varepsilon\right) \equiv \frac{m\left(sg\left(\varepsilon\right), a\left(\varepsilon\right)\right)}{sg\left(\varepsilon\right)} = m\left(1, \theta\left(\varepsilon\right)^{-1}\right) = \left[1 + \theta\left(\varepsilon\right)^{\alpha}\right]^{-\frac{1}{\alpha}} \tag{3}$$

$$f\left(\varepsilon\right) \equiv \frac{m\left(sg\left(\varepsilon\right), a\left(\varepsilon\right)\right)}{a\left(\varepsilon\right)} = m\left(\theta\left(\varepsilon\right), 1\right) = \theta\left(\varepsilon\right) q\left(\varepsilon\right) = \theta\left(\varepsilon\right)\left[1 + \theta\left(\varepsilon\right)^{\alpha}\right]^{-\frac{1}{\alpha}} \tag{4}$$

While the purchase probability decreases in tightness, the sale probability is increasing in tightness. This matching function ensures that both the purchase and sale probabilities are between 0 and 1.[8]

[8] This is an advantage of this specification compared to a Cobb-Douglas. This CES matching function has been used to model search and matching in labor markets by Den Haan, Ramey, and Watson (2000), Hagedorn and Manovskii (2008) and Schaal and Taschereau-Dumouchel (2019), among others.

Consumption. The quantity consumed by a member of a specific consumer type is given by $c(\varepsilon)$:

$$c(\varepsilon) = sq(\varepsilon) \tag{5}$$

Consumption is given by the quantity of goods demanded per consumer times the purchase probability. If consumers of a specific type receive a high quantity of ads relative to their demand of goods, market tightness will be low. Since the purchase probability is decreasing in market tightness, a low market tightness implies a high purchase probability: shoppers of this type will find a good to buy with a high probability and consume more goods. Next, I define aggregate consumption C:

$$C \equiv \int_{\varepsilon} \varepsilon c(\varepsilon) g(\varepsilon) d\varepsilon \tag{6}$$

The total consumption aggregates by type the mass of consumers $g(\varepsilon)$, the quantity of goods consumed $c(\varepsilon)$ and their valuation ε.

Price by Consumer Type. Goods sold to a specific consumer type are traded at a price $p(\varepsilon)$, which results from a price negotiation under Nash Bargaining between firms and consumers. An additional good consumed in that market increases aggregate consumption by ε, providing additional utility to the representative household of $u'(C)\varepsilon$. The bargaining proceeds as follows: the firm obtains the price, both the firm and the consumer have a zero outside option and β captures the bargaining power of the firm. As shown in Appendix A.1, the outcome of the Nash Bargaining is the following market price, where Q captures the change in utility for the household after a change in income:

$$p(\varepsilon) = \frac{\beta u'(C)\varepsilon}{Q} \tag{7}$$

The model features price discrimination because the price depends on the type of consumer. Following the previous example, consumers with a strong taste for traveling would be offered a higher price for a hotel room. While price discrimination was difficult to implement in the past, recent technologies have increased the possibilities to segment consumers into different demographic types to offer personalized prices or promotions.[9]

Search Effort. In order to obtain a demand of goods of s, the representative household exerts a shopping effort of $\frac{l_h}{\zeta}$, where higher values of ζ increase the shopping effort needed to buy the same quantity of goods. The household chooses first the search effort exerted by consumers, and then every consumer draws her realized type. The household cannot direct the search for firms, since the quantity of goods demanded s is the same across consumer types.

Utility and Budget Constraint. The household chooses labor supplied to firms L and shopping effort l_h, both in labor units. Aggregate consumption gives utility $U(C)$ and both shopping effort and labor give linear disutility to the household. The household faces the following budget constraint:

$$\int_{\varepsilon} p(\varepsilon) c(\varepsilon) g(\varepsilon) d\varepsilon = wL + \Pi$$

On the left hand side, she buys $c(\varepsilon)$ units at price $p(\varepsilon)$ for the mass of consumers $g(\varepsilon)$, where the quantity consumed is given by Equation (5). On the income side, the earnings are composed by the wages earned from the labor supplied to firms and profits.

[9] A recent report by the US Federal Trade Commission explains how firms group consumers into different types according to demographic and behavioral data, offering them customized prices or different promotions depending on the consumer profile (see FTC (2025)). Hannak et al. (2014) found anecdotal evidence of price discrimination in different businesses, while a recent experiment by a US magazine found that hotel room prices depend on the origin of the travelers (see SFGATE, 2025).

Household Problem. The household chooses labor L and shopping effort l_h to maximize utility:

$$\max_{l_h, L} \quad U\left(\int_\varepsilon \varepsilon c(\varepsilon) g(\varepsilon)\, d\varepsilon\right) - l_h - L \tag{8}$$

$$\text{s.t.} \quad \int_\varepsilon p(\varepsilon)\, c(\varepsilon) g(\varepsilon)\, d\varepsilon = wL + \Pi; \quad s = \frac{l_h}{\zeta}$$

The household is subject to the aggregate budget constraint and to the definitions of the quantity of goods consumed and demanded in every market. The purchase probabilities are taken as given. Finally I denote the multiplier on the household constraint by Q, which captures the change in utility from a marginal increase in income.

To summarize, the household is composed of different demographic groups, where high-valuation types are willing to pay a higher price and consume more when the quantity of goods posted to them is higher. Income is earned from labor and profits, and a higher search effort increases consumption but has a utility cost. The household cannot direct the search for goods and exerts a general effort that increases consumption equally across types. Since Big Data is used mainly by firms and not by consumers, I focus on how it affects the search decisions of firms, which will be able to target specific consumer types.

2.2 Production and Targeted Advertising

There is a distribution of firms that differ in two characteristics: productivity z and cost of processing data λ, which can take two values (λ^D for Big Data firms and λ^T for traditional). Big Data technologies have a lower processing cost: $\lambda^D < \lambda^T$. Firms obtain an exogenous draw of productivity z, but the processing cost is endogenously chosen. For the moment, I analyze firm decisions taking as given the productivity and the processing cost, and I describe the choice of a data processing cost in Section 2.3.

Production Function. First, productivity z is drawn, and then firms choose production labor l_p, where they need to pay a wage w per employee. Firms produce output using labor and the following production function with decreasing returns to scale, where productivity follows a log-normal distribution:

$$y = e^z l_p^{1-\nu}; \quad z \sim N\left(0, \sigma_z^2\right); \quad 0 < 1 - \nu < 1;$$

Targeting Strategies. Firms can post one ad per good produced, which they can target to the different markets of consumers. Firms choose a set of targeting probabilities $t(\varepsilon)$ over the different consumer types. This captures the probability of sending an ad to a consumer with valuation ε. Every time a firm sends an ad, they draw a market from this distribution, which needs to be a probability distribution function:

$$\int_\varepsilon t(\varepsilon)\, d\varepsilon = 1$$

The fact that firms can target different consumer types is one of the key characteristics of recent practices in advertising. Tadelis et al. (2023) use an experiment with data from Facebook and Instagram to provide a detailed analysis of the returns to advertising. They provide a description of the prevailing targeting practices: *"(...) digital ads allow advertising businesses to target relevant consumers, and reach 'niche' markets at relatively low costs. For example, an advertiser could target a user audience consisting of females between the ages 30-40 who have clicked on sneaker ads on Facebook."* The ability to reach specific consumer types is precisely the mechanism that the targeting probabilities capture in the model.

Firm Price. The firm faces an average price $\tilde{p}$ across consumer types. Firms have 1 ad per good, which they can post to a consumer type.[10] For every good, they draw a consumer type from the targeting distribution $t(\varepsilon)$. The ad posted finds a match with the sale probability $f(\varepsilon)$, and in that case the traded price is equal to $p(\varepsilon)$. With a continuum of goods, a law of large numbers applies and the firm faces the following price with certainty:[11]

$$\tilde{p} = \int_{\varepsilon} t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon$$

Given that the price for a consumer type depends on the valuation in Equation (7), a targeting strategy that posts most ads to high valuation consumers would imply a higher price for the firm.

Targeting Cost. The set of targeting probabilities will be chosen, but firms incur in targeting costs l_i:

$$\underbrace{l_i}_{\text{Targeting Cost}} = \overbrace{\lambda}^{\text{Processing Cost}} \underbrace{\int_{\varepsilon} t(\varepsilon) \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon}_{\text{Targeting Data}} \tag{9}$$

Targeting costs, paid in labor units, increase with the quantity of data collected. When the firm chooses a random advertising strategy, the targeting probability is equal to the mass of consumers in a market $t(\varepsilon) = g(\varepsilon)$. In that case, there is no need to process information and targeting costs are zero. In order to send ads with a high precision to specific consumer types, firms need to process information: a targeting probability that deviates a lot from the quantity of consumers is very costly $(t(\varepsilon) >> g(\varepsilon))$. The total targeting data processed by the firm aggregates across consumer types the deviations of the targeting probabilities from random search.

A functional form that captures these intuitions is the relative entropy in Equation (9), which measures the distance between the targeting distribution and the mass of consumers in every market. This specification has been used in targeted or partially directed search in other frameworks: by Cheremukhin, Restrepo-Echavarria, and Tutino (2020) for marriage markets and by Wu (2024) for labor markets.[12] Galdon-Sanchez, Gil, and Uriz-Uharte (2025) provide empirical evidence of how more information about consumers improves targeting: in an experiment where firms are given new consumer data, firms shift their sales towards specific gender-age customer groups. This is in line with the above definition of the targeting costs, where a precise targeting strategy to certain consumer types requires more consumer data.

How does Big Data affect advertising? The first term in Equation (9) is the processing cost λ, which is chosen in a previous step by firms as explained in Section 2.3. Firms that have Big Data technologies have a lower processing cost and it is cheaper for them to process large quantities of information about consumer types. As discussed by Tadelis et al. (2023), today firms can process information about consumer demographics mainly using social media data. Section 3 uses an ICT survey to identify firms that use Big Data from social media, finding that it is used mainly for advertising: this model suggests that consumer data is used to target consumer types.

[10] For tractability, I fix the intensive margin of ads to focus on the distribution across consumer types. Depending on the parameters, and the equilibrium selling probabilities, some goods can remain unmatched. In the calibration in Section 4, 87% of goods are matched.

[11] As discussed by Veldkamp and Chung (2024), abstracting from uncertainty can be an appropriate assumption for models of data and matching, while uncertainty can be key for macro-finance frameworks as in Eeckhout and Veldkamp (2023).

[12] Wu (2024) follows Matějka and McKay (2015) to micro-found the relative entropy cost using a choice of information structures that provide information about an unknown state. Here I do not micro-found the signal structure underlying the firm targeting problem, but I take a reduced form approach where relative entropy captures the difference between the targeting distribution and the underlying distribution of agents.

Firm Problem. The firm chooses labor and a targeting distribution across markets to maximize profits:

$$\pi(z,\lambda) \equiv \max_{l_p,\{t(\varepsilon)\}_\varepsilon} \overbrace{e^z l_p^{1-\nu}}^{y} \overbrace{\int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon}^{\tilde{p}} - w\left[l_p + \lambda \int_\varepsilon t(\varepsilon) \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon\right] \tag{10}$$

$$\text{s.t.} \quad \int_\varepsilon t(\varepsilon)\, d\varepsilon = 1$$

Firms obtain revenue from the quantity of goods sold at the average price across types, and incur in both the production and the targeting cost. The constraint faced by the firm states that the targeting strategy is a proper distribution function over markets. So far, the processing cost λ has been taken as given, and the choices of production labor and targeting probabilities result in profits given by Equation (10). I describe next how firms face a technology choice in a previous stage, where they can select a data processing cost.

2.3 Technology

Big Data. Big Data is often defined as large volumes of information, coming from a wide variety of formats and updated at a high velocity. It can be used for different purposes, with recent research focusing on how Big Data improves innovation or advertising.[13] Here, I focus on the use of information for marketing, as a lower processing cost decreases the targeting cost of firms and affects their advertising decisions. Firms choose between two data processing costs: Big Data (λ^D) and traditional (λ^T). Choosing a Big Data technology ensures a lower processing cost $(\lambda^D < \lambda^T)$, but it involves the random adoption cost e^ψ, paid in labor units. It captures a variety of costs related to new information technologies, including hiring new workers or complying with regulation. The cost of adopting Big Data is stochastic, with ψ given by the following distribution:

$$\psi \sim f(\psi) \equiv N\left(\mu_\psi, \sigma_\psi^2\right)$$

Firms select the processing costs with the higher profits, trading off the benefits of a cheaper processing of consumer data with the adoption cost:

$$\lambda(z,\psi) = \max_{\lambda^T,\lambda^D}\left\{\pi\left(z,\lambda^D\right) - e^\psi, \pi\left(z,\lambda^T\right)\right\} \tag{11}$$

The endogenous technology choice generates a mass $\mu(z,\lambda^D)$ of firms with a given productivity and low processing costs, and a mass $\mu(z,\lambda^T)$ of firms with high processing costs. Firms that adopt Big Data with measure $\mu(z,\lambda^D)$ pay on average a cost in labor units denoted by $l_d(z,\lambda^D)$.

2.4 Competitive Equilibrium

This section defines the competitive equilibrium of this economy. I characterize a set of aggregate outcomes before, including the total quantity of ads received by consumers of a specific type and the aggregate values of profits and labor.

Ads by Consumer Type. For consumers of a specific type, the number of matches in Equation (1)

[13] Jones and Tonetti (2020) and Wu, Hitt, and Lou (2020) provide, respectively, a theoretical framework and empirical evidence about the use of Big Data for innovation. Tadelis et al. (2023) show that advertisers who engage in more data collection in Meta improve their advertising. Baslandze, Greenwood, Marto, and Moreira (2023) find that digital advertising is linked to an increase in the number of varieties produced.

increase in the demand and the supply of goods $a(\varepsilon)$:

$$a(\varepsilon) \equiv \int_z \sum_\lambda \overbrace{a(\varepsilon; z, \lambda)}^{\text{Ads from } z,\lambda} dz; \quad a(\varepsilon; z, \lambda) = \underbrace{\mu(z, \lambda)}_{\text{Firms}} \overbrace{y(z, \lambda)}^{\text{Goods}} \underbrace{t(\varepsilon; z, \lambda)}_{\text{Prob. to } \varepsilon} \tag{12}$$

The measure of ads in a market aggregates the amount of ads sent from different firm types, taking into account the mass of firms, the quantity of goods produced from a specific firm type and the probability that an ad is posted in market ε, which is given by the endogenous targeting probability $t(\varepsilon; z, \lambda)$ of firms with a given productivity and data processing cost.

Profits. Firm profits in Equation (10) were defined before subtracting the cost of adopting technology $l_d\left(z, \lambda^D\right)$. I define the profits net of the adopting cost as $\pi^N(z, \lambda)$, including the technology cost only for the firms that adopt Big Data technologies:

$$\pi^N\left(z, \lambda^D\right) = \pi(z, \lambda^D) - l_d\left(z, \lambda^D\right); \quad \pi^N\left(z, \lambda^T\right) = \pi(z, \lambda^T)$$

The household receives the profits Π, which aggregate the profits net of technology costs:

$$\Pi = \int_z \sum_\lambda \pi^N(z, \lambda)\, dz$$

Labor. Aggregate labor L is defined by aggregating the three types of labor hired by firms: production (l_p), targeting cost (l_i), and the adoption costs associated with firms that use Big Data (l_d):

$$L = \int_z \sum_\lambda \left[l_p(z, \lambda) + l_i(z, \lambda)\right] \mu(z, \lambda)\, dz + \int_z l_d\left(z, \lambda^D\right) \mu\left(z, \lambda^D\right) dz \tag{13}$$

Definition. The *competitive equilibrium* is given by i) a Household choice of search effort l_h and labor L ii) production labor l_p and targeting probabilities for a firm with z productivity and λ processing cost, over the different consumer types $\{t(\varepsilon)\}_\varepsilon$; iii) a firm distribution $\mu(z, \lambda)$ and iv) market tightness $\{\theta(\varepsilon)\}_\varepsilon$ such that:

1. The Household chooses search effort l_h and labor L to maximize utility in (8), taking as given the equilibrium tightness for every consumer type $\{\theta(\varepsilon)\}_\varepsilon$.

2. Firms choose technology, targeting probabilities, and production labor to maximize profits, taking as given equilibrium tightness for every consumer type $\{\theta(\varepsilon)\}_\varepsilon$:

   - Given productivity z and a data processing cost λ, firms choose employment l_p and a targeting strategy $\{t(\varepsilon)\}_\varepsilon$ across consumer types to maximize profits in (10).

   - At the stage of Big Data Adoption, firms choose a data processing cost to maximize profits in (11), which results in firm distribution $\mu(z, \lambda)$.

3. For every consumer type, tightness must be equal to the quantity of goods demanded by the Household divided by the equilibrium quantity of goods advertised $a(\varepsilon)$, which is given by Equation (12):

$$\theta(\varepsilon) = \frac{sg(\varepsilon)}{a(\varepsilon)}, \quad \forall \varepsilon$$

The competitive equilibrium is given by the amount of goods supplied and demanded across consumer types such that the Household exerts search effort and supplies labor optimally, firms choose production and targeting to maximize individual profits, and firms decide to adopt Big Data based on the expected profits of the technology given their productivity. In the competitive equilibrium, an individual firm is assumed to be small compared to the market, such that it cannot affect the aggregate outcomes for any consumer type. Therefore, firms take as given the market tightness and the sale probability across consumer types in the competitive equilibrium. This is the key difference compared to the problem of the Social Planner in Section 2.6, where the impact of firm decisions on aggregate outcomes is internalized.

2.5 Solution to the Competitive Equilibrium

Household Problem. The problem in Equation (8) can be rewritten as a choice of the quantity of goods demanded by consumer and the labor supply. Taking the first order condition with respect to the demand s:

$$u'(C)\frac{\partial C}{\partial s} = \zeta + QI$$

The left hand side shows that an increase in the quantity of goods demanded leads to higher consumption by the household, which provides additional utility. On the right hand side, more demand decreases utility from higher search effort. Since Q captures the utility value of income, the second term captures the disutility from the additional working hours associated with an increase in expenditure (I). I assume a log utility and normalize $Q = 1$, and from the labor supply of the household I obtain that this normalization sets the wage as the numeraire $(w = 1)$. The Appendix A.2 shows that the normalization is without loss of generality for both firm and household decisions, and shows the derivation of the household problem.

Household Choices in Competitive Equilibrium. The household chooses a search effort such that the following quantity of goods per consumer are demanded :

$$s = \frac{1}{\zeta + I}; \quad I = \int_{\varepsilon} p(\varepsilon)\, q(\varepsilon)\, g(\varepsilon)\, d\varepsilon \tag{14}$$

There is an infinitely elastic labor supply at the normalized wage of $w = 1$.

The household demand for goods is decreasing in the cost of exerting search effort (ζ) and in the income associated with additional consumption (I). This last term depends on how additional search effort by the household translates into more consumption. If more search effort has a high impact on consumption, this is associated with a large change in income (I is high), which decreases demand. The reason is that higher consumption also requires more income in equilibrium and the household needs to work additional hours, which decrease utility.

Firm Problem: Targeted Advertising. After setting the wage as the numeraire, the firm problem is given by:

$$\pi(z,\lambda) \equiv \max_{l_p, \{t(\varepsilon)\}_{\varepsilon}} e^z l_p^{1-\nu} \int_{\varepsilon} t(\varepsilon)\, f(\varepsilon)\, p(\varepsilon) d\varepsilon - \left[l_p + \lambda \int_{\varepsilon} t(\varepsilon) \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon\right] \tag{15}$$

$$\text{s.t.} \quad \int_{\varepsilon} t(\varepsilon)\, d\varepsilon = 1; \quad (\kappa)$$

Taking the first order condition with respect to targeting $t(\varepsilon)$, and setting it equal to zero:

$$yf(\varepsilon)\, p(\varepsilon) = \lambda \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) + \lambda + \kappa$$

The left hand side shows the marginal benefit of targeting. The benefit of targeting increases in the amount of goods produced y, the selling probability $f(\varepsilon)$ and the price for a given consumer type $p(\varepsilon)$. First, a large firm benefits more from a precise targeting strategy, as it has more goods to allocate. Second, if a specific type is targeted intensively by competitors, the selling probability for an individual firm will be lower. Finally, consumer types with a high valuation are more attractive to target, as they are willing to pay more for the good. The marginal benefit needs to be equal to the marginal cost on the right hand side, which increases in the processing cost λ: targeting high-valuation types will be cheaper for Big Data firms, compared to traditional ($\lambda^D < \lambda^T$). The cost of targeting increases in the quantity of data processed, which is higher when the firm deviates substantially from random advertising ($t(\varepsilon) >> g(\varepsilon)$): in order to target a specific demographic group, firms need to process information about them. There is a continuum of first order conditions, one for each of the consumer types ε, which are combined in Appendix A.3.

Targeted Advertising in Competitive Equilibrium. The targeting probability to market ε for a firm with productivity z and processing cost λ is given by the following expression:

$$t(\varepsilon) = \frac{g(\varepsilon)\exp\left(\frac{yf(\varepsilon)p(\varepsilon)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{yf(\varepsilon')p(\varepsilon')}{\lambda}\right)d\varepsilon'} \tag{16}$$

Production is given by $y = e^z l_p^{1-\nu}$, and the price in a specific market is given by: $p(\varepsilon) = \beta u'(C)\varepsilon$.

The targeting strategy takes a logit form, where the term inside the exponential function increases in the targeting benefit and decreases in the processing cost. Cheremukhin, Restrepo-Echavarria, and Tutino (2020) and Wu (2024) also obtain a logit form in their targeted and partially directed search frameworks. Compared to these two papers, my framework abstracts from some dimensions: targeting is not two-sided as in Cheremukhin, Restrepo-Echavarria, and Tutino (2020) and prices are set through Nash bargaining instead of a dynamic game as in Wu (2024). However, I use the same notion of targeted search to study search frictions in product markets where firms use advertising to target consumers, adding an intensive margin of production and an endogenous choice of processing costs.

Firm Problem: Employment. Taking the first order condition with respect to labor l_p in Equation (15), and setting it equal to zero:

$$\overbrace{\int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon}^{\tilde{p}} = \overbrace{\frac{l_p^\nu}{e^z(1-\nu)}}^{\text{MgC}}$$

The left-hand side shows the price faced by the firm, which depends on the targeting strategy, the selling probability, and the bargained price across consumer types. The right-hand side is the marginal cost, which increases in firm size due to the decreasing returns to scale and decreases in productivity. Firms choose the level of employment such that the price is equal to the marginal cost.

Labor in Competitive Equilibrium. The equilibrium labor for a firm with productivity z and processing cost λ is given by the following expression:

$$l_p = \left[e^z(1-\nu)\tilde{p}\right]^{\frac{1}{\nu}} \tag{17}$$

The firm price is given by:

$$\tilde{p} = \int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon$$

Production employment increases in productivity and in the average firm price across consumer types.

This price is determined as a function of the targeting probabilities, the sale probabilities and the market prices. To obtain the production and advertising decisions in Equations (16) and (17), I solve a fixed point problem on the targeting strategies across all consumer types. Appendix C.1 discusses in more detail the numerical solution of the model.

Big Data Adoption. When adopting Big Data λ^D, firms pay the adoption cost e^{ψ}, where ψ has the following normal distribution:

$$\psi \sim f(\psi) \equiv N\left(\mu_{\psi}, \sigma_{\psi}^2\right)$$

Firms choose Big Data when the profits of the lower processing cost are higher than the profits in the traditional case, taking into account the adoption cost:

$$\pi\left(z, \lambda^D\right) - e^{\psi} \geq \pi\left(z, \lambda^T\right); \quad \pi\left(z, \lambda^D\right) - \pi\left(z, \lambda^T\right) \geq e^{\psi}$$

Given the stochastic adoption shock e^{ψ}, and characterizing its cdf with F_{ψ}, I can obtain the share of firms for which the adoption cost is compensated by larger profits. Then, for firms with productivity z, there is the following share of adopters:

Big Data Adoption. The endogenous share of firms with productivity z that adopt Big Data technologies λ^D is given by:

$$\text{Adoption rate with } z \text{ productivity} = \frac{\mu\left(z, \lambda^D\right)}{\mu\left(z, \lambda^D\right) + \mu\left(z, \lambda^T\right)} = F_{\psi}\left(\pi\left(z, \lambda^D\right) - \pi\left(z, \lambda^T\right)\right) \tag{18}$$

F_{ψ} characterizes the cdf of technology adoption costs. Profits are defined as in Equation (10), under the optimal decisions of employment and targeting strategies.

Across the productivity distribution, adoption increases when the differential in profits for adopting Big Data λ^D is higher. When computing profits net of technology adoption costs, I redistribute the average adoption cost within adopting firms of the same productivity.[14]

2.6 Social Planner

Problem of the Social Planner. I focus on a social planner that cannot affect the patterns of technology adoption but can choose how many goods are produced and demanded, and how firms target consumers of different valuations. The social planner takes as given the adoption rates across productivities $\mu(z, \lambda)$ from the competitive equilibrium, and chooses goods demand s, employment in production for all firms $l_p(z, \lambda)$ and targeting across markets $t(\varepsilon; z, \lambda)$. The choice is made for all firms in the productivity (z) and data cost distribution (λ). The social planner obtains utility from consumption and incurs in utility costs for search effort and aggregate labor, including both production and targeting costs. The labor associated to the adoption costs is not part of the problem because the adoption costs are sunk. The only constraint is that the targeting probabilities across consumers add up to 1 for all firm types. Consumption can be expressed as a function of the selling probability and the amount of ads by type, given by Equation (12):

[14]The average adoption cost for firms of productivity z depends on the profit difference associated with technology adoption:

$$l_d^T\left(z, \lambda^D\right) = \int_0^{\pi\left(z, \lambda^D\right) - \pi\left(z, \lambda^T\right)} e^{\psi} f\left(e^{\psi}\right) de^{\psi}$$

.

$$\max_{s,\{l_p(z,\lambda),t(\varepsilon;z,\lambda)\}_{z,\lambda}} U\left(\int_\varepsilon \varepsilon a(\varepsilon) f(\varepsilon)\, d\varepsilon\right) - \zeta s - \left(\int_z \sum_\lambda \left[l_p(z,\lambda) + \lambda \int_\varepsilon t(\varepsilon;z,\lambda) \log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right) d\varepsilon\right] \mu(z,\lambda)\right)$$

$$\int_\varepsilon t(\varepsilon;z,\lambda)\, d\varepsilon = 1; \quad \forall (z,\lambda)$$

Matching Elasticity. Before solving for the optimal choices by the social planner, it is useful to define the following matching elasticity with respect to demand:

$$\phi(\varepsilon) \equiv \frac{\partial m(sg(\varepsilon), a(\varepsilon))}{\partial (sg(\varepsilon))} \frac{sg(\varepsilon)}{m(sg(\varepsilon), a(\varepsilon))} \tag{19}$$

It captures the percentage change in the number of matches for a consumer type, after a percentage change in the quantity of goods demanded. The quantity of goods demanded depends on the extensive margin of consumption, given by the distribution of consumers; and the intensive margin, given by the quantity of goods demanded per consumer.

Targeting Distortion. Next, I drop the firm states of productivity and data processing costs for simplicity, and solve for the optimal advertising strategy. The first order condition with respect to the targeting probability can be expressed as follows:

$$yf(\varepsilon)\, u'(C)\, \varepsilon - yf(\varepsilon)\, \phi(\varepsilon)\, u'(C)\, \varepsilon = \lambda \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) + V;$$

The left hand side shows the marginal benefit of the targeting probability, while the right hand side shows the marginal cost, which is similar to the competitive equilibrium with the addition of a constant V. The targeting benefit for the social planner is similar to the competitive equilibrium with two exceptions. First, prices are not part of the problem, since the social planner only allocates resources from firms to consumers. Second, there is an additional term in the social planner problem, which arises because the social planner internalizes the impact of targeting on the aggregate amount of ads received by a specific consumer type. If a firm sends more ads to a specific market, it reduces the sale probability $f(\varepsilon)$ for the rest of the firms. In the competitive case, this effect is not present because firms take as given the aggregate amount of ads and the market tightness. Therefore, when comparing the competitive equilibrium with the social planner trategy, there is a *targeting distortion.* As shown in Appendix A.4, the distortion depends on the matching elasticity as defined in Equation (19). Combining across consumer types, the choices of targeted advertising can be expressed as a logit form.

Social Planner: Targeted Advertising. The targeting probability chosen by the social planner is given by:

$$t(\varepsilon) = \frac{g(\varepsilon) \exp\left(\frac{yf(\varepsilon)(1-\phi(\varepsilon))u'(C)\varepsilon}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon') \exp\left(\frac{yf(\theta(\varepsilon'))(1-\phi(\varepsilon'))u'(C)\varepsilon'}{\lambda}\right) d\varepsilon'} \tag{20}$$

Production Distortion. The optimal choice with respect to employment compares the marginal benefit of increasing firm size on the left hand side with the marginal cost of production on the right hand side.

$$\int_\varepsilon t(\varepsilon) f(\varepsilon)\, \varepsilon u'(C)\, d\varepsilon - \int_\varepsilon t(\varepsilon) f(\varepsilon)\, \phi(\varepsilon)\, \varepsilon u'(C)\, d\varepsilon = \frac{l_p^\nu}{e^z (1-\nu)};$$

Increasing the number of employees has two effects working on opposite directions. The first term shows that more output increases consumption, which is distributed across consumer types according to the targeting and the selling probabilities. The second term captures the fact that increasing production rises the number of ads distributed across consumer types, reducing the probability to find consumers for the rest of the firms. As in the targeting choice, this term introduces a *production distortion* compared to the competitive equilibrium, which depends on the matching elasticity.

Social Planner: Labor. The social planner chooses the following quantity of labor:

$$l_p = \left[e^z (1 - \nu) \int_\varepsilon t(\varepsilon) f(\varepsilon) (1 - \phi(\varepsilon)) \varepsilon u'(C) d\varepsilon \right]^{\frac{1}{\nu}} \tag{21}$$

Equation (21) solves for the optimal choice of firm size for the social planner. The demand of goods and the rest of the derivations of the social planner problem are discussed in more detail in Appendix A.4.

Proposition 1. *The social planner solution is obtained in competitive equilibrium if the bargaining power of the firm varies by consumer type and is equal to the following expression, where $\phi(\varepsilon)$ is defined in Equation (19):*

$$\beta(\varepsilon) = 1 - \phi(\varepsilon)$$

The proof of Propostion 1 is shown in Appendix A.5. In general, the competitive equilibrium is inefficient because the bargaining power is the same across consumer types and is not a function of the matching elasticity. In the competitive equilibrium, firms do not internalize the targeting or the production externalities. When deciding how much to produce or how to target different consumer types, firms ignore the impact on the sale probability of competitors. The competitive equilibrium achieves the solution of the social planner when the bargaining power for consumers $(1 - \beta(\varepsilon))$ is equal to the matching elasticity. In this case, the value that firms extract from a match is equal to the social value that includes the crowding-out effects of targeting and production decisions.

For the competitive equilibrium to be efficient, the contribution by consumers to the matching process must be equal to the bargaining power. This is related to the Hosios (1990) condition, which is also part of the condition for efficiency in the targeted search framework by Cheremukhin, Restrepo-Echavarria, and Tutino (2020).[15] A recent literature has explored different externalities in the use of data.[16] Farboodi and Veldkamp (2024) propose a dynamic framework where data is used to improve quality, finding an inefficient level of trade only when data reduces the quality of competitors, and propose marketing as a mechanism behind this result. Here, I propose a targeted advertising framework where production and marketing decisions result in an inefficient amount of targeting data collected by firms. However, the direction of the inefficiency depends on model parameters, as characterized by Proposition 1.

Next, I present empirical evidence about the use of consumer data by firms, which will be used in Section 4 to calibrate the model to France in the late 2010s. Section 5 revisits the efficiency problem for the calibrated model and explores the need for policy intervention.

3 Firms and Big Data: Evidence from France

Firms and Big Data: Model and Empirics. The previous section described a model where Big Data gives firms an advantage in targeting consumers. To understand the aggregate consequences of Big Data use by firms, we would need to answer the following empirical questions. Which firms decide to

[15] The condition for efficiency in Cheremukhin, Restrepo-Echavarria, and Tutino (2020) is also based on the Hosios rule, with some differences related to the two-sided search of their model of marriage markets. In the partially directed search model by Wu (2024), the equilibrium is efficient under some conditions for the search cost, but the dynamic game that sets the terms of trade is different from the Nash bargaining assumed here. He discusses an extension of his model to Nash bargaining, finding also that the equilibrium is inefficient and that a constant bargaining power cannot resolve the inefficiency.

[16] Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022) suggest that the lack of concern for others' privacy leads to excessive data sharing, while Jones and Tonetti (2020) show that firms do not share enough data when it is used for innovation.

use Big Data for targeting? Are the improvements in firm performance quantitatively relevant? This section presents a set of empirical facts related to the use by firms of Big Data from social media. I merge balance-sheet information for French firms with an ICT survey on the use of Big Data from different sources. I obtain a sample of firms for which I observe: i) standard balance-sheet information such as revenue or employment, and ii) a dummy variable for the use of Big Data from social media. I use this variable to find the empirical counterpart of firms that adopt Big Data in the model. Section 3.1 describes the firm-level datasets used from France and characterizes the main patterns in adoption. Section 3.2 provides suggestive evidence of the use of social media data for marketing and Section 3.3 describes the connection with firm size. Section 3.4 provides the main implications for firm-level outcomes, while Section 3.5 provides a series of robustness checks.

3.1 Sources and Adoption of Social Media Data

Firm Technology Survey. The *Enquête sur les technologies de l'information et de la communication dans les entreprises* (Enterprise ICT Survey) is a business survey elaborated by the French Statistical Institute (INSEE). It asks firms about the use of different technologies with an annual frequency. The unit addressed is the legal unit in France (*siren*) of at least 10 employees in non-financial sectors. Half of the sample of firms is renewed every year, while the other half is maintained: the sample contains a repeated cross-section of firms, with some companies appearing more than once. The use of survey weights implies that all statistics and regressions are representative for firms above 10 employees in France. For the years of 2015, 2017 and 2019, firms are asked about the use of big data technologies, defined as information available in a large volume, in a wide variety of formats and generated at a high speed. Firms are asked about whether the big data analysis is done internally or externally, and about the use on the extensive margin of Big Data coming from different sources:

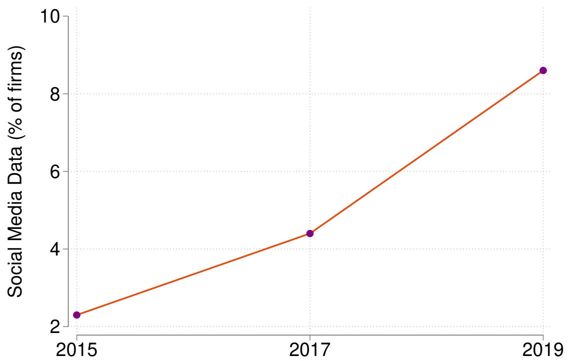
- *During the previous year, has your company analyzed big data from the following sources?*

    1. *The company's own data from intelligent or connected objects, or sensors.*
    2. *Geolocation data from portable devices*
    3. *Data generated by social media.*
    4. *Other sources of Big Data.*

To identify firms using Big Data from consumers, I create a dummy *Social Media Data*, taking the value of 1 if a firm analyzes big data internally and uses data generated from social media (3.). For comparison with other information sources, the variable *Other Sources* takes the value of 1 if a firm performs the analysis internally and uses data from objects (1.), geolocation (2.) or other sources (4.). While objects or sensors provide information about the firm, social media data can inform about consumer characteristics. For this reason, I use this variable as a proxy for the use of Big Data about consumers. I show below that firms using this type of information tend to use it for marketing, in line with the use of information for targeting in the model. Appendix B.2 provides more details on the Enterprise ICT Survey.

Balance-Sheet Data. INSEE also provides detailed information at the firm-level for French firms. The source used for balance-sheet data is FARE, which compiles data from fiscal declarations of firms in France and provides the same firm identifier that allows for the merge with the ICT Enterprise Survey. For the empirical analysis, the notion of firm is the legal unit in France *(siren)*, since it is the level at which the ICT Survey is conducted for the years of interest. From this source, I obtain various balance-sheet outcomes such as sales, employment, capital and costs. All financial variables are deflated to 2018 euros using sector-year price indexes from EU-KLEMS. Further details of the cleaning of the balance-sheet data and the variable definitions are provided in Appendix B.3.

Figure 2: Adoption of Social Media Data



(a) By Year

Sector	Social Media Data (% of firms)	Other Sources (% of firms)
IT & Telecom	14.5	16.1
R&D and Adv	13.6	15.6
Tourism	7.1	4.1
Business Services	7.1	13.7
Consulting	6.3	12.9
Utilities	5.9	21.8
Food	5.5	6.0
Trade	5.2	7.4
Chemicals	4.9	8.2
Household Services	4.7	9.6
Electronics	4.7	12.5
Textiles	3.4	10.3
Construction	3.3	12.4
Transport	3.1	27.4
Materials	2.7	7.2
Machinery	2.6	9.2

(b) By Sector

Note. Subfigure (a) plots the adoption of Big Data from social media for 2015, 2017 and 2019. Subfigure (b) shows the adoption rate across sectors of Big Data from social media and other sources (objects, geolocation and other sources). Appendix Table B.5 shows the detailed description of the 16 sectors based on 2-digit NACE codes. Source: FARE & ICT Enterprise Survey, INSEE.

Big Data Adoption. The sample includes approximately 20,000 firm-year observations over the years 2015, 2017 and 2019, for which I have information on balance-sheet outcomes and the use of social media data. Over the sample period, only 5.1% of firms in France are using data from social media. In line with the increasing availability of information in the economy, the left panel in Figure 6 shows that the adoption rate is increasing over time from 2.3% in 2015 to 8.6% in 2019. The right panel shows the uses of different sources of data by sector, pooling across all the years in the sample. The highest adoption rate of social media data is observed in *IT & Telecom*, where some firms may produce data-intensive services. While the specialization of firms in data production is an interesting avenue for research, I analyze instead the use of Big Data from consumers across all sectors.

Social media data is used more frequently by firms in services: the adoption rate varies from 7% to 13% for firms in Tourism, R&D and Advertising, or Business Services. The sectors with the lowest adoption rates are in manufacturing: less than 4% of firms use this technology in Transport, Machinery and Materials. Firms in those manufacturing sectors use other sources more often, which capture information generated by machines or by geolocation devices. Therefore, the use of Big data from consumers is growing over time, and the highest adoption rates are observed in services.

3.2 Social Media Data and Advertising

Data Uses. Apart from the type of data, firms are asked about whether they use information for marketing, for product development or for process improvement. I create a dummy *Marketing*, taking the value of 1 if a firms uses data to improve marketing. To understand the correlation between data sources and uses, I explore the following logit specification, regressing the probability to use data for marketing by firm i in sector s, on a logistic function F of a series of variables, including the uses of the two different types of data: social media and other sources.

$$P\left(Marketing_{is} = 1\right) = F\left(\alpha_s + \beta_0 \text{Social Media Data}_{is} + \beta_1 \text{Other Sources}_{is} + \beta_2 k_{is} + \beta_3 \text{Age}_{is} + \gamma' \mathbf{\Theta_{is}} + \epsilon_{is}\right)$$

I include sector fixed effects α_s, capital k_{is}, firm age and a set of controls for other technologies $\mathbf{\Theta_{is}}$, which include the use of a website, the access to high speed internet and an e-commerce dummy that captures whether the firm sells part of the output online.[17] Given the constraints on the availability of the variables related to the use of data, the sample is restricted to the firms that use any type of information in 2015. Firms in the survey can use social media data but also other sources at the same time: the coefficient on social media captures the effect of using this type of information, conditional on using other sources. Table 1 shows the marginal effects for both social media data and other sources.

Table 1: Data Sources and Use for Marketing

	(1) Marketing	(2) Marketing	(3) Marketing
Social Media Data	0.361***	0.344***	0.294***
	(0.04)	(0.04)	(0.05)
Other Sources	0.089*	0.062	0.060
	(0.05)	(0.05)	(0.05)
Observations	1012	1012	994
Sector FE	Yes	Yes	Yes
Capital & Age	No	Yes	Yes
Other Technologies	No	No	Yes
Estimate	Logit	Logit	Logit

Note. This table shows the marginal effect of the dummies on social media data and other sources in the logit specification. The sample is restricted to firms that use any type of Big Data in 2015 (social media or other sources), due to the survey design. The use of other technologies includes dummies on the use of high speed internet, if the firm has a website or sells using e-commerce. Standard errors in parentheses: $*(p < 0.10)$, $**(p < 0.05)$, $***(p < 0.01)$. Source: FARE & ICT Enterprise Survey, INSEE.

Fact 1: Use of Social Media Data is Associated with Marketing Improvement. Column (3) shows the preferred specification with all the controls. The use of social media data increases the probability to use data for marketing by 29.4%, while the use of other sources is not correlated with advertising. The first two columns show that the effect is larger before including sector or firm-level controls, suggesting that part of the effect is captured by those variables. However, the effect is significant at the 99% confidence level in the preferred specification. This provides empirical evidence of a key mechanism in the model, where social media data is used to improve marketing through precise consumer targeting. In addition, this finding is in line with the use of social media data from Facebook and Instagram described in Tadelis et al. (2023), where data on consumer demographics from the platforms improves advertising strategies by targeting specific consumer types.

Appendix Table B.3 shows the correlation of social media and other sources with the three uses: marketing, process improvement or product development. The use of other sources is correlated with innovation, either for process improvement or for product development. As shown in Figure 6, this type of information, related to the production processes, is used more often by firms in manufacturing. The connection between data and innovation is an interesting avenue for future research, and it has recently been explored from an empirical (Wu, Hitt, and Lou, 2020) and a theoretical perspective (Jones and Tonetti (2020), Cong, Xie, and Zhang (2021)). The association between digital advertising and product development is explored by Baslandze, Greenwood, Marto, and Moreira (2023), and Table B.3 finds evidence of this mechanism as social media data is also correlated with product development. In the model, I provide a different mechanism where the use of consumer data improves advertising, which is supported by the empirical evidence from the technology survey.

[17] High speed internet takes the value of 1 if the firm has access to internet with a speed above 100 Mb/s, and the e-commerce variable takes the value of 1 the firm sells the output online through the own website or using external platforms.

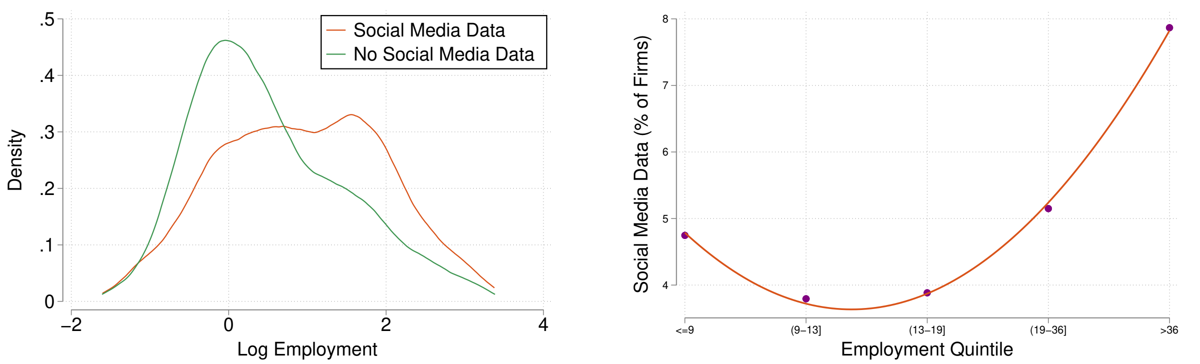
3.3 Social Media Data and Firm Size

Fact 2: Firm Size is Positively Correlated with the Use of Social Media Data. Figure 3a shows the employment distribution separately for firms that use social media data and for firms that do not use it, pooling across the sample years. The value of employment is residualized of sector-year fixed effects, capital, age and a vector of other technology controls, which include the use of a website, the access to high speed internet and the use of e-commerce. On average, Figure 3a shows that firms that use social media data tend to be larger. Figure 3b shows the use of social media data over the employment quintiles in the sample. With the exception of the first size quintile, the adoption is increasing in firm size: around 4% of firms below 20 employees use social media data, but it increases to around 8% for firms in the top quintile of the size distribution. While the positive correlation of data use with firm size is a consistent empirical finding in the literature, I document that this fact extends to the use of information from social media.[18]

Figure 3b will be key to calibrate the theoretical framework of data use for targeted advertising in Section 4. The average adoption rate of 5.1% and the differences in the use of social media data across employment quintiles will be targeted, while the slope of the relationship between size and adoption will be an untargeted moment used to validate the main mechanisms of the model. The heterogeneity in adoption across the size distributions and the implications for firm outcomes will be key to characterize the aggregate consequences of the use of social media data.

Figure 3: Social Media Data and Firm Size

(a) Size Distribution (b) Size Quintiles



Note. Subfigure (a) shows the distribution of the log of employment separately for firms using social media data and firms that do not adopt. The value of employment is first residualized of sector-year fixed effects, capital, age and other technology controls, and the variable is trimmed at the top and bottom 1%. Subfigure (b) shows the adoption rate across different size quintiles. Source: FARE & ICT Enterprise Survey, INSEE.

The use of social media data has no clear correlation with firm age. Appendix Table B.1 shows that the adoption rates are similar for firms with different ages: the adoption rate is approximately 5% for firms below and above 15 years. The use by employment quintile varies by larger magnitudes, from 4% to 8%. In addition the positive correlation between social media data and size holds for both young and old firms separately, as shown in Appendix Table B.2. Therefore, age does not seem to be a relevant factor driving the use of social media data, which suggests that abstracting from the firm life-cycle in the model in Section 2 can improve tractability without missing the main mechanisms at work.

[18] Among many others, Brynjolfsson and McElheran (2016b) show that adoption of data-driven decision making increases in firm size for US manufacturing establishments.

3.4 Social Media Data: High Productivity and Overhead Costs

Empirical Specification. By merging the Enterprise ICT Survey with balance-sheet information from FARE, I can analyze firm-level outcomes and the connection to the use of social media. With this purpose, I explore the following fixed-effect regression:

$$y_{ist} = \alpha_{st} + \beta_0 \text{Social Media Data}_{ist} + \beta_1 k_{ist} + \beta_2 \text{age}_{ist} + \gamma' \boldsymbol{\Theta}_{\mathbf{ist}} + \epsilon_{ist} \tag{22}$$

On the dependent variable, y_{ist} denotes a balance-sheet outcome of firm i, in sector s at year t, where the sectors are defined as in Figure 2b. The independent variable is a dummy that takes the value of 1 if a firm uses social media data in a specific period. The previous section documented substantial sector heterogeneity in adoption, which suggests that sector-year fixed effects (α_{st}) should be introduced in the regression. To capture both the life-cycle and the scale of the firm, I control for age and for the value of capital k_{ist}. Additionally, I include the technology controls $\boldsymbol{\Theta}_{\mathbf{ist}}$ of website use, the access to high speed internet and the use of e-commerce.

The cross-sectional comparison of firms is the preferred specification for the baseline empirical facts because it allows for a clean comparison with the static model in Section 2. In addition, the firm life-cycle does not seem to be correlated with the adoption of social media data, as discussed in Appendix Table B.1. This empirical strategy compares firms from the same sector and year, with similar size, age and use of other technologies, where some use social media data and others do not. This specification does not control for unobservable characteristics at the firm level that can be correlated with both social media and firm-level outcomes. Since adoption can be endogenous to those unobservable variables, the coefficient on the use of social media data should not be interpreted as the treatment effect of an exogenous assignment to the use of social media data. Through the lens of the model, unobservable drivers of firm-performance are captured by the exogenous productivity. In addition, the adoption of social media is endogenous to this productivity in the targeted advertising framework, which suggests that the current empirical specification can be informative about the model outcomes, as social media use is endogenous in both the data and the model. Section 3.5 provides an alternative specification that analyzes the within-firm dynamic impact of social media data for a subsample of firms.

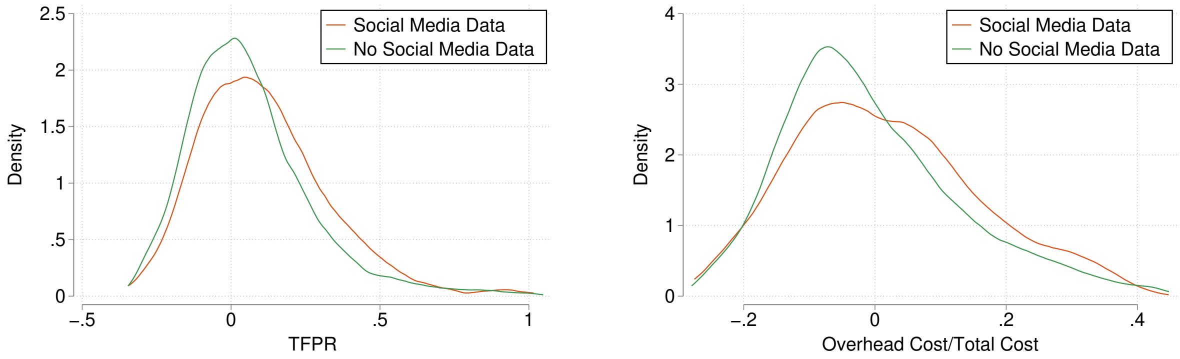
Fact 3: Firms with Social Media Data are More Productive. If Big Data provides an advantage in targeted advertising, we should observe that firms that use this technology have better performance. Here, I focus on Total Factor Revenue Productivity (TFPR), an estimate of the residual of revenue on employment and capital. There is a long literature addressing estimation issues in TFPR, including the bias driven by the fact that productivity is observed for the firm but not for the econometrician. I follow a standard production function estimation methodology from Ackerberg, Caves, and Frazer (2015), who updated the seminal work by Levinsohn and Petrin (2003). The parameters of the production function are estimated separately by sector, and for this estimation I extend the balance-sheet sample to the entire decade of the 2010s. Further details are provided in Appendix B.4.

Figure 4a shows the distribution of TFPR for firms that use social media data and firms that do not use it, residualized of the same set of controls as in Equation (22). The distribution of TFPR is shifted to the right for firms that use social media data. Table 2 shows the results of estimating the baseline empirical specification in Equation (22), comparing TFPR for firms with and without social media data. The preferred specification in column (2) shows that TFPR is 4.5% higher for firms that use social media data, compared to firms in the same sector with similar characteristics. This result is in line with previous results in the literature on the impact of Big Data, extending the existing findings to the use of information from social media.[19]

[19]Brynjolfsson and McElheran (2016a) and Galdon-Sanchez, Gil, and Uriz-Uharte (2025) show that revenue increases from 4-

Figure 4: Social Media Data in France: Productivity and Costs

(a) Total Factor Productivity Revenue (TFPR) (b) Share of Overhead Costs



Note. Subfigure (a) shows the distribution of TFPR from the ACF estimation, while subfigure (b) shows the distribution of the share of overhead costs over total costs. Both plots show the distribution separately for firms using social media data and firms that do not adopt; I residualize from sector-year fixed effects, capital, age and other technology controls; and the variable is trimmed at the top and bottom 1%. Source: FARE & ICT Enterprise Survey, INSEE.

TFPR can be higher for firms with social media data for two reasons. First, social media data can allow firms to improve targeted advertising to sell their products on high valuation markets. This would allow firms to set higher prices and increase their revenues. Second, firms with ex-ante higher productivity can endogenously decide to use social media data with a higher probability. Both mechanisms are present in the model, where the adoption patterns of social media data and the cross-sectional variation in TFPR will be key to quantify the advertising advantage driven by Big Data in Section 4. In addition, Section 4.2 will use the model outcomes to understand whether the observed differences in TFPR are driven by prices or by productivity.

Table 2: Social Media Data and Total Factor Productivity

	(1)	(2)	(3)	(4)
	TFPR	TFPR	$\frac{c^F}{c^T}$	$\frac{c^F}{c^T}$
Social Media Data	0.065***	0.045***	0.039***	0.028***
	(0.02)	(0.01)	(0.01)	(0.01)
R^2	0.545	0.549	0.344	0.354
Observations	21709	20911	21740	20940
Sector × Year FE	Yes	Yes	Yes	Yes
Capital & Age	Yes	Yes	Yes	Yes
Other Technologies	No	Yes	No	Yes
Estimate	OLS	OLS	OLS	OLS

Note. This table shows the results of the empirical specification in Equation (22). Columns (1) and (2) show the results for the TFPR estimation following Ackerberg, Caves, and Frazer (2015), while columns (3) and (4) show the share of overhead costs to total costs (overhead plus variable), taking values from 0 to 1. The use of other technologies includes dummies on the use of high speed internet, if the firm has a website or sells using e-commerce. Standard errors in parentheses: $*(p < 0.10), **(p < 0.05), ***(p < 0.01)$. Source: FARE & ICT Enterprise Survey, INSEE.

Fact 4: Firms with Social Media Data Show Higher Overhead Costs. Since social media data tends to be used for marketing, it can lead to changes in advertising costs for French firms. I follow the literature that uses overhead costs to capture advertising expenditures.[20] Overhead costs are measured using the

9%. The marketing literature documents that firms in specific settings increase revenues when they obtain data about consumers: Anderson, Chintagunta, Kaul, and Vilcassim (2020), Berman and Israeli (2022), Jin and Sun (2019).

[20] Among many others, Chiavari (2024) and Afrouzi, Drenik, and Kim (2023) use the Selling, General and Administrative (SGA) expenses in Compustat to capture the marketing expenditures in US firms. Further details on the variable construction are provided in Appendix B.3.

Other Operating Expenses variable available in the balance-sheet of firms. I focus on the ratio of overhead costs (c^F) to total costs (c^T), which takes values from 0 to 1. Total costs include the sum of overhead costs and *Cost of Goods Sold (COGS)*, which contains the expenditures in materials and wages.

Figure 4b shows the distribution of the share of Overhead costs for firms that use social media data and for firms that do not adopt, residualized of the same set of controls as in Equation (22). The distribution of the overhead cost share is shifted to the right for firms using social media data. Column (4) in Table 2 uses the baseline empirical specification to quantify the impact of social media data on the overhead share. Firms that use social media data have a ratio of fixed to total cost that is 2.8 percentage points higher. A reasonable concern with this finding would be that the use of social media captures the use of general information technologies, which also tend to increase overhead costs.[21] However, the results are robust to the use of other technologies as controls, which include high internet speed, the use of a website and e-commerce. Therefore, there is empirical evidence of higher expenditure in overhead costs for firms using social media data, conditional on the use of other information technologies. I will return to this result on the cost structure of firms in Section 4.2.

3.5 Robustness

Profits and Market Shares. Fact 3 characterizes the higher Total Factor Revenue Productivity for firms using social media data. Apart from TFPR, Appendix B.5 explores the correlation of social media data with profits and employment market shares at a disaggregated 4-digit NACE sector. Using a similar empirical specification as in Equation (22), I find that profits are 19% higher while the employment market share at the 4-digit sector are 25% higher for firms using social media data. I focus on TFPR because it allows for a clean calibration of the model, but there is evidence that other measures of firm performance and profitability are positively correlated with the use of social media data.

Within-Firm Impact. To study the impact of social media data, the empirical specification in Equation (22) performs a cross-sectional comparison of firms in the same sector, with similar size, age and use of other technologies. Alternatively, I explore here the within-firm variation in balance-sheet outcomes before and after starting to use social media data. For this analysis, I restrict the sample to the set of firms that are surveyed more than once, and also report switching from not using social media data to using it. I also include the firms that never adopt as controls. I estimate the following regression, where β^h captures the impact on a balance-sheet outcome h periods ahead.

$$y_{i,s,t+h} = \gamma_i + \alpha_{st} + \beta^h \mathcal{I}\left\{\text{Switch to Social Media Data at t}\right\} + \beta_1 k_{ist} + \gamma' \mathbf{\Theta_{ist}} + \epsilon_{i,s,t+h};$$

Apart from capital, the technology controls and the sector-year fixed effect, this specification allows for a firm fixed effect γ_i. Switching from not using to adopting social media data has a positive and significant impact on Total Factor Productivity Revenue. However, the impact is lower in magnitude than in the cross-sectional regression from Table 2: in the within-firm specification, TFPR increases by 2.9% one period after switching to social media, while TFPR is 4.5% higher for adopters in the cross-sectional comparison. For the overhead cost share, the impact two and three years ahead seems to be positive but not significant. For the model calibration, I select the cross-sectional results, which are the closest to the model. Appendix B.6 provides further details about these results.

[21] De Ridder (2024) finds that overhead costs are correlated with higher software expenditure in both France and the US.

4 Big Data, Firm Size and Productivity

Section 4.1 uses the previous empirical findings to calibrate the model of data and targeted advertising to France in the late 2010s. Facts 1 and 2 showed that Big Data is used for advertising and it is positively correlated with firm size. Section 4.2 revisits these empirical findings explaining the key mechanisms of the model. Facts 3 and 4 showed that Big Data is correlated with higher productivity and with higher overhead costs. Section 4.3 analyzes the implications of the use of data and targeted advertising for productivity and the cost structure of firms.

4.1 Calibration

Model Solution. To solve the model numerically, I discretize the number of markets, each characterized by a valuation ε. Valuations are assumed to be uniform between 0 and 1, such that all markets are equally likely ex-ante with a large continuum of consumers $(g(\varepsilon) = g(\varepsilon'))$. Firms are characterized by their productivity and their processing cost λ. Given these two variables, they choose both targeting over different markets in Equation (16) and employment in Equation (17). The profit difference associated with lower information processing costs allows to compute the schedule of Big Data adoption as a function of productivity in Equation (18). To obtain the solution, I guess for a level of aggregate consumption and a vector of tightness for all markets, compute the equilibrium and evaluate the equilibrium consumption and tightness. Then, I iterate both values until convergence. Appendix C.1 describes in more detail the steps to obtain the numerical solution of the model.

Parameters. I calibrate the model to firms with more than 10 employees in France in the late 2010s, based on the empirical evidence in Section 3. There are 9 parameters to calibrate: the span of control in the production function $(1-\nu)$, the cost of shopping effort for the household ζ, the matching function parameter α, the dispersion of productivity σ_z, the bargaining power of firms β, the data processing cost for Big Data λ^D and traditional firms λ^T, the average cost of adopting Big Data μ_ψ and its dispersion σ_ψ. First, the ratio between revenue and variable cost is a direct function of the returns to scale in the model.[22] I set the returns to scale to 0.59, which matches the observed average ratio of 1.69 for the sample of French firms.

Simulated Method of Moments. I use the Simulated Method of Moments to calibrate internally the remaining 8 parameters. I search for the parameter vector Θ that minimizes the following loss function, where m_m is the vector of moments in the model and m_d is the vector of empirical moments:

$$\mathcal{L}(\Theta) = \min_{\Theta} \left(\frac{m_m(\Theta)}{m_d} - 1 \right)' W \left(\frac{m_m(\theta)}{m_d} - 1 \right); \quad \Theta = \left(\zeta, \alpha.\sigma_z, \beta, \lambda^T, \frac{\lambda^D}{\lambda^T}, \mu_\psi, \sigma_\psi \right)$$

I minimize the distance between the vector of empirical moments and the vector of model moments, with the weighting matrix W.[23] While most parameters move several moments, I provide below a summary of the targets that are more relevant for each parameter, while a more detailed description is provided in Appendix C.2.

Household Moments and Productivity. To match the cost of shopping effort (ζ), I target the time spent shopping relative to the time at work in France. I use additional data from the Harmonized European Time Use Survey (HETUS) to obtain a ratio of shopping time to total labor time of 6.4%, a similar value to Gourio and Rudanko (2014). The parameter on the matching function (α) relates to

[22] See Appendix C.2 for more details.

[23] W is a diagonal matrix except for the moments associated with parameters σ_ψ and β, with a weight of 2, and the moments associated with parameters σ_z and μ_ψ, with a weight of 3. These parameters are key to match the adoption rate of Big Data, its impact on TFPR and the size distribution of firms.

the complementarity of ads and the shopping effort to generate trades in product markets. I combine data from e-commerce sales from a French business association (FEVAD (2024)), and compare it with information about digital advertising from Statista. By comparing different points in time, I obtain that the elasticity of online sales to digital advertising is equal to 0.62. While there is no time variation in the model, I obtain a similar elasticity of sales to advertising by comparing markets with different valuations. To match the productivity dispersion (σ_z), I obtain the employment share of the top 10% firms in the sample, which is equal to 67%. This value is in line with the literature: the same statistic for the US is equal to 65% in Morazzoni and Sy (2022).

TFPR and Social Media Data. The bargaining power of firms is captured by β, which can be related to the increase in revenue from using Big Data. If β is very high, firms that target high valuation consumers can extract a high surplus. To capture the increase in revenue, I turn to Fact 3 from the previous section, which documented that using Big Data is correlated with an increase in TFPR of 4.5%. I explore a similar regression in the model. First, I define TFPR as the residual of revenue on employment:[24]

$$TFPR \equiv e^z \tilde{p}; \quad \tilde{p} = \int_{\varepsilon} t(\varepsilon) f(\varepsilon) p(\varepsilon) \tag{23}$$

TFPR depends both on the exogenous productivity and the endogenous price faced by the firm. The following specification projects a model outcome x on a dummy that captures whether the firm uses Big Data, controlling for the size quintile.

$$\log(x(z,\lambda)) = \alpha_0 + \gamma \mathcal{I}\left[\lambda = \lambda^D\right] + \sum_{k=1}^{5} \alpha_k \mathcal{I}\{\text{Quintile k}\} + \epsilon_{z,\lambda} \tag{24}$$

I set the dependent variable to the log of TFPR, and I compare the coefficient γ in this regression with the coefficient from Table 2 that captured the impact of social media data on TFPR for French firms.

Data Processing Costs. Two parameters are related to the processing cost for Big Data firms (λ^D) and for non-adopters (λ^T). To capture the advantage in information processing costs for Big Data $\left(\frac{\lambda^D}{\lambda^T}\right)$, I use the share of advertising through digital media in France, which according to Statista is equal to 46%. For the processing cost of non-adopters, I use the share of workers in advertising in the French economy. To obtain this moment, I use the measures of exposure to digital advertising by occupation, as provided by Prytkova et al. (2024). By combining these measures with the occupation employment shares from the Labor Force Survey for France and the digital advertising share in the economy, I estimate that 1.9% of the workforce is allocated to advertising. A similar value is obtained by Cavenaile and Roldan-Blanco (2021) for the US, where advertising activities comprise 2.2% of GDP.

Big Data Adoption. I use the findings from Section 3.3 to calibrate the remaining parameters, which are related to the first and second moment of the adoption cost of Big Data. To match the average adoption cost (μ_ψ), I use the adoption rate of Social Media data across all years (5.1%). The volatility of the stochastic adoption cost (σ_ψ) is calibrated using changes in adoption across the firm-size distribution: if the benefits of adoption vary with firm size, a cost with low volatility will create a sharp threshold in adoption, while high volatility will be associated with a flat adoption rate across the size distribution. I use the values from Figure 3b to compute the standard deviation of the adoption rate across different size quintiles, which is equal to 1.5.

Estimation. Table 3 summarizes the set of parameters and targets and compares the moments in the model with the empirical counterparts. The span of control is set externally but the remaining parameters are estimated by SMM. Both the employment share of the top decile of firms and the increase in

[24] TFPR in the model is analogous to its empirical counterpart ω_{it} in Appendix B.4: $\log(y\tilde{p}) = (1-\nu)\, l_p + \log(\text{TFPR})$

Table 3: Calibration

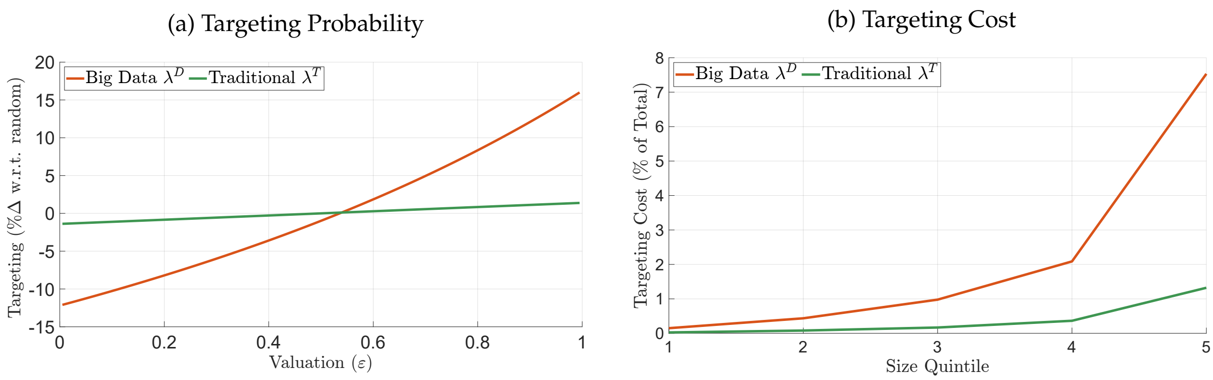
Parameter	Value	Description	Target	Data	Model	Source
$1-\nu$	0.59	Span of Control	Average Revenue/Variable Cost	1.69	1.69	FARE & ICT Survey
ζ	0.013	Shopping Cost	Shopping Time / Labor Time	6.43%	3.61%	HETUS
α	1.86	Matching Function	Elasticity of Sales to Advertising	0.62	0.75	Statista & FEVAD
σ_z	0.65	Productivity Dispersion	Employment Share Top 10 % Firms	68%	50%	FARE & ICT Survey
β	0.98	Firm Bargaining Power	%Δ in TFPR by Social Media	4.5%	1.1%	FARE & ICT Survey
λ^T	64.1	Processing Cost (T)	Advertising Employment Share	1.9%	2.3%	Prytkova et al. (2024) & LFS
$\frac{\lambda^D}{\lambda^T}$	0.16	Reduction in Processing Cost (D)	Digital Advertising Share	46%	50%	Statista
μ_ψ	24.9	Average Adoption Cost (D)	Adoption Rate of Social Media	5.1%	6%	FARE & ICT Survey
σ_ψ	19.9	Dispersion Adoption Cost (D)	Std. Dev. of Adoption per Quintile	1.5	2.3	FARE & ICT Survey

TFPR from social media are slightly underestimated, but the rest of the moments are close to the empirical counterparts, including the adoption of social media data and the key patterns in advertising. The next section explores the competitive equilibrium of the model under this calibration, and connects it to the key empirical facts about the use of Big Data by firms in France.

4.2 Big Data, Targeted Advertising and Firm Size

Fact 1: Use of Social Media Data is Associated with Marketing Improvement. Figure 5a shows the targeting probabilities for Big Data and traditional firms. The targeting probabilities are shown as a deviation from random advertising, where ads are sent with a probability equal to market size $(t(\varepsilon) = g(\varepsilon))$.

Figure 5: Targeted Advertising, Technology and Firm Size



Note. Subfigure (a) shows the equilibrium targeting probabilities from Equation (16), expressed as a deviation from random targeting: $t(\varepsilon)/g(\varepsilon)$. I group firms according to their technology, and I average the targeting probabilities separately for Big Data and traditional firms. Subfigure (b) plots the average targeting cost from Equation (9) for Big Data and traditional firms for different size quintiles, expressed as a % of total costs that include production, targeting and adoption costs.

All firms target high valuation consumer types with a higher probability. Since these consumers have a higher willingness to pay, firms obtain a higher price if high valuation types are targeted more intensively. However, there is heterogeneity in the advertising strategies: the most valuable consumer types are targeted with a higher probability by Big Data firms than by traditional firms. The reason is that a lower data processing cost allows them to process more targeting data and design a more precise strategy, compared with traditional firms. The targeted advertising strategies relate to Fact 1, which showed that firms using Big data coming from social media tend to use it for marketing. The model proposes a mechanism in which consumer data is used for advertising: firms using Big Data have a lower cost of processing information, and endogenously choose to obtain more consumer data to target high valuation consumers. Recent research provides empirical evidence supporting the hypothesis that data availability improves targeting.[25]

Figure 5b shows targeting costs for firms with both technologies across the firm-size distribution. Targeting costs depend on both the cost of processing information and the quantity of consumer data processed. In the current calibration, Big Data firms have a cost of processing information that is 16% lower than the cost for traditional firms. However, the quantity of consumer data processed is much higher and Big Data firms have higher targeting costs for all size quintiles. Comparing across size quintiles, firms at the top of the size distribution spend more resources in targeting, conditional on a given technology. To explain the connection with firm-size, I return below to the second empirical fact.

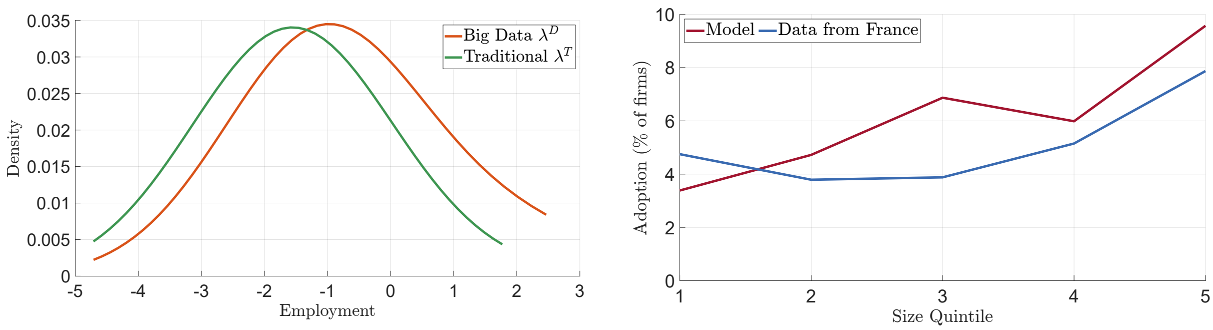
Targeted Advertising and Consumer Valuation. Firm decisions about targeted advertising have consequences for consumers. Appendix Figure C.1 plots the share of ads posted by Big Data firms for every consumer type, showing that high-valuation consumers are more likely to see ads from Big Data firms. There is substantial heterogeneity across consumer valuation: low-valuation types receive around 7% of the ads from Big Data firms, while high-valuation types receive around 14%. Consumer types with high valuation and high willingness to pay can reflect either specific demographic groups or high income. The definition of aggregate consumption in Equation (6) suggests that the preferred interpretation is the first, since high-valuation types contribute more to aggregate household utility, compared to low-valuation types. Galdon-Sanchez, Gil, and Uriz-Uharte (2025) provide empirical evidence in line with this hypothesis, as firms with more data shift their sales towards specific demographic groups. However, Arvai and Mann (2025) show that high-income households consume a higher share of digital goods: disentangling whether data allows for targeting on taste or income would be an interesting avenue for future research. I return to the implications of Big Data use for consumption in Section 5.

Fact 2: Firm Size is Positively Correlated with the Use of Social Media Data. Figure 6a plots the distribution of the log of employment in the model, which shows a similar pattern than Figure 3a for French firms: Big Data firms tend to be larger. Figure 6b shows the adoption rate of Big Data across employment quintiles in the French economy (blue) and in the model (red). The calibration of the adoption costs targeted the average adoption rate and the differences in adoption across size quintiles. However, the slope of adoption with firm size is not targeted, and the model successfully generates the upward sloping adoption rate of social media data in France.

Figure 6: Big Data and Firm Size

(a) Employment Distribution

(b) Big Data Adoption By Size Quintile



Note. Subfigure (a) plots the marginal distribution of the log of employment for Big Data and traditional firms. The blue line in subfigure (b) shows the adoption by employment quintile from Figure 3b, while the red line shows the model counterpart.

Targeting Data and Firm Size. The model proposes a mechanism why Big Data firms are larger, based on a two-way feedback loop between targeting data and firm size. First, firms at the top of the size dis-

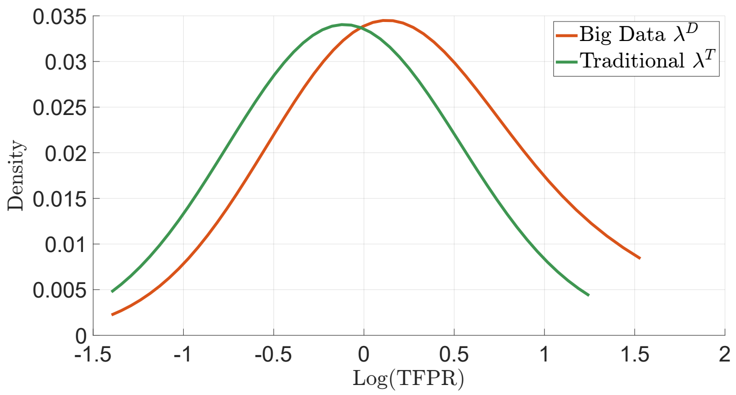
[25] Tadelis et al. (2023) show that data-intensive firms improve their advertising strategies.

tribution spend more resources in targeting high-valuation consumers, as shown by Figure 5b. Large firms have incentives to design a very precise advertising strategy: targeting costs are part of the overhead costs of the firm, while the targeting benefit grows with firm size and the quantity of goods sold. Then, high-productivity firms with more output are more likely to adopt, since the adoption costs of Big Data are independent of size. This generates the increasing adoption with firm size in Figure 6b. Second, more targeting data increases firm size because a precise targeting to high-valuation consumer types leads to a higher price faced by the firm, which increases optimal hiring. The literature proposes different mechanisms explaining the association between large firms and Information and Communication Technologies (ICT).[26] I propose here a new feedback loop between targeting data and firm size: the overhead targeting costs are more valuable for large firms, and the use of targeting data makes firms even larger through a higher price that increases their optimal employment.

4.3 Big Data, Productivity and Costs

Fact 3: Firms with Social Media Data are More Productive. Figure 7 compares the distribution of TFPR in the model for Big Data and traditional firms. The distribution of TFPR is shifted to the right, as in Figure 4a for French firms. While the difference in TFPR is a targeted moment, the model can provide additional insights to understand why Big Data firms have higher revenues. Are they more productive or do they sell their goods at a higher price? This is an interesting outcome of the model, since the balance-sheet information in France only contains revenue data and does not have information on prices and quantities.

Figure 7: TFPR and Processing Cost



Note. This figure plots the marginal distribution of the log of TFPR for Big Data and traditional firms, as defined as in Section 4.1: TFPR $= e^z \tilde{p}$. The price $\tilde{p}$ characterizes the average price across consumer types: $\tilde{p} = \int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon$.

I use model outcomes in the same regression used to calibrate TFPR in Equation (24). I run two separate regressions changing the dependent variable to the log of firm price and the log of productivity.[27] This regression captures the difference in prices and productivity driven by Big Data, conditional on size. I obtain that Big Data firms in the model have prices that are 4.1% higher, while their productivity is 2.9% lower. Prices are higher because they can target high-valuation consumers with a higher probability. However, firms in the same size quintile that use social media data are less productive: the higher prices increase their incentives to scale up and push them upwards in the firm distribution, beyond the size level that would correspond to a traditional firm with the same productivity. Therefore, top size

[26] Among many others, Lashkari, Bauer, and Boussard (2024) attribute the higher use of software and hardware by large firms to production non-homotheticities, while Farboodi and Veldkamp (2024) explore a two-way interaction between production and data, where more output generates more information, which improves firm predictions and quality, increasing production. Asriyan and Kohlhas (2025) observe that large firms make better predictions, and rationalize it in a model of product choice and firm scale optimization under uncertainty.

[27] I set the dependent variable $x(z, \lambda)$ to $\log(\tilde{p})$ and z.

quintiles include Big Data firms with low productivity but higher prices. The model suggests that the observed differences in TFPR for firms using social media data are driven by higher prices and not by higher productivity.

Fact 4: Firms with Social Media Data Show Higher Overhead Costs. Finally, I analyze the cost structure for adopting firms, which is an untargeted moment of the model. I run the same regression in Equation (24), with the share of overhead to total costs as a dependent variable.[28] Big Data firms in the model have a share of overhead to total costs that is 2.9 percentage points higher, while Table 2 in the previous section showed that the ratio of overhead to total costs is 2.8 percentage points higher for firms that use social media data in France. The difference in overhead costs for firms with different patterns is not targeted, but the model generates a similar increase in overhead costs for firms that use social media data. Firms with Big Data spend more resources in targeting, as shown in Figure 5b.

5 Implications for Firms and Consumers

The previous section showed that the model of data and targeted advertising captured relevant features about the use of Big Data in France, related to the adoption across the firm size distribution and the cost structure of firms. This section provides the main results of the paper: Section 5.1 explores the implications of technological change, Section 5.2 discusses the efficiency properties of the competitive equilibrium, and Section 5.3 presents a set of policy counterfactuals.

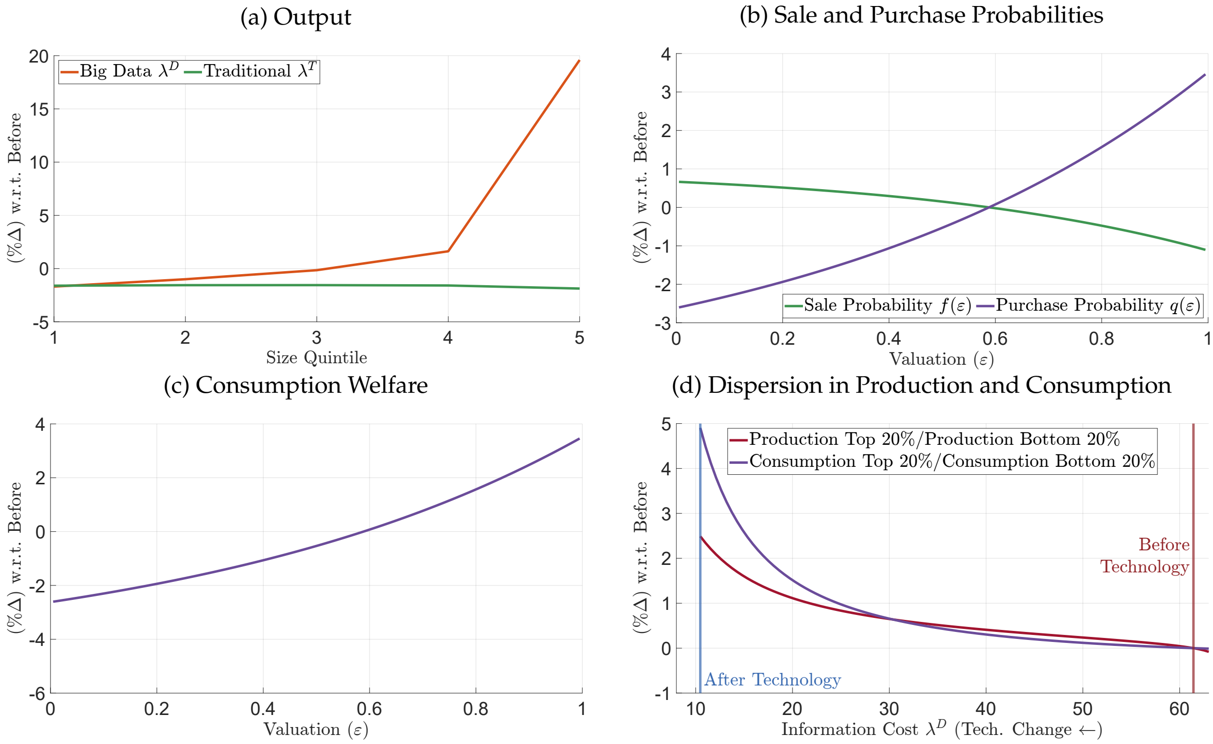
5.1 Technological Change

Decline in Data Processing Costs. Recent technological developments have decreased the costs of processing data: Appendix Figure B.1a shows an exponential decrease in computing costs over time. This section explores the aggregate consequences of this decline in processing costs for different firms and consumers. I capture technological change in the model as a decline in data processing costs for firms that use Big Data (λ^D). The economy described above was calibrated to France in the late 2010s, and the data processing cost for firms that use Big Data was 84% lower than the cost for traditional firms. I compare the baseline economy *(after technology)* to a counterfactual where the only change is that the processing cost for firms using Big Data is higher and closer to the cost for traditional firms. More precisely, I solve the model holding all parameters constant but setting the processing cost for Big Data firms to be only 4% lower than the value for traditional firms. This economy *(before technology)* represents France at the onset of social media expansion.

Large Big Data Firms Benefit. Figure 8a shows the impact of a decline in processing costs on output, where the value in the baseline calibration is shown as a percentage deviation from the model before technological change. The effect is shown separately for firms across the five quintiles of the size distribution and for Big Data and traditional firms. A decline in the cost of processing information only increases output for firms that use Big Data and that are at the top of the size distribution. It benefits this group disproportionately because of the feedback loop between targeting and firm size, which explains why large firms spend more resources in consumer targeting. A decrease in processing costs for Big Data firms facilitates targeting to high valuation consumers, increasing the aggregate quantity of ads posted to them, and reducing the equilibrium probability to sell to high valuation types in Figure 8b. This general equilibrium effect explains why technological change generates a reduction in output for traditional firms and small adopters of Big Data in Figure 8a. At the aggregate level, Appendix Figure C.2 shows that technological change increases the share of production employment and targeting expenditures by Big Data firms.

[28] Overhead costs add targeting cost $l_i(z, \lambda)$ and adoption cost $l_d(z, \lambda)$. Total costs include overhead costs plus labor $l_p(z, \lambda)$.

Figure 8: Technological Change, Firms and Consumers



Note. Subfigure a) shows the average output produced by firms in the same size quintile and with the same technology. Subfigure (b) shows the purchase probability from Equation (3) and the sale probability from Equation (4). Subfigure (c) shows consumption welfare: $\varepsilon c(\varepsilon) g(\varepsilon)$. Subfigure (d) shows two measures of dispersion in production and consumption: the share of production by the top quintile over the production by the bottom quintile of firms, and the share of consumption by the top quintile over the bottom quintile of consumers. All figures show the value under the baseline calibration as a % change with respect to the model before technological change.

A recent literature studies how production is affected by a technological change that increases data availability.[29] I propose a new mechanism that explains why large firms benefit more from Big Data: the intensive use of targeting data shifts their sales towards high-valuation consumer types and increases prices. This is consistent with the empirical evidence from France, where social media data is used for advertising, it is positively correlated with firm-size and is associated with higher overhead costs.

High-Valuation Consumers Benefit. In the model, technological change brings a consumption increase of 0.8%. This increase is partially compensated with the disutility of higher labor supplied to production and advertising. The net effect is positive, with a small increase in welfare of 0.1%. However, the aggregate increase in welfare masks important heterogeneities across consumer types. Technological change increases the quantity of ads posted to high valuation types, which increases their purchase probability in Figure 8b. Figure 8c shows the increase in consumption welfare resulting from an increase in the purchase probability for high-valuation types. Groups with a low valuation receive less ads after technological change and decrease their consumption. This section complements existing analyses of targeting technologies in the literature, focusing on the impact of a decline in targeting costs for different firms and consumers: I find that high-valuation consumers and large firms that use Big Data technologies are the main winners of technological change.[30]

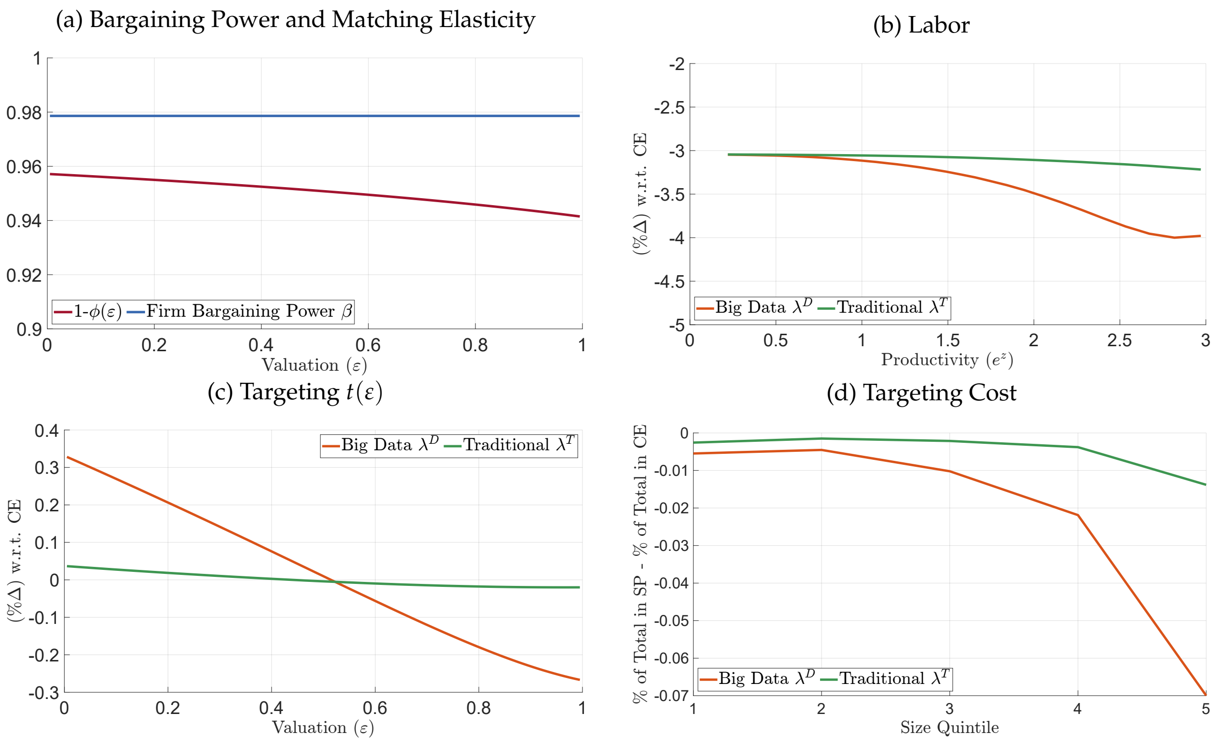
[29] Abis and Veldkamp (2024) document a decline in the labor share, while large firms in Begenau, Farboodi, and Veldkamp (2018) benefit more from technological change because they produce more information, which reduces investors' uncertainty about firm assets and ultimately decreases financing costs.

[30] Cavenaile, Celik, Perla, and Roldan-Blanco (2023) consider a counterfactual exercise of a decrease in targeting costs within their framework of product life-cycle with consumer awareness and targeted advertising. They find that an improvement in targeting technology increases the quality of matches between firms and consumers and provide additional results on the effects of targeting costs on product creation and markups.

5.2 Efficiency

Distortions in Production and Targeting. This section characterizes the efficiency of the competitive equilibrium under the baseline calibration. Proposition 1 in Section 2.6 showed that the competitive equilibrium is inefficient unless the bargaining power of firms (β) is equal to the following function of the matching elasticity: $\beta = (1 - \phi(\varepsilon))$. While the results depend on the model parameters, Figure 9a shows these values under the current calibration. The bargaining power is close to 1 and it is higher than the function of the matching elasticity. The difference is larger for high-valuation consumer types. Figure 9b plots the social planner choice of Big Data and traditional labor across the productivity distribution, relative to the competitive equilibrium. As discussed in Section 2.6, there is a *distortion in production*: the social planner would reduce employment for all firms, and the reduction should be larger for Big Data firms at the top of the size distribution.

Figure 9: Social Planner and Competitive Equilibrium



Note. Subfigure (a) shows the firm bargaining power β under the current calibration, along with the matching elasticity $1 - \phi(\varepsilon)$, defined as $\phi(\varepsilon) \equiv \frac{\partial m(sg(\varepsilon),a(\varepsilon))}{\partial(sg(\varepsilon))}\frac{sg(\varepsilon)}{m(sg(\varepsilon),a(\varepsilon))}$. Subfigure (b) shows the social planner choice of labor, conditional on technology and productivity, and relative to the competitive equilibrium. Subfigure (c) shows the social planner choice of targeting across consumer types, taking the average value by firms with the same technology, and relative to the competitive equilibrium. Subfigure (d) shows the share of targeting cost in total labor for Big Data and traditional firms, focusing on the social planner choice as a % difference with respect to the competitive equilibrium.

Figure 9c shows the targeting choice of the social planner as a percentage deviation with respect to the targeting strategy in the competitive equilibrium. The social planner would reduce targeting to high-valuation types, especially for Big Data firms, which are posting too many ads to these groups of consumers. With less targeting to high-valuation types, targeting costs decline for all firms in Figure 9d, which shows the change in advertising expenditures for firms with different technologies across the firm size distribution. The social planner would reduce targeting more by large Big Data firms, which are the most intensive users. These results suggest that there is a *distortion in targeting*. When firms decide which consumers to target, they do not internalize the impact on competitors. The private value of meeting consumers depends on the bargaining power β, which is above the social value. The inefficiency is especially relevant for large firms using Big Data, since the social planner would reduce especially their targeting expenditures and their firm size.

5.3 Policy Analysis: Digital Advertising Tax and GDPR

Digital Advertising Tax. There has been a recent discussion on the regulation of targeting: among others, Acemoglu and Johnson (2024) recently proposed a 50% tax on digital advertising. If targeted advertising allocates goods towards high-valuation consumers, and is used differently across the firm-size distribution, this tax could have significant consequences on the reallocation of production and the consumption distribution. I use the model to explore the impact of this regulation: I introduce a tax on advertising that uses Big Data - a proxy for digital advertising. The intervention is assumed to be announced after firms have made their adoption decisions, which implies that the adoption of Big Data is identical to the equilibrium without the tax. Tax revenues are rebated to the household, and the decisions under the competitive equilibrium are identical to Section 2.5, with the exception of firms that adopt Big Data technologies. These firms would maximize profits in the following problem:

$$\pi\left(z,\lambda^{D}\right) \equiv \max_{l_p,\left\{t(\varepsilon)\right\}_{\varepsilon}} \; e^{z} l_p^{1-\nu} \int_{\varepsilon} t\left(\varepsilon\right) f\left(\varepsilon\right) p(\varepsilon) d\varepsilon - \left[l_p + \lambda^{D} \int_{\varepsilon} t\left(\varepsilon\right) \log\left(\frac{t\left(\varepsilon\right)}{g\left(\varepsilon\right)}\right) d\varepsilon + \tau \int_{\varepsilon} t\left(\varepsilon\right) \log\left(\frac{t\left(\varepsilon\right)}{g\left(\varepsilon\right)}\right) d\varepsilon\right]$$

$$\text{s.t.} \quad \int_{\varepsilon} t\left(\varepsilon\right) d\varepsilon = 1$$

The problem is identical to the competitive equilibrium without the tax with the exception of the last term: firms pay a tax τ on the quantity of targeting data. Appendix A.6 shows that the advertising decision is distorted, as the tax increases the cost of targeting.

Targeted Advertising with Tax on Big Data. The targeting probability to market ε for a firm with productivity z and processing cost λ^{D} is given by the following expression:

$$t\left(\varepsilon\right) = \frac{g\left(\varepsilon\right) \exp\left(\frac{y f(\varepsilon) p(\varepsilon)}{\lambda^{D}+\tau}\right)}{\int_{\varepsilon'} g\left(\varepsilon'\right) \exp\left(\frac{y f(\varepsilon') p(\varepsilon')}{\lambda^{D}+\tau}\right) d\varepsilon'} \tag{25}$$

Production is given by $y = e^{z} l_p^{1-\nu}$, and the price in a specific market is given by: $p(\varepsilon) = \beta u'\left(C\right)\varepsilon$.

The solution to the labor choice is identical to the problem without the tax, but the value of employment can change, as the targeting decisions are different and the price faced by the firm can vary. I solve the competitive equilibrium for values of the tax from 1% to 100%, finding that the welfare-maximizing tax is 30%. I compare the digital advertising tax with a general tax on advertising, where the tax affects all firms instead of only Big Data firms as in Equation (25). The value that maximizes welfare is also 30%.

Lifting the GDPR. The General Data Protection Regulation (GDPR) was approved in 2016, and started to be enforced in 2018. Since the calibration targeted France in the late 2010s, the GDRP was in place for most of the period in the baseline model. I study an alternative calibration to explore the aggregate effects of lifting the GDPR. The GDPR increased the costs of using Big Data technologies, since firms had to increase their security protocols, hire new workers to comply with the regulation and provide careful reports about the use of data.[31] I introduce a lift of the GDPR as a decline in the costs of using Big Data (μ_{ψ}). To discipline the exercise, I rely on the results by Demirer, Hernández, Li, and Peng (2024), who use detailed information on the intensive margin use of data by firms to document that data storage decreased by 26% after the implementation of the GDPR. A back of the envelope calculation based on

[31] See Johnson (2023) for a detailed review on the impacts of the GDPR on data use.

these results suggests that lifting the GDPR would increase the quantity of data processed by 35%.[32] Therefore, this policy counterfactual introduces a decline of 16% in average adoption costs, which lead to a 35% increase in the quantity of consumer data processed by Big Data firms.

Table 4: Policy Counterfactuals and Aggregate Outcomes

	Lift GDPR	**Digital Advertising Tax**	**Advertising Tax**
Welfare	0.05%	0.03%	0.05%
Consumption	0.3%	-0.26%	-0.5%
Labor	0.4%	-0.51%	-0.98%
Shopping Effort	0%	0%	0%
Prod. Top 20% / Prod. Bottom 20%	1.35%	-0.6%	-1.2%
Cons. Top 20% / Cons. Bottom 20%	1.75%	-1.65%	-2.86%

Note. This table shows the impact of lifting the GDPR, a digital advertising tax of 30% and a general advertising tax of the same rate. All values are shown as a % change with respect to the competitive equilibrium in the baseline calibration. Equation (8) shows household welfare and its components. I compute the ratio between the production by the top quintile of the size distribution and the production by the bottom quintile, and the ratio between the consumption by the top quintile of consumers in the consumption distribution and the consumption by the bottom quintile.

Policy Counterfactuals: Aggregate Effects. Table 4 shows the aggregate effects of lifting the GDPR, a 30% tax on digital advertising and a 30% tax on general advertising. I will focus on GDPR and the digital advertising tax, as the results for general advertising show some quantitative differences but are qualitatively similar to the latter policy. Lifting the GDPR and the digital advertising tax have small positive impacts on welfare. However, the mechanisms are different. Lifting the GDPR increases the adoption of Big Data to 9%, pushing more resources into targeted advertising: the employment share in targeted advertising increases from 1.2% to 1.7%.[33] This decreases household utility because it increases advertising and production labor. However, more targeted advertising shifts goods towards high-valuation consumers, increasing utility from consumption. The net effect is positive, with a small increase in welfare. The tax has also positive welfare effects: holding fixed the adoption of Big Data, a tax on targeted advertising improves welfare, reducing innefficient targeting and bringing the competitive equilibrium closer to the social planner problem. The mechanisms are opposite to lifting the GDPR: the digital advertising tax decreases labor in advertising but also reduces consumption.

Policy Counterfactuals: Production Reallocation and Consumption Welfare. Next, I explore the implications of these policies for the distribution of consumption and production.

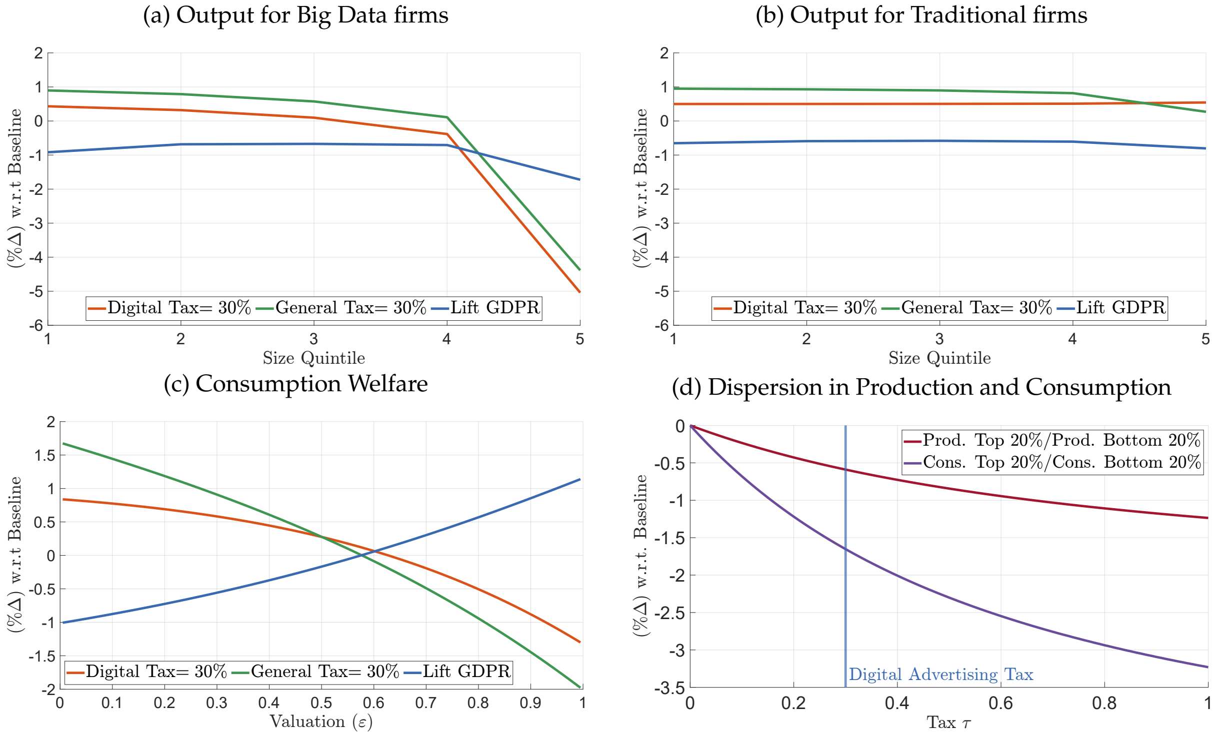
Figure 10a shows that output falls by approximately 5% for Big Data firms at the top size quintile, following a 30% tax on digital advertising. Large Big Data firms have the largest expenditure on targeted advertising, and they are the most vulnerable to the tax. On average, they would face lower valuation consumer types, which decreases the average price and reduces production. The digital advertising tax induces positive general equilibrium effects on small Big Data firms and Traditional firms in Figure 10b, which increase output. Relaxing consumer data regulation has different effects on production: it introduces new Big Data competitors and reduces production across the size distribution for firms with both technologies. The impact on consumer welfare is shown in Figure 10c. The digital advertising

[32] If x is the data with the GDPR, and by y is data without the GDPR:

$$x = 0.74y; \quad y = \frac{1}{0.74}x; \quad y = 1.35x$$

[33] Appendix Table C.1 shows how the policy counterfactuals change the distribution of employment between advertising and production and firms with different technologies.

Figure 10: Policy Counterfactuals: Production Reallocation and Consumption Welfare



Note. Subfigures a) and b) shows the average output produced by firms in the same size quintile and with the same technology (Big Data firms shown in a) while b) shows Traditional firms). Subfigure (c) shows consumption welfare: $\varepsilon c(\varepsilon) g(\varepsilon)$. Subfigure (d) shows two measures of dispersion in production and consumption: the production by the top quintile over the production by the bottom quintile of firms, and the consumption by the top quintile over the bottom quintile of consumers. All figures show the value under the counterfactual, as a % change with respect to the baseline model

tax reduces targeting, decreasing consumption by high-valuation types. Lifting the GDPR has opposite effects: it benefits high-valuation consumers because it decreases the adoption costs of Big Data and increases targeted advertising.

Figure 10d shows, for different values of the tax, output for firms at the top quintile over output for the bottom quintile, and the same ratio for the distribution of consumption. First, the digital advertising tax reduces dispersion in production, since it reduces output for large firms at the top of the size distribution. Second, the tax reduces dispersion in consumption because high-valuation types, at the top of the consumption distribution, become less targeted. However, lifting the GDPR increases Big Data adoption and targeting, reallocating production towards the largest firms that use targeting more intensively. This increases the existing dispersion in production, increasing the share of output by firms in the top size quintile, as shown in Table 4. By increasing targeted advertising expenditures, this policy also increases consumption by high-valuation types, further increasing dispersion in consumption.

While both policies have small positive welfare impacts, they have different reallocation effects across firms and consumers. The digital advertising tax reduces dispersion in production and consumption, while lifting the GDPR has opposite effects.

6 Conclusion

Big Data technologies have dramatically increased the availability of consumer information, enabling firms to engage in more precise targeted advertising. Which firms and consumers benefit from this technological change? Should governments regulate or tax the use of consumer data? I develop a theoretical framework where heterogeneous firms produce and target their ads to different consumer types. Firms endogenously adopt Big Data technologies that lower the cost of targeting. Using a technology survey matched with French balance sheet data, I provide empirical evidence that firms analyzing Big Data from social media—a proxy for the use of consumer data in advertising—are larger, more productive, and devote a higher share of costs to overhead expenditures. I use these findings to calibrate the model to France in the late 2010s.

The model replicates the positive correlation between Big Data use, firm size, and overhead costs. Comparing this baseline with an economy facing higher information-processing costs—proxying the onset of the social media expansion—I find that only large firms benefit from technological change. Smaller firms and traditional advertisers lose due to stronger competition from large firms that use targeted advertising more intensively. Aggregate welfare rises modestly, but high-valuation consumers capture most of the gains.

The social planner would reduce targeting expenditures in equilibrium to internalize the crowding-out distortions from excessive targeting. Finally, I evaluate two data policies: a digital advertising tax and the lifting of the GDPR. Both generate small welfare gains but distinct reallocations. A 30% digital advertising tax reduces targeting and production by large adopters, benefiting smaller firms, while lifting the GDPR increases Big Data adoption, production concentration, and welfare for high-valuation consumers. These results suggest that policies regulating consumer data can have relevant implications for the reallocation of production across firms and the consumption distribution.

References

Abis, S., & Veldkamp, L. (2024). The changing economics of knowledge production. *The Review of Financial Studies*, *37*(1), 89–118.

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340.

Acemoglu, D., & Johnson, S. (2024). The urgent need to tax digital advertising. *Network Law Review, Spring*, *5*.

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–256.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*(6), 2411–2451.

Afrouzi, H., Drenik, A., & Kim, R. (2023). *Concentration, market power, and misallocation: The role of endogenous customer acquisition* (tech. rep.). National Bureau of Economic Research.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2023). A theory of falling growth and rising rents. *Review of Economic Studies*, *90*(6), 2675–2702.

AI-Impacts. (2025). Wikipedia history of gflops. *Retrieved July 29, 2025 from https://aiimpacts.org/wikipedia-history-of-gflops-costs/*.

Anderson, S., Chintagunta, P., Kaul, R., & Vilcassim, N. (2020). Impact of marketing analytics 1.0 on entrepreneur and firm performance: Field experiment evidence from rwanda.

Arvai, K., & Mann, K. (2025). Consumption inequality in the digital age.

Asriyan, V., & Kohlhas, A. (2025). The macroeconomics of data: Scale, product choice, and pricing in the information age.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745.

Baley, I., & Veldkamp, L. (2025). *The data economy: Tools and applications*. Princeton University Press.

Baslandze, S., Greenwood, J., Marto, R., & Moreira, S. (2023). The expansion of varieties in the new age of advertising. *Review of Economic Dynamics*, *50*, 171–210.

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, *97*, 71–87.

Bergemann, D., Bonatti, A., & Gan, T. (2022). The economics of social data. *The RAND Journal of Economics*, *53*(2), 263–296.

Berman, R., & Israeli, A. (2022). The value of descriptive analytics: Evidence from online retailers. *Marketing Science*, *41*(6), 1074–1096.

Brand, J. M., Demirer, M., Finucane, C., & Kreps, A. A. (2024). Firm productivity and learning in the digital economy: Evidence from cloud computing.

Brynjolfsson, E., & McElheran, K. (2016a). Data in action: Data-driven decision making in us manufacturing. *US Census Bureau Center for Economic Studies Paper No. CES-WP-16-06, Rotman School of Management Working Paper*, (2722502).

Brynjolfsson, E., & McElheran, K. (2016b). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133–139.

Calderón, J. B. S., & Rassier, D. G. (2025). Valuing the us data economy using machine learning and online job postings. In S. Basu, L. Eldridge, J. Haltiwanger, & E. Strassner (Eds.), *Technology, productivity, and economic growth* (pp. 245–284). University of Chicago Press.

Cavenaile, L., Celik, M. A., Perla, J., & Roldan-Blanco, P. (2023). A model of product awareness and industry life cycles.

Cavenaile, L., Celik, M. A., Roldan-Blanco, P., & Tian, X. (2025). Style over substance? advertising, innovation, and endogenous market structure. *Journal of Monetary Economics*, *149*, 103683.

Cavenaile, L., & Roldan-Blanco, P. (2021). Advertising, innovation, and economic growth. *American Economic Journal: Macroeconomics*, *13*(3), 251–303.

Cheremukhin, A., Restrepo-Echavarria, P., & Tutino, A. (2020). Targeted search in matching markets. *Journal of Economic Theory*, *185*, 104956.

Chiavari, A. (2024). Customer accumulation, returns to scale, and secular trends.

Cong, L. W., Xie, D., & Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, *67*(10), 6480–6492.

Corrado, C., Haskel, J., Iommi, M., Jona-Lasinio, C., & Bontadini, F. (2025). Data, intangible capital, and productivity. In S. Basu, L. Eldridge, J. Haltiwanger, & E. Strassner (Eds.), *Technology, productivity, and economic growth* (pp. 193–244). University of Chicago Press.

De Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review*, *114*(1), 199–251.

Demirer, M., Hernández, D. J. J., Li, D., & Peng, S. (2024). Data, privacy laws and firm production: Evidence from the gdpr.

Den Haan, W. J., Ramey, G., & Watson, J. (2000). Job destruction and propagation of shocks. *American economic review*, *90*(3), 482–498.

Eeckhout, J., & Veldkamp, L. (2023). Data and markups: A macro-finance perspective.

Farboodi, M., & Veldkamp, L. (2024). A growth model of the data economy.

Fedyk, A., & Hodson, J. (2023). Trading on talent: Human capital and firm performance. *Review of Finance*, *27*(5), 1659–1698.

FEVAD. (2024). Chiffres clés e-commerce. *Fédération E-Commerce et Vente à Distance*.

Freeman, R. B., Yang, B., & Zhang, B. (2023). Data deepening and nonbalanced economic growth. *Journal of Macroeconomics*, *75*, 103503.

FTC. (2025). Surveillance pricing 6(b) study: Research summaries. *Federal Trade Commission Staff*.

Fudenberg, D., & Villas-Boas, J. M. (2012). Price discrimination in the digital economy. *Oxford Handbook of the Digital Economy*, *254*.

Galdon-Sanchez, J. E., Gil, R., & Uriz-Uharte, G. (2025). The value of information in competitive markets: Evidence from small and medium-sized enterprises. *Journal of Political Economy*, *133*(1), 252–305.

Goodridge, P., Haskel, J., & Edquist, H. (2022). We see data everywhere except in the productivity statistics. *Review of Income and Wealth*, *68*(4), 862–894.

Gourio, F., & Rudanko, L. (2014). Customer capital. *Review of Economic Studies*, *81*(3), 1102–1136.

Greenwood, J., Ma, Y., & Yorukoglu, M. (2025). "you will:" a macroeconomic analysis of digital advertising. *Review of Economic Studies*, *92*(3), 1837–1881.

Hagedorn, M., & Manovskii, I. (2008). The cyclical behavior of equilibrium unemployment and vacancies revisited. *American Economic Review*, *98*(4), 1692–1706.

Hagiu, A., & Wright, J. (2023). Data-enabled learning, network effects, and competitive advantage. *The RAND Journal of Economics*, *54*(4), 638–667.

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

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, *57*(2), 279–298.

Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, *110*(2), 569–595.

IDC, & Statista. (2024). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023, with forecasts from 2024 to 2028. *Retrieved from https://www.statista.com/statistics/871513/worldwide-data-created/, November 28, 2024.*

Jiang, X. (2024). Information and communication technology and firm geographic expansion.

Jin, Y., & Sun, Z. (2019). Information acquisition and the return to data.

Johnson, G. (2023). Economic research on privacy regulation: Lessons from the gdpr and beyond.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*(9), 2819–58.

Kehoe, P. J., Larsen, B. J., & Pastorino, E. (2020). Dynamic competition in the era of big data.

Lashkari, D., Bauer, A., & Boussard, J. (2024). Information technology and returns to scale. *American Economic Review*, *114*(6), 1769–1815.

Lentz, R., Maibom, J., & Moen, E. R. (2022). Competitive or random search?

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, *70*(2), 317–341.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, *61*(1), 226–273.

Matějka, F., & McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review*, *102*(3), 24–29.

Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*(1), 272–298.

Maurin, L., Kolev, A., & Revoltella, D. (2023). Investment report 2022/2023: Resilience and renewal in europe. *EIB*.

Morazzoni, M., & Sy, A. (2022). Female entrepreneurship, financial frictions and capital misallocation in the us. *Journal of Monetary Economics*, *129*, 93–118.

Prüfer, J., & Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, *69*(4), 967–1008.

Prytkova, E., Petit, F., Li, D., Chaturvedi, S., & Ciarli, T. (2024). The employment impact of emerging digital technologies.

Rachel, L. (2024). Leisure-enhancing technological change.

Schaal, E., & Taschereau-Dumouchel, M. (2019). Aggregate demand and the dynamics of unemployment.

Schmidt, J., Pilgrim, G., & Mourougane, A. (2023). What is the role of data in jobs in the united kingdom, canada, and the united states?: A natural language processing approach. *OECD*.

SFGATE. (2025). Hotel booking sites show higher prices to travelers from bay area. *Retrieved July 10, 2025, from https://www.sfgate.com/travel/article/hotel-booking-sites-overcharge-bay-area-travelers-20025145.php*.

Shen, S. (2025). Search frictions, customer allocation and business dynamism.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665–690.

Statista. (2024). Statista market insights: Digital advertising in europe. *Retrieved November 28, 2024, from https://www.statista.com/outlook/dmo/digital-advertising/europe*.

Tadelis, S., Hooton, C., Manjeer, U., Deisenroth, D., Wernerfelt, N., Dadson, N., & Greenbaum, L. (2023). Learning, sophistication, and the returns to advertising: Implications for differences in firm performance. *National Bureau of Economic Research*.

Veldkamp, L., & Chung, C. (2024). Data and the aggregate economy. *Journal of Economic Literature*, *62*(2), 458–484.

Weiss, J. (2020). Intangible investment and market concentration.

Wong, H. C. (2025). Understanding high-wage firms: Monopoly, monopsony, and bargaining power.

Wu, L. (2024). Partially directed search in the labor market. *Journal of Economic Theory*, 105858.

Wu, L., Hitt, L., & Lou, B. (2020). Data analytics, innovation, and firm productivity. *Management Science*, *66*(5), 2017–2039.

A Model Appendix

A.1 Price Bargaining

This section explains the Nash Bargaining game between the firm and the household.

1. **consumer of valuation ε:** this consumer values the good by ε, increasing C by ε and therefore contributing by an additional $u'(C)\varepsilon$ to the utility of the representative household. In exchange for this utility, the consumer pays the price $p(\varepsilon)$, which in utility value corresponds to $Qp(\varepsilon)$, since Q captures the change in utility value after changes in income. It is assumed that upon matching the household does not have any other outside option, so the net payoff in utility value is given by: $u'(C)\varepsilon - Qp(\varepsilon)$.

2. **Firm:** Upon matching, the firm obtains an income equal to the price $p(\varepsilon)$, which corresponds to $Qp(\varepsilon)$ in utility value. Given the assumptions on the timing of production, goods have already been produced at the time of the bargaining. The firm cannot recover the production cost, so it also has an outside option of zero. In utility value, the firm obtains the following payoff: $Qp(\varepsilon)$.

3. **Bargaining problem:** the price $p(\varepsilon)$ between a firm and a consumer of valuation ε is the solution to the following Nash Bargaining problem, where $\beta \in (0,1)$ captures the bargaining power of the firm:
$$\max_{p(\varepsilon)} (Qp(\varepsilon))^{\beta} \left(u'(C)\varepsilon - Qp(\varepsilon)\right)^{1-\beta}$$
Taking the first order condition with respect to $p(\varepsilon)$, and solving for the price yields the following result:
$$p(\varepsilon) = \frac{\beta u'(C)\varepsilon}{Q}$$

A.2 Household Choices and Wage Normalization

Household Problem. Substituting the definitions of goods demanded s and consumption $c(\varepsilon)$, I can rewrite the household problem as a choice of demand s and labor L in Equation (8):

$$\max_{s,L} \quad U(C) - \zeta s - L$$
$$\text{s.t.} \quad C = wL + \Pi$$

Since Q is the multiplier on the budget constraint, the FOC with respect to goods demand s and labor L:
$$\underline{s}: \quad u'(C)\frac{\partial C}{\partial s} - \zeta - Q\frac{\partial C}{\partial s} = 0;$$
$$\underline{L}: \quad -1 + wQ = 0; \quad Q = \frac{1}{w}$$
Taking the FOC of the demand from goods and plugging the FOC for Labor:
$$u'(C)\frac{\partial C}{\partial s} - \zeta - \frac{1}{w}\frac{\partial C}{\partial s} = 0;$$

Next, I assume a log utility function $U(C) = \log(C)$, and notice that:
$$\frac{\partial C}{\partial s} = \int_{\varepsilon} p(\varepsilon)\, q(\varepsilon)\, g(\varepsilon)\, d\varepsilon$$

Back to the FOC of the demand, using the definition of C and solving for s:

$$\frac{\int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon}{s\int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon} - \zeta - \frac{1}{w}\int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon = 0$$

$$\frac{1}{s} - \zeta - \frac{1}{w}\int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon = 0;$$

$$s = \frac{w}{\zeta w + \int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon}$$

Notice that the demand of goods is the same irrespective of the chosen normalization, given the price definition:

$$s = \frac{w}{\zeta w + \int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon} = \frac{w}{\zeta w + \int_\varepsilon \frac{\beta u'(C)\varepsilon}{Q} q(\varepsilon) g(\varepsilon) d\varepsilon} = \frac{1}{\zeta + \int_\varepsilon \beta u'(C)\varepsilon q(\varepsilon) g(\varepsilon) d\varepsilon}$$

Therefore, I normalize $Q = 1$, where it emerges from the household decision that $w = 1$. The household supplies an infinitely elastic labor supply at the wage $w = 1$. Then, the demand is given by

$$s = \frac{1}{\zeta + I}; \quad I \equiv \frac{\partial C}{\partial s} = \int_\varepsilon p(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon$$

Firm Choices. Next, I show that setting $w = 1$ does not affect firm-level decisions, as both prices and costs scale with wages. Taking Equation (10):

$$\pi(z,\lambda) = y\int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon - w\lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon - wl_p$$

Next, I plug in the market price in Equation (7):

$$\pi(z,\lambda) = y\int_\varepsilon t(\varepsilon) f(\varepsilon) \frac{\beta u'(C)\varepsilon}{Q} d\varepsilon - w\lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon - wl_p$$

$$\pi(z,\lambda) = \beta u'(C)\frac{1}{Q} y\int_\varepsilon t(\varepsilon) f(\varepsilon)\varepsilon d\varepsilon - w\lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon - wl_p$$

Plugging in the Household FOC with respect to labor $Q = \frac{1}{w}$:

$$\pi(z,\lambda) = \beta u'(C) wy\int_\varepsilon t(\varepsilon) f(\varepsilon)\varepsilon d\varepsilon - w\lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon - wl_p$$

$$\pi(z,\lambda) = w\left[y\int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon - \lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon - l_p\right]$$

Profits are scaled by the wage, so firm decisions would be the same regardless of the normalization, where the market price after the normalization is $p(\varepsilon) = \beta u'(C)\varepsilon$

A.3 Targeting Strategy in the Competitive Equilibrium

The firm chooses a targeting strategy $t(\varepsilon)$ to maximize profits taking as given the selling probability $f(\varepsilon)$, which is a function of the market tightness $\theta(\varepsilon)$:

$$\pi(z,\lambda) \equiv \max_{l_p, \{t(\varepsilon)\}_\varepsilon} y\int_\varepsilon t(\varepsilon) f(\varepsilon) p(\varepsilon) d\varepsilon - \left[l_p + \lambda\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) d\varepsilon\right] \tag{26}$$

$$\text{s.t.} \quad \int_{\varepsilon} t(\varepsilon)\, d\varepsilon = 1$$

Denoting by κ the multiplier on the constraint, and taking the FOC with respect to $t(\varepsilon)$:

$$yf(\varepsilon)\, p(\varepsilon) - \lambda \left[\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) + 1\right] - \kappa = 0;$$

$$yf(\varepsilon)\, p(\varepsilon) - \lambda - \kappa = \lambda \log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right)$$

$$t(\varepsilon) = g(\varepsilon) \exp\left(\frac{yf(\varepsilon)\, p(\varepsilon) - \lambda - \kappa}{\lambda}\right) \tag{27}$$

Using the constraint that the probabilities across consumer types ε' add up to 1:

$$\int_{\varepsilon'} g(\varepsilon') \exp\left(\frac{yf(\varepsilon')\, p(\varepsilon') - \lambda - \kappa}{\lambda}\right) d\varepsilon' = 1;$$

Taking the common part across ε':

$$\int_{\varepsilon'} g(\varepsilon') \frac{\exp\left(\frac{yf(\varepsilon')p(\varepsilon')}{\lambda}\right)}{\exp\left(\frac{\lambda+\kappa}{\lambda}\right)} d\varepsilon' = 1;$$

$$\int_{\varepsilon'} g(\varepsilon') \exp\left(\frac{yf(\varepsilon')\, p(\varepsilon')}{\lambda}\right) d\varepsilon' = \exp\left(\frac{\lambda+\kappa}{\lambda}\right);$$

I plug in this expression into the FOC of a generic targeting strategy in Equation (27):

$$t(\varepsilon) = \frac{g(\varepsilon) \exp\left(\frac{yf(\varepsilon)p(\varepsilon)}{\lambda}\right)}{\exp\left(\frac{\lambda+\kappa}{\lambda}\right)}$$

$$t(\varepsilon) = \frac{g(\varepsilon) \exp\left(\frac{yf(\varepsilon)p(\varepsilon)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon') \exp\left(\frac{yf(\varepsilon')p(\varepsilon')}{\lambda}\right) d\varepsilon'}$$

Taking the second derivative shows that the problem is concave in the targeting strategy:

$$\frac{\partial^2 \pi(z,\lambda)}{\partial^2 t(\varepsilon)} = -\frac{\lambda}{t(\varepsilon)} < 0;$$

A.4 Problem of the Social Planner

The social planner chooses

$$\mathcal{L} = \max_{s,\{l_p(z,\lambda),t(\varepsilon;z,\lambda)\}_{z,\lambda}} \log\left(\int_{\varepsilon} \varepsilon a(\varepsilon)\, f(\varepsilon)\, d\varepsilon\right) - \zeta s - \left(\int_z \sum_{\lambda} \left[l_p(z,\lambda) + \lambda \int_{\varepsilon} t(\varepsilon;z,\lambda) \log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right) d\varepsilon\right] \mu(z,\lambda)\right)$$

With the following definitions of the ads per market and the constraint on the selling probability:

$$a(\varepsilon) \equiv \int_z \sum_{\lambda} \mu(z,\lambda)\, e^z l_p(z,\lambda)^{1-\nu} t(\varepsilon;z,\lambda)\, dz; \quad f(\varepsilon) = \frac{m(sg(\varepsilon), a(\varepsilon))}{a(\varepsilon)}; \quad \int t(\varepsilon;z,\lambda)\, d\varepsilon = 1; \quad \forall (z,\lambda)$$

A.4.1 Labor Choice

- **Step 1: FOC.** Taking the FOC with respect to $l_p(z, \lambda)$:

$$u'(C)\left[\int \varepsilon\left(\frac{\partial a(\varepsilon)}{\partial l_p(z,\lambda)} f(\varepsilon) + a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)}\frac{\partial a(\varepsilon)}{\partial l_p(z,\lambda)}\right) d\varepsilon\right] - \mu(z,\lambda) = 0;$$

$$\int \varepsilon u'(C)\frac{\partial a(\varepsilon)}{\partial l_p(z,\lambda)}\left(f(\varepsilon) + a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)}\right) d\varepsilon = \mu(z,\lambda);$$

- **Step 2: Matching Elasticity.** As an intermediate step, we define the matching elasticity:

$$\phi(\varepsilon) \equiv \frac{\partial m(sg(\varepsilon), a(\varepsilon))}{\partial(sg(\varepsilon))}\frac{sg(\varepsilon)}{m(sg(\varepsilon), a(\varepsilon))}$$

Taking the first term:

$$\frac{\partial m(sg(\varepsilon), a(\varepsilon))}{\partial(sg(\varepsilon))} = \frac{\partial[a(\varepsilon)m(\theta(\varepsilon),1)]}{\partial(sg(\varepsilon))} = a(\varepsilon)m_1(\theta(\varepsilon),1)\frac{\partial\theta(\varepsilon)}{\partial(sg(\varepsilon))} = m_1(\theta(\varepsilon),1)$$

Back into the definition of the elasticity:

$$\phi(\varepsilon) = m_1(\theta(\varepsilon),1)\frac{sg(\varepsilon)}{m(sg(\varepsilon), a(\varepsilon))} = \frac{m_1(\theta(\varepsilon),1)}{m\left(1, \theta(\varepsilon)^{-1}\right)}$$

- **Step 3: Ads, Selling Probability and Matching Elasticity.** Next, I show the following equality:

$$\frac{\partial f(\varepsilon)}{\partial\theta(\varepsilon)} = q(\varepsilon)\phi(\varepsilon)$$

First, we know that:

$$f(\varepsilon) = q(\varepsilon)\theta(\varepsilon); \quad q(\varepsilon) = \frac{f(\varepsilon)}{\theta(\varepsilon)} = \frac{m(\theta(\varepsilon),1)}{\theta(\varepsilon)} \tag{28}$$

Taking the derivative w.r.t. $\theta(\varepsilon)$ in Equation (28), and using the purchase probability definition:

$$\frac{\partial f(\varepsilon)}{\partial\theta(\varepsilon)} = \frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}\theta(\varepsilon) + q(\varepsilon) = q(\varepsilon)\left[1 + \theta(\varepsilon)\frac{\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}}{q(\varepsilon)}\right] =$$

$$= q(\varepsilon)\left[1 + \frac{\theta(\varepsilon)}{\frac{m(\theta(\varepsilon),1)}{\theta(\varepsilon)}}\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}\right] = q(\varepsilon)\left[1 + \frac{\theta(\varepsilon)}{m\left(1,\theta(\varepsilon)^{-1}\right)}\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}\right] \tag{29}$$

Next, show the derivative of the purchase probability with respect to tightness in Equation (29):

$$\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)} = \frac{m_1(\theta(\varepsilon),1)\theta(\varepsilon) - m(\theta(\varepsilon),1)}{\theta(\varepsilon)^2} = \frac{m_1(\theta(\varepsilon),1)}{\theta(\varepsilon)} - \frac{m(\theta(\varepsilon),1)}{\theta(\varepsilon)^2}$$

Plugging back this last step into Equation (29):

$$\frac{\partial f(\varepsilon)}{\partial\theta(\varepsilon)} = q(\varepsilon)\left[1 + \frac{\theta(\varepsilon)}{m\left(1,\theta(\varepsilon)^{-1}\right)}\left(\frac{m_1(\theta(\varepsilon),1)}{\theta(\varepsilon)} - \frac{m(\theta(\varepsilon),1)}{\theta(\varepsilon)^2}\right)\right] =$$

$$= q(\varepsilon)\left[1+\frac{1}{m\left(1,\theta(\varepsilon)^{-1}\right)}\left(m_1(\theta(\varepsilon),1)-m\left(1,\theta(\varepsilon)^{-1}\right)\right)\right] = q(\varepsilon)\frac{m_1(\theta(\varepsilon),1)}{m\left(1,\theta(\varepsilon)^{-1}\right)}$$

Using the matching elasticity from Step 2:

$$\frac{\partial f(\varepsilon)}{\partial \theta(\varepsilon)} = q(\varepsilon)\phi(\varepsilon)$$

- **Step 4: Externality.** Next, we work out the following component from the FOC, using steps 2 and 3:

$$f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)} = f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial \theta(\varepsilon)}\frac{\partial \theta(\varepsilon)}{\partial a(\varepsilon)} = f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial \theta(\varepsilon)}\left(-\frac{\theta(\varepsilon)}{a(\varepsilon)}\right) =$$
$$= f(\varepsilon)-\theta(\varepsilon)\frac{\partial f(\varepsilon)}{\partial \theta(\varepsilon)} = f(\varepsilon)-\theta(\varepsilon)q(\varepsilon)\phi(\varepsilon) = f(\varepsilon)(1-\phi(\varepsilon))$$

- **Step 5: Labor Choice for Social Planner.** Re-writting the FOC: Returning to the FOC:

$$\int \varepsilon u'(C)\frac{\partial a(\varepsilon)}{\partial l_p(z,\lambda)}\left(f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)}\right)d\varepsilon = \mu(z,\lambda);$$

The only component left is the derivative of ads with respect to labor:

$$\frac{\partial a(\varepsilon)}{\partial l_p(z,\lambda)} = \mu(z,\lambda)e^z(1-\nu)l_p(z,\lambda)^{-\nu}t(\varepsilon;z,\lambda)$$

Plugging this derivative and Step 4 into the FOC gives the expression in Section 2.6:

$$\mu(z,\lambda)\int \varepsilon u'(C)e^z(1-\nu)l_p(z,\lambda)^{-\nu}t(\varepsilon;z,\lambda)f(\varepsilon)(1-\phi(\varepsilon))d\varepsilon = \mu(z,\lambda);$$

$$l_p(z,\lambda)^{-\nu}e^z(1-\nu)\int t(\varepsilon;z,\lambda)f(\varepsilon)(1-\phi(\varepsilon))\varepsilon u'(C)d\varepsilon = 1;$$

$$l_p(z,\lambda) = \left[e^z(1-\nu)\int t(\varepsilon;z,\lambda)f(\varepsilon)(1-\phi(\varepsilon))\varepsilon u'(C)d\varepsilon\right]^{\frac{1}{\nu}}$$

A.4.2 Targeting Choice

- **Step 1: Targeting FOC.** Taking the FOC with respect to $t(\varepsilon;z,\lambda)$, where $\kappa_{z,\lambda}$ is the multiplier on the probabilities constraints:

$$u'(C)\varepsilon\left(\frac{\partial a(\varepsilon)}{\partial t(\varepsilon;z,\lambda)}f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)}\frac{\partial a(\varepsilon)}{\partial t(\varepsilon;z,\lambda)}\right)-\mu(z,\lambda)\lambda\left[\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right)+1\right]-\kappa_{z,\lambda}=0;$$

$$u'(C)\varepsilon\frac{\partial a(\varepsilon)}{\partial t(\varepsilon;z,\lambda)}\left(f(\varepsilon)+a(\varepsilon)\frac{\partial f(\varepsilon)}{\partial a(\varepsilon)}\right)-\mu(z,\lambda)\lambda\left[\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right)+1\right]-\kappa_{z,\lambda}=0;$$

Combining Steps 2,3 and 4 from the previous derivation:

$$u'(C)\varepsilon\frac{\partial a(\varepsilon)}{\partial t(\varepsilon;z,\lambda)}f(\varepsilon)(1-\phi(\varepsilon))-\mu(z,\lambda)\lambda\left[\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right)+1\right]-\kappa_{z,\lambda}=0;$$

Plugging the derivative of aggregate ads as a function of targeting gives the FOC shown in Section 2.6:

$$u'(C)\,\varepsilon\mu(z,\lambda)\,e^{z}l_p(z,\lambda)^{1-\nu}f(\varepsilon)\,(1-\phi(\varepsilon))-\mu(z,\lambda)\,\lambda\left[\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right)+1\right]-\kappa_{z,\lambda}=0;$$

$$u'(C)\,\varepsilon\mu(z,\lambda)\,e^{z}l_p(z,\lambda)^{1-\nu}f(\varepsilon)\,(1-\phi(\varepsilon))-\mu(z,\lambda)\,\lambda-\kappa_{z,\lambda}=\mu(z,\lambda)\,\lambda\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right);$$

$$g(\varepsilon)\exp\left(\frac{u'(C)\,\varepsilon e^{z}l_p(z,\lambda)^{1-\nu}f(\varepsilon)\,(1-\phi(\varepsilon))-\mu(z,\lambda)\,\lambda-\kappa_{z,\lambda}}{\lambda}\right)=t(\varepsilon;z,\lambda);$$

- **Step 2: Logit.** Using the constraint that the targeting probabilities add up to 1:

$$\int g(\varepsilon')\exp\left(\frac{u'(C)\,\varepsilon' e^{z}l_p(z,\lambda)^{1-\nu}f(\theta(\varepsilon'))\,(1-\phi(\varepsilon'))-\mu(z,\lambda)\,\lambda-\kappa_{z,\lambda}}{\lambda}\right)d\varepsilon'=1;$$

$$\int g(\varepsilon')\exp\left(\frac{u'(C)\,\varepsilon' e^{z}l_p(z,\lambda)^{1-\nu}f(\theta(\varepsilon'))\,(1-\phi(\varepsilon'))}{\lambda}\right)d\varepsilon'=\exp\left(\frac{\kappa_{z,\lambda}+\mu(z,\lambda)\,\lambda}{\mu(z,\lambda)\,\lambda}\right);$$

Then, re-writing for a generic market:

$$t(\varepsilon;z,\lambda)=\frac{g(\varepsilon)\exp\left(\frac{e^{z}l_p(z,\lambda)^{1-\nu}f(\varepsilon)(1-\phi(\varepsilon))u'(C)\varepsilon}{\lambda}\right)}{\int g(\varepsilon')\exp\left(\frac{e^{z}l_p(z,\lambda)^{1-\nu}f(\theta(\varepsilon'))(1-\phi(\varepsilon'))u'(C)\varepsilon'}{\lambda}\right)d\varepsilon'};$$

A.4.3 Demand Choice

First, notice that from the household side we can write consumption and purchasing probabilities:

$$C=\int_{\varepsilon}\varepsilon sg(\varepsilon)\,q(\varepsilon)\,d\varepsilon;\quad q(\varepsilon)=\frac{m(sg(\varepsilon),a(\varepsilon))}{sg(\varepsilon)}$$

We can rewrite the maximization problem as follows:

$$\mathcal{L}=\max_{s,\{l_p(z,\lambda),t(\varepsilon;z,\lambda)\}_{z,\lambda}}\log\left(\int_{\varepsilon}\varepsilon sg(\varepsilon)\,q(\varepsilon)\,d\varepsilon\right)-\zeta s-\left(\int_z\sum_{\lambda}\left[l_p(z,\lambda)+\lambda\int_{\varepsilon}t(\varepsilon;z,\lambda)\log\left(\frac{t(\varepsilon;z,\lambda)}{g(\varepsilon)}\right)d\varepsilon\right]\mu(z,\lambda)\right)$$

Taking the first order condition with respect to the goods demand s:

$$\frac{1}{\int_{\varepsilon}\varepsilon sg(\varepsilon)\,q(\varepsilon)\,d\varepsilon}\int_{\varepsilon}\varepsilon g(\varepsilon)\left[q(\varepsilon)+s\frac{\partial q(\varepsilon)}{\partial s}\right]d\varepsilon-\zeta=0;$$

The term inside the brackets, where the last derivation follows from Step 3 in the labor choice.

$$q(\varepsilon)+s\frac{\partial q(\varepsilon)}{\partial s}=q(\varepsilon)+s\frac{\partial\theta(\varepsilon)}{\partial s}\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}=q(\varepsilon)+\theta(\varepsilon)\frac{\partial q(\varepsilon)}{\partial\theta(\varepsilon)}=q(\varepsilon)\,\phi(\varepsilon)$$

Solving for s:

$$\frac{1}{s\int_{\varepsilon}\varepsilon g(\varepsilon)\,q(\varepsilon)\,d\varepsilon}\int_{\varepsilon}\varepsilon g(\varepsilon)\,q(\varepsilon)\,\phi(\varepsilon)\,d\varepsilon=\zeta;$$

$$s=\frac{\int_{\varepsilon}\varepsilon\phi(\varepsilon)q(\varepsilon)g(\varepsilon)d\varepsilon}{\zeta\int_{\varepsilon}\varepsilon q(\varepsilon)g(\varepsilon)d\varepsilon};$$

A.5 Proof of Proposition 1

Here I show that the solution of the competitive equilibrium is equal to the social planner if the bargaining power varies by consumer type and is equal to the following function of the matching elasticity:

$$\beta(\varepsilon) = 1 - \phi(\varepsilon) \tag{30}$$

Variables in the competitive equilibrium are denoted with CE, while variables of the social planner solution are denoted with SP. First, I conjecture that if Equation (30) holds, the consumption level, the equilibrium market tightness, the sale and purchase probabilities are the same in the competitive and the social planner problem. This will be verified later.

$$f^{CE}(\varepsilon) = f^{SP}(\varepsilon) = f(\varepsilon); \quad q^{CE}(\varepsilon) = q^{SP}(\varepsilon) = q(\varepsilon); \quad C^{CE} = C^{SP}$$

Demand. I start with the demand for goods in the competitive case, using the price definition in Equation (7), with the variation that now the bargaining power changes by consumer type:

$$s^{CE} = \frac{1}{\zeta + \int_{\varepsilon} p(\varepsilon) q\left(\varepsilon\right) g\left(\varepsilon\right) d\varepsilon} = \frac{1}{\zeta + \int_{\varepsilon} \beta(\varepsilon)\varepsilon u'(C) q\left(\varepsilon\right) g\left(\varepsilon\right) d\varepsilon} = \frac{1}{\zeta + \int_{\varepsilon} (1-\phi(\varepsilon))\varepsilon u'(C) q\left(\varepsilon\right) g\left(\varepsilon\right) d\varepsilon}$$

We can rewrite this equality as follows:

$$s^{CE}\zeta + u'(C) \int_{\varepsilon} \varepsilon s^{CE} q\left(\varepsilon\right) g(\varepsilon) d\varepsilon - s^{CE} u'(C) \int_{\varepsilon} \varepsilon \phi(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon = 1$$

Using the definition of consumption in Equation (5) and the log utility, we have that the second term is equal to 1:

$$s^{CE}\zeta + 1 - s^{CE} u'(C) \int_{\varepsilon} \varepsilon \phi(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon = 1$$

For the third term, we use the fact that $s^{CE} u'(C) = \frac{1}{\int_{\varepsilon} \varepsilon q(\varepsilon) g(\varepsilon) d\varepsilon}$:

$$s^{CE}\zeta = \frac{\int_{\varepsilon} \varepsilon \phi(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon}{\int_{\varepsilon} \varepsilon q(\varepsilon) g(\varepsilon) d\varepsilon}$$

Then, the demand for goods in the competitive equilibrium is equal to the demand of the social planner:

$$s^{CE} = \frac{\int_{\varepsilon} \varepsilon \phi(\varepsilon) q(\varepsilon) g(\varepsilon) d\varepsilon}{\zeta \int_{\varepsilon} \varepsilon q(\varepsilon) g(\varepsilon) d\varepsilon} = s^{SP}$$

Labor Choice. Plugging Equation (30) into the labor choice in the competitive equilibrium, I obtain the employment choice of the social planner.

$$l_p^{CE} = \left[e^z \left(1-\nu\right) \int_{\varepsilon} t\left(\varepsilon\right) f\left(\varepsilon\right) p\left(\varepsilon\right) d\varepsilon \right]^{\frac{1}{\nu}} = \left[e^z \left(1-\nu\right) \int_{\varepsilon} t\left(\varepsilon\right) f\left(\varepsilon\right) \beta(\varepsilon) \varepsilon u'(C) d\varepsilon \right]^{\frac{1}{\nu}} =$$

$$= \left[e^z \left(1-\nu\right) \int_{\varepsilon} t\left(\varepsilon\right) f\left(\varepsilon\right) \left(1-\phi(\varepsilon)\right) \varepsilon u'(C) d\varepsilon \right]^{\frac{1}{\nu}} = l^{SP}$$

Targeted Advertising. Similarly, using Equation (30) into the targeting choice in competitive equilibrium:

$$t(\varepsilon) = \frac{g(\varepsilon)\exp\left(\frac{yf(\varepsilon)p(\varepsilon)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{yf(\varepsilon')p(\varepsilon')}{\lambda}\right)d\varepsilon'} = \frac{g(\varepsilon)\exp\left(\frac{yf(\varepsilon)\beta(\varepsilon)\varepsilon u'(C)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{yf(\varepsilon')\beta(\varepsilon')\varepsilon' u'(C)=}{\lambda}\right)d\varepsilon'}$$

$$= \frac{g(\varepsilon)\exp\left(\frac{yf(\varepsilon)(1-\phi(\varepsilon))\varepsilon u'(C)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{yf(\varepsilon'(1-\phi(\varepsilon'))\varepsilon' u'(C)}{\lambda}\right)d\varepsilon'}$$

Finally, I verify that the purchase and sale probabilities and the consumption level are equal in the competitive and social planner cases under Equation (30). First, the market tightness depends on the supply and demand of goods by consumer type:

$$\theta(\varepsilon) \equiv \frac{sg(\varepsilon)}{a(\varepsilon)}; \quad a(\varepsilon) \equiv \int_z \sum_\lambda \mu(z,\lambda)\, y(z,\lambda)\, t(\varepsilon; z, \lambda)$$

The numerator is equal in the competitive and the social planner cases, as the demand on the intensive margin is the same. In the denominator, the quantity of ads by consumer type depends on the firm distribution, the production of goods and targeted advertising. Since the firm distribution is the same and firms choose the same labor and targeting strategies, the quantity of ads is the same in the social planner problem and the competitive equilibrium. Then, the market tightness is the same for all consumer types and the consumption level and the purchase and sale probabilities are also identical, given that they are a function of the market tightness in Equations (3), (4) and (5).

A.6 Problem of Big Data Firms with Tax

This section describes the solution to the targeted advertising and labor choice of a Big Data firm with a tax. We can rewrite the problem as follows:

$$\pi\left(z, \lambda^D\right) \equiv \max_{l_p, \left\{t(\varepsilon)\right\}_\varepsilon} e^z l_p^{1-\nu} \int_\varepsilon t(\varepsilon)\, f(\varepsilon)\, p(\varepsilon) d\varepsilon - \left[l_p + (\lambda^D + \tau)\int_\varepsilon t(\varepsilon)\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right)d\varepsilon\right]$$

$$\text{s.t.} \quad \int_\varepsilon t(\varepsilon)\, d\varepsilon = 1$$

Denoting by κ the multiplier on the constraint, and taking the FOC with respect to $t(\varepsilon)$:

$$yf(\varepsilon)\, p(\varepsilon) - (\lambda^D + \tau)\left[\log\left(\frac{t(\varepsilon)}{g(\varepsilon)}\right) + 1\right] - \kappa = 0;$$

Following the same derivation as in Section A.3, I obtain the targeting probability:

$$t(\varepsilon) = \frac{g(\varepsilon)\exp\left(\frac{yf(\varepsilon)p(\varepsilon)}{\lambda^D+\tau}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{yf(\varepsilon')p(\varepsilon')}{\lambda^D+\tau}\right)d\varepsilon'}$$

Taking the first order condition with respect to labor:

$$\int_\varepsilon t(\varepsilon)\, f(\varepsilon)\, p(\varepsilon) d\varepsilon - \frac{l_p^\nu}{e^z(1-\nu)} = 0;$$

Solving for labor gives the same solution as in the competitive equilibrium without the tax:

$$l_p = \left[e^z(1-\nu)\int_\varepsilon t(\varepsilon)\, f(\varepsilon)\, p(\varepsilon) d\varepsilon\right]^{\frac{1}{\nu}}$$

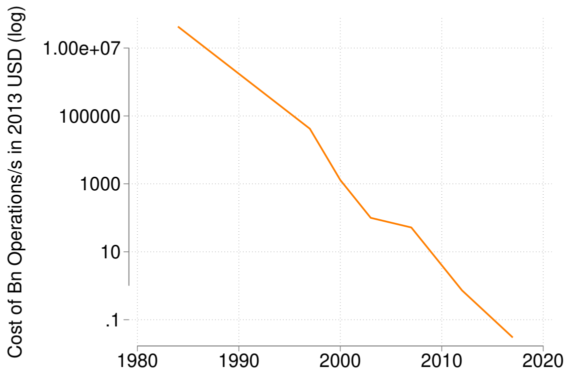
B Data Appendix

B.1 Trends in Computing Cost and Big Data

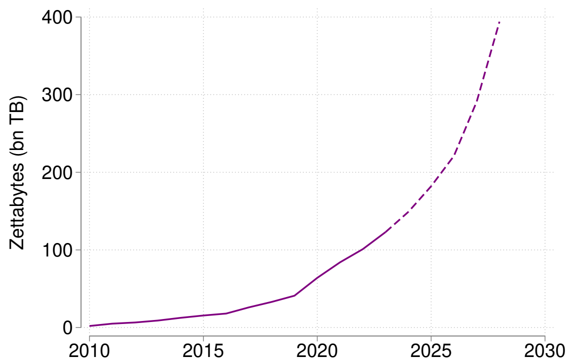
Figure B.1a shows a decline in the computing cost over time from the 1980s to the 2010s, while Figure B.1b shows an exponential growth in the quantity of data available over the past 15 years.

Figure B.1: Technological Change

(a) Decline in Processing Cost



(b) Global Increase in Big Data



Note. Subfigure (a) shows estimates of the computing cost, measured as a GFLOP: the cost in 2013 USD of performing one billion computing operations per second, as collected from different sources by AI-Impacts (2025). The scale of the y-axis is logarithmic. Subfigure (b) shows an approximation of the amount of Big Data available worldwide, as estimated by IDC and Statista (2024). The unit is a zettabytye, corresponding to one billion terabytes, and the dotted line shows forecasted values.

B.2 Business ICT Survey

Survey. The *Enquête sur les technologies de l'information et de la communication dans les entreprises* is elaborated by INSEE every year since 2007, asking firms about the adoption of different technologies. The unit addressed is the legal unit *(siren)* of at least 10 workers in the non-financial sectors France, and the sample is collected through simple stratified random sampling, by last period sector, employees bracket and turnover bracket. Firms are surveyed from the following sectors (NACE rev. 2): Manufacture (C), Electricity, Gas, Steam and Air Conditioning Supply (D), Water and Waste Management (E), Construction (F), Wholesale and Retail Trade, Repair of Motor Vehicles and Motorcycles (G), Transportation and Storage (H), Accommodation and Food Services (I), Information and Communication (J), Real Estate Activities (L), Professional, Scientific and Technical Activities (M, only divisions 69-74), Administrative and Support Service Activities (N) and Other Service Activities (S, group 951 only). Firms from the Real Estate sector have been dropped from the analysis.

The surveys contain between 8,000 and 12,000 observations, depending on the year, though the final number of observations used is slightly reduced due to the cleaning procedure after the merge with FARE. In general, the survey elaborated at year t asks about technology adoption during the year $t-1$, though some questions refer to other periods.

Big Data Definition. First, firms are provided with the following definition of this technology:

1. It has a large *volume* resulting from a large quantity of data generated over time
2. There is a *variety* concerning the different formats of complex data, structured or unstructured
3. Its *velocity*, due to the high speed at which data is generated and becomes available

Data Questions. Firms are asked the following question about the years 2015, 2017 and 2019:

- *During the previous year, has your company analyzed big data from the following sources?*
    1. *The company's own data from intelligent or connected objects, or sensors.*
    2. *Geolocation data from portable devices*
    3. *Data generated by social media.*
    4. *Other sources of Big Data.*

While the question does not change across years, some survey years include external analysis of these data sources and some do not. To homogenize the meaning of the variable, I use another survey question about the use of internal or external data. By combining both survey questions, the final *Social Media Data* variable takes the value of 1 if big data is analyzed internally within the firm, and firms use data from social media. Firms are also asked about the use of Big Data, where the following question is only answered by firms that use Big Data from at least one source in 2015:

- *In 2015, the firm has used Big Data analysis to:*
    1. *Improve marketing or sales management?*
    2. *Develop or improve goods or services?*
    3. *Optimize internal processes for producing goods or providing services?*

Social Media Data and Age. Table B.1 shows the adoption of social media data by age. There is no clear correlation, as adoption rates do not change by large magnitudes depending on firm age. However, adoption rates vary from 4% to 8% depending on the employment quintile. Table B.2 shows the adoption rate depending by employment quintile separately for young firms (¡15 years) and old (¿=15 years). The positive correlation between social media data and size holds for firms irrespectively of their age.

Age	Social Media Data (% of firms)
Below 5 Years	4.87
5 to 15 Years	5.72
15 to 30 Years	4.80
30 to 50 Years	4.77
Above 50 Years	5.50
Total	5.10

Table B.1: Social Media Data by Age

Employees	Social Media Data (% of young)	Social Media Data (% of old)
<=9	4.99	4.58
(9-13]	4.43	3.42
(13-19]	5.22	3.23
(19-36]	7.14	4.43
>36	8.10	7.82

Table B.2: Social Media Data, Age and Size

Data Types and Uses. Table B.3 shows the main logit specification for the different uses of data: Marketing, Process Improvement and Product Development. *Social Media Data* is correlated with marketing, as in the model in Section 2, and with product development as in Baslandze, Greenwood, Marto, and Moreira (2023). *Other Sources* seem to be correlated with innovation, in line with the theoretical frameworks in Jones and Tonetti, 2020 and Cong, Xie, and Zhang, 2021.

B.3 Firm-level Balance-Sheet Data

Variable Definition. FARE is part of ESANE (*Élaboration des Statistiques Annuelles d'Entreprise*), which compiles detailed fiscal information for firms in France. I follow the definitions of balance-sheet variables used by De Ridder (2024) and Wong (2025):

Table B.3: Data Sources and Uses

	(1) Marketing	(2) Process Improvement	(3) Product Development
Social Media Data	0.294***	-0.037	0.236***
	(0.05)	(0.04)	(0.05)
Other Sources	0.060	0.252***	0.250***
	(0.05)	(0.05)	(0.04)
Observations	994	994	994
Sector FE	Yes	Yes	Yes
Capital & Age	Yes	Yes	Yes
Other Technologies	Yes	Yes	Yes
Estimate	Logit	Logit	Logit

Note. This table shows the marginal effect of the dummies on social media data and other sources in the logit specification. Column (1) replicates the results for marketing in the body of the paper, column (2) shows the results for a dummy on whether firms use data to improve processes and column (3) focuses on whether data is used to develop new products. The sample is restricted to firms that use any type of Big Data in 2015 (social media or other sources), due to the survey design. The use of other technologies includes dummies on the use of high speed internet, if the firm has a website or sells using e-commerce. Standard errors in parentheses: $*(p < 0.10), **(p < 0.05), ***(p < 0.01)$. Source: FARE & ICT Enterprise Survey, INSEE.

- *Revenue* (R): revenue is obtained from the item `redi_r310` in FARE.
- *Capital* (k): tangible fixed assets, corresponing to `immo_corp`.
- *Employment* (l): number of employees from variable `redi_e200`.
- *Wagebill* (wb): cost of employees corresponds to the sum of wage payments (`redi_r216`) and social security contributions (`redi_r217`).
- *Material Costs* (m): merchandise purchases (`redi_r210`)-change in merchandise stock (`redi_r211`)+purchase of raw materials (`redi_r212`)-change in raw materials stock (`redi_r213`).
- *Cost of Goods Sold* (v): here I add costs of employees and material costs, to obtain a measure of variable costs $v = wb + m$.
- *Fixed Costs* (f): other operating expenses, from variable `redi_r214`.
- *Total Costs* (t): I add variable and fixed costs, $t = v + f$.
- *Profits* (π): revenue minus total costs, $\pi = R - t$.

Summary Statistics. Table B.4 provides the average value for a set of balance-sheet outcomes. All values except employment are expressed in thousand 2018 EUR. Users of Social Media data tend to be larger in terms of both sales and employment.

Variable	Full Sample	Social Media Data	No Social Media Data
Sales	18053	56790	15972
Employees	53	137	48
Capital	9754	54497	7350
Variable Cost	12352	34469	11164
Fixed Cost	4611	18264	3877

Table B.4: Summary Statistics

Cleaning Procedure. Below I describe the main cleaning and deflation procedure for the firm-level balance-sheet data.

- Select the FARE firms that can be merged to the Business ICT Survey, using the *(siren)* identifier, and select the modal industry for every firm.

- Drop missing and negative firm-year observations of: *Sales*, *Capital*, *Wagebill*, *Material Costs* and *Fixed Costs*.

- Drop observations at the top and bottom 0.1 % of the following rations: *Employment/Capital*, *Employment/Sales* and *Sales/Capital*.

- Deflate all financial variables to 2018 euros, using sector deflators for France from EU_KLEMS :
  - For financial variables, use the sector value-added price index (VA_PI).
  - For capital, use the sector price of capital (IP_GFCF)

Sector Definition. Table B.5 shows the classification of 16 sectors based on 2-digit NACE rev.2 codes.

	NACE codes (2-digit)	Description
Food	10,11,12	Manufacturing of food and beverages
Textiles	13,14,15,31,32,33	Manufacturing of textiles, leather and others (furniture, games, jewellery...)
Chemicals	20,21	Manufacturing of chemicals and pharmaceutical products
Materials	16,17,18,19,22,23,24,25	Manufacturing of wood, paper, refined petroleum, metals, cement..
Electronics	26,27	Manufacturing of consumer electronics and electrical equipment
Machinery	28,29,30	Manufacturing of motor vehicles, transport equipment and other machinery
Utilities	35,36,37,38,39	Electricity, water and waste collection
Construction	41,42,43	Construction of buildings, civil engineering and associated activities
Trade	45,46,47	Wholesale of vehicles, food, household goods and equipment
Transport	49,50,51,52,53	Transport by land, water or air
Tourism	55,56	Hotels and other accommodation, restaurants and other food services
IT and Telecom	58,59,60,61,62,63	Data processing, software, programming, telecommunications and TV
Consulting	69,70,71	Legal services, consulting services and architecture
R&D & Adv	72,73,74,75	Research, advertising and market research, design
Business Services	77,78,79,80,81,82	Rental, HR, travel agencies, office support and services to businesses
Household Services	94,95,96	Repair of goods and services to households

Table B.5: Classification of Sectors

B.4 Production Function Estimation

Methodology. For this section, I extend the panel of firms to include firm-year observations from the year 2010. The estimation of the production function is done separately for each of the sectors in Table B.5. The methodology is based on Ackerberg, Caves, and Frazer (2015), who updated the original approach by Levinsohn and Petrin (2003). The following value-added production function is assumed, with revenue y_{it} depending on labor l_{it} and capital k_{it}, with all variables in logs:

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it}$$

Revenue also depends on the productivity observed to the firm ω_{it} and an unexpected error ϵ_{it}. As in Ackerberg, Caves, and Frazer (2015), the gross production function is leontieff in materials m_{it}, which react to inputs and productivity according to the function $\tilde{f}_t$:

$$m_{it} = \tilde{f}_t\left(k_{it}, l_{it}, \omega_{it}\right); \quad \omega_{it} = \tilde{f}_t^{-1}\left(k_{it}, l_{it}, m_{it}\right)$$

We can obtain productivity as a function of materials, labor and capital by inverting the reaction function of materials into $\tilde{f}_t^{-1}$, which is estimated in the data.

First Stage: The first stage obtains expected output for the firm, residualized of the unexpected error

ϵ_{it}. Output y_{it} is regressed on a second order polynomial on labor, capital and materials:

$$y_{it} = \phi_t (l_{it}, k_{it}, m_{it}) + \epsilon_{it}$$

Given a set of parameters β_0, β_l and β_k, we can back up estimated productivity ω_{it}:

$$\omega_{it} = \hat{\phi}_t (l_{it}, k_{it}, m_{it}) - \beta_0 - \beta_l l_{it} - \beta_k k_{it} \tag{31}$$

Second Stage: We assume that productivity follows a first-order Markov process:

$$\omega_{it} = \rho \omega_{i,t-1} + \zeta_{it}$$

Next, the parameters of the production function β_0, β_l and β_k and the persistence of productivity ρ are estimated using the following GMM condition:

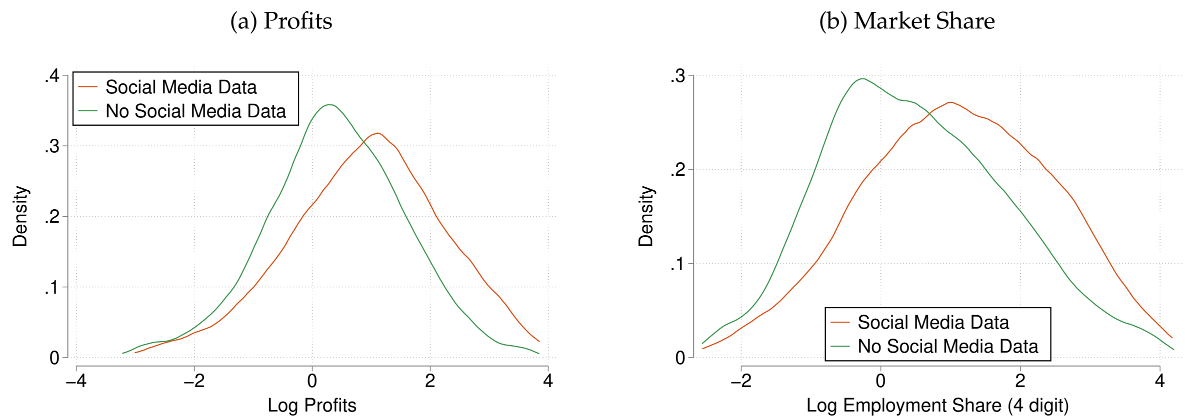
$$E \left[\zeta_{it} (\beta_0, \beta_l, \beta_k, \rho) \times \underbrace{\begin{bmatrix} 1 \\ k_{it} \\ l_{it-1} \\ \hat{\phi}_{t-1} (l_{it-1}, k_{it-1}, m_{it-1}) \end{bmatrix}}_{\mathbf{z}_{it}} \right] = 0$$

The GMM condition uses the set of instruments $\mathbf{z}_{it}$ from Ackerberg, Caves, and Frazer (2015), and it relies on the fact that the innovations to the productivity process should be orthogonal to current capital and labor and expected output from the previous period. The four parameters $(\beta_0, \beta_l, \beta_k, \rho)$ are estimated using GMM for each of the sectors. Given the final parameters of the production function, I obtain productivity ω_{it} using Equation (31), which I use to analyze TFPR at the firm-level. To take into account the sectoral differences, TFPR is always shown with respect to the sectoral average.

B.5 Profits and Market Shares

Profits. This section shows that other measures of profitability are positively correlated with the use of social media data.

Figure B.2: Distribution of Alternative Variables



Note. Subfigure (a) shows the distribution of profits, while subfigure (b) shows the distribution of employment share at the 4-digit NACE. Both plots show the distribution separately for firms using social media data and firms that do not adopt; I residualize from sector-year fixed effects, capital, age and other technology controls; and the variable is trimmed at the top and bottom 1%. Source: FARE & ICT Enterprise Survey, INSEE.

Table B.6 shows the results of the empirical specification in Equation (22) using profits on the dependent variable. In the specification with all controls, profits are approximately 19% higher for firms using social media data. In addition, Figure B.2a shows that the profit distribution is shifted to the right for firms using social media data, after residualizing for the set of controls used in Equation (22).

Table B.6: Profits and Social Media Data

	(1)	(2)	(3)	(4)
	π	π	Empl. Share	Empl. Share
Social Media Data	0.257^{***}	0.188^{***}	0.350^{***}	0.256^{***}
	(0.06)	(0.05)	(0.09)	(0.08)
R^2	0.433	0.441	0.488	0.496
Observations	17053	16421	21709	20911
Sector $\times$ Year FE	Yes	Yes	Yes	Yes
Capital & Age	Yes	Yes	Yes	Yes
Other Technologies	No	Yes	No	Yes
Estimate	OLS	OLS	OLS	OLS

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note. This table shows the results of the empirical specification in Equation (22). Columns (1) and (2) show the results for profits, while columns (3) and (4) show the employment share at the 4-digit NACE. The use of other technologies includes dummies on the use of high speed internet, if the firm has a website or sells using e-commerce. Standard errors in parentheses: $*(p < 0.10), **(p < 0.05), ***(p < 0.01)$. Source: FARE & ICT Enterprise Survey, INSEE.

Employment Market Share. I compute an alternative variable to measure the market share within narrow sectors. Using the universe of firms in the FARE dataset, I compute the employment market share at the 4-digit NACE sector in a given year. Figure B.2b shows the residualized distributions of the employment share, where firms using social media data show on average a higher market share. Table B.6 confirms these findings using the main empirical strategy, where the employment share is 25.6% higher for social media data firms. Similar results are obtained when using revenue market share instead of employment market share.

B.6 Within-Firm Impact

Empirical Strategy. Equation (22) compares firms that use social media data with similar firms in the same year that do not adopt. While this cross-sectional comparison is informative about the differences in performance correlated with technology adoption, the information about French firms allows for a further exploration of the dynamic impact within a firm. Since the Business ICT Survey combines a panel sample with new sampled observations, for some firms we can characterize the patterns of technology adoption over time. In this section, I restrict the analysis to firms that are surveyed about the use of social media data more than once. In particular, from the firms that have been surveyed multiple times I select firms that switch from not adopting to using social media data, and those that never adopted. For this subset of firms, I know the year in which they started using this type of information, apart from the balance-sheet information for a selection of years before and after switching to adoption. I obtain a total of 422 events of firms switching from not using social media data to adopting it.

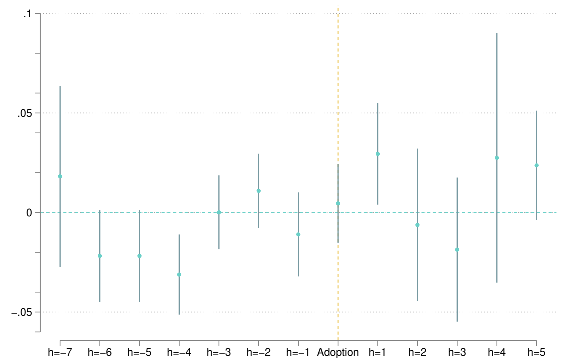
Using this sample, I estimate the following regression to explore the evolution of firm-level outcomes around the timing of "switching" to using Social Media data. The coefficient of interest is β^h, which captures the impact on a balance-sheet outcome h periods ahead.

$$y_{i,s,t+h} = \gamma_i + \alpha_{st} + \beta^h \mathcal{I}\Big\{\text{Switch to Social Media Data at t}\Big\} + \beta_1 k_{ist} + \gamma' \mathbf{\Theta_{ist}} + \epsilon_{i,s,t+h};$$

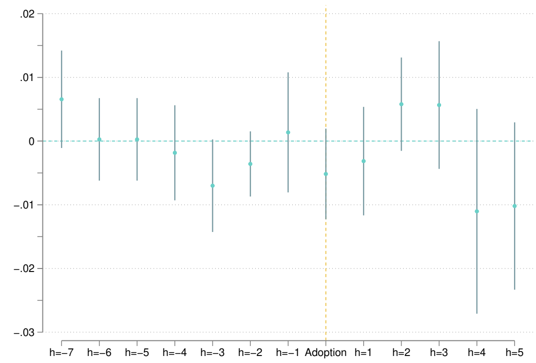
Firms "switch" when the survey response changes from not using social media data to using it. The outcome $y_{i,t+h}$ is a firm-level metric of firm i at time $t+h$ in levels, while the right hand side includes a firm-level fixed effect γ_i, and a dummy equal to 1 for the year t, when the firm "switches" to using social media data. Apart from capital, the technology controls and the sector-year fixed effect, I control for firm fixed effect γ_i. Exploring the time series dimension of the balance-sheet variables, I explore the impact of technology adoption h periods ahead, where negative values of h capture periods before adoption.

Figure B.3: Within-Firm Impact: Results

(a) TFPR



(b) Overhead Share



Note. This figure shows the results of estimating the within-firm specification for TFPR and the overhead share.

Figure B.3a shows that the impact on TFPR of switching to social media data is positive and statistically significant one period after adoption. However, the magnitude of the impact is slightly lower than for the cross-sectional regression: TFPR is 2.9% higher in the within firm specification, while it is 4.5% higher in the main cross-sectional comparison. Figure B.3b shows the results for the the share of overhead costs, where the results are positive for two and three years ahead but not statistically significant.

C Quantitative Appendix

C.1 Numerical Solution

1. Discretize the space of preferences ε and productivity z.

2. Set j=1 and guess a level of consumption and a vector of market tightness for all markets ε: $C_g^j, \{\theta_g(\varepsilon)^j\}$

3. Given $\theta_g(\varepsilon)^j$ and C_g^j, solve for production labor $l_p(z,\lambda)$ and targeted advertising $\{t(\varepsilon;z,\lambda)\}_\varepsilon$:

$$l_p(z,\lambda) = \left[e^z(1-\nu)\int_\varepsilon t(\varepsilon;z,\lambda)\, f(\varepsilon)\, p(\varepsilon) d\varepsilon\right]^{\frac{1}{\nu}}$$

$$t(\varepsilon;z,\lambda) = \frac{g(\varepsilon)\exp\left(\frac{y(z,\lambda) f\left(\theta_g(\varepsilon)^j\right) p(\varepsilon)}{\lambda}\right)}{\int_{\varepsilon'} g(\varepsilon')\exp\left(\frac{y(z,\lambda) f\left(\theta_g(\varepsilon)^j\right) p(\varepsilon)}{\lambda}\right) d\varepsilon'};$$

The market price is given by $p(\varepsilon) = \beta u'\left(C_g^j\right)\varepsilon$.

4. Given profits, obtain the endogenous mass of firms with different processing costs $\mu(z,\lambda)$:

$$\frac{\mu(z,\lambda^D)}{\mu(z,\lambda^D)+\mu(z,\lambda^T)} = F_\psi\left(\pi\left(z,\lambda^D\right) - \pi\left(z,\lambda^T\right)\right)$$

5. Obtain ads per market $a(\varepsilon)$, as a function firm decisions and the measure of firms $\mu(z,\lambda)$:

$$a(\varepsilon) \equiv \int_z \sum_\lambda a(\varepsilon;z,\lambda)\, dz; \quad a(\varepsilon;z,\lambda) = \mu(z,\lambda)\, y(z,\lambda)\, t(\varepsilon;z,\lambda)$$

6. Compute aggregate prices and household intensive demand s:

$$P \equiv \int_\varepsilon P(\varepsilon)\, q(\varepsilon)\, g(\varepsilon)\, d\varepsilon; \quad s = \frac{1}{\zeta + P}$$

7. Compute equilibrium consumption level and market tightness for all markets ε:

$$C_e^j = s\int_\varepsilon \varepsilon q(\varepsilon)\, g(\varepsilon)\, d\varepsilon; \quad \theta_e(\varepsilon)^j \equiv \frac{sg(\varepsilon)}{a(\varepsilon)}$$

8. If $\begin{cases} \max \|\theta_e(\varepsilon)^j - \theta_g(\varepsilon)^j, C_e^j - C_g^j\| > \eta & \rightarrow \quad j=j+1, \quad \text{update} \quad C_g^{j+1}, \left\{\theta_g(\varepsilon)^{j+1}\right\} \rightarrow \text{back to 3.} \\ \max \|\theta_e(\varepsilon)^j - \theta_g(\varepsilon)^j, C_e^j - C_g^j\| < \eta & \rightarrow \quad \text{stop, solution found} \end{cases}$

C.2 Moments

- **Average Revenue/Variable Cost.**

  Data. Average sales/variable cost in the balance-sheet is 1.69.

  Model. The ratio of revenue to variable cost is a function of returns to scale in the model:

$$\frac{e^z l_p(z,\lambda)^{1-\nu}\tilde{p}(z,\lambda)}{l_p(z,\lambda)} = e^z \tilde{p}\left(z,\lambda\right)\left[e^z\left(1-\nu\right)\tilde{p}\left(z,\lambda\right)\right]^{\frac{-\nu}{\nu}} = \frac{1}{1-\nu}$$

  Setting $1-\nu$ to 0.59 makes revenue to variable cost equal to 1.69 in the model.

- **Shopping Time/ Labor Time**

  Data. I use the Eurostat Harmonised European Time Use Surveys (HETUS), Wave 2 for 2010.[34] People between 45 and 64 in France in 2010 spend on average 30 minutes in *Shopping and Services.*, while for people with ages between 25 to 44, the shopping time is around 24 minutes. Taking the average of 27 minutes between the two age groups, and dividing by the 7 hours of working time for people that participate in the labor market, I obtain a ratio of shopping time to labor market time of 6.43%. The corresponding statistic in Gourio and Rudanko (2014) would be around 5%, though in their work they take a third of this shopping time which is allocated to new matches. However, I am interested in the whole shopping time, given the static nature of the model.

  Model. The ratio of shopping labor to production labor is given by:

$$\frac{\zeta s}{L}$$

- **Elasticity of Sales to Advertising.**

  Data. I compute the following elasticity of sales to advertising, focusing on the change between 2017 and 2022. To obtain this estimate, I focus on online markets in France. The value of online sales in France is obtained from FEVAD (2024), while the value of advertising is obtained from Statista:[35]

$$\frac{(\text{Sales}_{2022} - \text{Sales}_{2017})}{\text{Sales}_{2017}} / \frac{(\text{Advertising}_{2022} - \text{Advertising}_{2017})}{\text{Advertising}_{2017}} = \left(\frac{144.7-80.5}{80.5}\right)\left(\frac{10.8-4.7}{4.7}\right) = 0.62$$

  Model. In the model, I compute the elasticity of matches with respect to advertising in the top and bottom values of the valuation $\varepsilon_{\min}, \varepsilon_{\max}$.

$$\frac{m\left(sg\left(\varepsilon_{\max}\right),a\left(\varepsilon_{\max}\right)\right) - m\left(sg\left(\varepsilon_{\min}\right),a\left(\varepsilon_{\min}\right)\right)}{m\left(sg\left(\varepsilon_{\min}\right),a\left(\varepsilon_{\min}\right)\right)} / \left(\frac{a\left(\varepsilon_{\max}\right)-a\left(\varepsilon_{\min}\right)}{a\left(\varepsilon_{\min}\right)}\right)$$

- **Employment Share Top 10% Firms.**

  Data. For every year, I look at the firms in the top decile of the employment distribution, using survey weights. Across years, these firms represent an average of 68% of total employment.

[34] Source, accessed 4 September 2024: https://ec.europa.eu/eurostat/databrowser/view/tus_00age__custom_12751915/default/table?lang=en&page=time:2010

[35] Source, accessed 24 March 2025: https://www.statista.com/outlook/amo/advertising/france?currency=USD#digital-ad-spending

Model. I compute the share of total employment by firms in the top employment decile:

$$\frac{\int_z \sum_\lambda \mathbb{I}\{\text{Top Employment Decile}\} l_p(z,\lambda)\,\mu(z,\lambda)\,dz}{\int_z \sum_\lambda l_p(z,\lambda)\,\mu(z,\lambda)\,dz}$$

- **%Δ in TFPR by Social Media.**

  Data. The coefficient of TFPR on data from consumers is equal to 0.045, suggesting an increase in 4.5% in TFPR.

  Model. I repeat the same model regression using model TFPR, which combines productivity and targeting strategy. The difference in TFPR given by data is captured by α^{TFPR}:

$$\log\left(TFPR(z,\lambda)\right) = \alpha_0 + \alpha^{TFPR}\mathcal{I}\left[\lambda = \lambda^D\right] + \sum_{k=1}^{5} \alpha_k \mathcal{I}\{\text{Quintile k}\} + \epsilon_{z,\lambda}$$

$$TFPR(z,\lambda) \equiv e^z \int_\varepsilon t(\varepsilon; z,\lambda)\, f(\varepsilon)\, p(\varepsilon)$$

- **Advertising Employment Share**

  Data. To compute the denominator I decompose it in the following steps:

$$\frac{\text{Advertising Labor}}{\text{Total Labor}} = \frac{\frac{\text{Advertising Labor}}{\text{Digital Advertising Labor}}}{\frac{\text{Total Labor}}{\text{Digital Advertising Labor}}}$$

  The numerator is the inverse of the digital advertising share, which is computed below (0.46). All that is left is to denominator, which is the inverse of the share of employment allocated to digital advertising. With this purpose, I use the exposure to digital technologies from Prytkova et al. (2024), who map patents from detailed digital technologies to the tasks associated to jobs.[36] They build an index of exposure to different technologies by occupation, including *digital advertising*. I take the following steps:

  1. *Relative Exposure to Digital Advertising.* For every 2-digit ISCO-08 (i), I divide the exposure index to digital advertising x_i^{da} by the total exposure to digital technologies (k), obtaining the relative exposure to digital advertising $\tilde{x}_i^{da}$:

$$\tilde{x}_i^{da} = \frac{x_i^{da}}{\sum_k x_i^k}$$

     Top occupations in relative exposure to digital advertising are *Administrative and commercial managers*, *Business and administration professionals* and *Sales workers*.

  2. *Employment Share in Digital Advertising.* I combine this measure with employment shares per 2-digit occupation for the period 2015-2022 in France from the Labor Force Survey, and with

[36] Their database is publicly available at https://github.com/FabienPetitEconomics/TechXposure.

the exposure to overall digital technologies for French workers.[37]

$$da_t = 0.315 \sum_i s_{it} \tilde{x}_i^{da}$$

I combine the relative exposure to digital advertising with the employment share by occupation s_{it}, which gives an aggregate measure of the relevance of tasks related to digital advertising in the overall economy, relative to other digital technologies. Using the LFS ad-hoc module for 2022, I obtain that 31.5% of workers in France use digital technologies intensively across tasks, so I multiply by this number to obtain the share of total labor allocated to digital advertising per year. According to this measure, the share of employment exposed to digital advertising has increased from 0.82% in 2015 to 0.89% in 2022, where the average for the whole period is equal to 0.85%. Therefore, using the evidence from Prytkova et al. (2024), I obtain the following:

$$\frac{\text{Digital Advertising Labor}}{\text{Total Labor}} = 0.0085$$

Returning to the previous ratios:

$$\frac{\text{Advertising Labor}}{\text{Total Labor}} = \frac{\frac{\text{Advertising Labor}}{\text{Digital Advertising Labor}}}{\frac{\text{Total Labor}}{\text{Digital Advertising Labor}}} = \frac{\frac{1}{0.46}}{\frac{1}{0.0085}} = \frac{0.0085}{0.46} = 0.0185$$

Model.

$$\frac{\int_z \sum_\lambda \lambda \int_\varepsilon t\left(\varepsilon; z, \lambda\right) \log\left(\frac{t(\varepsilon; z, \lambda)}{g(\varepsilon)}\right) d\varepsilon dz}{L}$$

- **Digital Advertising Share.**

  Data. From Statista Market Insights.[38] In the section of Digital Advertising, Statista computes the share of Digital Advertising per year in France, starting from 33% in 2017 to 58% in 2022. The cross-year average from 2017 to 2022, yields a share of digital advertising equal to 46%.

  Model.

$$100 \times \frac{\int_z \lambda^D \int_\varepsilon t\left(\varepsilon; z, \lambda^D\right) \log\left(\frac{t\left(\varepsilon; z, \lambda^D\right)}{g(\varepsilon)}\right) d\varepsilon dz}{\int_z \sum_\lambda \lambda \int_\varepsilon t\left(\varepsilon; z, \lambda\right) \log\left(\frac{t(\varepsilon; z, \lambda)}{g(\varepsilon)}\right) d\varepsilon dz}$$

- **Adoption Rate of Social Media.**

  Data. The average adoption rate of data from consumers is 5.1%.

  Model. I compute the adoption rate across the marginal distribution of productivity realizations,

[37] See https://ec.europa.eu/eurostat/databrowser/view/lfsa_egai2d__custom_12764446/default/table?lang=en for the employment shares and https://ec.europa.eu/eurostat/databrowser/view/lfso_22aswt01/default/table?lang=en&category=labour.employ.lfso.lfso_22.lfso_22aswt for the use of digital technologies at work, which is only available in the LFS 2022 ad-hoc module for the whole workforce in France, where I focus on workers that use all or most of the working time using digital devices across different tasks. I exclude occupations related to armed forces, as the digital exposure variable is not available.

[38] Source, accessed 24 March 2025: https://www.statista.com/outlook/amo/advertising/france?currency=USD#digital-ad-spending

where $\mu^z(z)$ is the marginal distribution of productivity

$$100 \times \left[\int_z \frac{\mu\left(z,\lambda^D\right)}{\mu\left(z,\lambda^D\right)+\mu\left(z,\lambda^T\right)}\mu^z(z)\,dz\right]$$

- **Std. Dev. of Adoption Rate per Quintile.**

  Data. The adoption rate across the five employment quintiles: 4.4%, 4.1%, 5.4%, 8.4% and 14.1%. The standard deviation of these five numbers is equal to 3.7.

  Model. I obtain the five quintiles of the equilibrium employment distributions, and for each quintile k I compute the following adoption rate for firms within that quintile:

$$\mathrm{R(k)} = 100\left(\frac{\int_z \mu\left(z,\lambda^D\right)\mathcal{I}\left[l_p\left(z,\lambda^D\right)\in \text{Quintile k}\right]dz}{\int_z \mu\left(z,\lambda^D\right)\mathcal{I}\left[l_p\left(z,\lambda^D\right)\in \text{Quintile k}\right]dz+\int_z \mathcal{I}\left[l_p\left(z,\lambda^T\right)\in \text{Quintile k}\right]dz}\right)$$

  Then, I compute the standard deviation of the adoption rate in the 5 quintiles: R(1), R(2),...,R(5).

C.3 Additional Quantitative Results

Baseline model: share of ads from Big Data. Figure C.1 shows the share of ads received that come from Big Data firms by consumer valuation. While the adoption of Big Data is around 6% in the model, adopters are more productive than the average, and represent a larger share of aggregate production. There is substantial heterogeneity across consumer valuation: low-valuation types receive around 7% of the ads from Big Data firms, while high-valuation types receive around 14%.

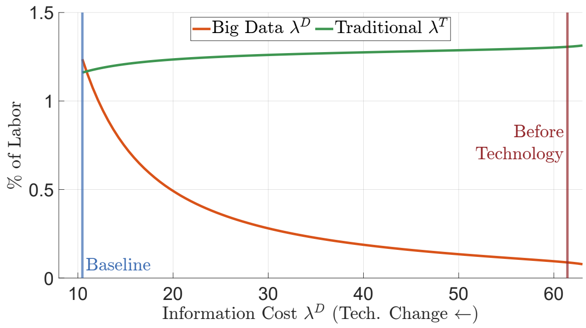
Figure C.1: Consumer Valuation and Ads from Big Data firms

Note. This figure shows the share of ads received $a(\varepsilon)$ that come from Big Data firms in the baseline calibration.

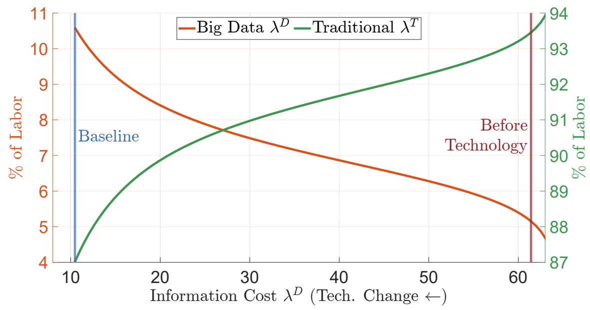
Technological change and employment distribution. I solve the model for different values of the cost of processing Big Data, ranging from the baseline calibration to the version of the model before technological change. For each of the versions of the model, I compute the share of total labor from

Figure C.2: Technological Change and Employment

(a) Targeting Cost and Adoption of Big Data



(b) Employment in Production (l_p)



Note. For each of the values of the information cost for Big Data λ^D, Subfigure (a) shows the share of labor into targeting or adoption costs for firms with both technologies. Subfigure (b) shows the share of production employment.

targeting and adoption costs in Figure C.2a, and the share of employment in production in Figure C.2b. When the cost of processing information decreases, both production and targeting expenditures increase for Big Data firms.

Policy counterfactuals and employment distribution. Table C.1 shows the distribution of employment in the baseline model, joint with each of the policy counterfactuals. Digital advertising is reduced if any tax is in place, while it is increased if the GDRP is lifted.

Table C.1: Policy Counterfactuals and Labor Decomposition

	Baseline	**Lift GDPR**	**Digital Advertising Tax**	**Advertising Tax**
Digital Advertising (Big Data)	1.2%	1.7%	0.7%	0.7%
Traditional Advertising	1.1%	1%	1.2%	0.7%
Total Advertising	2.3%	2.7%	1.9%	1.4%
Production Big Data	10.6%	14.7%	9.9%	10%
Production Traditional	87%	82.6%	88.2%	88.6%
Total Production	97.6%	97.3%	98.1%	98.6%
Total	100%	100%	100%	100%

Note. This table shows the impact of lifting the GDPR, a digital advertising tax of 30% and a general advertising tax of the same rate. For each counterfactual, total labor is split between production and advertising, and within each type of labor, further into Big Data and Traditional firms. Digital advertising labor includes the cost of adopting Big Data.