

# Data and Markets

Maryam Farboodi\* and Laura Veldkamp<sup>†</sup>

November 22, 2022

## Abstract

Big data is changing every corner of economics and finance. The largest firms in the U.S. economy are valued chiefly for their data. Yet, this data is largely excluded from macroeconomic and finance research. We review work and relevant tools for measuring economic activity, market power, data markets and the role of data in financial markets. We also highlight areas where future work is needed.

Data is changing every corner of economics. It changes how goods and services are provided and priced, how firms compete, how financial assets are priced and traded, how firms grow, as well as notions of privacy and the structure of markets themselves. We summarize insights from macro, industrial organization and finance about how data is used, valued and traded.

There are a multitude of dimensions where big data creates value for the economy. Data is a tool for prediction, thus, it helps reduce the risk exposure of investors. Analyzing big data can lead to creating different customer profiles and result in customizing products to fit specific market segments, which benefits both firms and customers. Algorithms that analyze big data can replace manual decision-making, thereby optimizing processes and improving accuracy through automation. Finally, by analyzing the data related to user behavior, companies can discover patterns that can identify the need for a new product or an upgrade of an existing one. Thus, big data fosters innovation as well.

---

\*MIT Sloan School of Business, NBER, and CEPR; farboodi@mit.edu.

<sup>†</sup>Columbia Graduate School of Business, NBER, and CEPR, 3022 Broadway, New York, NY 10027; lv2405@columbia.edu.

# 1 What Is Big Data?

Big data refers to the large volumes of data, often from multiple sources, and the ability to gather, store and process it to produce new kinds of observations, measurements and predictions about individual customers. [NIST \(National Institute of Standards and Technology\)](#) defines big data as: “Big Data consists of extensive datasets – primarily in the characteristics of volume, velocity, variety, and/or variability – that require a scalable architecture for efficient storage, manipulation, and analysis.” Importantly, big data not only refers to the rapidly growing digitized datasets, but also to the accompanied technological innovation that is necessary to process, analyze and manage them.

The past decade has witnessed enormous growth in the size of the big data market. [Kolanovic and Krishnamachari \(2017\)](#) report the value of the global market for big data and related technology and analytics to be \$130bn in 2017 and expected to grow to over \$200bn by 2020. Importantly, these resources are not only spent on big datasets but also on complementary technology and skilled labor.

Big data is associated with various types of data providers and skill laborers. [Kolanovic and Krishnamachari \(2017\)](#) report that in this market, data providers can be classified into three groups, based on the type of data that they provide: raw data providers, providers of semi-processed data, and providers of signals and reports. They also identify three different types of intermediaries in the financial data market, given the services that they provide. First, there are “data intermediaries” who collect aggregate data from the numerous alternative data providers and channel them to investors through a centralized portal. Second, there are “technology intermediaries” who offer technology solutions to clients. These solutions include public, private or hybrid cloud architecture and computational services. Lastly, many consultants have specialized to advise firms in the process of onboarding big data and its legal issues.

[Martin \(2020\)](#) provides an industry-based categorization of the costs and benefits of big data and big data technologies. Despite the fact that benefits and costs differ across industries, most benefits involve some form of improvement in quality of products and services and customer access, while most costs can be regarded as some form of discrimination or privacy violation. We will use this simplified classification in what follows.

What companies learn through big data can be used to design products and services that deliver more value to the individual consumer and/or enable the consumer

to more effectively find his desired products. At the same time, without sufficient competition, a seller's ability to better predict consumers' willingness-to-pay can lead to price discrimination and decrease consumer surplus. Alternatively, consumers might view their privacy violated when firms collect a lot of data about their behavior.

## 2 Measuring GDP in a Data Economy

The rise of the data economy brings with it new challenges for the measurement of economic activity. Lots of online activity generates consumer surplus that is not priced. Time-saving convenience is typically not accounted for in quality adjustments. Most importantly, many physical goods and services may be under-priced relative to their consumer value because they are paid for, in part, with data.

Most of us have a phone with apps that were "free." Weather apps, clocks, flashlight apps, games, and thousands of other software apps cost money to develop and to advertise, but are given away to customers at a monetary price of zero. But such apps are not truly free. The reason these businesses profit and persist is that they collect customers' data and use and sell that data. App users are not the client; the businesses that want user data or user attention are the paying client. Still, app users get a valuable digital service, in exchange for letting the app record and sell their data. Paying for a service with data is a barter trade. It is two non-monetary instruments being exchanged at a zero monetary price. [Brynjolfsson, Collis, Diewert, Eggers and Fox \(2019b\)](#) introduce the concept of GDP-B, which measures the welfare contributions of these goods that are bartered for data at zero price. Using choice experiments, they estimate that Facebook adds 0.05-0.11 percentage points to U.S. GDP-B growth per year. Similarly, experiments by [Brynjolfsson, Collis and Eggers \(2019a\)](#) reveal that losing access to all search engines or all email services for one year is equivalent to earning \$500-\$1,000 less per year, on average.

Another category of mismeasurement is digital service innovations. [Byrne and Corrado \(2019\)](#) propose a framework for measuring improvements in consumers content provision. They find that content delivery services increased consumer surplus by almost \$1,800 per connected user per year and contributed over one-half percentage point to US real GDP growth in the last 10 years.

Not only are there free digital goods, but digital commerce has offered consumers a more efficient way to acquire traditional goods. This is an unmeasured improvement in quality. This increased efficiency in provision is bundled with the physical good or

service. [Hulten and Nakamura \(2017\)](#) model information goods as “output-saving.” Information technology like e-commerce reduces the need for transportation, for example. That transportation would otherwise be captured in GDP. However, since the electronic component of e-commerce itself has no explicit price, it is not counted in GDP. The authors propose expanded GDP (EGDP) that combines the conventional GDP measure with a measure of willingness-to-pay for output-saving innovations.

A free app is an obvious example of trading data for a good or service. But many instances of data barter are hidden. Most transactions generate some data as a by-product. For digital transactions, this might involve linking current transactions to past ones, matching them to a delivery address and using this information to build a user profile. Even for transactions in physical stores, credit card companies harvest the transaction data and sell marketing services based on it. Those same credit card companies may offer cash back or other rewards that effectively lower the price of a purchase. Even a cash transaction in a physical retail outlet generates data for the store owner about what local demand is for her products. That store owner can use this information to stock the right goods and operate more efficiently. One clear example of this sort of partial data barter is the discount offered to Whole Foods customers who link their purchases to their Amazon Prime accounts. A customer who produces a QR code at checkout will have their grocery cart purchases linked to their Amazon profile. In return, they receive a 5 or 10 percent discount on their groceries. That grocery discount is paid for with customer data.

In each of these transactions, a good was sold, in exchange for a monetary payment and data. In each of the transactions, valuable data was transferred from the purchaser to the seller. A partial barter trade is a good or service that is provided in exchange for monetary payment and another valuable asset. Almost every transaction in our modern economy is therefore a partial barter trade. This idea has enormous implications for measurement as it implies that the price of goods and services does not capture their value. Price is no longer consumers’ willingness to pay. Consumers pay for goods and services with a payment, plus data. When we construct measures of aggregate activity by multiplying the quantity times the price of goods and services, we are systematically underestimating their value because we miss the value of the data barter part of the transaction. For some given transaction, that may be a small error. However, the enormous valuations placed on the value of firms that collect and monetize customer data suggests that the aggregate value of all this data being transacted is not small at all.

If every grocery book-purchase and book-sale is paid for partly with data, then

firm earnings are also not a reliable indicator of firm value. Correcting the value of transactions for the unmeasured data barter component may help us measure aggregate activity and value firms more accurately. [Farboodi and Veldkamp \(2022\)](#) build a recursive model where valuable data is a by-product of every transaction, and all trades of physical goods are partial barter trades. That model gives rise to a value function for each firm that values the earnings and the data of a firm. Quantifying that model could provide insight about the magnitude of the missing GDP.

But how does one correct GDP or firm valuations for this data barter? One approach has been to survey consumers about their willingness-to-pay for digital goods. For example, [Allcott et al. \(2020\)](#) asked people how much they would need to be paid to give up access to Facebook for four weeks. That approach is a substantial improvement over assigning Facebook's service zero GDP weight, corresponding to its monetary price of zero.

However, even this survey approach may not fully capture the part of Facebook's value that is bartered for one's data. Suppose a user values keep in touch with friends using Facebook at \$1,000, but perceives the cost of losing their privacy at \$995. The user would report a \$5 willingness-to-pay for Facebook. However, if we think of data as a form of payment, this captures only the value of Facebook, net of the data payment. It is the residual surplus, being counted as if it were the entire value of the service. Instead of counting Facebook as if it has a low \$5 value for this consumer, a measure of economic value created like GDP should count this as provision of a high-value service, at a high cost, paid for with \$995 worth of data.

In order to fully measure the value of goods and services partially bartered for data, we need tools to value data. Most of these tools involve using a structural model to map observable variables such as sales, firm equity values or the hiring of data workers, into unobserved quantities of interest, such as a firm's data stock. That is a topic we return to in [Section 7](#).

### 3 Data Privacy

While data offers economic value, it can also be used in ways that violate consumer privacy. Recent research explores whether it is optimal to give consumers more direct control over sharing the data that they generate through their economic transaction, or allow them to require direct compensation for sharing their data. The ownership rights of transaction data is an important topic of debate and lies at the heart of a

number of recent data regulations, such as the CCPA enacted by the state of California and the GDPR enacted by the European Union.

The California Consumer Privacy Act of 2018 (CCPA) is a state statute intended to enhance privacy rights and consumer protection for residents of California, US. It gives consumers more control over the personal information that businesses collect about them. The CCPA regulations also provide guidance on how to implement the law.

EU General Data Protection Regulation (GDPR) is a regulation that went into effect on May 25, 2018 in the European Union. It is aimed at guiding and regulating the way companies across the world will handle their customers' personal information and creating strengthened and unified data protection for all individuals within the EU.

[Ali et al. \(2022\)](#) consider the trade-off between product personalization and price discrimination through personalized pricing, and find that giving consumers control over data sharing can be beneficial to consumers in both monopolistic and competitive markets. However, giving consumers the option to share data might be insufficient if consumers can be tempted as defined in [Gul and Pesendorfer \(2001\)](#). These consumers get tempted by the products that they observe and buy it only because they suffer a mental cost from resisting temptation. [Liu et al. \(2020\)](#) use this framework to examine the welfare implications of different data sharing schemes in an economy populated by agents who have a preference for privacy but have heterogeneous degrees of self-control. They find that when the temptation problem of consumers without self-control is sufficiently severe, a policy with an option to opt-in and opt-out might not be adequate enough to correct the externality and no data sharing can improve welfare.

The empirical evidence for the impact of privacy concerns on consumers' online behavior is ambiguous. On the one hand, [Goldfarb and Tucker \(2011\)](#) use data from a large-scale randomized field experiment on 2,892 distinct web advertising campaigns to explore what influences the effectiveness of online advertising. They find that matching an ad to website content and increasing its obtrusiveness independently increase purchase intent. However, the two strategies are ineffective in combination, and the negative effect is strongest for people with the highest preference for privacy. On the other hand, there is more recent empirical evidence of the "privacy paradox", the dichotomy between how a person intends to protect their online privacy versus how they actually behave online — and how they don't protect their information online. [Chen et al. \(2021\)](#) combine the results of a survey conducted by Alipay about users'

concerns regarding data sharing with Alipay’s mini-programs with their actual visits to the mini-programs and their authorized data sharing activity, both before and after the survey. They do not find any statistically significant difference in the number of data authorizations across users with different levels of privacy concerns regarding data sharing.

## 4 Increasing Returns Market Power

In this section, we explore the origin and nature of the increasing returns created by data economics. We discuss how the data feedback loop creates increasing returns and superstar firms, its consequences for market competition and the effect of data on how that competition is measured.

### 4.1 Data Feedback Loop

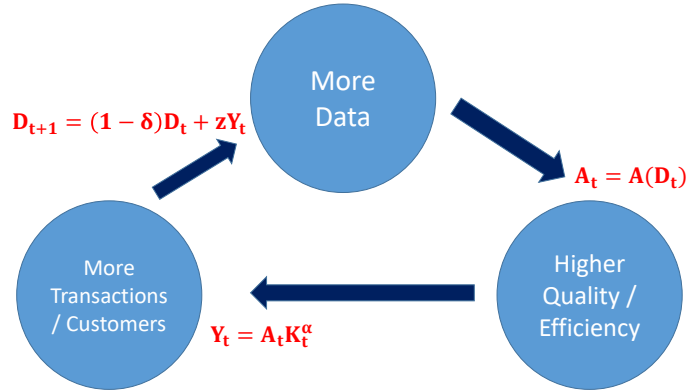
Like other forms of information, data has returns to scale. That means big firms extract more value from the same data than small firms could. The big firm can use a data set to reach many customers, optimize many departments or tweak many products.

At the same time, data encourages firm growth. [Brynjolfsson and McElheran \(2016\)](#) design the first-ever Management and Organizational Practices Survey to investigate the impact of data-driven decision making. They find that increasing the usage of data-driven decision making is linked to a statistically significant boost in productivity of 3% or more on average. But then, if firms grow and big firms benefit more from data, this dynamic is self-reinforcing. In other words, there are increasing returns to firm growth and data growth. This self-reinforcing dynamic is known as the *data feedback loop*.

The data feedback loop can be summarized in three intuitive equations. The first equation is the simplest form of a production economy, an AK model, where output is produced with a productivity level  $A_t$  and capital  $K_t$ :

$$Y_t = A_t K_t^\alpha.$$

Figure 1: Data Feedback Loop



The second equation links the data that a firm already owns,  $D_t$ , to its productivity:

$$A_t = A(D_t).$$

The third equation says that new data is generated when goods are produced and sold. This new data is added to the stock of existing data, which depreciates at a rate  $\delta$ :

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

The parameter  $z$  simply scales the number of units of sales  $Y_t$  into the right measure of data. This simple structure is embedded in the richer dynamic models of [Farboodi and Veldkamp \(2022\)](#) and [Jones and Tonetti \(2020\)](#).

This incredibly simple three-equation model is sufficient to generate superstar firms. A firm that starts out slightly larger will generate more data in the first period. That allows this firm to grow more quickly, generate even more additional data, relative to the smaller firm, and take a divergent growth path. Of course, data may well have diminishing returns. The first few data points are eye-opening, while the millionth adds little insight. This can be built in as a concave  $A(\cdot)$  function that maps data to productivity. Yet, as [Farboodi and Veldkamp \(2022\)](#) show, the increasing returns force may still dominate while firms are small, causing firms to diverge in size, giving rise to superstars.

The upside of diminishing returns to data is that, eventually, other firms may accumulate enough data to catch up. If data has diminishing returns, there *should* be convergence. This is the same logic as to why there should be convergence among countries' income if there are diminishing returns to capital. In the cross-country growth context, the models teach us that there should be convergence. And yet, there is not. Growth theorists debate what features keep poor countries poor. A similar



debate may arise about what keeps small companies small. Initially, data may be a cause. But in the long run, diminishing returns to data mean that other factors which work to preserve large size, like market power or incumbency advantage, are likely to work.

Pieces of this three-equation model appear in many corners of economics. The idea that information or data is generated from economic activity appears in many models. More specifically, work on information frictions in business cycles ([Caplin and Leahy, 1994](#); [Veldkamp, 2005](#); [Lorenzoni, 2009](#); [Ordonez, 2013](#); [Ilut and Schneider, 2014](#); [Fajgelbaum et al., 2017](#)) has versions of a data feedback loop that operate at the level of the aggregate economy: more data enables more aggregate production, which, in turn, produces more data. But these models do not capture the idea that data is an asset of a firm. [Wilson \(1975\)](#) writes about informational economies of scale. His simple model captures the part of the data feedback loop where large firms grow larger through benefiting more from the data. Furthermore, while the simple model sketched here embodies the idea that large firms generate more data, it is not necessarily the case that they choose to purchase more data. In fact, they might specialize in selling data and data services, an idea explored in [Farboodi \(2022\)](#).

Related ideas in management preceded this in sketching out the logic of the data feedback loop. [Agrawal et al. \(2018\)](#) describe the dynamics of the new digital economy as follows: “By launching sooner, the Amazon flywheel can get ahead as better predictions will attract more shoppers and more shoppers will generate more data to train the AI, more data will lead to better predictions and so on, creating a virtuous cycle.” This is also related to the idea of tournament-like or winner-take-all payoffs that can also arise in a data economy, where everyone has access to and a preference for the highest-quality good. [Agrawal et al. \(2019\)](#) take this logic further and describe a data economy in which a slight head start in data or product quality enables a firm to take over the entire market for a good or service.

**A financial data feedback loop.** A related form of feedback loop works through financial markets. In [Begenau et al. \(2018\)](#), financial investors use data about a firm to forecast that firms’ future earnings more accurately. More accurate forecasts make investing in that firm less risky (uncertain) to the investors. Thus, more data about a firm lowers the risk premium on the firm and lowers its cost of capital. Such firms can expand more cheaply and grow faster. As a larger firm generates more data, firm growth makes learning about the firm more valuable. Hence growth spawns data, which lowers risk premia and prompts more growth. This is similar to the product

market data feedback loop, with all the same implications of increasing returns. But this mechanism works through the pricing of risk in financial markets. Both feedback loops may operate concurrently.

**Directions for research: Heterogeneity.** The effect of big data on firm dynamics and the emergence of data feedback loop in the presence of firms that are heterogeneous in different dimensions is an unexplored area of research. In most real world situations, data that is produced by one firm is only imperfectly correlated with other firms' productivity, even for firms who are in the same market. In other words, different firms can use the same piece of data to varying degrees. Moreover, firms also use data that is produced in markets different from the one that they directly operate in. This cross-usage of data can also interact with the competitive structure of different markets, a topic that we discuss in more detail in Section 4.

## 4.2 Superstar Firms

This increasing returns dynamic that fuels the growth of the largest firms offers one potential rationale for the phenomenon of the superstar firms. "Superstar" is a term that was first applied to companies by the late University of Chicago economist Sherwin Rosen to describe a phenomenon in which relatively few people earn enormous amounts of money and dominate the activities in which they engage (Rosen, 1981). The same idea applies to firms. The revenues of large companies often rival those of national governments. In a list combining both corporate and government revenues for 2015, ten companies appear in the largest 30 entities in the world, and only eight countries in the world have national governments that generate as much revenue as Walmart does (Zingales, 2017).

There is also enormous heterogeneity in the amount of digital capital that companies own. Tambe et al. (2020) find evidence of striking firm-to-firm heterogeneity in digital capital value, with most of the value concentrated in a small group of superstar firms with market values in the top decile. They write that, "By 2016, the stock of digital capital accounted for about 25% of total capital stock for firms in our sample." They further conclude that inequality in digital capital among firms is growing. The top firms' digital capital is growing at a faster rate than the rest.

According to Mitchell and Brynjolfsson (2017), data and data technology are the most important factors in explaining the superstar firm phenomenon. To formalize

this idea, [Benzell and Brynjolfsson \(2019\)](#) developed a model to explain why most workers and capital owners have not reaped the benefits of digitalization, while a few superstars have. These firms own most of the data and have benefitted from increasing returns.

An alternative key open question for research is how much of the superstar firm phenomenon might possibly be due to firms' accumulation of data. Since firm size and data stocks clearly have two-way causality, a structural model or a clever identification strategy will be needed to identify the causal effect of data on the size of large firms.<sup>1</sup>

### 4.3 Data and Market Competition

As big firms grow bigger, many economists fear that market competition is suffering. [Gutierrez and Philippon \(2017\)](#) find that investment has been low relative to profitability since the early 2000's and argue that decreased competition is a key driver of this investment gap. Others simply point to the growing market superstars we discussed above as evidence of declining competition.

[De Loecker et al. \(2020\)](#) find that evidence of trends in competition depend on how one measures product markups. Sales-weighted firm markups are skyrocketing. The largest firms with the most sales seem to be making larger profit margins. On the other hand, cost-weighted markups rise much more slowly. Firms that produce high-cost products seem to have similar competitive margins now, as they did in the past. They conclude that the rising markup phenomenon arises from a shift in the composition of production to large, low-cost firms.

Core digital services exhibit strong network effects as well as (partly data driven) strong economies of scale. Network effects and economies of scale shield those dominant firms from competition, enabling them to extract a significant proportion of the surplus that their presence in the market generates ([Crémer et al., 2021](#)).

Others dispute the idea that data undermines competition. [Lambrecht and Tucker \(2015\)](#) argue that in order for data to provide a lasting and harmful barrier to competition, it would have to be difficult to replicate, rare and difficult to substitute. Most transaction data do not satisfy these features. Furthermore, they provide examples of data enabling firms like Uber or Airbnb to enter and disrupt traditional markets where rents were high.

---

<sup>1</sup>One particularly interesting form of a superstar firm is a data platform. We discuss data platforms separately in Section 5.

The open research question is a quantitative one: How much does firms' use of data contribute to this trend? While a complete answer to this question will likely span many future papers, one starting point is to calibrate one of these existing models, to see how much divergence in firm size might be plausibly caused by data.

**Data and imperfect competition: Frameworks for measurement.** In order to answer questions about how much data contributes to trends in market competition, new frameworks are needed to explore how data and market competition relate. Simply examining empirical trends, even with well-identified shocks of some kind, doesn't reveal the answers if we do not know what evidence of data's effects to look for.

One reason that data's competitive effects might be difficult to detect or tease out is that data affects three aspects of a firm, all at once: i) size, ii) efficiency and iii) risk. From the discussion of the data feedback loop, we know that data can encourage firms to grow larger, and larger firms can generate more data. Such large firms may have the power to distort markets in ways that harm consumer surplus. A review article by [Goldfarb and Tucker \(2019\)](#) explored the second aspect, efficiency. It details ways in which data helps firms optimize business processes. This optimization helps firms to improve their efficiency and reduce their costs, which may show up as a higher markup, but is not detrimental to consumers.

Risk reduction, the third piece of data's effect, is perhaps the least explored. The financial feedback loop in [Begenau et al. \(2018\)](#) exhibits how big data favors large firms in investors' investment choices by disproportionately reducing their risk, as large firms generate more data. As the effects of risk and uncertainty are typically the domain of finance, they are often neglected in models of market competition. But data is a tool for reducing risk, by making the future less uncertain. As such, an alternative risk reduction channel works through how firms use data themselves: Data helps firms make better forecasts, which reduces their uncertainty about the future.

Why is firm's risk consideration an important feature to consider when measuring market competition? The effects of risk on firm behavior can mimic anti-competitive behavior. A firm that is uncertain about the future may behave more conservatively and scale back production. A monopolist would also scale back production, for different reasons. An uncertain firm should only make a new investment or launch a new product if the profits were sufficient to compensate for the risk being borne. This is not just a hypothetical consideration. It is a bedrock principle of corporate finance. Nearly every MBA program in the country teaches prospective managers to do this risk-return trade-off. The field of empirical corporate finance has documented repeatedly that such risk

is indeed priced (Eckbo, ed, 2008). This means that firms facing uncertainty should earn higher margins that are actually compensation for risk, but could mimic the high margins earned by a monopolist.

Eeckhout and Veldkamp (2022) investigate how data affects competition, through all three of these channels. They introduce data into an imperfect competition framework, with firms that price risk. Data makes firms choose to be large; it allows them to produce more efficiently by producing the goods consumers want most; and it reduces firms' risk. When all firms acquire more data at the same rate, the net effect is unambiguously good for welfare. Data improves efficiency and resolves uncertainty that is a deadweight loss. However, when one firm has much more data than others, allowing this firm access to even more data can be good or bad for welfare. We learn that data is not a problem for competition, but data asymmetry may be.

Another mechanism by which firms can use data to become superstars is through advertising. Prat and Valletti (2021) explore how firms use data for advertising. Platforms with monopoly power can extract large rents, they argue, by preventing firms that do not pay them from matching with customers. The idea that data is used for matching and that holders of such data may act as gatekeepers for firms' access to customers is a very different use of data than firms who use data to innovate, choose products or improve business processes. This is more similar to the grim view of data's use in inhibiting competition introduced in Morton and Dinielli (2020).

Alternatively, Gan and Riddiough (2008) model a lender's incentives to protect their competitive data advantage in a home mortgage market. They describe how an incumbent could deter new entrants by quoting loan rates that do not reveal much of what they know about the potential borrower. This use of data is quite different from a platform acting as a gatekeeper, which, in turn, is quite different from firms using data to forecast well and produce lower-cost goods than others. Each of these data strategies likely has its own unique competitive effects. This raises the possibility that we will need different frameworks to analyze competitive behavior, depending on the way in which data is used.

**Directions for research.** So far, the research on data and market competition has assembled many facts and ideas. However, this agenda has yet to paint a complete picture of the role of data in markets. That is to say, there is lots more research for economists to do. One open question is how firms optimally produce, acquire and use data to improve their competitive position when data is long-lived. The insights above are mostly static. However, lots of valuable business data is about the characteristics

of customers or suppliers, features that do not change rapidly over time. Such data can be accumulated, stored and gradually depreciated. Just like the study of investment and firm growth requires dynamic, recursive models, so too does the study of data investments and its effect on a firm's competitive position.

In particular, one phenomenon that a dynamic approach could speak to is firms that sell goods or often digital services, below cost. When we discussed GDP measurement, we explained how data is being bartered or partially bartered for goods. Adding this possibility to a dynamic model with imperfect competition could be important for two reasons. First, selling cheaply or giving goods away in exchange for data could be a competitive strategy for a firm that wants to build their data stock quickly. Second, if customers are paying for goods or services partly with their data, the nominal price may be low. Traditional tests for anti-competitive behavior that look for high margins may not detect firms' rents and market power. Conversely, these low monetary prices might look like predatory pricing, which is a well established form of barrier to entry and a failure of competition (Demsetz, 1982). But if goods and services are paid for with data, as well as money, the price may not be predatory after all. The question then arises: Where should economists look to see if competition is strong? What does anti-competitive behavior look like in a data economy? Many legal scholars opine on this topic. But formal models, with markets and equilibrium effects, would add to the debate and enable new forms of measurement.

## 5 Data Markets

Large firms with a lot of transactions, such as digital platforms, have a comparative advantage in data production, which they can then use internally or sell to other firms, giving rise to an active market for data among firms, the *inter-firm data market*.

In this market, two critical characteristics make data different from other intangibles, even those that are a by-product of production such as learning-by-doing. First, data is tradable. Second data is non-exclusive (or non-rival). The firm who produces the data can sell it and still partially use it, thus, each unit of data is worth more than one unit. It can help multiple firms improve their products. As such, a benevolent social planner always wants some degree of data sharing among firms. Farboodi (2022) explores the data markets among firms with these features. It finds that when the data market is frictionless and perfectly competitive, identical firms always choose to trade data even in a stationary steady state.

Furthermore, the paper illustrates how *data platforms*, platforms used for collecting and managing data, emerge endogenously in a data economy. When some firms have a comparative advantage in data collection, they are larger and support more transactions. These firms specialize in making profits from monetizing data services to other firms on the data market, rather than selling high quality products on the goods market, and data platforms emerge.

An important dimension of data production and inter-firm data sharing is that the information advantage that the firms get from data collection can lead to non-competitive behavior, given the market structure. Firm non-competitive behavior introduces new forces in the data market. It can lead to market frictions or even a break down of the data market, which in turn necessitates regulating these markets. This is an under-explored area of research.

A second potential market for data is the one between firms and consumers, *consumer-firm data markets*. A number of papers have investigated this market by using frameworks that feature buyers, sellers, and intermediaries.

Consider how firms and consumers divide the surplus from their economic activity. The wealth of data that is generated as a by-product of consumer-firm transactions generates a lot of surplus. Part of the surplus accrues to consumers and improves their welfare if firms use data to improve their products and services and/or firms and intermediaries target consumers more effectively. On the other hand, data can hurt consumer and even the aggregate welfare if it gives firms too much market power, or gives the intermediary excessive advantage versus firms or consumers themselves. On the firm side of the platform, consumer data gives an information advantage to the intermediary and hurts the outside option of the sellers. On the consumer side, a given consumer's data can be used to price discriminate against him and/or against other consumers, using personalized pricing.

A prime example is the utilization of consumer analytics for sales and marketing. Customer analytics is the process through which customer data is collected and processed to identify, captivate, and retain customers by creating personalized products, services and experiences. Consider e-commerce: analytics provide the company with data-driven insights into how shoppers interact with its website. When shopping or looking for a movie online, results are often displayed in an order that reflects personal preference. We have also all encountered numerous advertisements that have a high degree of correlation with our previous online activity such as our shopping history, websites that we have visited, or even our emails. This can reduce the time that we spend shopping and improve our efficiency. At the same time, many users feel that this

degree of personalization violates their privacy. Furthermore, e-commerce platforms can engage in price discrimination against their users.

In an empirical study, [Shiller \(2013\)](#) provides examples of online platforms using consumer data non-competitively. He uses a structural model to study the pricing behavior of Netflix and finds that using data on user-specific web browsing to price subscriptions substantially increases Netflix's profits.

A number of recent papers have analyzed the welfare effects of digital platforms theoretically. In [Kirpalani and Philippon \(2020\)](#), sellers and consumers can either match directly (offline) or use a two-sided platform to match (online). In sharing their data with the intermediary, consumers face a trade-off: consumer data sharing on the platform improves the quality of the match but it gives the intermediary market power facing the sellers. Specifically, if the intermediary uses consumer data for goods production itself, consumer data sharing hurts the sellers. They show that consumers do not internalize this externality, which in turn leads to excessive information sharing with the platforms.

[Bergemann and Bonatti \(2022\)](#) explore a related trade-off: more data on the platform improves the match quality, but creates room for price discrimination. They use a model of a two-sided platform, with both an online and an offline market. The offline market limits the scope of price discrimination on the platform in that framework. [Bergemann et al. \(2022\)](#) study an economy where a monopolistic intermediary acquires data about customers with correlated preferences and (re)sells it to firms, who in turn use the data to personalize consumer prices in the product market. Consumers benefit from complete data sharing only if they learn about their own preferences from others' data, while they are hurt due to price discrimination. Thus, the data intermediation is inefficient unless it helps customer. However, producers always benefit from data sharing and pay the intermediary to share consumer data with them.

At the same time, strategic consumers can adjust their demand over time understanding that firms use their data to price discriminate against them. [Bonatti and Cisternas \(2020\)](#) explore the implications of price discrimination by a monopolistic firm when the firm uses the history of prior consumer transaction data for pricing. They find that price discrimination based on purchase histories unambiguously harms naive consumers but can benefit strategic consumers, as the latter consumers can influence the future prices that they face manipulating their demand.

Another predominant concern about data sharing and data market design for regulators has been consumer privacy, investigated in a number of recent papers. [Jones](#)



and Tonetti (2020) consider a model where data is a by-product of economic activity and is non-rival, thus some level of data sharing improves welfare. However, they show that when firms own the data, they engage in excessive data sharing and undermine consumer privacy. Alternatively, Acemoglu et al. (2021) show that when consumer valuations are correlated, data about one consumer depresses the price of other consumers' data, violates others' privacy and leads to excessive data sharing.

## 6 Financial Data

Big data and big data technology have permeated the financial industry more than most. Kolanovic and Krishnamachari (2017) estimate the spending of the investment management industry on big data to be in the \$2-3bn range, with double digit annual growth. One of the most important uses of big financial data is forecasting returns and risks. Investors use data to reduce the uncertainty of asset returns, adjust their portfolio holdings accordingly and increase their return. In a decision theoretic framework, Frankel and Kamenica (2019) show that the information content of a piece of data is equivalent to the corresponding expected reduction in uncertainty. Alternatively, Farboodi et al. (2022a) propose a structural approach to measure the quantity of investors' private data about different assets.

**Growth of big data technology.** When data technology improves, investors do not just process more of the same types of data. Instead, they seek out new types of data. Specifically, there has been a proliferation of trading strategies that make use of the order flow data of other investors. Farboodi and Veldkamp (2020) rationalize this trend. In their model, unbiased technological growth leads investors to gradually shift from processing data about fundamental asset valuation to data about investor demand or sentiment. The reason is that when data becomes more abundant, the value of competing with other informed investors falls but the value of seeking out ill-informed trades (“dumb money”) grows.

In the cross-section, firms do not benefit equally from big data technology growth. Farboodi et al. (2022a) uses a structural model to demonstrate that investors have acquired more data about large-growth firms in the past few decades. Alternatively, Begenau et al. (2018) shows that improvement in data technology can contribute to the disproportionate growth of large firms.

**Income inequality.** There is a large literature about increasing income inequality. A piece of the increase in income inequality comes from growth in financial income. One reason that rich investors might earn higher financial returns is because wealthier investors are likely to acquire more data (Peress, 2009; Kacperczyk et al., 2019). Of course, one way to overcome this disadvantage is for smaller investors to band together to collectively acquire data. This is the essence of a mutual fund. However, even in the presence of mutual funds, poorer investors still cannot catch up with their richer counterparts because finding a skilled fund requires search (Mihet, 2021).

Another strand of the literature focuses on labor income inequality and considers the effect of skilled-biased technological change on the skill premium (Krusell et al., 2000; Aghion et al., 1999; Acemoglu, 2002). Machine learning and AI are also new technologies that require new skills and could make older data skills obsolete. Abis and Veldkamp (2020) use job posting and wage data to track financial management firms' hiring of financial analysts with and without new data skills. They find that AI and machine learning have reduced the labor share of financial income. Just like the industrial revolution replaced artisanal labor with industrial equipment, new data technologies have replaced artisanal crafters of statistical models with optimized, computer-generated forecasting models. The magnitude of these changes has been similar.

**Data-driven credit analytics.** Customer analytics is also actively used for mortgage origination and credit extension. Lenders aggregate data on credit history with data on customers' online footprint such as purchasing behavior, social media profiles, job history to predict customer credit worthiness. It helps them figure out which customers are at risk of default and adjust mortgage/credit extension and rates accordingly. Berg et al. (2020) use data from an e-commerce company selling furniture in Germany to study the impact of using customer analytics for credit extension. The e-commerce platform uses both sources of information to predict creditworthiness of the buyers because goods are shipped first and invoiced for later. They find that customer analytics complements rather than substitutes for credit bureau information and affects access to credit and reduces default rates.

Big data in credit could result in more financial inclusion. Fintech promises to expand financial services to a broader set of customers in two ways: First, improvements in data availability and data processing technologies enable credit provision to customers who have traditionally been excluded. Second, algorithmic analysis that is completely based on the financial activity of the customer might someday offer equal

treatment regardless of race, sexual orientation, or gender. [Frost et al. \(2019\)](#) document that BigTech firms lend more in countries with less competitive banking sectors and less stringent regulation, and find that they serve unbanked borrowers in Argentina. However, they do note that this is early stage evidence and is not conclusive for the longer term impact. [Hau et al. \(2019\)](#) use data from Ant Financial in China and show that FinTech credit expands the extensive margin of credit to borrowers of lower credit scores, improving credit inclusion. Lastly, [Ouyang \(2021\)](#) finds that adoption of cashless payment considerably improves access to credit especially for the less educated and the elderly, on Alipay platform in China. A lot more data and research is necessary to make progress on the topic.

**Open Banking.** Customer transactions within a bank provides that particular institution with valuable data which then gives it a comparative advantage in pricing, marketing, and customizing financial services. In order to enable customers to reap more of the benefits from the data generated by their financial transactions and promote competition between traditional banks and Fintech entrants, several governments across the world have launched an initiative called open banking. Open Banking is a regulatory framework to make it easier for consumers to share their financial data with banks, Fintechs, and third-party providers (TPP) using application programmable interfaces (APIs). Put differently, it is designed to enable voluntary customer data sharing by themselves.

These initiatives are springing up globally, including the United Kingdom’s Open Banking Implementation Entity, the European Union’s second payment services directive, Australia’s new consumer protection laws, and Brazil’s drafting of open data guidelines. In the United States, the Consumer Financial Protection Bureau aims to facilitate a consumer-authorized data-sharing market, while the Financial Data Exchange consortium attempts to promote common, interoperable standards for secure access to financial data. The [official open banking website of the UK government](#) states that: “Open banking helps you move, manage and make more of your money. Opt-in to a world of secure apps and services for more clarity and control over your finances.”

[He et al. \(2020\)](#) use a theoretical model of credit market competition to study the impact of open banking with a focus on Fintech entry. They find that open banking regulation can help or harm welfare of borrowers even if it increases total welfare. In particular, data sharing can give new entrants an excessive information advantage, which in turn hurts borrower surplus. However, we are still in early stages of

open banking and there is scarcely any reliable data available about the corresponding outcomes among households in terms of financial access. Thus, very little is understood about the impact of open banking on customer welfare even in the short run.

Big data can also be used by financial institutions to better tailor their services and boost client retention or for fraud detection and increased security. Digital lending and Blockchain are two other areas of Fintech industry that rely extensively on AI, big data and big data technologies and have been active areas of research. In particular, due to the fast growth of cryptocurrency and blockchain technology in the financial sector, the term “Tokenomics” has been coined to describe the economic incentives governing crypto assets.<sup>2</sup>

**Directions for research: Insurance.** One segment of the financial industry which relies heavily on customer analytics is insurance. Insurance companies use big data to analyze a customer’s risk and determine which client is trustworthy or may default, and price accordingly. Similar to what we discussed in Section 5, customers in the insurance industry can be hurt by price discrimination. Note that the underlying premise of the insurance industry is risk sharing, which can be undermined by big data analysis. As such, the negative impact on insurance customers can be large. Assessing the effect of big data technologies on insurance outcomes is a fruitful area of

## 7 Data Valuation

Given the scale of global investment in big data and big data technologies, big data is a valuable asset. According to a 2019 McKinsey report ([Gottlieb and Weinberg, 2019](#)), companies with the greatest overall growth and revenue earnings are three times more likely than other companies to say that their data and analytics initiatives have contributed at least 20 percent to earnings before interest and taxes (EBIT) over the past three years. Nevertheless, data valuation has proven to be a difficult task. One reason is that disaggregated publicly available data is severely lacking.

Some researchers proxy for the value of data by quantifying IT investments. For instance, the Bureau of Economic Analysis reports that IT accounts for about 30% of all business investment ([Saunders and Brynjolfsson, 2016](#)). [Brynjolfsson et al. \(2002\)](#) and [Saunders and Brynjolfsson \(2016\)](#) use firm data to show that for the periods of

---

<sup>2</sup>*The Review of Financial Studies*, Volume 32, Issue 5, May 2019 focuses on the research in academic finance community on the topic.

1987-97 and 2003-06, each \$1 of computer hardware is correlated with more than \$10 of market value. They argue that a broadened definition of IT to include intangibles such as capitalized software, internal IT services, IT consulting, and IT-related training accounts for the missing 90% of market value.

Work on valuing financial data mostly builds on frameworks designed to explore information asymmetry in financial markets. The Noisy Rational Expectations Equilibrium (NREE) literature started with the seminal work of [Grossman and Stiglitz \(1980\)](#). [Admati \(1985\)](#) enriched this model with multiple assets and correlated information, and [Kyle \(1989\)](#) added imperfect competition. However, these frameworks have information structures that are too limited to accommodate a realistic representation of data. Some papers ([Cabrales et al., 2013](#); [Epstein et al., 2014](#)) have theoretically explored the value of reducing uncertainty by extending such settings. [Savov \(2014\)](#) quantifies the value of private information of mutual funds by calibrating a dynamic dispersed information model with constant relative-risk aversion agents. [Kadan and Manela \(2019\)](#) decompose the value of information to investors into its consumption/investment value versus its value for early resolution of uncertainty. They use their framework to quantify the value of aggregate macroeconomic indicators such as GDP and unemployment. [Farboodi et al. \(2022a\)](#) allow for asset heterogeneity and quantify the value of data about various types of equities. [Kadan and Manela \(2020\)](#) allow for price impact and quantify the value of asset-specific information to a representative trader who is strategic.

All of the aforementioned papers focus on valuing data for a representative investor. However, one reason that data is so challenging to value is that it is worth different amounts to different investors, depending on how they will use the data. In other words, data has a large private value component. That makes data similar to many consumer goods, but quite different from a typical financial asset. [Peress \(2004\)](#) begins to examine heterogeneous demands for information; he considers the information acquisition choice of investors with different levels of wealth and finds that larger investors have a higher demand for information. [Farboodi et al. \(2022b\)](#) add many more degrees of heterogeneity, propose sufficient statistics for measurement and quantify data values. They find that valuations of the same data by investors with different characteristics differ by many orders of magnitude. Using a noisy rational expectations model augmented with rich heterogeneity among investors in wealth, investment style and existing data, and price impact, they derive sufficient statistics to value data, for a particular investor. Their statistics depend only on conditional moments of stock returns and the investor's own characteristics. This approach decomposes the value of data to investors into two pieces. The first is the individual investor's Sharpe ratio

given his data, and the second piece is the variance reduction that he can achieve using this data. This approach bypasses the need to know others' information sets and characteristics.

**Directions for research.** Knowing how different investors value data differently is a first step in estimating a demand curve for data. Combining these individual-level valuations with estimates of the distributions of the most data-value-relevant characteristics allows one to estimate a distribution of data valuations in the investor population. The area under this distribution is demand at each price.

The next step would be to model and measure the supply side of data markets. Supply considerations might include questions, such as: How can firms enter this market? What are the entry barriers? What are the optimal pricing schemes? Armed with a demand curve and an understanding of how data is supplied, economists would be equipped to analyze markets for data.

## 8 Conclusion

This article summarized a modernized set of tools to describe and measure the data economy. While progress has been made, much work remains to understand the modern knowledge economy, measure digital economic activity and value data assets. This is fertile ground for more research and innovation.

## References

- Abis, Simona and Laura Veldkamp**, “The Changing Economics of Knowledge Production,” *Available at SSRN 3570130*, 2020.
- Acemoglu, Daron**, “Technical change, inequality, and the labor market,” *Journal of economic literature*, 2002, 40 (1), 7–72.
- , **Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar**, “Too much data: Prices and inefficiencies in data markets,” *Forthcoming, American Economic Journal: Microeconomics*, 2021.
- Admati, Anat R**, “A noisy rational expectations equilibrium for multi-asset securities markets,” *Econometrica: Journal of the Econometric Society*, 1985, pp. 629–657.
- Aghion, Philippe, Eve Caroli, and Cecilia Garcia-Penalosa**, “Inequality and economic growth: the perspective of the new growth theories,” *Journal of Economic literature*, 1999, 37 (4), 1615–1660.
- Agrawal, Ajay, Joshua Gans, and Avi Goldfarb**, *Prediction machines: the simple economics of artificial intelligence*, Harvard Business Press, 2018.
- , – , and – , “Economic policy for artificial intelligence,” *Innovation Policy and the Economy*, 2019, 19 (1), 139–159.
- Ali, S Nageeb, Greg Lewis, and Shoshana Vasserman**, “Voluntary disclosure and personalized pricing,” *The Review of Economic Studies*, 2022.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow**, “The Welfare Effects of Social Media,” *American Economic Review*, March 2020, 110 (3), 629–76.
- Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp**, “Big data in finance and the growth of large firms,” *Journal of Monetary Economics*, 2018, 97, 71–87.
- Benzell, Seth G and Erik Brynjolfsson**, “Digital Abundance and Scarce Genius: Implications for Wages, Interest Rates, and Growth,” Working Paper 25585, National Bureau of Economic Research February 2019.
- Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri**, “On the rise of fintechns: Credit scoring using digital footprints,” *The Review of Financial Studies*, 2020, 33 (7), 2845–2897.

- Bergemann, Dirk, Alessandro Bonatti, and Tan Gan**, “The economics of social data,” *RAND Journal of Economics*,, 2022.
- and – , “Data, Competition, and Digital Platforms,” Technical Report 2022.
- Bonatti, Alessandro and Gonzalo Cisternas**, “Consumer scores and price discrimination,” *The Review of Economic Studies*, 2020, *87* (2), 750–791.
- Brynjolfsson, Erik and Kristina McElheran**, “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 2016.
- , **Avinash Collis, and Felix Eggers**, “Using Massive Online Choice Experiments to Measure Changes in Well-being,” *Proceedings of the National Academy of Sciences*, April 2019, *116* (15), 7250–7255.
- , – , **W. Erwin Diewert, Felix Eggers, and Kevin J Fox**, “GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy,” Working Paper 25695, National Bureau of Economic Research March 2019.
- , **Lorin M Hitt, and Shinkyu Yang**, “Intangible assets: Computers and organizational capital,” *Brookings papers on economic activity*, 2002, *2002* (1), 137–181.
- Byrne, David and Carol Corrado**, “Accounting for Innovation in Consumer Digital Services: IT Still Matters,” Working Paper 26010, National Bureau of Economic Research June 2019.
- Cabrales, Antonio, Olivier Gossner, and Roberto Serrano**, “Entropy and the value of information for investors,” *American Economic Review*, 2013, *103* (1), 360–77.
- Caplin, Andrew and John Leahy**, “Business as Usual, Market Crashes, and Wisdom After the Fact,” *American Economic Review*, 1994, *84* (3), 548–565.
- Chen, Long, Yadong Huang, Shumiao Ouyang, and Wei Xiong**, “The data privacy paradox and digital demand,” Technical Report, National Bureau of Economic Research 2021.
- Crémer, Jacques, Gregory S Crawford, David Dinielli, Amelia Fletcher, Paul Heidhues, Monika Schnitzer, Fiona M Scott Morton, and Katja Seim**, “Fairness and Contestability in the Digital Markets Act,” *Available at SSRN*, 2021.



- Demsetz, Harold**, “Barriers to entry,” *The American economic review*, 1982, 72 (1), 47–57.
- Eckbo, B. Espen, ed.**, *Handbook of Empirical Corporate Finance SET* number 9780444532657. In ‘Elsevier Monographs.’, Elsevier, 2008.
- Eeckhout, Jan and Laura Veldkamp**, “Data and Market Power,” Working Paper 30022, National Bureau of Economic Research May 2022.
- Epstein, Larry G, Emmanuel Farhi, and Tomasz Strzalecki**, “How much would you pay to resolve long-run risk?,” *American Economic Review*, 2014, 104 (9), 2680–97.
- Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel**, “Uncertainty Traps,” *The Quarterly Journal of Economics*, 2017, 132 (4), 1641–1692.
- Farboodi, Maryam**, “Data markets and digital platforms,” Technical Report, Working Paper 2022.
- , **Adrien Matray, Laura Veldkamp, and Venky Venkateswaran**, “Where has all the data gone?,” *The Review of Financial Studies*, 2022, 35 (7), 3101–3138.
- **and Laura Veldkamp**, “Long-run growth of financial data technology,” *American Economic Review*, 2020, 110 (8), 2485–2523.
- **and** – , “A Model of the Data Economy,” Working Paper 28427, National Bureau of Economic Research 2022.
- , **Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran**, “Valuing financial data,” Technical Report, National Bureau of Economic Research 2022.
- Frankel, Alexander and Emir Kamenica**, “Quantifying information and uncertainty,” *American Economic Review*, 2019, 109 (10), 3650–80.
- Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden**, “BigTech and the changing structure of financial intermediation,” *Economic Policy*, 2019, 34 (100), 761–799.
- Gan, Jie and Timothy J Riddiough**, “Monopoly and information advantage in the residential mortgage market,” *The Review of Financial Studies*, 2008, 21 (6), 2677–2703.
- Goldfarb, Avi and Catherine Tucker**, “Online display advertising: Targeting and obtrusiveness,” *Marketing Science*, 2011, 30 (3), 389–404.

- and –, “Digital economics,” *Journal of Economic Literature*, 2019, 57 (1), 3–43.
- Gottlieb, J and A Weinberg**, “Catch them if you can: How leaders in data and analytics have pulled ahead,” *McKinsey Analytics*, 2019, pp. 1–8.
- Grossman, Sanford J and Joseph E Stiglitz**, “On the impossibility of informationally efficient markets,” *The American economic review*, 1980, 70 (3), 393–408.
- Gul, Faruk and Wolfgang Pesendorfer**, “Temptation and self-control,” *Econometrica*, 2001, 69 (6), 1403–1435.
- Gutierrez, German and Thomas Philippon**, “Investmentless Growth: An Empirical Investigation,” *Brookings Papers on Economic Activity*, 2017, 48 (2), 89–190.
- Hau, Harald, Yi Huang, Hongzhe Shan, and Zixia Sheng**, “How FinTech enters China’s credit market,” in “AEA Papers and Proceedings,” Vol. 109 2019, pp. 60–64.
- He, Zhiguo, Jing Huang, and Jidong Zhou**, “Open banking: credit market competition when borrowers own the data,” Technical Report, National Bureau of Economic Research 2020.
- Hulten, Charles and Leonard Nakamura**, “Accounting for Growth in the Age of the Internet: The Importance of Output-Saving Technical Change,” Working Paper 23315, National Bureau of Economic Research April 2017.
- Ilut, Cosmin and Martin Schneider**, “Ambiguous Business Cycles,” *American Economic Review*, August 2014, 104 (8), 2368–99.
- Jones, Charles I and Christopher Tonetti**, “Nonrivalry and the Economics of Data,” *American Economic Review*, 2020, 110 (9), 2819–58.
- Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens**, “Investor sophistication and capital income inequality,” *Journal of Monetary Economics*, 2019, 107 (C), 18–31.
- Kadan, Ohad and Asaf Manela**, “Estimating the value of information,” *The Review of Financial Studies*, 2019, 32 (3), 951–991.
- and –, “Liquidity and the Strategic Value of Information,” Working paper 2020.
- Kirpalani, Rishabh and Thomas Philippon**, “Data sharing and market power with two-sided platforms,” Technical Report, National Bureau of Economic Research 2020.

- Kolanovic, Marko and Rajesh T Krishnamachari**, “Big data and AI strategies: Machine learning and alternative data approach to investing,” *JP Morgan Global Quantitative & Derivatives Strategy Report*, 2017.
- Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante**, “Capital-skill complementarity and inequality: A macroeconomic analysis,” *Econometrica*, 2000, *68* (5), 1029–1053.
- Kyle, Albert S**, “Informed speculation with imperfect competition,” *The Review of Economic Studies*, 1989, *56* (3), 317–355.
- Lambrecht, Anja and Catherine E Tucker**, “Can big data protect a firm from competition?,” *Available at SSRN 2705530*, 2015.
- Liu, Zhuang, Michael Sockin, and Wei Xiong**, “Data privacy and temptation,” Technical Report, National Bureau of Economic Research 2020.
- Loecker, Jan De, Jan Eeckhout, and Gabriel Unger**, “The Rise of Market Power and the Macroeconomic Implications,” *The Quarterly Journal of Economics*, 01 2020, *135* (2), 561–644.
- Lorenzoni, Guido**, “A Theory of Demand Shocks,” *American Economic Review*, December 2009, *99* (5), 2050–84.
- Martin, Kirsten E**, “Ethical issues in the big data industry,” in “Strategic Information Management,” Routledge, 2020, pp. 450–471.
- Mihet, Roxana**, “Financial Technology and the Inequality Gap,” Swiss Finance Institute Research Paper Series 21-04, Swiss Finance Institute January 2021.
- Mitchell, Tom and Erik Brynjolfsson**, “Track how technology is transforming work,” *Nature*, 2017, *544* (7650), 290–292.
- Morton, Fiona M Scott and David C Dinielli**, “Roadmap for a digital advertising monopolization case against google,” *Omidyar Network*, 2020.
- Ordóñez, Guillermo**, “The Asymmetric Effects of Financial Frictions,” *Journal of Political Economy*, 2013, *121* (5), 844–895.
- Ouyang, Shumiao**, “Cashless Payment and Financial Inclusion,” *Available at SSRN 3948925*, 2021.
- Peress, Joel**, “Wealth, Information Acquisition and Portfolio Choice,” *The Review of Financial Studies*, 2004, *17*(3), 879–914.

- , “The tradeoff between risk sharing and information production in financial markets,” *Journal of Economic Theory*, 2009, *forthcoming*.
- Prat, Andrea and Tommaso M Valletti**, “Attention oligopoly,” *American Economic Journal: Microeconomics*, *Forthcoming*, 2021.
- Rosen, Sherwin**, “The economics of superstars,” *The American Economic Review*, 1981, *71* (5), 845–858.
- Saunders, Adam and Erik Brynjolfsson**, “Valuing information technology related intangible assets,” *Mis Quarterly*, 2016, *40* (1), 83–110.
- Savov, Alexi**, “The price of skill: Performance evaluation by households,” *Journal of Financial Economics*, 2014, *112* (2), 213–231.
- Shiller, Benjamin Reed**, *First degree price discrimination using big data*, Brandeis Univ., Department of Economics, 2013.
- Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson**, “Digital Capital and Superstar Firms,” Working Paper 28285, National Bureau of Economic Research December 2020.
- Veldkamp, Laura**, “Slow Boom, Sudden Crash,” *Journal of Economic Theory*, 2005, *124*(2), 230–257.
- Wilson, Robert**, “Informational Economies of Scale,” *Bell Journal of Economics*, 1975, *6*, 184–95.
- Zingales, Luigi**, “Towards a political theory of the firm,” *Journal of Economic Perspectives*, 2017, *31* (3), 113–30.