CONTENTS

1

The Data Economy

The most valuable firms in the global economy are valued largely for their data. Amazon, Google, Apple, and other highly prized technology firms have market caps that exceed the total assets of many small countries. Historically, high-value firms were ones with high-value physical assets: factories, offices, or equipment. Today, a firm with high-precision equipment and prime office real estate can still reap value from those assets. But neither confers the competitive advantage of a good data set. As data storage has become cheaper and data science has improved, the value of data has risen.

Data helps firms find customers, select new products, manage inventory, and choose suppliers. A firm that consistently makes these decisions well has a competitive advantage over those that do not. Such a data-rich firm raises its revenues and cuts its costs. It will profit more, grow more, sell more, and accumulate more data. More data begets more data. This is the data feedback loop. The self-reinforcing nature of data in a competitive market can explain trends in market power and the rise of superstar firms.

Despite the growing importance of data as a strategic asset, modern textbook theories of macroeconomics and finance neglect its role. The tools of macroeconomics describe an industrial economy: physical capital and labor combine to make goods or services that can be used by only one consumer at a time. These models do not describe a modern knowledge economy. An economy where workers use data to add value behaves quite differently from one in which value added comes from combining work with physical capital. Firms built on data compete, grow, and price goods and services differently. In the aggregate, the data economy grows and fluctuates differently. When vast amounts of data can be collected, mined, and traded, what agents know may be less important than where they direct their attention or what data they decide to analyze. When data-intensive firms offer products or services for free or at a discount in exchange for data—bartering for data—then gross domestic product (GDP) will fail to capture an important share of economic activity. Adopting new tools that reflect modern uses of data can offer fresh insights into every corner of macroeconomics and new ways of measuring economic activity.

In finance, data is an important asset to price: it changes firm valuation, and it is a key consideration for entrepreneurs starting new firms. The rise of the data economy is changing sources of both revenue and risk. The industrial-age measurement and valuation tools commonly used in finance need updating for a new era. Furthermore, data may constitute a significant barrier for entrepreneurs and an impediment to business dynamism.

At its essence, data is digitized information. According to information theory pioneer Claude Shannon, information is that which reduces randomness or uncertainty. New data science tools, like machine learning and artificial intelligence, use data to make more precise predictions, making the future more predictable and less uncertain. This book starts from the premise that data is something that facilitates prediction and thereby reduces uncertainty. This definition distinguishes data from ideas and technologies, from knowledge and human capital, and from physical capital. Data helps firms forecast or predict uncertain outcomes, such as which products, suppliers, investments, and advertisements would be most profitable. A hallmark of data-informed choices is that they covary with outcomes like demand, returns, and sales.

Since data is information, the tools of information economics are valuable inputs for data economy research. Therefore, this book collects many information-related tools from macroeconomics and finance that can be used for modeling and measuring data economies. The following chapters teach a mix of old and new tools, many of which were not originally intended to describe a data economy. When discussing existing tools from other topics, we describe how they might be repurposed to model or measure data economies. The tools are primarily theoretical, with careful attention to how these theories can inform or enable measurement.

Data economics research is in its infancy. It will take decades for competing theories to be developed, measurements to be executed, and controversies to be resolved. Answering the many questions posed by the data economy will be a research journey. The role of this book is to facilitate new research about the data economy, its welfare consequences, and optimal data policy.

1.1 Why Theorize About Data?

Research on the data economy is exploding. This book is not a comprehensive summary of that research. Instead, this book teaches a set of modeling tools and describes how the resulting models can inform measurement and policy. Many of the questions in this area cannot be answered without theory. Empirical evidence alone cannot teach us what will happen if we enact a data privacy policy that has never been tried. Often the interpretation of a covariance or parameter estimate is not clear. Theories guide measurement and the interpretation of what is measured. Even within the set of data economy theories, we are not comprehensive. Our focus is on aggregate theories. These are not theories about how one firm uses data, although that is often an ingredient in the model. These are theories designed to think about a market or an economy as a whole. Most models have many agents and balance supply with demand. While many books teach the reader the microeconomics of data, which is a worthy field of study, this book takes a macro perspective to study the data economy.

Macro models always miss important details. They are like maps of the world that are missing most roads. They do not have the detail one would need to drive somewhere. At the same time, a local map is useful for a commute but lacks the perspective of a globe. Both are imperfect. Each is more useful for some purpose. As we build macro models, or maps, of the data economy, a reader is sure to feel dissatisfied at some point with an ingredient that feels essential but is missing. This is the point at which we sincerely hope you will take out a pencil or open a laptop and get to work adding the important details to build out these models. These models are simple structures. Think of them like an artist's canvas on which to project creative ideas. They are not an end product but rather a beginning.

1.2 Essential Features of Data

The modern data economy arose from computing, data storage, and data science innovations. Recorded data has been around for millennia. Data, accounting, and bookkeeping are some of the most basic building blocks of modern human civilization. However, while this data was essential for organizing human activity, it was not a valuable, traded asset or a key input into firm production until recently. Headlines like "Data Is the New Oil" appeared in response to a surge in the storage, use, and value of economic data.

Data moved to the forefront of economic debates because we became capable of storing large data sets at low cost and using them to compute estimates and, most of all, because of new data science algorithms that used big data to make better predictions. Artificial intelligence (AI) and machine learning are two names for these new prediction algorithms. While AI research has been going on since the 1950s, it is only recently that breakthroughs in deep neural networks have enabled these algorithms to be useful for most prediction tasks. In 2012, AI algorithms could finally classify photos of cats and reliably identify them as such. A few years later, these algorithms could identify most human faces or species of birds. Of course, identifying images is not what most businesses do with data. But, because that task is challenging, it is an indicator of the progress of algorithms that take a set of data, like the pixels in a large collection of images, and make predictions about what these data points might mean. In the last decade, data scientists have made immense progress in using large data sets to make predictions.

1.2.1 An Input into Predictions

Since the data science advances that have made data a valuable asset are inherently prediction technologies, most of the tools we introduce in this book treat data as something used to predict uncertain outcomes. Data is an input into this prediction. Data might need to be combined with labor input; capital input, such as computer equipment; or other complementary inputs. But the output is a prediction with higher accuracy than was previously possible. More data enables better, meaning more accurate, predictions.

Thus, one feature distinguishing data from technology, human capital, and other intangible capital is that it improves prediction. Because firms with bigger data sets can predict which ads, products, or procurement sources will be most profitable, they can make choices that covary more with the realized return on each choice. Firms with more data can place ads that are more likely to result in sales, produce goods that are more likely to be in high demand, or choose suppliers that are more likely to offer consistently low costs. By facilitating better predictions, data improves firms' decision-making and lowers their uncertainty about the payoffs of their decisions.

Chapter 2 describes statistical tools used for prediction. Many are applications of Bayes law. Survey evidence and "information-provision experiments" confirm that economic agents revise their beliefs in response to new information in ways broadly consistent with Bayes' Law.[1] Bayes' Law is the foundational tool of most models in this book.

Other prediction tools in Chapter 2 are more like machine learning and artificial intelligence in that they are non-Bayesian. None of the estimation tools we present are as sophisticated as modern data science algorithms. This book contains tools for models. Models are caricatures or simplifications of reality. Our objective is not to estimate outcomes as precisely as possible but to characterize types of estimates firms might make with data. We distill the approaches to their essence in order to gain economic understanding.

[1] See, for instance, Coibion, Gorodnichenko, and Weber (2019); and Coibion et al. (2021) for the case of households and, for firms, Coibion, Gorodnichenko, and Kumar (2018); and Chapter 14 of the *Handbook of Economic Expectations* (Bachmann, Topa, and van der Klaauw, 2022).

1.2.2 A By-Product of Economic Activity

Most new prediction technologies require vast troves of data. The tool becomes powerful only when combined with a large quantity of training data. While data can come from many sources, large data sets arise naturally as the digital footprint is left by the multitude of daily economic transactions that firms, consumers, suppliers, and governments do. Almost every economically relevant action leaves a digital trace in the modern economy. That action could be buying something, making a payment with a credit or debit card, driving over a bridge, parking in a lot, searching the web, browsing a store either online or in person, or even checking the weather on one's phone. All of these activities generate data as a by-product of the action, even if the data was simply that someone drove, visited, or parked. If every action leaves a trace, then the volume of data generated in the aggregate is stunningly large.

Because digital footprints of economic activity naturally generate so much data that is relatively easy to harvest, it is the primary source of most business data. This sort of data is also naturally relevant for businesses. Evidence of demand, search behavior, traffic patterns, or attention all help firms forecast demand, target advertising, manage supply chains, and predict costs. Thus, data is not produced separately from goods or services. Unlike physical capital and most intangible capital, and unlike new technologies, data does not require a separate investment to produce. Unlike human capital, data does not require a person's time to cultivate it. Of course, an investment might be required to structure, store, and analyze data. But the data itself is a by-product of economic activity. It is generated in the process of buying or selling goods and services. This form of data production is important because large firms that produce and sell lots of goods and services will naturally acquire more data assets. This gives a natural advantage to large firms.

1.2.3 Nonrivalry

Consider a physical good, like a pencil. If Alice sells Bob a pencil, Bob now has the pencil. Alice cannot use that pencil, as only one person can write with the pencil simultaneously. Data is very different. Data fundamentally differs from physical goods because of its nonrival nature. Nonrivalry means that one party can use data at the same time as another party is using the same data. A seller can sell data and still keep the same data.

The nonrival nature of data distinguishes it from human capital acquired through learning by doing, for example. However, the nonrivalry of data is similar to the nonrivalry of ideas. Recognition of the nonrival nature of ideas or technologies revitalized growth theory in the 1980s and 1990s (e.g., Romer,

1990) because models with the production of nonrival ideas could produce sustained growth in a way that goods production could not.

Nonrivalry does not mean that sharing data is without cost. When competing firms use the same data set, they may reduce the value of the data to each other. Both firms can make use of the same data at the same time. But doing so can still make both parties worse off. The idea that one firm's use of data reduces the value of that data to other firms is central to the competitive equilibrium in the data market and the goods market.

1.2.4 Increasing and Decreasing Returns

Data can have both increasing returns to scale and diminishing marginal returns. Returns to scale in data refers to the idea that when an economy is scaled up, meaning that all its inputs or endowments are multiplied, the value of data rises. For example, data has returns to scale in a portfolio problem because it can be used to evaluate one share of an asset or many shares of that asset. When a decision-maker has lots of an asset, data about that asset's payoff is more valuable.

The increasing return to scale in data has important aggregate consequences. There are returns to scale in data because information is expensive to discover but cheap to replicate. If the first copy of a piece of data is much more expensive than the hundredth, there is an incentive to buy the data that others are buying because it is less expensive. This is complementarity in data acquisition. When agents acquire the information others acquire, it leads them to take similar actions. Such mimicking actions resemble fads, herds, or frenzies. Returns to scale also means that big firms extract more value from data than small ones. Evidence from Eckert, Ganapati, and Walsh (2022) supports the finding that large firms benefit more. This force creates gains from merging firms and their data sets. Thus, the data economy may favor large firms and disadvantage small ones.

Data also has decreasing marginal returns. A small amount of data greatly improves prediction. The millionth piece of data has a minute effect. When a firm has an enormous data stock, it can make close to the best possible prediction. At that point, the gains from additional data are small.

While the data economy naturally generates increasing returns to firm size, the decreasing data value for prediction produces decreasing returns. These two forces compete. Increasing returns may dominate early in a firm's life while it is still small and data-poor. Unless the prediction problem changes over time, decreasing returns will dominate and slow growth in the long run.

1.2.5 An Asset of a Firm

The value of economic data has exploded in the past decade. The global market for big data and related technology and analytics was valued at $130 billion in 2017 and grew to over $200 billion by 2020 (Kolanovic and Krishnamachari, 2017). Data is an enormously valuable resource with some durable value, but it also depreciates. Because data used in one period can also be used in the next period, we will need the tools of recursive macroeconomics to model and value it. At the same time, data is used to forecast outcomes in a changing world, so old data is typically less relevant than new data. The loss of relevance shows up as rapid depreciation.

As of today, US data is primarily owned by firms. It is possible to enact a legal framework so that consumers have ownership and control over their data. However, since the current legal framework makes that challenging, in practice, firms own data. Therefore, we model data as an asset that firms own, buy, sell, and extract value from, recognizing that it is possible that another regime could arise, requiring new models.

Because it is difficult to quantify, not typically observed, and often unreported on firms' balance sheets, data is an especially challenging asset to measure and value. The book will discuss these challenges and offer various tools to overcome them.

1.3 How Do Firms Profit from Data?

1.3.1 Buying and Selling Data

The data economy has brought about new business models. At some large firms like Facebook and Google, most revenue does not come from their primary products—search and information sharing. Instead, they harvest the data of product users and earn most of their revenue from selling that data or selling services derived from it.

Other firms acquire data by intermediating trades. Amazon and eBay allow sellers to offer products on their digital platforms. They earn revenue partly through seller fees and partly through data. This business model is similar to that of financial intermediaries or dealers, who have long benefited from seeing the information in clients' order flow.

Not all data is sold directly. Some data is monetized through advertising services. Data-rich firms use their data to target ads to customers with particular characteristics that make them likely to purchase the advertised product. Placing ads is like investing capital in a risky portfolio of assets. The firm is investing advertising funds in a risky portfolio of ads. Just like mutual fund managers (should) use data to invest capital in high-return assets on behalf of their clients, data intermediaries use data to invest a firm's advertising budget in purchasing ads that are predicted to have a high customer conversion rate. The primary difference between these two problems is that a high-return asset conveys a high return to all who invest in it, but a promising customer for one firm is not necessarily promising for another. So, while financial assets are risky common-value assets, advertising invests in an asset portfolio with a large private-value component.

Chapter 3 explores sources of data. It offers tools to capture data revealed in the production process and to describe data choices. Data choices might involve the choice of what to buy or sell or the choice of what data to structure, process, or analyze. Such choices are subject to costs or constraints. Modeling costs or constraints requires measures of data. Thus, the chapter explores different measures and cost functions for data in order to enable the investigation of data choices.

1.3.2 Prediction and Data-Driven Decisions

Firms use data to enhance profits by predicting their costs and revenue shocks and acting to anticipate or mitigate their effect on profits. Brynjolfsson and McElheran (2016) and Goldfarb and Tucker (2019a) give examples of and describe the mechanics of such data-driven decision-making. They orient their analysis around what sorts of firm shocks the data is used to predict, such as sales, advertising, procurement, or hiring.

For a given firm, what to predict and how to use the prediction are central. Whether firms better anticipate costs, revenues, or inventories may not matter as much for aggregate economic outcomes. What matters for the economy as a whole is the way in which data is used to make the prediction. Chapter 4 catalogs and dissects types of prediction problems. Firms could use data to track and forecast a changing economic state, like consumer tastes or the price of raw materials. They might also use data to distinguish short-term fluctuations from long-term economic trends. Alternatively, a firm might use data to decipher what part of an economic shock or trend is firm-specific and what part is aggregate. Is demand generally surging, or did the firm produce a hot product? When solving these problems, firms might buy or choose what data to use in their forecast. The chapter explores data choice and traces its consequences for the aggregate economy.

Data has strategic uses as well. It can be useful to infer what others know or forecast what others will do. Chapter 5 explores the use of data and data choices in settings where firms' or agents' choices depend on what they believe others will do. Since the book focuses on aggregate outcomes, the strategic games it analyzes are ones with many players, often called “mean-field games.” A subset of such models is global games. Data can reduce inertia or dampen volatility depending on the strategic motives in the mean-field game. The chapter classifies economic settings, describes the effect of data in each class, and offers economic examples, highlighting the role and effect of data and data choice.

1.3.3 Resolving Risk

Data also reduces risk. Data, at its core, is digitized information, and information is a technology used to reduce uncertainty. Firms use prediction technologies such as machine learning and AI. These technologies aim to predict whether an observation belongs to one set or another. Machine learning can be used in making better predictions about uncertain consumer demands for various products, costs to make certain products, or returns to a portfolio of assets one will buy. Thus, not only can data increase returns, but it can also decrease uncertainty.

Chapter 6 introduces the benefits of risk reduction in the context of an investment allocation problem. Data is often used to allocate resources in an uncertain environment. A salient example of this is the choice of a financial asset portfolio. An investor who can use data to predict the future returns of risky assets can choose to buy more of the profitable assets and earn more. Portfolio problems typically include an aversion to risk or a price of risk. Data not only increases the expected profit but also helps to reduce the utility and financial costs associated with bearing risk. The chapter explores varieties of investor choice problems, the prices that result when many investors solve a data choice and investment choice problem, and the price inelasticity that can arise from investors’ use of data.

A simple calculation suggests that risk matters enormously for firm and data valuations. Of the 10% expected returns on firms, about 3% is the riskless return, and 7% is the risk premium, the compensation for risk. This suggests that, for the average firm, risk matters twice as much as the expected value that most economists model and measure. For financial data, Farboodi et al. (2022b) compute the value originating from each component. They break out the part of data value that comes from increasing the expected return and the part that comes from reducing uncertainty. In most cases, far less than half of the value comes from a higher expected return. Most of the value of financial data comes from its ability to resolve risk, making forecasts less uncertain.

Risk also matters in firms’ decisions. Corporate finance classes teach potential firm managers to price risk and scale back investment in the face of risk. Corporate finance research finds abundant evidence that this risk adjustment does, in fact, take place throughout the economy. Thus, firms making real output decisions price risk. This implies that they should value data for its risk-reduction properties and its use as an asset. Neglecting the risk component of data's value could lead to a substantial undervaluation of its financial value and its welfare benefit.

1.3.4 Market Power

One of the greatest concerns about using big data is the potential market power it engenders. Data could allow firms to grow larger and take a larger market share. It could function as an entry barrier. And, data could allow firms to compete aggressively on exactly the new products offered by new entrants. All of these are uses of data that will maximize the long-run revenue of the incumbent firm. Chapter 7 explores ways in which firms may use data to create market power. Much of the chapter builds on the investment choice problem. Instead of considering what risky assets to buy, firms choose what products to produce when those products have uncertain demand. A portfolio of assets becomes a portfolio of attributes in a product the firm might design. Investment research is similar to using data for product innovation. The end of the chapter explores the ways in which customer relationships, often called customer capital, might function as data to create and sustain market dominance.

1.4 Distinguishing Features of a Data Economy

The uses of data described above are ways that individual firms use data to profit. Most of these uses have been described in other texts. This book contributes to the formal mathematical modeling tools that can enable structural measurement and inform policy choices. These models can structure our thinking about the aggregate consequences of firms' data choices. While many others have explored the microeconomics of data, this book is about the macro view. It considers equilibrium. Adding the macroeconomic perspective is essential for understanding economic growth, business cycle fluctuations, price-setting, investment, measurement, and sound policymaking.

1.4.1 Platforms and Data Brokers

Intermediaries who connect buyers and sellers have been around for a long time. But data has given them a new business model. Traditional intermediaries were typically compensated by commissions. Either the buyer, seller, or both paid the matchmaker a fee. Household investors paid stockbrokers, property sellers paid real estate agents, and hiring firms paid for job postings in the newspaper. Today, many intermediaries do not earn most of their revenues from fees. Instead, their revenue comes from the data they observe. The intermediary sees the buyers' and the sellers' actions. They have data about multiple parties in the market that no other single agent can know.

Some intermediaries now offer their services at a zero monetary price. For example, Robinhood, the digital platform for trading financial assets, offers free trading. However, they will sell information about your trades to other financial market participants. Order execution is being bartered for order flow data. But because intermediation is data-rich—it exposes both sides of the trade—the barter trade is particularly beneficial to the data platform or intermediary.

Furthermore, some intermediaries are also market participants. Amazon sees which products are profitable and introduces its own brands in that space. Zillow used its house price data to bid for and sell houses. Many regulatory questions surround this type of behavior. Structural models can assist policymakers in assessing its welfare consequences.

Chapter 8 explores models of data platforms. It focuses on equilibrium models with market clearing prices because they are compatible with the macro orientation of the book. Some of the models build on matching tools taken from macroeconomics of labor markets. Some of the models are from industrial organization. But some of the models are an extension of the framework of production allocation choice with market power, built up in the previous two chapters. It adds an intermediary who observes the trades of agents who use their platform. But in return, the intermediary may advise the buyer and/or seller about what to produce, what to purchase, or what other platform participants are buying or selling. Buyers and sellers who choose to trade through the platform barter data for data: they barter the data about their trades for aggregate data provided by the platform. While data platforms are typically associated with sales of goods, they also describe the behavior of financial intermediaries who observe client order flow and provide information or just execution services in return. The data platform model for the goods market is a novel framework, not previously published, written for this book.

1.4.2 A Data Feedback Loop

The data feedback loop refers to the self-reinforcing growth dynamic that arises when firms produce data as a by-product of economic activity. Suppose that having more transactions or getting more customers generates more data. Firms find out a wealth of information about their customers, such as what they like to buy, what kind of credit card they have, where they live, their zip code, and so forth. Firms use this data to generate higher-quality or better-matched goods for customers, and they become more efficient. Firms may use data to appropriately stock their shelves and inventory or hire the right workers to be more profitable. Becoming more efficient or having higher-quality goods allows a firm to attract more customers and transact more. Higher efficiency also incentivizes the firm to invest more and grow larger. Thus, a firm with more data has greater efficiency and more customers, and it gathers even more data.

Amazon's notion of a flywheel captures the spirit of this dynamic. Agrawal, Gans, and Goldfarb (2018) describe Amazon's strategy: 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 prediction algorithm. As more data improves predictions, the product will be more successful, creating a virtuous cycle.

Chapter 9 builds a dynamic, recursive model where firms produce goods, taking account of the data that will result from their future production. The chapter shows how a slight lead in data allows a company to collect more and better data, reinforcing its lead and generating market dominance over time. Data feedback from production can also change the shape of business cycles, speeding up recessions and slowing booms. Finally, data feedback can make economies and markets fragile, when agents do not know the structure of the economic environment.

1.4.3 Data Barter

A new feature of the modern data economy is that many digital goods and services are given away at zero price. These products were costly to develop. This sort of behavior is difficult to rationalize in a classical production economy. However, when data is a valuable asset generated by economic transactions, this pricing behavior makes sense. Firms develop products to attract users. Selling the products at zero price is profitable because the value of the data that is generated from each sale of the product is valuable enough to compensate the firm for the product development.

In short, these zero-price goods are not free; they are part of barter trade. The digital product is bartered for the user's data at zero monetary price. Some economists and policymakers have called for users to be paid for using their data. In a data-barter view of the world, consumers are paid: they are paid with the zero-price digital service.

The data feedback model formalizes the logic, the value, and the consequences of this barter trade. Structural models of this barter phenomenon are useful because the magnitude of the barter trades has yet to be carefully measured. This barter value is missing in GDP and will likely grow over time. It could lead us to underestimate aggregate growth.

Pure barter trades with zero-price services are still relatively rare. What is potentially much more common is partial barter trades. Consider a firm that recognizes that data from transactions is quite valuable. Such a firm should be eager to do lots of transactions. How does the firm achieve more transactions? They lower their price to lure more customers. A firm that values its customers' data should charge a price below the no-data optimal price. The difference between the price with data and what the price would be without data is the data discount. Now, when a customer buys a product, they are paying a fraction of the true cost with money and a fraction with their data. The true value of the transaction is the posted price plus the value of the data transferred.

Another version of the barter trade is bartering a product for the consumer's attention. For a long time, this was the business model of commercial television. Viewers got zero-price entertainment. In return, they were served with ads. Viewers bartered the entertainment for their attention. The attention allocation tools in Chapter 3 could be used to speak to this slightly different form of barter.

Measuring the value of data barter is not easy. Data-free prices are rarely posted. Perhaps customer loyalty card discounts represent payments for data, although some data can still be collected without using the loyalty card. But it is likely that economists will want structural models to help fill in for counterfactual prices that are never observed.

1.4.4 Superstar Firms

Data favors large firms. Returns to scale in data and the data feedback loop are two separate forces, both of which advantage large firms over smaller ones. Because of returns to scale, a large firm values a piece of data more than a smaller version of the same firm would. The large firm can use the information in that data to produce many units more efficiently or profitably. The small firm can only use the data at a limited scale. Thus, large firms are more likely to make the investments needed to gather, process, and act on data efficiently.

Because of the data feedback loop, not only does a large firm benefit more from a given piece of data, but it also gets more data from its customers. Working together, these forces make large firms more efficient data collectors and more efficient producers than small firms.

Data encourages and enables large firms to grow larger. It makes mergers or acquisitions more valuable. A merger not only merges two firms but also merges two data sets. With a larger data set, both original businesses can become more profitable.

Thus, one of the hallmarks of the data economy is a change in the optimal size of firms. Large firms dominating their market is to be expected when the returns to size increase with the advent of new data technologies. This is not proof that the increase in market concentration observed in the US economy is because of data technology. It simply means that improvements in data technology should explain some increase in concentration, either now or in the future. The magnitude of this effect remains to be seen. Another open question is whether these large firms represent mostly a gain for consumers from greater efficiency or a loss from the greater use of market power to extract consumer surplus.

1.4.5 Growing Covariance of Payoffs and Choices

When a given firm has more data, it uses that data to align its choices more closely with profit opportunities. In other words, data increases the covariance between actions and the uncertain states that data is used to predict. This growing covariance can take many forms and have important aggregate consequences.

For example, Chapter 4 considers firms that are uncertain about monetary policy and are choosing prices for their goods. More data would allow these firms to align their prices more precisely with monetary policy. This would reduce the efficacy of monetary policy and bring us closer to monetary neutrality. Chapter 5 considers settings where agents want to coordinate. Perhaps they have a preference to behave like others. Perhaps these are firms in a supply chain that want to coordinate; for example, they may want to produce more laptop keyboards when the makers of laptop screens produce more screens for them to connect their keyboards to. In any case, data allows them to coordinate more effectively. More data might show up as more synchronized supply chains. As a final example, if firms can predict which product will be in high demand, they can produce more of these products. In an imperfectly competitive market, high-demand products are typically high-markup products. If data helps firms predict demand, they can skew their product mix toward high-markup goods to generate more profit.

Chapter 10 uses this insight and many others to propose ways of measuring and valuing data. While the book is primarily about presenting theoretical tools, the success of these theoretical tools will depend, in large part, on their ability to make contact with data and policy. Measuring data is a challenging task, and different tools are needed for different contexts.

1.4.6 Privacy Concerns

Another feature that makes data different from other intangible assets is that data is personal. Data is often about people, their characteristics, and their actions. Many people would prefer that their actions and characteristics were not widely known. This gives rise to a concern for data privacy. Privacy preferences could encompass concerns about financial theft, identity theft, physical safety, harassment, price discrimination, or a feeling of being violated. Until the last chapter, this book will neglect these concerns. This is not to say they are not important. But the tools of macroeconomics and finance are not yet integrated with tools to model or measure privacy costs. We acknowledge that privacy is a serious concern. The little attention given to the topic reflects that work on the aggregate value of privacy is in its infancy.

Chapter 11 studies the welfare implications of the data economy and points to policy challenges. It describes data externalities, labor market effects, public and private data, risk-sharing, as well as privacy.

1.5 Goals

This research area is nascent. The models are incomplete descriptions and imperfect tools; they need improvement. The data measures should be refined and applied to suitable data sets. Since new papers on this topic continuously appear, our literature reviews are surely incomplete. We apologize to the authors we have omitted. This book aims to quickly bring the reader to a frontier of this research area, to enable them to advance this frontier, and to facilitate new theory, measurement, and policy analysis of the growing data economy.

INDEX