Journal of Economic Literature 2024, 62(2), 458–484
https://doi.org/10.1257/jel.20221580

Data and the Aggregate Economy†

Laura Veldkamp and Cindy Chung*

Recent data technology innovations, such as artificial intelligence and machine learning, have transformed the production of knowledge and increased the importance of data. This review explores how data—digitized information—has been modeled within classic macroeconomic frameworks. It compares the economics of data to other concepts such as ideas, patents, and learning-by-doing. This paper also shows potential ways to model applications for data, including innovation, process optimization, and matching. Because this research area is nascent, much of the article is devoted to open questions and directions for future data economy research. (*JEL* C80, D21, D83, E23, E24, J23)

1. *Introduction*

While the data economy has changed the way people shop and businesses operate, it has only just begun to permeate economists' thinking about aggregate economic phenomena. In the early twentieth century, economists like Schultz (1943) analyzed agrarian economies and land-use issues. As agricultural productivity improved, production shifted toward manufacturing. Modern macroeconomics adapted with models featuring capital and labor, markets for goods, and equilibrium wages (Solow 1956).

Once again, productivity improvements have shifted the nature of production. In the information age, production increasingly revolves around information and, specifically, data. The largest publicly listed firms in the world—Apple, Google, Microsoft, Amazon, and Facebook—have a combined market value that is nearly 20 percent of the market capitalization in the United States. These are digital services firms whose valuation reflects the value of their data. As for the amount of digital data produced, the United Nations (2019) estimates that every second, over 275,000 gigabytes of data are produced on digital platforms, which translates to nearly 1 billion 200-page books generated per second. These large quantities and valuations raise the question of how macroeconomists might incorporate data as a service, an input or as an asset.

This article explores the various ways that the growth of data interacts with classic macroeconomic questions concerning topics such as GDP measurement, monetary neutrality, growth, and firm dynamics; it describes tools we already have to understand the data

*Veldkamp: Columbia Graduate School of Business, NBER, and CEPR. Chung: Stanford Graduate School of Business.

† Go to https://doi.org/10.1257/jel.20221580 to visit the article page and view author disclosure statement(s).

economy; and it compares and contrasts frameworks that integrate data in a standard macroeconomic environment.

What do we mean by data? Data are information that can be encoded as a binary sequence of zeroes and ones. Of course, that includes an enormous spectrum of knowledge, not all of which is the subject of this article. For example, music, poetry, technological breakthroughs, patents, and ideas about efficient management or operations can all be digitally encoded. These are not the types of digital information typically accumulated by the big data revolution. Instead, much of big data is records or evidence of transactions. It is personal information about online buyers, satellite images of traffic patterns near stores, textual analysis of user reviews, click-through data, and other evidence of economic activity that is used to forecast sales, earnings, and the future value of firms and their product lines.

Many types of data are generated from economic activity. Some, such as Internal Revenue Service tax records, census demographic data, or import-export records, are produced or collected by governments. Other types of data are produced as a result of consumer and business activity. Within this category, some data are intentionally collected by firms as part of controlled experiments to identify causality. But most are produced as a digital byproduct of firm transactions. We will focus on this last group for our discussion.

Data are interlinked with technology; however, the two concepts are distinct. We define technology as innovation that allows firms, with a given set of inputs, to produce more outputs. As will be elaborated in more detail in section 3, data and technology have different methods of production, dispersion, and regulations. Data are often used to power technology, such as artificial intelligence (AI), which we will refer to as data technology. But the technology is distinct from the data that powers it. Our focus is data. We will touch on data technologies when they affect how data are used or valued.

We begin in section 2 by describing recent innovations in data technology that prompted much big data discussion. Section 3 describes theoretical tools and ideas from related literatures that we can use to understand the aggregate consequences of data. It explores the ways in which data are similar to and different from technology. It also describes tools from the information frictions literature we can use to understand the mechanisms at work in the data information economy. Section 4 embeds tools from both the growth and information frictions literatures in equilibrium frameworks. It presents and compares three frameworks that shed light on the conceptual questions surrounding the measurement of data and of aggregate economic activity. Finally, section 5 catalogs evidence about the importance of such digital information and how it affects various facets of the macroeconomy. But precise measurement of the data economy is beset by conceptual challenges: How should we value data? How should we account for digital services, many of which are offered at a zero price, in exchange for the user's data? Should the prevalence of digital goods and services change the way we measure GDP? These questions cannot be answered with empirics. They are conceptual challenges that require a conceptual framework to sort out. We then expand on our presented frameworks, to show how the macro frameworks could be used to resolve the data economy questions.

2. *Innovations in Data Technology*

The concept of utilizing data is not new. Why then has the focus on data been so prominent lately? The answer lies in improvements in computing power that help companies process data faster, innovations in infrastructure that allow more data to be stored more cost effectively, and most of all, progress in machine-learning techniques. Combined with the increasing availability of larger volumes and varieties of data, these advances have led to the popularity of using "big data" to generate insights. Business has adopted big data and machine-learning techniques for a variety of applications from advertising, to making faster and more accurate business decisions, to automating manual processes. In the asset management industry, such computational techniques have made investment decisions that rival those of humans (Abis 2020).

One form of data processing is AI, which can be defined as machines that display responses to stimulation, including cognitive decision-making processes, that are similar to conventional responses from humans (Shukla and Jaiswal 2013). Recently, artificial intelligence has been in the news for outperforming humans in games, image recognition, and investment (Knight 2017).

In an economic context, data are starting to play an important role in improving firms' decision-making. In the empirical micro literature, Goldfarb and Tucker (2019) discuss the reduction of economic costs in the following five activities: search, replication, transportation, tracking, and verification. These five activities can be studied in the broader scope of macroeconomic topics, which we discuss in section 3. Namely, search can be studied in a macroeconomic labor context; replication can be analyzed through the study of economies of scale and the share of the same pool of data among several firms; transportation and tracking can be studied in the context of productivity improvements; and lastly, verification can be understood in the context of banking. Bajari et al. (2019) study one particularly intensive data user, Amazon. They use Amazon's data to examine the accuracy of forecasts by firms in two dimensions: the number of products (N), and the number of time periods for which a product is available for sale (T). Their empirical results indicate gains in forecast improvement in the T dimension, but a flat N effect. The firm's overall forecast performance, controlling for N and T effects across product lines, has improved over time, suggesting gradual improvements in forecasting from the introduction of new models and improved technology. Firms have also benefited from the decreased cost of storing, computing, and transmitting data.

The influx of digital data is also playing a role in influencing tools for empirical analysis in economics research. Specifically on the topic of digital text data, Gentzkow, Kelly, and Taddy (2019) discuss how tools for analyzing text data can be incorporated into current statistical methods for economics research. In general, the rise of data has led to improvements in technology and methods for analysis. As a result, it is important to understand how these advancements have improved productivity from a macroeconomic perspective.

While the microeconomics of data, the operations implications of data, and the use of data in economics are all important topics, they are not the focus of this article. Our focus is on how these types of changes in data accumulation and data use affect economic aggregates.

3. *Existing Tools for Understanding the Data Economy*

Since the literature on data and macroeconomics is nascent, we consider which ideas from related literatures are relevant. Three literatures in particular guide much of the existing theory work: the literature on growth theory, technological progress and learning-by-doing; the literature on information frictions; and the literature on intangible capital. By comparing data to ideas, human capital, signals, and intangible capital, we can see what existing knowledge is portable to this new domain.

3.1 *Ideas and Tools from the Literature on Growth and Technology*

Comparing Data and Technology.—Data and technology have some important similarities. Data are used by firms to make strategic decisions that enhance their productivity and profitability. Modern firms use data to decide which locations to close or open, which product lines to cut, and to forecast which new goods will enjoy high demand. Like technology or total factor productivity (TFP) in standard economic models, firms use data to take a given set of inputs and produce more valuable outputs.

Another key similarity between data and technology is that both are non-rival. Just as many firms can make use of an idea at the same time, many firms can use the same data. Of course, the value of data is affected by how intensively others use it, just as the value of widget production is affected by the number of widget producers in the economy. But if a widget is a physical good, one person's use of it typically precludes another's simultaneous use of the same widget. Widgets are rival. Since data can be freely copied, this is not the case for data. The non-rivalry of data creates a force for increasing returns. These increasing returns are important because they may favor larger firms, larger investors, and larger economies.

However, data and technology are not the same. One important difference is that data and technology are produced in different ways. Creating new technology requires resources: skilled labor, perhaps a laboratory, new good prototypes, and perhaps many failures before a successful technology is discovered. In contrast, data is a by-product of economic activity. Producing and selling generates data about the volume of sales, the means of payment, and characteristics of buyers. Sometimes collecting and processing data to extract knowledge are costly. But data are not typically produced in a lab. More data comes from more economic activity. This difference in production matters. One of the fundamental insights of Romer (1990) is that monopolies are necessary to incentivize idea production. This is not true of data production. Because data are a by-product of economic transactions, no extra incentives are needed for its production.

Other differences between data and technology are the ways in which they leak and their patent protection. We know that ideas or technologies leak (Easterly 2002). Workers at garment factories in Bangladesh take their ideas to start their own firms. When they are hired away by competitors, Silicon Valley workers take their technological knowledge with them. But data are not embodied in one's mind. It is too complex, too nuanced, too extensive for that. A worker might steal data from their firm. But that is a crime. The data are not embodied in their human capital. This feature is also what distinguishes data from human capital or forms of learning by doing, defined as understanding how a process works by interactively engaging with it. Conversely, some ideas are illegal to take from one firm to another. These are ideas that are protected by patents. But patents do not protect data. Aside from datasets that can be protected as a trade secret, data do not have a set of legal institutions designed specifically to ensure the exclusive right of one entity to use a particular set of data. At the same time, because data do not easily leak, and because patent protection is not needed to incentivize the creation of data, perhaps a patent system for data is not needed.

Lower leakage of knowledge encoded as data is important for growth. It also could explain one of the most important trends in the US macroeconomy, the decrease in business dynamism. In a recent working paper, Akcigit and Ates (2023) study the sources of this declining dynamism and find that the primary cause is the decline in knowledge diffusion from the largest to the smallest firms. Thus, if data lends itself to less diffusion than traditional technologies, then the growing data economy could be responsible for the decline in firm dynamism.

Finally, the ability to monetize data and the widespread sale of data also distinguish it from technology. A few features of data lend themselves well to such transactions. First, a seller can clearly describe the contents of a dataset without revealing its information content. A buyer can know exactly how many users or clicks or transactions from which stores or websites, how much revenue is involved, which images are analyzed or text parsed, and still not know what the data will say. Second, data can be easily split. One can sell 1,000 data points or 999, or however much the buyer is willing to pay for. As mentioned above, data are less likely to leak. That makes it a more desirable purchase.

Similar to technologies, data can be sold both directly and indirectly. A data vendor can sell you a dataset directly by transferring the binary code that constitutes the data. But they can also offer data services, which are an indirect sale of data. Such a service might entail using their data to place your ad on the screen of a particular user type, it could entail using their data to choose assets to invest your money in, or it could involve using proprietary data to provide a firm with strategic business consulting advice. These services monetize data without transferring the underlying data used in the service.

Several authors have also assessed the effects of big data and AI on growth, including Agrawal, McHale, and Oettl (2018); Jones (1995); Lu (2021); Aghion, Jones, and Jones (2018); and Jones and Tonetti (2020). We will describe these papers in more detail when we discuss frameworks that treat data as technology and productivity in section 4.2.

3.2 *Ideas and Tools from Information Frictions*

Data are information. We often refer to the transmission of information as a signal. The following are ideas from the information frictions literature that use the language of signals, but can easily be adapted to think about data.

Rational Inattention as a Data Processing Allocation Problem.—In a way, models of data processing are already prevalent in strands of the macroeconomics literature. For example, one of the leading explanations for the real effects of monetary policy is costly information processing. Often referred to as rational inattention, following Sims (2003) and Maćkowiak, Matějka, and Wiederholt (2018), such models consider what types of information or data are most valuable to process, subject to a constraint on the mutual information of the processed information and the underlying uncertain economic variables. The idea of using mutual information as a constraint or the basis of a cost function comes from the computer science literature on information theory. The mutual information of a signal and a state is an approximation to the length of the bit string or binary code necessary to transmit that information (Cover and Thomas 1991). While the interpretation of rational inattention in economics has been mostly as a cognitive limitation on processing information, the tool was originally designed to model computers' processing of data. Quantification of attention to information by individuals can be likened to quantification of data by computer scientists. As a result, the process of rational inattention used to understand trade-offs in information can be directly applied to understanding the trading off of data used by firms to make business decisions. When we reinterpret rational inattention theories as data processing theories, we gain insights such as the notion that when firms have limited or costly data processing, they allocate their processing power optimally among data on various types of shocks. If they use the resulting knowledge to set prices, the aggregate response of prices to monetary shocks looks realistic (Maćkowiak, and Wiederholt 2009). Similarly, when investment managers allocate their data processing ability optimally across various types of financial data and then choose portfolios of equities to invest in, this can explain features of portfolio returns and equilibrium asset prices (Kacperczyk, Van Nieuwerburgh, and Veldkamp 2016). Rational inattention has been applied to many more problems, all of which could be reinterpreted as data processing problems. Taking these models and growing the data processing capacity (as in Begenau, Farboodi, and Veldkamp 2018) can give us new insights into the ways in which abundant data is reshaping macroeconomic forces.

Data as a By-product of Economic Activity.—Veldkamp (2005); Ordoñez (2013); and Fajgelbaum, Schaal, and Taschereau-Dumouchel (2017) all consider the idea that information is a by-product of economic activity and that firms use this information to reduce uncertainty and guide their decision-making. These authors did not call the information data. But it has all the hallmarks of modern transactions data. The data produced was used by firms to forecast the state of the business cycle. Better forecasting enabled the firms to invest more wisely and be more profitable. As such, these papers have early versions of a data feedback loop whereby more data enables more production, which in turn produces more data. These models restrict attention to data about aggregate productivity. In the data economy, that is not primarily what firms are using data for. But such modeling structures can be adapted so that production can also generate firm- or industry-specific information. As a result, they provide useful equilibrium frameworks on which to build a data economy.

Matching Frictions.—Another form of information friction is a matching friction. One way to model the data economy is as an economy with reduced matching frictions. For example, works on data platforms, such as Kirpalani and Philippon (2020) or Bergemann, Bonatti, and Gan (2022), also typically use a matching technology to embody the role of data on the platform. While much data is undoubtedly used to match workers with jobs, customers to products, and producers with suppliers, both matching models and noisy signal models predict that reduced frictions, that is, better data, will improve such matches.

What then is the difference between modeling data as more precise signals or higher-quality matches? The difference is risk. With matching models, there is rarely a role for uncertainty. Matches are high quality or low quality, but not typically risky. With noisy signals, agents typically explicitly consider the uncertainty of an outcome when they choose an action. Noisy signals capture the idea that data decrease uncertainty. In many circumstances, uncertainty may be important. For example, in financial markets, the expected return on equity is about two-thirds risk premium and one-third expected return. In such settings, capturing the effect of data on risk would be important.

3.3 *Ideas from Intangible Capital*

Data are also a form of intangible capital, though distinct from other forms of intangible capital because data pose different measurement issues. Intangible capital measures generally use the cost of investment to value the intangible capital stock. But data are often a by-product of activity, with little or no creation cost. Though there is some cost associated with warehousing and processing, the data aren't created in a costly way, so they are often not counted as having positive value in the intangible capital stock.

There is a possible category on firms' balance sheets for data, allowing data to be considered intangible capital. If a firm buys data, it is clear that it should be valued at its market price. But if the firm produces its own data, the value may be determined at the firm's discretion. Intangible capital is a key component of the debates about investment stagnation, long-run increasing markups, and productivity measurement.

In the investment literature on Q theory, Crouzet and Eberly (2018) build on Hayashi (1982) by adding intangibles as a different form of capital with a similar role, but not perfectly substitutable with tangible capital. They attribute the rise of intangible capital as the reason for the retail sector trend of increased investment in technology-driven business practices and therefore higher productivity despite weak investment relative to strong cash flow and valuation. Belo et al. (2022) incorporate two types of intangible capital, knowledge capital and brand capital, into their neoclassical model of investment in order to study determinants of a firm's market value. They also note the declining importance of physical capital, in contrast with the growing importance of knowledge capital, over the past few decades.

Similarly, Brynjolfsson, Hitt, and Yang (2002) argue that information technology (IT) is associated with indirect and direct measures of intangible assets. They argue that the financial markets treat the organizational assets that complement information technology similarly to other assets associated with long-term growth. This is helpful because we can measure information technology expenditure, accumulation, and depreciation more accurately than we can measure data. Byrne, Corrado, and Sichel (2018) construct measures to quantify the service prices, quantities, and capital investment in the US cloud services industry. They highlight the growth of capital expenditure by large cloud service providers. This suggests an explosion in the value of data and other intangible digital capital.

In contrast to the growth literature on data, the intangible investment literature often assumes that data, with diminishing returns, function like physical capital. Instead, this literature focuses more on measurement than the other two. As such, it offers measurement ideas that could be important for the macro-data agenda.

4. *Modeling the Data Economy*

The previous sections explored data facts and tools. What is only starting to emerge is a clear idea of how data are an integral part of the economy, both as an input into economic activity and as a by-product of economic activity. An equilibrium framework with both features is valuable because it can explain the difference between price, cost, and value of data and data-related goods.

One reason to use theory is that the economy is in the midst of a transition. Empirical work takes past trends or covariances and extrapolates to form predictions. But in transitions, the future often looks systematically different from the past. Another reason to model data is for policy analysis. Policy questions about trade and firm competition all involve potential regulation of data. Such policy changes have equilibrium effects. We have no data from such alternative policy regimes with which to estimate these effects. Instead we need structural models, which we can estimate and then alter to perform policy counterfactual experiments.

Specifically, we compare three classes of models. The first set of models thinks of data as a by-product of transactions that can help businesses optimize processes. The next class of models considers data as an input to research and knowledge creation. In both types of models, the use of data is equated with the production of data. Data production depends on the macroeconomy because it is a by-product of economic activity; it influences the macroeconomy because firms use it to improve their productivity. The final class of models posits an intermediate step between data production and data use. It treats data as an input that can be combined with labor to produce actionable knowledge.

4.1 *Models of Data for Business Processes*

The first model framework focuses on transactional information that aids firms in improving their business processes. When data are information used to forecast an unknown state, it will not sustain perpetual growth because there is a natural limit to how much a forecast can improve. Variance (forecast errors) can never be less than zero. While models of data as both information and technology argue that there can be a region of increasing returns, data in the information-based model will eventually have returns that decrease, bringing data-driven growth to a halt.

The below model is a simplified version of Farboodi and Veldkamp (2022).

Model Setup.—Note that time is infinite and discrete. We index a unit continuum of competitive firms by i. Every firm i uses $k_{i,t}$ units of capital to produce $k_{i,t}^{\alpha}$ units of goods, each of which have quality $A_{i,t}$. We denote P_t as the equilibrium price of these goods with varying amounts of quality. The aggregate quality-adjusted supply is: $Y_t = \int_i A_{i,t} k_{i,t}^{\alpha} di$. The inverse demand function describes the market clearing price that declines as supply Y_t grows: $P_t = \bar{P} Y_t^{-\gamma}$. Firms treat the aggregate price P_t as given.

The good's quality is important because firms will use data to improve quality. Quality $A_{i,t}$ is dependent on the firm's choice of a production technique $a_{i,t}$, and each firm has one optimal technique. That optimal technique has a persistent and a transitory component: $\theta_{i,t} + \epsilon_{a,i,t}$, neither of which can be separately observed. The persistent component $\theta_{i,t}$ follows an AR(1) process: $\theta_{i,t} = \bar{\theta} + \rho\left(\theta_{i,t-1} - \bar{\theta}\right) + \eta_{i,t}$, which firms can use data to learn about. $\eta_{i,t}$, the AR(1) innovation is i.i.d. over time. For simplicity, we assume here that $\eta_{i,t}$ is also independent across firms.[1] Meanwhile, the transitory shock $\epsilon_{a,i,t}$ is also i.i.d. across firms and time, but cannot be learned. The optimal technique is significant to a firm because it determines the quality of a firm's good. Quality $A_{i,t}$ depends on the squared difference between the firm's chosen production technique $a_{i,t}$ and the optimal technique $\theta_{i,t} + \epsilon_{a,i,t}$:

$$A_{i,t} = g\left(\left(a_{i,t} - \theta_{i,t} - \epsilon_{a,i,t}\right)^2\right), \tag{1}$$

where $g(\cdot)$ is a decreasing function. The role of data in this model is to help firms select better production techniques. One interpretation of this role is that data can, for example, inform a firm whether blue or purple shoes or gas or electric cars will be more valued by their consumers. This allows the firm to produce accordingly. Transactions reveal these preferences. However, preferences are continually changing and firms must constantly adapt to them. Another interpretation is that data can help a firm to optimize its inventory or transport operations.

Data provide information on $\theta_{i,t}$. The purpose of the temporary shock ϵ_a in this model

[1] Farboodi and Veldkamp (2022) relax this assumption. When the optimal technique is correlated across firms, then firms can learn from each others' data. In such an environment, firms choose to buy and sell data to other firms.

is to prevent firms from learning $\theta_{i,t}$ through observation of their revenue at the end of period t. Without the inclusion of ϵ_a, the collection of past data would not be considered a valuable asset since the firm could maximize the quality of the good upon learning the current value of $\theta_{i,t}$ by setting $a_{i,t} = \theta_{i,t}$. The next assumption embodies the notion that data are a by-product of economic activity. We define $n_{i,t}$, the number of data points observed by firm i at time t, as a function of their production in the prior period $k^{\alpha}_{i,t-1}$:

$$n_{i,t} = z_i k^{\alpha}_{i,t-1}, \tag{2}$$

where z_i is the parameter that denotes the "data-savviness" of a firm. Here, a data-savvy firm harvests lots of data per unit of output. For example, a firm that produces a simple mobile application but harvests all the data on its consumers' phones has high z_i because it obtains many data points per unit of output. On the other hand, firms that do not track transactions have low or zero z_i. Each data point $m \in [1 : n_{i,t}]$ reveals the following signal:

$$s_{i,t,m} = \theta_{i,t} + \varepsilon_{i,t,m}, \tag{3}$$

where $\varepsilon_{i,t,m}$ is i.i.d. across firms, time, and signals. For ease of use, we assume a normal distribution for all the shocks in the model: fundamental uncertainty is $\eta_{i,t} \sim N(\mu, \sigma^2_{\theta})$, unlearnable quality transitory shock is $\epsilon_{a,i,t} \sim N(0, \sigma^2_a)$, and signal noise is $\varepsilon_{i,t,m} \sim N(0, \sigma^2_{\varepsilon})$.

Firm Problem.—A firm makes a series of production and quality choices $k_{i,t}, a_{i,t}$ to maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(P_t A_{i,t} k^{\alpha}_{i,t} - r k_{i,t}\right). \tag{4}$$

Belief updating for $\theta_{i,t}$ is done using Bayes's law. Firms observe the previous period's revenues and data each period, and then decide on capital level k and production technique a. The information set of firm i when selecting $a_{i,t}$ is denoted as $\mathcal{I}_{i,t} = \left[\{A_{i,\tau}\}^{t-1}_{\tau=0}; \left\{\{s_{i,\tau,m}\}^{n_{i,\tau}}_{m=1}\right\}^{t}_{\tau=0}\right]$. For simplicity, we assume the rental rate of capital is given. However, one can also embed an equilibrium context into this setup where capital markets are cleared or where one can add labor markets and endogenize the demand for goods. This model is only a theoretical sketch. It allows us to focus on the data-relevant mechanisms.

Solution.—The state variables of the model's recursive problem are (1) the beliefs about $\theta_{i,t}$, and (2) the precision of those beliefs. Solving the first-order condition with respect to the technique choice, we arrive at the following optimal technique: $a^*_{i,t} = E_i[\theta_{i,t} | \mathcal{I}_{i,t}]$. We then define posterior precision of beliefs as $\Omega_{i,t} := E_i\left[\left(E_i[\theta_{i,t}|\mathcal{I}_{i,t}] - \theta_{i,t}\right)^2\right]^{-1}$ to get the expected quality $E_i[A_{i,t}] = \bar{A} - \Omega^{-1}_{i,t} - \sigma^2_a$. Substituting for the optimal technique choice a^* allows one to eliminate the firm's expected value of $\theta_{i,t}$ as a state variable. We can then express expected firm value recursively.

LEMMA 1: *The optimal sequence of capital investment choices $\{k_{i,t}\}$ solves the following recursive problem:*

$$V_t(\Omega_{i,t}) = \max_{k_{i,t}} P_t\left(\bar{A} - \Omega^{-1}_{i,t} - \sigma^2_a\right)k^{\alpha}_{i,t} - r k_{i,t} + \beta V_{t+1}(\Omega_{i,t+1}), \tag{5}$$

where $n_{i,t+1} = z_i k^{\alpha}_{i,t}$ and

$$\Omega_{i,t} = \left[\rho^2\left(\Omega_{i,t-1} + \sigma^{-2}_a\right)^{-1} + \sigma^2_{\theta}\right]^{-1} + n_{i,t}\sigma^{-2}_{\epsilon}. \tag{6}$$

Refer to Farboodi and Veldkamp (2022) for the proof. Note that equation (1) is the value of a firm with data Ω_{it}. For an expected profit maximizer, this is the equity value. Equation (6) describes the depreciation of data. Yesterday's stock of data $\Omega_{i,t-1}$ is augmented by the observation of a firm's own output, which adds precision σ_a^{-2}. But then two forces make yesterday's data less relevant for forecasting today. The fact that the state has persistence ρ less than one means that information about yesterday is less relevant for today. Also, the variance of the innovations in the state process σ_θ^2 makes information about yesterday's state less relevant. This result allows us to reduce our setup to a deterministic problem with only one state variable, $\Omega_{i,t}$, because expected quality $A_{i,t}$ depends on the conditional variance of $\theta_{i,t}$ and because our information structure can conform to the structure of a Kalman filter where the conditional variance sequence is deterministic.

4.2 *Models of Data for Ideas and Research*

While the previous class of models looked at transaction data for business process improvements, the next type of model studies data used as input to idea generation and research.

Jones and Tonetti (2020) explore a growth economy where data is a by-product of economic activity and an input into productivity. Units of data translate into firm total factor productivity. They explore how different data ownership models affect the rate of economic growth. These model differences are essential for their main question: How should policymakers regulate data ownership? Jones and Tonetti (2020) show that in equilibrium, firms undershare data. Because of non-rivalry, there may be large social gains to sharing data across firms, even in the presence of privacy considerations. However, fearing creative destruction, firms may choose to hoard data they own, leading to the inefficient use of non-rival data. They conclude that giving the data property rights to consumers can generate allocations that are close to optimal. While this model covers data privacy considerations, we are primarily interested in the characteristics of how it relates data to productivity.

Below is a simplified version of Jones and Tonetti (2020).

Model Setup.—Time is infinite and continuous. There are N_t varieties of consumer goods produced by N_t firms at each time period t. There is also a representative consumer with the following log flow utility at time period t:

$$(7)\quad u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\kappa}{2}\frac{1}{N_t^2}\int_0^{N_t} x_{it}^2\, di - \frac{\tilde{\kappa}}{2}\frac{1}{N_t}\int_0^{N_t} \tilde{x}_{it}^2\, di,$$

where c_t is the individual's consumption, x_{it} is the proportion of data on an individual's consumption of consumer good variety i that is used by firm i, and $\tilde{x}_{it}$ is the proportion of the consumer's data on good variety i that is shared with other firms that aren't the one producing variety i. Note that privacy costs are incorporated as a quadratic loss function with weights κ and $\tilde{\kappa}$ to represent the tradeoff experienced between privacy and consumption. The individual gains utility from consumption of the good but loses utility from their data being directly used by the firm and shared with other firms. Firm i produced good variety i according to:

$$(8)\quad Y_{it} = \Omega_{it}^{\eta} L_{it}, \quad \text{with} \quad \eta \in (0,1),$$

where Ω_{it} is the amount of data used to produce good variety i and L_{it} is labor. Note that there are constant returns to scale in the competing input, labor, and increasing returns to scale when considering both labor and data together. The increasing returns arise from the non-rival property. All the workers at the firm are able to access the same pool of data available to make good variety i without depleting it. As in the previous model, data is created as a by-product of consumption. We define n_{it} as the amount of data created about variety i.

$$(9) \qquad n_{it} = c_{it} L_t = Y_{it}.$$

The amount of data, Ω_{it}, used by the firm is:

$$(10) \qquad \Omega_{it} \leq \alpha x_{it} n_{it} + (1-\alpha) B_t,$$

where the first term represents the amount of data on variety good i that firm i can use to produce, while B_t is the aggregated bundle of data shared by other firms on other varieties of goods.

$$(11) \qquad B_t = \left(N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} \Omega_{sit}^{\frac{\epsilon}{\epsilon-1}} di \right)^{\frac{\epsilon}{\epsilon-1}},$$

$$\text{with} \quad \epsilon > 1,$$

where $\Omega_{sit} = \tilde{x}_{it} n_{it}$ is the amount of data on good variety i that is shared with other firms for the production of their good varieties. Since data is non-rival, the bundle can be simultaneously used by other firms without being depleted. Note that the above expression is an inequality because in this setup, if consumers own the data, then they may restrict the amount of data that can be used by the firm (i.e., $x_{it} < 1$).

Firms exit through a random death process. A firm gets hit by an exit shock $\{0,1\}$, where the probability of exit δ depends on how much of its data the firm shares with others. The idea is that firms who sell all their data lose some of their competitiveness and that makes them less likely to survive. This assumption means that not all firms are willing to share all of their data.

The solution presented by Jones and Tonetti (2020) shows that the social output per person is proportional to size of the economy, raised to a power greater than one, due to both the standard preference of more variety and the result that the non-rival property of data increases returns. Larger economies that produce more data experience large gains because this data gets reused by firms to increase production. As data's importance in the economy increases, more resources are devoted to activities that create more data (i.e., production) as opposed to activities that don't (i.e., firm entry).

Model Comparisons.—Both models have similar production structures that treat data as a by-product of consumption and use data to augment productivity. The key difference is how data is mapped into productivity (A). In Farboodi and Veldkamp's (2022) model, if we restrict the quality function to be linear, then the expected quality of a good depends on a constant $\bar{A}$, the forecast variance Ω^{-1} and the variance of the unlearnable shock σ_a^2:

$$(12) \qquad A(\Omega) = \bar{A} - \Omega^{-1} - \sigma_a^2.$$

The model portrays data as a way to reduce forecasting error to zero. Reducing a forecast error has bounded value because forecast variance is bounded below by zero. If there is an infinite amount of data, the forecast variance Ω^{-1} falls to zero. But quality is still finite.

Meanwhile, in Jones and Tonetti's (2020) model, the mapping from data to productivity is:

$$(13) \qquad A(\Omega) = \Omega^{\eta}.$$

This models data as something that contributes to productivity in an unbounded way. The implication is that data accumulation is akin to idea accumulation.

In other words, Farboodi and Veldkamp (2022) treat data as contributing to prediction. Jones and Tonetti (2020) consider data as something that advances ideas. Some of the authors who model AI consider data to be an input into the production of ideas. It is true that machine-learning algorithms are ultimately prediction algorithms. They are designed to spot patterns in data and use it to predict likely outcomes. At the same time, artificial intelligence has impacted idea creation. In the health and medical industries, it is revolutionizing the way research and development (R&D) is performed in the lab by predicting which directions of research are likely to be fruitful. In areas where advances in drug discovery, chemistry, and material sciences are slowing and becoming more costly, the ability of deep learning algorithms to uncover complex insights and predictions more quickly and affordably has been key to pushing forward the research frontier and making investments in these areas more viable (Rotman 2019).

A similarity of the first two frameworks, Jones and Tonetti (2020) and Farboodi and Veldkamp (2022), is that allowing the sale of data matters. Both models make assumptions to guarantee that not all data in the economy can be used to produce a given good. In Jones and Tonetti (2020), firms that share too much data are more likely to exit. In Farboodi and Veldkamp (2022), firms that sell data lose some of that data. The reason for these costs and restrictions is that, otherwise, a competitive market equilibirum for data does not exist. In a competitive environment, where data do not deplete and any one firm's actions do not affect the market prices, all firms would choose to sell an infinite amount of data for any positive price. This would lead to a lack of a well-defined finite selling policy and consequently, in the absence of any price impact, there is no cost to selling one's data. In order to model meaningful data-sharing decisions by firms, a firm's decision must affect its market conditions so that the notion that selling more data will reduce profits can be internalized. That implies some form of imperfect competition.

So, just like growth models with idea production, models of data production require some market power. But the reasons are quite different. Growth models require market power to provide benefits that induce firms to produce ideas. Data models do not need incentives to produce data because data is a by-product of economic activity. Instead, data models require market power to create costs that limit data sharing.

Another important difference between the two models is data depreciation. In Jones and Tonetti (2020), data depreciates completely at the end of every period. That simplifies the problem by removing data as a state variable. This allows one to use the model to infer a flow value of data, but not to value data as an asset. Meanwhile, Farboodi and Veldkamp (2022) allow for the accumulation of data and introduce the idea of data depreciation. This provides a more traditional accounting approach to measuring and valuing data as an intangible asset.

The Interaction of AI and Data-driven Growth.—One objection to the conclusion of diminishing returns to data is that perhaps AI technology allows data accumulation to create new ideas, which can sustain growth. That is likely. But how much growth is created depends largely on how data contributes to knowledge and ideas.

Cong, Xie, and Zhang (2021) construct an endogenous growth economy, where data is an input into research. Data is neither equivalent to technology, nor is productivity held fixed. Instead, intermediate good firms use data to contribute to product variety in a way that distorts the allocation of labor. However, because of diminishing returns to variety, an economy with a higher level of knowledge accumulation experiences less benefit from an additional marginal unit of data. That diminishing returns to data have a different origin, but are similar to the diminishing returns in the Farboodi and Veldkamp (2022) model. However, as in Jones and Tonetti (2020), data are dynamically non-rival and do continue contributing to long-run growth because data contribute to the creation of new goods.

The conclusion that data exhibit diminishing returns is similar to that of Nordhaus (2021), but for different reasons. He explores the question: Does AI bring us close to economic *singularity*—an increase in economic growth at an accelerating pace resulting from growth in computation and AI? Nordhaus considers a production economy with information capital K, which one might call data, and information technology A that augments the information capital. Output Y is a Cobb–Douglas function of AK and labor L,

$$Y_t = (A_t K_t)^{\alpha} L_t^{1-\alpha_t}, \tag{14}$$

but with one twist: the exponent α_t can be time varying. The key question for Nordhaus is whether data technology raises the income share α, leaving labor earning less and less of aggregate income, over time, but allowing the economy to grow at an ever-expanding rate.

To answer this question, he evaluates the substitutability between information (data) and conventional inputs along seven different metrics and concludes that singularity is not near. Several nonroutine tasks remain a bottleneck.

In a similar vein, but more explicitly about the AI technology rather than about data, Aghion, Jones, and Jones (2018) explore the role of AI in the growth process and its reallocative effects. Using Baumol's (1967) cost disease insight, the authors argue that Baumol's cost disease leads to the traditional industries' declining share of GDP as they become automated, which is offset by the growing fraction of automated industry. This explains the observed stability in the capital share and per capita GDP growth over the past century, despite evolving automation. In their model, labor share remains substantial because labor represents a bottleneck for growth. The authors also suggest that AI can increase growth of new ideas and potentially obviate the role of population growth in generating exponential economic growth. Nevertheless, even though AI can theoretically generate runaway technological growth and propel infinite income in finite time, growth may remain limited due to essential areas of production that are hard to improve. Moreover, AI may discourage future innovation for fear of imitation, undermining incentives to innovate in the first place.

Agrawal, McHale, and Oettl (2018) develop a combinatorial-based knowledge production function and embed it in the classic Jones (1995) growth model to explore how breakthroughs in AI could enhance discovery rates and economic growth. Lu (2021) embeds self-accumulating AI in a Lucas (1988) growth model and examines growth transition paths from an economy without AI to an economy with AI and how employment and welfare evolve. The authors show that AI can increase economic growth, and the evolution of AI increases household lifetime welfare in the long run.

Thus, Nordhaus concludes that data/data technology have diminishing marginal returns by looking at how they are used in the production process and then assuming a functional form, consistent with his observations. Farboodi and Veldkamp argue for diminishing returns, based on the properties of information used in forecasting. Aghion and Jones argue that the existence of AI technology is unlikely to overturn this conclusion. For some purposes, modeling data exactly like capital is simple and transparent. For others, a nuanced understanding of why the returns to data decline and what circumstances lead to faster or slower decline could advance thinking.

4.3 *Distinguishing Data Production From Data Use*

Previously discussed models equated data production with data use. However, without computers, skilled analysts, data warehousers, and other complementary workers, data are not useable. The next class of models show how data production is mapped to data use. It explores how data are combined with other inputs to create knowledge.

The empirical literature linking IT equipment to firm efficiency is active. Section 3.1 discussed research that links IT investment and expenditures, which may include data, to productivity (e.g., Brynjolfsson and Yang 1996). But knowing that these are correlated is not the same as understanding how data might be used to produce knowledge. That understanding requires some structure.

Abis and Veldkamp (2024) explore how new data technologies have changed the production function for knowledge. Instead of mapping capital and labor to production of goods, they consider an economy where data and labor are combined to produce knowledge. They model the new-technology knowledge production function as:

$$P_{it}^{AI} = A_t^{AI} D_{it}^{\alpha} L_{it}^{1-\alpha},$$

where P^{AI} is knowledge produced using these new technology advances, A^{AI} is a time-varying productivity parameter, D is structured data, and L is the labor input for data analysts trained with machine-learning skills.

The old-technology production function represents knowledge produced without the new big data and machine-learning advances. It takes the same form, with potentially different productivity and income shares:

$$P_{it}^{OT} = A_t^{OT} D_{it}^{\gamma} l_{it}^{1-\gamma},$$

where P^{OT} is knowledge produced using old technology existent before these data advances, A^{OT} is a time-varying productivity parameter, D is structured data, and l is the labor input for data analysts trained with traditional analysis skills. They use hiring and wage data to structurally estimate the difference between α and γ. The large difference they find suggests that the data revolution is as large a change in technology as the industrial revolution.

This approach is similar to Joseba Martinez (2018), who focuses on a specific type of data-related asset, automation capital, and studies the production function associated with it. While automation and data are distinct, the idea that production functions are changing is similar. Automation technology is modeled as a form of innovation that can replace labor with capital. The constant elasticity of subsitution (CES) production function parameters depend on automation technology adoption. The difference is that data are not directly an input into good production. Data are used to create knowledge, which in turn is used to produce goods and services more efficiently. But like automation, new data technologies are likely changing the optimal capital–labor mix.

One could extend these models by incorporating IT equipment along with data as inputs to knowledge production. Furthermore, one could also model data as a by-product of economic production. Incorporating that property might facilitate data measurement, since transactions are observable. Overall, this framework gives us a macroeconomics approach to understanding data, along with other inputs such as capital labor and IT equipment, in the context of economic growth and knowledge production.

5. *Directions for Macro Data Research*

The technological progress in data processing has not gone unnoticed by economists.

The data economy has impacted every corner of macroeconomic research. This section catalogs facts that speak to these changes. It then discusses the different ways data have impacted the macroeconomy through the lens of the tools and model frameworks we previously presented. These lines of research are still in their early stages. The following facts motivate the discussed theoretical frameworks.

5.1 *Measuring GDP and the Size of the Economy*

The data economy is large. However, not all the gains in well-being arising from digital goods and services are captured by measures of GDP.

One missing component from national accounts is the value of zero-price goods, which are prevalent in the digital economy. Brynjolfsson et al. (2019) introduce a new metric, GDP-B, which quantifies and captures the welfare contributions of these goods. Through incentive compatible choice experiments, they show that welfare gains from Facebook add about 0.05–0.11 percentage points to GDP-B growth per year in the United States. Brynjolfsson, Collis, and Eggers (2019) use online choice experiments to measure consumers' willingness to accept compensation for losing access to various digital goods and show that losing access to all search engines or all email services for one year, for instance, generates disutility equivalent to earning $500–$1,000 less per year.

Often, such zero-price digital goods are not truly free. They are services offered in exchange for data. This is a barter trade, where the service is being bartered for personal data. Barter is not measured by GDP.

Digital service innovations are also not fully captured by traditional GDP measures. Byrne and Corrado (2019) propose a framework for measuring digital services improvements, specifically in the context of advancements made in providing consumers content, by incorporating capital service flows that improve existing GDP measures of personal consumption. With their new measure, Byrne and Corrado (2019) show that these content delivery services increased consumer surplus by almost $1,800 per connected user per year and contributed over one-half of a percentage point to US real GDP growth in the last 10 years.

A framework that attempts to capture the digital value that GDP misses is Hulten and Nakamura's (2017) model of information goods as "output saving." Information technology like e-commerce reduces the need for traditional consumer goods and services, such as transportation, that would otherwise be captured in GDP. Since the electronic component of e-commerce itself has no explicit price, it is not captured in GDP. Since these goods reduce our need for value that would otherwise be measured in GDP and are not themselves counted, they bypass GDP and create consumer surplus directly. The authors propose a concept of expanded GDP (EGDP) that combines the conventional GDP measure with a willingness-to-pay metric of the value of output-saving innovation to consumers. This new metric suggests that living standards, as measured by EGDP, rose at a faster rate than real GDP growth.

Data also increase measured GDP of non-data firms by increasing productivity. Data-driven decision-making (DDD) is defined as the process of making business decisions based on the collection and analysis of external and internal data. Brynjolfsson, Hitt, and Kim (2011) find that firms that adopt DDD have output and productivity that is 5–6 percent higher than what would be expected given their other investments and information technology usage. The relationship between DDD and performance also appears in other performance measures such as asset utilization, return on equity, and market value.

One active area of debate is whether the difficulty of measuring the digital economy can explain the productivity slowdown. Brynjolfsson, Rock, and Syverson (2018) postulate four data-related explanations for the productivity slowdown, which has exceeded 1 percent per year for more than a decade: false hopes about AI, mismeasurement, redistribution, and implementation lags. Of the four hypotheses, the last one—the fact that AI has not yet diffused widely—has been the biggest contributor, according to the authors. Data advancements have also led to novel ways to measure various macroeconomic variables. Improved image and data processing of satellite night lights data have notably been useful in improving upon GDP measures in countries where traditional data tend to be of poor quality or generally unavailable (Henderson, Storeygard, and Weil 2012).

Data and digital goods and services certainly present multiple measurement challenges. These challenges are both practical and conceptual. Research and practice will have to adapt for our measures of aggregate economic activity to remain accurate.

The above frameworks can be used to express the value of data. For example, in Farboodi and Veldkamp (2022), the value function, $V(\Omega)$, expresses the value of the data stock Ω. Because the model has data accumulation, the distinction between the stock and flow of data lends itself well to adopting measures like the investment and stock of capital. On the other hand, in Jones and Tonetti (2020) there is no accumulation of data, since data depreciates fully in each period. As a result, the model cannot cannot use national income and product accounts (NIPA) methodologies to value the stock of data. Furthermore, in Jones and Tonetti (2020), production output is zero without data, similar to how output is zero without capital or without labor. Thus, just like the value of capital and labor can be computed by considering the capital share and labor share of income, we can consider a similar valuation for data. Both models suggest that the inflow of data, the analog to investment in an intangible asset, is determined by the sales of the firm.

5.2 *Valuing Data*

The literature discussed suggests a variety of empirical approaches that can be used to estimate data's value.

In Farboodi and Veldkamp (2022), $\Omega_{i,t}$ can be thought of as the amount of data accumulated by the firm. While $\Omega_{i,t}$ is technically defined as the precision of the firm's posterior belief, it is also a sufficient statistic for the history of data observed, up until date t. Thus, the marginal value of an additional piece of data with precision 1 would be $\partial V_t / \partial \Omega_{it}$. Similar to how macroeconomists have calibrated traditional recursive general equilibrium models with capital, a similar approach with data-related moments could be used to calibrate this model. The calibrated model would reveal the aggregate value of data.

A second approach would be to add markets for buying and selling data. The data price would represent the firm's demand, that is, its marginal willingness to pay for—or to sell—data. Consequently, one could infer parameters of the model using data sales prices.

A third approach to valuing data is by looking at the labor market. In Cong, Xie, and Zhang (2021) and Abis and Veldkamp (2024), the wages and employment of workers who work with data reveal how much the firm values the data they work with.

Similar to how a large asset pricing literature used a variety of approaches to detemine the value of an equity asset, many future studies will be needed to develop robust data valuation techniques. All three approaches should be explored. More work remains to be done to uncover datasets, refine measurement approaches, and determine what is the most accurate valuation method.

5.3 *Valuing Zero-Price Goods*

The optimal price of a firm's good in this setting may be close to zero. The reason is that the firm wants to sell many units in order to accumulate data that will boost the productivity of future production. It could also be that the firm wants to accumulate data in order to sell it. For example, social media platforms may be incentivized to set the optimal price of their service to zero. By allowing individuals to use their product for free, they are able to collect more data. Such data is valuable for third-party advertising and marketing services. So, the collected data become the primary source of the firm's profit.

5.4 *Measuring Data*

The frameworks suggest two possible ways of measuring data. One is to measure output or transactions. If we think data are a by-product of economic activity, then a measure of that activity should be a good indicator of data production. Of course, that could be different from data use if data are traded. But one can adjust data production for data sales and purchases to get a flow measure of data. Then, a stock of data is a discounted sum of data flows. The discount rate depends on the persistence of the market. If the data are about demand for fashion, then rapidly changing tastes imply that type of data has a short longevity and a high discount rate. If the data are mailing addresses, that market is quite persistent. An AR(1) coefficient of the variable being forecasted is sufficient to determine the discount rate.

The second means of measuring data is to look at what actions it allows firms to choose. A firm with more data can respond more quickly to market conditions than a firm with little data to guide it. To use this measurement approach, one needs to take a stand on what actions firms are using data to inform and what variable firms are using the data to forecast, and then measure both the variable and the actions. One example is portfolio choice in financial markets. Farboodi et al. (2022) use the covariance between prices and future earnings to infer the covariance between investment choices and future earnings. That covariance between choices and unknown states reveals how much data investors have about the future unknown state. A similar approach could be to use the correlation between consumer demand and firm production across a portfolio of goods to infer a firm's data about demand.

Which approach is better depends mostly on the data available. One difference between the two is the units. Measuring correlation gives rise to natural units, in terms of the precision of the information contained in the total dataset. The first approach of counting data points measures the data points more directly. But not all data are equal. Some data are more useful for forecasting a particular variable. The usefulness or relevance of the data is captured in how it is used to correlate decisions and uncertain states.

5.5 *Growth, Development, and Trade*

The rise of the data economy has promoted global economic fluidity, reducing language and geographical barriers. As a result, this digital age could present opportunities for new countries to emerge as potential key players in the economy. At the same time, the increased connectedness between countries has led to concerns about how best to regulate sharing data across borders.

The rise of a data-driven economy has raised fundamental questions about exporting and tariff policy that are at the center of current trade negotiation. Control of data has become a significant issue in trade negotiations. Given that consumers' data are not priced, every transaction with a foreign company results in an asset, consumers' data, being given away to foreign firms. One new argument for tariffs that arises in the modern economy is that they compensate the home country for the value of the unpriced data being transferred. Another rationale for tariffs is that the economies of scale in data and the economies of scope and knowledge externalities in AI innovation could create the opportunity for country-level rents and strategic trade policy (Goldfarb and Trefler 2018). Furthermore, the rise of data complicates international law. Data is similar to intellectual property in the sense that it is non-rival and can be duplicated when purchased and resold. As a result, research on intellectual property protections in trade may become relevant in understanding protection of data in trade policy.

In international trade, data can reduce border frictions by decreasing language frictions and by lowering transportation costs. Both could help poor, remote countries benefit more from data. Using eBay, the American e-commerce company, as the study sample, Brynjolfsson, Hui, and Liu (2019) find that the introduction of a machine translation system has significantly increased international trade, raising exports by 17.5 percent through a reduction in translation-related search costs. It is also clear that transportation costs of data and digital goods are quite different from those of traditional goods. Transporting data requires a large, fixed investment of IT infrastructure at the country level. However, even in many of the poorest countries, privately owned mobile phone networks already exist to carry data. Once some data infrastructure is in place, it may not be free to use, but its cost is nowhere near that of the road, sea, or air transport of a stream of physical goods.

Many remote countries have been held back by the difficulty of physically transporting goods to markets where the goods are valued. If these countries can develop the human capital to produce digital goods and services, they may be able to overcome their locational and physical infrastructure impediments. Indeed, IT infrastructure has significantly impacted many countries. For example, Hjort and Poulsen (2019) study the impact of fast internet on employment in 12 countries in Africa. They find large positive effects on employment rates with minimal job displacement and an overall increase in average income.

As a result of lower language barriers and the negligible marginal cost of transporting digital goods, many policymakers hope to capitalize on advantages offered by the digital economy to help their countries leapfrog. For instance, President Paul Kagame of Rwanda, a small, landlocked country, introduced his plan for the National Information and Communication Infrastructure policy by claiming: "In Africa, we have missed both the agricultural and industrial revolutions and in Rwanda we are determined to take full advantage of the digital revolution. This revolution is summed up by the fact that it no longer is of utmost importance where you are but rather what you can do." Pushing for policy that promotes development of knowledge, skills, and infrastructure for information and communications technology, President Kagame expressed his hopes that such innovations would help bring Rwanda to the forefront of global economic competition (Government of Rwanda 2010).

The ability to benefit from data surely depends on local human capital. While Mankiw, Romer, and Weil (1992) show how to augment the Solow growth model with human capital, a similar exercise could add human capital to (12) or (13).

The presented frameworks provide guidance about how unpriced data transferred during international trade might be valued and taxed. The valuation issue is the same issue as GDP measurement. In order to explore taxation, one needs a framework where taxation is optimal. Costinot, Lorenzoni, and Werning (2014) develop a theory of capital controls between two countries where controls are set to dynamically manipulate terms-of-trade (the price of exports relative to imports). They show that optimal taxation on capital flows for a country depends on how quickly that country is growing relative to the rest of the world. One might substitute a stock of data for the capital stock to explore the optimal set of tariffs or controls on cross-country data flows.

5.6 *Pricing and Monetary Neutrality*

One reason firms use data is to set prices. Online stores can do this easily with algorithms to automatically adjust prices. This shift in pricing technology therefore improves price flexibility. But since monetary policy efficacy depends on price rigidity, the digital economy might reduce monetary policy effectiveness.

Gorodnichenko and Talavera (2017) find that price changes occur more frequently in online stores (once every three weeks or more frequently) than in regular stores (once every 4-5 months or less frequently). Such an increase in price flexibility could alter the real effect of monetary policy.

Data are facilitating personalized pricing. Kehoe, Larsen, and Pastorino (2018) explore how data might be used for personalized pricing. The authors argue that personalized pricing will change firm competition and market dynamics. Additionally, Jin and Vasserman (2021) study one example of personalized pricing: a car insurance company that collects data from consumers and uses this information to personalize insurance costs for each individual, giving discounts to safer drivers and surcharges to riskier ones. This pricing strategy changes the relevant price elasticity that a firm faces. Thus, individualized prices could be more or less responsive to changes in monetary policy. Finally, the abundance of data is creating a new digital services sector. These services, such as on-demand computing, are being priced in new ways (Huang and Sundararajan 2011).

Another way data interact with monetary policy is through the banking sector. Vives (2019) explores how data technology has increased efficiency and service in the banking industry, but also how it has raised concerns about consumers' data protection. These competing forces will challenge regulators to balance the needs of innovation and privacy as they oversee the industry's transformation. Specifically, one form of data technology that has played a role is fintech, defined as the use of automation technology and innovative information in financial services. Vives (2017) describes the influence of fintech on the banking sector and capital markets in the following areas: efficiency, banking market structure, incumbent and entrant strategies, and financial stability. While fintech has potential for enhancing welfare, Vives (2017) calls for regulation to ensure financial stability is preserved.

Furthermore, the rise in the availability of data could give rise to pricing bias and new sources of monetary non-neutrality. When investors have access to thousands of observable forecasting variables, Martin and Nagel (2022) show that commonly used shrinkage estimators introduce systematic errors in equity prices. This same rationale could apply to firm price setting. By down-weighting new observations, the widespread use of shrinkage estimators could lower firms' sensitivity to monetary policy.

The idea that data is a by-product of economic activity could be integrated in an existing model of price-setting with imperfect information. Having information come from production, rather than from an information allocation or attention allocation choice, may change the aggregate and cross-sectional predictions of the theory. Embedding these ideas in a standard new-Keynesian framework might provide guidance about potential future changes in the efficacy of monetary policy.

5.7 *Firm Dynamics*

A prevalent trend in the macroeconomy is that firm sizes are increasing. In the last 30 years, the proportion of those employed at firms that have been in business for fewer than five years has dropped from 13 percent to less than 8 percent, and the share of those employed at firms with fewer than 100 employees has declined from 40 percent to 35 percent. Meanwhile, the percentage of employment at firms with more than 1,000 employees has increased from 25 percent to 33 percent. For firms in the top 5 percent of revenues, their revenues rose from 57 percent to 67 percent of the revenue share (Davis and Haltiwanger 2019).

While all firms can benefit from new data technologies, some may benefit more than others. In particular, larger, older firms seem to benefit more. This increases the competitiveness of such firms, making competition harder for smaller, newer firms. In Farboodi et al. (2019), this force causes the firm size distribution to diverge.

One reason that data may affect firm dynamics is that it can broaden the span of control in firms, which favors larger firms. Aghion et al. (2023) argue that information and communications technology (ICT) innovations in the 1990s allowed high-productivity firms to profitably expand. This expansion came at the cost of labor and small firms. The authors argue that it produced a decline in business dynamism—the process by which firms grow, shrink, enter, and fail. They also argue that IT reduces the labor share of income. Furthermore, Lashkari, Bauer, and Boussard (2018) find that firm size and demand for IT are positively correlated. They argue that this result implies that the relative marginal product of IT inputs rises with scale, since IT helps firms address organizational limits to scale.

Another reason that data interacts with firm size is that financial data may be changing the firm size distribution. Begenau, Farboodi, and Veldkamp (2018) argue that the use of big data in financial markets can significantly lower the cost of capital for large firms, relative to small ones. Cheaper financing enables large firms to grow larger. Large firms with more economic activity and a longer firm history offer more data to process, making large firms more valuable targets for data analysis. Once processed, that data can better forecast firm value, reduce the risk of equity investment, and thus reduce the firm's cost of capital. As big data technology improves, large firms attract a more than proportional share of the financial data processing, enabling large firms to invest cheaply and grow larger.

Another important macroeconomic trend, as documented in Autor et al. (2020), involves industries becoming dominated by superstar firms, firms with high productivity, that is, high markups and low labor shares; the dominance of firms has led to an increase in overall average markups and a rise in firm concentration. Crouzet and Eberly (2019) claim that large modern firms have high intangible investment levels, a property that is correlated with firms having high markups. Relatedly, Eeckhout and Veldkamp (2022) present a model that shows how the accumulation of customer data by firms can explain these trends in firm productivity and markups.

Evidence in support of the effect on firm dynamics comes from firms' investment in IT resources. Tambe and Hitt (2012) use data on IT productivity to show that IT returns are substantially lower in midsize firms than in Fortune 500 firms. However, IT returns also materialize more slowly in large firms, whereas in midsize firms the short-run contribution of IT to output is similar to the long-run output contribution. They also find that the measured marginal product of IT spending is higher from 2000 to 2006 than in any previous period, suggesting that firms, and especially large firms, have been continuing to develop valuable uses of IT.

Brynjolfsson and McElheran (2019) find that adoption of DDD, once again defined as the process of making business decisions based on the analysis of data, is earlier and more prevalent among larger, older plants belonging to multiunit firms. Smaller single-establishment firms adopt later but have a higher correlation with performance than similar non-adopters. The average value-added for later DDD adopters is 3 percent greater than for non-adopters. DDD-related performance differentials decrease over time for early and late adopters, consistent with firm learning and development of organizational complementarities.

To study the role of data in the phenomenon of large firms growing larger, one could embed the current firm structure from Farboodi and Veldkamp's (2022) model into a Hopenhayn (1992) industry equilibrium model that endogenously determines entry, exit, firm size, and general firm dynamics through a stationary equilibrium analysis. The firm's stock of data could replace the capital stock or complement it. Different data precision paths for different firms could reveal how the firms' overall forecast performance for product sales varies with size, firm age, and other firm characteristics.

5.8 *Labor Demand*

Another important question about the macroeconomy is how new data technologies will impact the allocation of and demand for labor. One possible effect to be explored is that data could facilitate employer–employee matching. Martellini and Menzio (2020) use a search model to show how reduced search frictions keep unemployment steady. They argue that the reason is that while there are more encounters between unemployed workers and hiring firms, the augmented search technology allows both sides to be more selective in match quality, since they have more access to alternative partners. Overall, they show that while improvements in search technology have no effect on unemployment, these advancements do increase match quality and play a role in the growth of the economy.

Arrieta-Ibarra et al. (2018) advocate treating users who trade their personal data in exchange for digital services as laborers and producers of digital goods, as opposed to simply consumers. They argue that counteracting the current data monopsonist-dominated system with features such as competition, a data labor union, or regulatory measures to encourage payment of users providing data will greatly help remedy labor market problems, political problems, and the potentially problematic interactions between the two.

A growing literature studies the labor market effect of AI and robotics.[2] This is a separate phenomenon from the accumulation of user data. But there are a few issues in common. For example, Acemoglu and Restrepo (2018) study AI adoption and document both a displacement effect, which reduces demand for labor and wages, and a productivity effect, which increases demand for labor in nonautomated tasks. The authors also highlight the risk of a mismatch between workers' skills and the skills required to use new technologies. Forecasting with user-generated data does not obviously displace labor. But it might reduce demand for management with old technology skills.

[2] Furman and Seamans (2019); Bughin et al. (2017); Bessen (2018); Brynjolfsson and Mitchell (2017).

Similarly, its benefits could be tempered by a shortage of data-skilled labor.

Any of the above frameworks could be adapted to treat data and labor as substitutes. The idea would be that data can be used to train AI algorithms, which in turn can substitute for human labor. Of course, different sectors have different degrees of substitutability. Making this change would allow us to study how wages, capital returns, and income inequality evolve as the data economy grows. Since the goal of AI is to perform cognitive capabilities similar to that of a human, AI technology is often used to automate tasks that were previously performed by humans. Subsequently, we can adjust Jones and Tonetti's (2020) model so that it treats data and labor as substitutes by using a linear production function. One could use data as a proxy for AI because AI's effectiveness tends to improve when it can learn from more data inputs.

Furthermore, the Cong, Xie, and Zhang (2021) and Abis and Veldkamp (2024) frameworks could be used to measure degrees of complementarity and substitutability between data and labor. Recall that this type of model compares the quantity and price of labor from data scientists to labor with more traditional analysis skills. One can look at a spectrum of roles that have varying degrees of complementarity and substitutability with data.

5.9 *Direct and Indirect Sale of Information*

Once we start considering data as an asset and the sale of data as the sale of an asset, it leads us to think about the booming information services or subscription economy. A small, older literature in finance explored the question of whether an owner of information should sell that information directly to investors or whether they should use their information to manage the investors' portfolios for them and charge a fee for the service. Admati and Pfleiderer (1990) consider this trade-off in the context of a noisy rational expectations financial market, not a goods market economy. Of course, a financial investment decision is not so different from firms' real investment decisions. Risky assets might be interpreted as product lines with uncertain profits. While many aspects of the model are finance specific, the tools and lessons bear revisiting and importing into new models of the data macroeconomy.

Applying these tools could be important for macroeconomists because they could help with GDP measurement. A service flow is measured. Data as an asset might not be, if the data are not sold or are bartered. Or it could be that when data is sold as an asset, the economic gains are booked immediately. Data sold as a service, gradually over time, might change the timing of the measurement of economic value. Just as economists impute housing services of owner-occupied housing to include it in GDP, one might think of imputing the value of a stream of data services as part of GDP measurement. To know how to do that, we need to know how to think about valuing such services.

5.10 *Using Data for Business Stealing*

It is also true that much data is used for advertising, which may not enhance aggregate output. Farboodi and Veldkamp (2022) show how to model data used for business stealing. They show that the firm dynamics results are unchanged. But of course, data that is simply a means for stealing others' business produces no aggregate economic growth.

Using a modeling structure from Morris and Shin (2002), we can model business-stealing activity by adding an externality term to the productivity process in Equation (1):

$$A_{i,t} = \hat{A} - \left(a_{i,t} - \theta_t - \epsilon_{a,i,t}\right)^2 + \int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj. \tag{15}$$

The integral term at the end embodies the idea that when one firm uses data to reduce the distance between their chosen technique a_{it} and the optimal technique $\theta + \epsilon$, all other firms lose a little bit. These gains and losses are such that, when added up to compute aggregate productivity, they cancel out: $\int A_{it} = \hat{A}$. In other words, data accumulation does not benefit aggregate productivity at all. It only benefits the individual firm that collects the data.

This type of modeling structure could be used to study data used for marketing. Along with empirical evidence on uses of data, future research might use benefits and distortions to assess welfare and inform policy. The payoff externality could dramatically change the types of data regulations that would maximize welfare.

5.11 *Biased Data*

If information is sold, the incentives of the data provider may influence what content is provided. Nimark (2014) explains that unusual events are more likely to be reported in the news than commonplace ones. The author shows that such news reporting tendencies can explain changes in macroeconomic uncertainty and volatility. Furthermore, an increased level of news focused on the economy can not only increase an agent's uncertainty but also its sensitivity to the information, too.

Gentzkow and Shapiro (2010) find that newspapers have an incentive to align the political leaning of their news with that of their readers, as opposed to that of the companies' owners. One reason is that Bayesian consumers tend to consider information to be higher quality when it comes from sources that confirm their prior expectations. Thus, news media may be incentivized to align their slant to the prior beliefs of their consumers to gain a reputation for quality (Gentzkow and Shapiro 2006). The same incentives that lead news outlets to choose some news to provide may apply to data providers. As we start to explore data sales, we should remember that data sellers are strategic actors.

5.12 *Uncertainty Shocks and the Value of Data*

Uncertainty could rise for two reasons: (i) The environment becomes more volatile and less predictable, or (ii) the data becomes less precise or relevant. Reason (i) makes data more valuable. Reason (ii) makes data less valuable. If prices of firms are tied to their data value, then firms' values could either rise or fall in an uncertainty shock episode.

If we can better measure the value of data, we could gain insight into the sources of uncertainty shocks. Furthermore, uncertainty could be injected into long-run models from section 4 to make data a source of cyclical fluctuation.

5.13 *Public Disclosures and the Value of Data*

Amador and Weill (2010) explore how central bank communications affect the value and use of private information in the realm of monetary policy. One could meld those tools with the frameworks presented in this paper to explore how the dissemination of government statistics affects firms' data value.

Modeling the effect of government disclosure might require an additional step—assessing how related government data and private information really are. One could access some private data, or construct a model of private and public data. Using linear projections (e.g., ordinary least squares), one could determine how much of the information content of the private data is contained in the public signal.

5.14 *Networks and Business-to-Business (B2B) Information Transmission*

The purchases of one's customers or prices of one's suppliers may convey information about one's economic forecasts. A network model could be applied to study how much information is conveyed, to whom, and in what patterns. Similar to how public disclosures reduce the value of private information, knowledge revealed through one's network also makes private data less valuable. To capture this, one could use models in the spirit of Herskovic and Ramos (2020) or Allen, Bilir, and Tonetti (2017). While these papers are not focused on data or B2B transactions, their tools could be adapted for this purpose. The nascent literature on data platforms that facilitate information sharing (Bergemann, Bonatti, and Gan 2022; Kirpalani and Philippon 2020) also provides tools that could be incorporated into models of macroeconomic phenomena.

6. *Conclusions*

Data are changing how we think about measuring economic value. They are reducing the importance of borders with lower transportation cost and ability to translate, changing price setting in ways that could affect monetary policy, and changing the functioning of labor markets and the distribution of the size of firms. We see evidence of all these changes in the literature studied here. But to predict how far-reaching these changes will be and how large they will eventually become requires a framework for prediction. In the midst of structural transformation, projection of past trends into the future is not a reliable guide. Instead, we need structural models and theories to guide our thinking about the nature of this change.

While data are affecting every corner of the economy, and evidence is accumulating that the effects are transformative, theories exploring the role and value of data are only just emerging. A variety of theoretical literatures are bringing different tools and perspectives to the debate. From the growth and patents literature come tools to think about non-rivalry, imperfect competition, and the ownership of data. The models may also equate data with productivity. That makes data a source of increasing returns and perpetual growth.

The information frictions literature brings the perspective that data are fundamentally information. Information is a tool to reduce uncertainty about unknown outcomes. Machine learning and AI are prediction algorithms, not invention algorithms. They are designed to detect complex patterns and use those patterns to forecast future outcomes. Resolving uncertainty has finite benefits. Infinite data may or may not bring perfect foresight. Foresight makes firms more profitable, perhaps much more profitable. But it does not allow a firm, and certainly not the aggregate economy, to produce arbitrarily large quantities or value of real goods. The bounded benefits to data imply that there must be decreasing returns at some point.

The knowledge production approach views data as input, typically with decreasing returns, consistent with the information literature. But this literature emphasizes the need for complementary investments in computing equipment and skilled labor.

Data are all three: data represents ideas, information used for forecasting, and input into the creation of knowledge. Each perspective has truth in it. But each is useful for a different purpose. The idea literature has lots of complementary tools to think about idea/data ownership. The information tool kit links data to risk and decision-making under uncertainty. The inputs literature can be combined with either of the other two and helpfully predicts observable correlates of data accumulation. Viewing data as input might be useful for measurement because it allows us to infer data from observing physical investments or skilled labor employment.

New work is needed to bring these perspectives closer, extract the best from each, and develop new frameworks that can help us to understand, to measure and to value the data economy.

References

Abis, Simona. 2020. "Man vs. Machine: Quantitative and Discretionary Equity Management." http://dx.doi.org/10.2139/ssrn.3717371.

Abis, Simona, and Laura Veldkamp. 2024. "The Changing Economics of Knowledge Production." *Review of Financial Studies* 37 (1): 89–118.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Artificial Intelligence, Automation and Work." In *The Economics of Artificial Intelligence: An Agenda*, 197–236. Chicago: University of Chicago Press.

Admati, Anat, and Paul Pfleiderer. 1990. "Direct and Indirect Sale of Information." *Econometrica* 58 (4): 901–28.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. 2023. "A Theory of Falling Growth and Rising Rents." *Review of Economic Studies* 90 (6): 2675–2702.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones. 2018. "Artificial Intelligence and Economic Growth." In *The Economics of Artificial Intelligence: An Agenda*, 237–82. Chicago: University of Chicago Press.

Agrawal, Ajay, John McHale, and Alexander Oettl. 2018. "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth." In *The Economics of Artificial Intelligence: An Agenda*, 149–74. Chicago: University of Chicago Press.

Akcigit, Ufuk, and Sina T. Ates. 2023. "What Happened to US Business Dynamism?" *Journal of Political Economy* 131 (8): 2059–2124.

Allen, Treb, Kamran Bilir, and Christopher Tonetti. 2017. "The Spatial Diffusion of Knowledge." https://ideas.repec.org/p/red/sed017/1498.html.

Amador, Manuel, and Pierre-Olivier Weill. 2010. "Learning from Prices: Public Communication and Welfare." *Journal of Political Economy* 118 (5): 866–907.

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E. Glen Weyl. 2018. "Should We Treat Data as Labor? Moving beyond 'Free.'" *AEA Papers and Proceedings* 108: 38–42.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135 (2): 645–709.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki. 2019. "The Impact of Big Data on Firm Performance: An Empirical Investigation." *AEA Papers and Proceedings* 109: 33–37.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review* 57 (3): 415–26.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp. 2018. "Big Data in Finance and the Growth of Large Firms." *Journal of Monetary Economics* 97: 71–87.

Belo, Frederico, Vito D. Gala, Juliana Salomao, and Maria Ana Vitorino. 2022. "Decomposing Firm Value." *Journal of Financial Economics* 143 (2): 619–39.

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan. 2022. "The Economics of Social Data." *RAND Journal of Economics* 53 (2): 263–96.

Bessen, James. 2018. "AI and Jobs: The Role of Demand." NBER Working Paper 24235.

Brynjolfsson, Erik, Avinash Collis, and Felix Eggers. 2019. "Using Massive Online Choice Experiments to Measure Changes in Well-being." *Proceedings of the National Academy of Sciences* 116 (15): 7250–55.

Brynjolfsson, Erik, Avinash Collis, W. Erwin Diewert, Felix Eggers, and Kevin J. Fox. 2019. "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy." NBER Working Paper 25695.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung H. Kim. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" http://dx.doi.org/10.2139/ssrn.1819486.

Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang. 2002. "Intangible Assets: Computers and Organizational Capital." *Brookings Papers on Economic Activity* 33 (1): 137–81.

Brynjolfsson, Erik, Xiang Hui, and Meng Liu. 2019. "Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform." *Management Science* 65 (12): 5449–60.

Brynjolfsson, Erik, and Kristina McElheran. 2019. "Data in Action: Data-Driven Decision Making in U.S. Manufacturing." Rotman School of Management Working Paper 2722502.

Brynjolfsson, Erik, and Tom Mitchell. 2017. "What Can Machine Learning Do? Workforce Implications." *Science* 358 (6370): 1530–34.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2018. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." In *The Economics of Artificial Intelligence: An Agenda*, 23–57. Chicago: University of Chicago Press.

Brynjolfsson, Erik, and Shinkyu Yang. 1996. "Information Technology and Productivity: A Review of the Literature." In *Advances in Computers*, Vol. 43, 179–214. London: Academic Press.

Bughin, Jacques et al. 2017. "Artificial Intelligence: The Next Digital Frontier?" https://www.mckinsey.

com/capabilities/quantumblack/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies.

Byrne, David, and Carol Corrado. 2019. "Accounting for Innovation in Consumer Digital Services: IT Still Matters." NBER Working Paper 26010.

Byrne, David, Carol Corrado, and Daniel E. Sichel. 2018. "The Rise of Cloud Computing: Minding Your P's, Q's and K's." NBER Working Paper 25188.

Cong, Lin William, Danxia Xie, and Longtian Zhang. 2021. "Knowledge Accumulation, Privacy, and Growth in a Data Economy." *Management Science* 67 (10): 6480–92.

Costinot, Arnaud, Guido Lorenzoni, and Iván Werning. 2014. "A Theory of Capital Controls as Dynamic Terms-of-Trade Manipulation." *Journal of Political Economy* 122 (1): 77–128.

Cover, Thomas, and Joy Thomas. 1991. *Elements of Information Theory.* New York: John Wiley and Sons.

Crouzet, Nicolas, and Janice Eberly. 2018. "Intangibles, Investment, and Efficiency." *AEA Papers and Proceedings* 108: 426–31.

Crouzet, Nicolas, and Janice Eberly. 2019. "Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles." NBER Working Paper 25869.

Davis, Steven J., and John C. Haltiwanger. 2019. "Dynamism Diminished: The Role of Housing Markets and Credit Conditions." NBER Working Paper 25466.

Easterly, William R. 2002. *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics.* Cambridge, MA: MIT Press.

Eeckhout, Jan, and Laura Veldkamp. 2022. "Data and Market Power." NBER Working Paper 30022.

Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel. 2017. "Uncertainty Traps." *Quarterly Journal of Economics* 132 (4): 1641–92.

Farboodi, Maryam, Adrien Matray, Laura Veldkamp, and Venky Vekateswaran. 2022. "Where Has All the Data Gone?" *Review of Financial Studies* 35 (7): 3101–38.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. "Big Data and Firm Dynamics." *AEA Papers and Proceedings* 109: 38–42.

Farboodi, Maryam, and Laura Veldkamp. 2022. "A Model of the Data Economy." NBER Working Paper 28427.

Furman, Jason, and Robert Seamans. 2019. "AI and the Economy." *Innovation Policy and the Economy* 19: 161–91.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–74.

Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114 (2): 280–316.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71.

Goldfarb, Avi, and Daniel Trefler. 2018. "AI and International Trade." NBER Working Paper 24254.

Goldfarb, Avi, and Catherine Tucker. 2019. "Digital Economics." *Journal of Economic Literature* 57 (1): 3–43.

Gorodnichenko, Yuriy, and Oleksandr Talavera. 2017. "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration." *American Economic Review* 107 (1): 249–82.

Government of Rwanda. 2010. *An Integrated ICT-Led Socio-Economic Development Plan for Rwanda, 2006–2010: The NICI-2010 Plan*. Kigali: Government of Rwanda.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica* 50 (1): 213–24.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Herskovic, Bernard, and João Ramos. 2020. "Acquiring Information through Peers." *American Economic Review* 110 (7): 2128–52.

Hjort, Jonas, and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review* 109 (3): 1032–79.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–50.

Huang, Ke-Wei, and Arun Sundararajan. 2011. "Pricing Digital Goods: Discontinuous Costs and Shared Infrastructure." *Information Systems Research* 22 (4): 721–38.

Hulten, Charles, and Leonard Nakamura. 2017. "Accounting for Growth in the Age of the Internet: The Importance of Output-Saving Technical Change." NBER Working Paper 23315.

Jin, Yizhou, and Shoshana Vasserman. 2021. "Buying Data from Consumers: The Impact of Monitoring Programs in US Auto Insurance." NBER Working Paper 29096.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–84.

Jones, Charles I., and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." *American Economic Review* 110 (9): 2819–58.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A Rational Theory of Mutual Funds' Attention Allocation." *Econometrica* 84 (2): 571–626.

Kehoe, Patrick J., Bradley J. Larsen, and Elena Pastorino. 2018. "Dynamic Competition in the Era of Big Data." https://www.ftc.gov/system/files/documents/public_events/1567421/kehoelarsenpastorino_updated.pdf.

Kirpalani, Rishabh, and Thomas Philippon. 2020. "Data Sharing and Market Power with Two-Sided Platforms." NBER Working Paper 28023.

Knight, Will. 2017. "AlphaGo Zero Shows Machines Can Become Superhuman without Any Help." *MIT Technology Review*, October 18. https://www.technologyreview.com/2017/10/18/148511/alphago-zero-shows-machines-can-become-superhuman-without-any-help/.

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard. 2018. "Information Technology and Returns to Scale." http://dx.doi.org/10.2139/ssrn.3458604.

Lu, Chia-Hui. 2021. "The Impact of Artificial Intelligence on Economic Growth and Welfare." *Journal of Macroeconomics* 69: 103342.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.

Mac´kowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99 (3): 769–803.

Mac´kowiak, Bartosz, Filip Mateˇjka, and Mirko Wiederholt. 2018. "Rational Inattention: A Disciplined Behavioral Model." European Central Bank Working Paper 2570.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407–37.

Martellini, Paolo, and Guido Menzio. 2020. "Declining Search Frictions, Unemployment and Growth." *Journal of Political Economy* 128 (12): 4387–4437.

Martin, Ian W. R., and Stefan Nagel. 2022. "Market Efficiency in the Age of Big Data." *Journal of Financial Economics* 145 (1): 154–77.

Martinez, Joseba. 2018. "Automation, Growth and Factor Shares." Society for Economic Dynamics 2018 Meeting Papers 736.

Morris, Stephen, and Hyun Song Shin. 2002. "Social Value of Public information." *American Economic Review* 92 (5): 1521–34.

Nimark, Kristoffer P. 2014. "Man-Bites-Dog Business Cycles." *American Economic Review* 104 (8): 2320–67.

Nordhaus, William D. 2021. "Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth." *American Economic Journal: Macroeconomics* 13 (1): 299–332.

Ordoñez, Guillermo. 2013. "The Asymmetric Effects of Financial Frictions." *Journal of Political Economy* 121 (5): 844–95.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–102.

Rotman, David. 2019. "AI is Reinventing the Way We Invent." *MIT Technology Review*, February 15. https://www.technologyreview.com/2019/02/15/137023/ai-is-reinventing-the-way-we-invent.

Schultz, Theodore W. 1943. *Redirecting Farm Policy.* New York: Macmillan.

Shukla, Shubhendu S., and Vijay Jaiswal. 2013. "Applicability of Artificial Intelligence in Different Fields of Life." *International Journal of Scientific Engineering and Research* 1 (1) 28–35.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70 (1): 65–94.

Tambe, Prasanna, and Lorin M. Hitt. 2012. "The Productivity of Information Technology Investments: New Evidence from IT Labor Data." *Information Systems Research* 23 (3 Part 1): 599–617.

United Nations. 2019. "Data Economy: Radical Transformation or Dystopia?" Frontier Technology Quarterly, January 2019. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf.

Veldkamp, Laura L. 2005. "Slow Boom, Sudden Crash." *Journal of Economic Theory* 124 (2): 230–57.

Vives, Xavier. 2017. "The Impact of FinTech on Banking." *European Economy* 2017 (2): 97–105.

Vives, Xavier. 2019. "Digital Disruption in Banking." *Annual Review of Financial Economics* 11: 243–72.