A Model of the Data Economy

Maryam Farboodi* and Laura Veldkamp†

September 6, 2021

Abstract

The rise of information technology and big data analytics has given rise to "the new economy." But are its economics new? This article constructs a dynamic equilibrium model where firms accumulate data, instead of capital. We incorporate three key features of data: 1) Data is a by-product of economic activity; 2) data is information used for prediction, and 3) uncertainty reduction enhances firm profitability. The model can explain why data-intensive goods or services, like apps, are given away for free, why many new firms are unprofitable and why some of the biggest firms in the economy profit primarily from selling data. While the transition dynamics of the data economy and a capital economy differ, the long-run dynamics are similar: Data has diminishing returns; only R&D can sustain growth, and comparative advantage dictates who produces what.

Does the new information economy have new economics? In the information age, production increasingly revolves around information and specifically, data. Many firms, particularly the most valuable U.S. firms, are valued primarily for the data they have accumulated. Collection and use of data is as old as book-keeping. But recent innovations in data-intensive prediction technologies allow us to use more data more efficiently. Our goal is to explore whether or not the accumulation of data used by these new technologies generates dynamics that are fundamentally different from capital or idea accumulation. We show that the long-run forces are mostly familiar. However, the short-run forces are new: The transition dynamics of data-accumulating firms differ from their capital-accumulating predecessors.

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

†Columbia Graduate School of Business, NBER, and CEPR, 3022 Broadway, New York, NY 10027; lv2405@columbia.edu.
Thanks to Rebekah Dix and Ran Liu for invaluable research assistance and to participants at the 2019 SED plenary, Kansas City Federal Reserve, 2020 ASSA and Bank of England annual research conference for helpful comments and suggestions. Keywords: Data, growth, digital economy, data barter.

Most of the discussion about data is about a particular type of digitized information: transaction-generated data, used by firms to optimize their business processes, by accurately predicting future outcomes. The hype about data economics has arisen because of breakthrough data technologies, like machine learning and artificial intelligence. These algorithms are for prediction. They require enormous amounts of data, which are naturally generated by transactions: 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. The predictions made with this big data are typically used for business process optimization, such as advertising, forecasting sales, earnings, inventories, shipping needs or the future value of firms and their product lines. The reason we focus on this type of data is because this is where the technological change took place (Goldfarb and Tucker, 2019), this is the type of data that concerns regulators, and this is the type of data that supposedly will spawn a new information age. This is what we mean when we talk about "data."

What we are not focusing on is digitized knowledge – patents, textbooks or art – or data used for research or innovation. This type of data is not what the claims of a new data economy are about. There are existing literatures on patents and on the role of ideas in growth. While we extend our model to include data as an input in the growth process, such research data or digitized knowledge is not what "data" refers to, in this paper.

Therefore, Section 1 proposes a model of a data economy where data is user-generated, is used to form predictions about uncertain future outcomes and helps firms to optimize their business processes. Because data is non-rival, increases productivity and is freely replicable (has returns to scale), previous thinking equated data growth with idea or technological growth. What is new in this model is that data is information, used for prediction.

Section 2 examines the data economy in the long run. We start with a thought experiment: Can data sustain growth, in the absence of any technological progress? This is an analogy to the question Solow (1956) asks about capital. We find that diminishing returns arises from data's role in improving predictions. Prediction errors can only be reduced to zero. That places a natural bound on how much prediction error data can possibly resolve. Unforecastable randomness is a second force that limits how much firms can benefit from better data and better predictions. Both of these forces ensure that when data is abundant, it must have diminishing returns. The presence of diminishing returns means data accumulated for process improvements cannot sustain growth.

Is it possible for data to play a role in long-run growth? Yes, if data is an input into research and development. Section 2.2 shows how data might feature in a quality-ladder model of endogenous growth. The key question for long-run growth then becomes whether transactions data – who bought what – can inform growth-sustaining technological innovation.

The use and sale of data is governed by a mostly familiar force – comparative advantage – with some novel predictions as well. Section 2.3 shows that data firms have returns to specialization. When high-efficiency data produces are scarce, they specialize more in selling data to others. The organization of the market resembles the large data platform companies we see currently dominating the U.S. market.

Finally, a data economy may feature specialization. Large firms that have a comparative advantage in data production may derive most of their profit from data sales. Meanwhile, small firms have a comparative advantage in high-quality goods production. Therefore, such large firms produce high volumes of low-price goods. The business model of these large firms is to do lots of transactions at a low price for goods and earn more revenue from data sales. While we know that many firms execute a strategy like this and the idea of comparative advantage is familiar, this is different from a capital accumulation economy. It surprising that, even with diminishing returns everywhere, this realistic market structure arises from simple economic properties of data as information.

In the long-run, data may be a source of inefficiency. The answer depends on whether data is used to create value or steal business from rivals. Despite its non-rival nature and the many externalities associated with non-rival technologies, data prices create the right incentives for efficient data trade when data is used to create value.

In Section 3, we focus on the short run, meaning the transition dynamics before a firm gets to steady state. When data is scarce, it may have increasing returns, which arise because of the way in which data is produced. Our model features what is referred to as a "data feedback loop." More data makes a firm more productive, which results in more production and transactions, which generate more data, further increasing productivity and data generation. This force is the dominant force when data is scarce, before the diminishing returns to forecasting set in and overwhelm it. This increasing returns force can generate a data poverty trap. Firms with low levels of data earn low profits, which makes little production optimal. But little production generates little data,

which keeps the firm poor.

Because data is a long-lived asset, firms may choose to produce goods with negative profits, because goods production will also produce data, which is an asset with long-lived value. This rationalizes the commonly-observed practice of data barter (Section 3.2). Many digital services, like apps, which were costly to develop, are given away to customers at zero price. This is not generosity. Firms are exchanging these services for their customers' data. The exchange of data for a service, at a zero monetary price, is a classic barter trade. Such trades can arise in our model: Firms give away their goods, as a form of costly investment in data.

The primary contribution of the paper is not the particular predictions we explore. Some of those predictions are more obvious, some more surprising. The larger contribution is a tool to think clearly about the economics of aggregate data accumulation. Because our tool is a simple one, many applications and extensions are possible. We explore some in subsections of the paper; others are suggested in the conclusion. While adding features to the main model is tempting, to allow the model to better address one question or another, the advantage of keeping the main model simple is that it allows the tool to be used in many different ways.

The model also offers guidance for measurement. Measuring and valuing data is complicated by the fact that frequently, data is given away, in exchange for a free digital service. Our model makes sense of this pricing behavior and assigns a private and social value to goods and data that have a zero transactions price. In so doing, it moves beyond price-based valuation, which often delivers misleading answers when valuing digital assets.

Related Literature. Compared to the existing literature on data and growth, the key difference in our model is that data is information, used to forecast a random variable. In Jones and Tonetti (2018) and related work Cong et al. (2020), data contributes directly to productivity. It is not information. A fundamental characteristic of information is that it reduces uncertainty about something. When we model data as information with an exogenous bound on technology, Jones and Tonetti (2018)'s conclusions about the benefits of data privacy may still hold. But the long-run growth conclusions do not.

In models of learning-by-doing (Jovanovic and Nyarko (1996), Oberfield and Venkateswaran (2018)) and organizational capital Atkeson and Kehoe (2005), firms also accumulate a form of

knowledge. But the economics differ. Unlike prediction data, this knowledge need not have long-run diminishing returns. Also, it is not a tradeable asset. Our short-run increasing return to data differs from growth models with increasing returns Farmer and Benhabib (1994), because those are based on positive spillovers between firms. Ours is a feedback loop within a firm.

Work on information frictions in business cycles (Caplin and Leahy (1994), Veldkamp (2005), Lorenzoni (2009), Ordonez (2013), Ilut and Schneider (2014) and Fajgelbaum et al. (2017)) have early versions of a data-feedback loop whereby more data enables more production, which in turn, produces more data. In each of these models, information was a by-product of economic activity; firms used this information to reduce uncertainty and guide their decision-making. But the key difference is that information was a public good, not a private asset. The private asset assumption in this paper changes firms' incentives to produce data. In these earlier models, firms use data to forecast business cycles, not optimal firm strategy. We model data that is industry or firm specific, and is private property of the firm.

Work exploring the interactions of data and innovation complements ours. For example, Agrawal et al. (2018) develop a combinatorial-based knowledge production function to explore how breakthroughs in AI could enhance discovery rates and economic growth.[1] Acemoglu and Restrepo (2018) explore similar questions to ours, but about the growth potential from robots. Although robots require data, they are rival capital goods, distinct from the data itself. Our work analyzes big data and new prediction algorithms, in the absence of technological change. Once we understand this foundation, one can layer these insights about data, innovation and automation on top.

In the finance literature, Begenau et al. (2018) explore how growth in the processing of financial data affects firm size. They do not model firms' use of their own data. There is also a literature on data-driven decision making, which explores how data matters at a microeconomic level. We add the aggregate effects of such activities.

Finally, the insight that the stock of knowledge can serve as a state variable comes from the five-equation toy model sketched in Farboodi et al. (2019). That was a partial-equilibrium numerical exercise, designed to explore the size of firms with heterogeneous data. This paper builds an aggregate equilibrium model, with goods markets, data markets, data non-rivalry, and analyt-

[1]Other work in the vein includes: Lu (2019) who embeds self-accumulating AI in a Lucas (1988) growth model and examines growth transition paths and welfare; Aghion et al. (2017) who explore the reallocative effects of AI, as Baumol (1967)'s cost disease leads to the declining share of traditional industries' GDP.

ical solutions. These new innovations shape the answers to our main questions about aggregate dynamics and long-run outcomes.

1 A Data Economy

We build a framework in which data is information, which helps forecast random outcomes. More accurate forecasts help firms optimize business processes. The model looks much like a simple Solow (1956) model. To isolate the effect of data accumulation, the model holds fixed productivity, aside from that which results from data accumulation. There are inflows of data from new economic activity and outflows, as data depreciates. The depreciation comes from the fact that firms are forecasting a moving target. Economic activity many periods ago was quite informative about the state at the time. However, since the state has random drift, such old data is less informative about what the state is today.

The key differences between our data accumulation model and Solow's capital accumulation model are three-fold: 1) Data is used for forecasting; 2) data is a by-product of economic activity, and 3) data is, at least partially, non-rival. Multiple firms can use the same data, at the same time. These subtle changes in model assumptions are consequential. They alter the source of diminishing returns, create increasing returns and data barter, and produce returns to specialization.

1.1 Model

Real Goods Production Time is discrete and infinite. There is a continuum of competitive firms indexed by i. Each firm can produce $k_{i,t}^{\alpha}$ units of goods with $k_{i,t}$ units of capital. These goods have quality $A_{i,t}$. Thus firm i's quality-adjusted output is

$$y_{i,t} = A_{i,t}k_{i,t}^{\alpha} \tag{1}$$

The quality of a good depends on a firm's choice of a production technique $a_{i,t}$. Each period firm i has one optimal technique, with a persistent and a transitory components: $\theta_t + \epsilon_{a,i,t}$. Neither component is separately observed. The persistent component θ_t follows an AR(1) process: $\theta_t = \bar{\theta} + \rho(\theta_{t-1} - \bar{\theta}) + \eta_t$. The AR(1) innovation $\eta_t \sim N(0, \sigma_\theta^2)$ is *i.i.d.* across time.[2] The transitory

[2]One might consider different possible correlations of $\eta_{i,t}$ across firms i. An independent θ processes

shock $\epsilon_{a,i,t} \sim N(0, \sigma_a^2)$ is *i.i.d.* across time and firms and is unlearnable.

The optimal technique is important for a firm because the quality of a firm's good, $A_{i,t}$, depends on the squared distance between the firm's production technique choice $a_{i,t}$ and the optimal technique $\theta_t + \epsilon_{a,i,t}$:

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

The function g is strictly decreasing. A decreasing function means that techniques far away from the optimum result in worse quality goods.

Data The role of data is that it helps firms to choose better production techniques. One interpretation is that data can inform a firm whether blue or green cars or white or brown kitchens will be more valued by their consumers, and produce or advertise accordingly. In other words, a technique could represent a resource allocation. Transactions help to reveal customers' marginal values, but these values are constantly changing. Firms must continually learn to catch up. Another interpretation is that the technique is inventory management, or other cost-saving activities. Observing production and sales processes at work provides useful information for optimizing business practices. For now, we model data as welfare-enhancing. We relax that assumption in Section 3.

Specifically, data is informative about θ_t. At the start of date t, nature chooses a countably infinite set of potential data points for each firm i: $\zeta_{it} := \{s_{i,t,m}\}_{m=1}^{\infty}$. Each data point m reveals

$$s_{i,t,m} = \theta_{t+1} + \epsilon_{i,t,m}, \tag{3}$$

where $\epsilon_{i,t,m}$ is *i.i.d.* across firms, time, and signals. For tractability, we assume that all the shocks in the model are normally distributed: fundamental uncertainty is $\eta_t \sim N(\mu, \sigma_\theta^2)$, signal noise is $\epsilon_{i,t,m} \sim N(0, \sigma_\epsilon^2)$.

The next assumption captures the idea that data is a by-product of economic activity. The

($corr(\eta_{i,t}, \eta_{j,t}) = 0,\ \forall i \neq j$) would effectively shut down any trade in data. Since buying and selling data happens and is worth exploring, we consider an aggregate θ process ($corr(\eta_{i,t}, \eta_{j,t}) = 1,\ \forall i, j$). It is also possible to achive an imperfect, but non-zero correlation.

number of data points n observed by firm i at the end of period t depends on their production $k_{i,t}^{\alpha}$:

$$n_{i,t} = z_i k_{i,t}^{\alpha}, \tag{4}$$

where z_i is the parameter that governs how much data a firm can mine from its customers. A data mining firm is one that harvests lots of data per unit of output. Thus, firm i's production uncovers signals $\{s_m\}_{m=1}^{n_{i,t}}$.

The temporary shock ε_a is important in preserving the value of past data. It prevents firms, whose payoffs reveal their productivity $A_{i,t}$, from inferring θ_t at the end of each period. Without it, the accumulation of past data would not be a valuable asset. If a firm knew the value of θ_{t-1} at the start of time t, it would maximize quality by conditioning its action $a_{i,t}$ on period-t data $n_{i,t}$ and θ_{t-1}, but not on any data from before t. All past data is just a noisy signal about θ_{t-1}, which the firm now knows. Thus preventing the revelation of θ_{t-1} keeps old data relevant and valuable.

Data Trading and Non-Rivalry Let $\delta_{i,t}$ be the amount of data traded by firm i a time t. If $\delta_{i,t} < 0$, the firm is selling data. If $\delta_{it} > 0$, the firm purchased data.[3] We restrict $\delta_{i,t} \geq -n_{i,t}$ so that a firm cannot sell more data than it produces. Let the price of one piece of data be denoted π_t.

Of course, data is non-rival. Some firms use data and also sell that same data to others. If there were no cost to selling one's data, then every firm in this competitive, price-taking environment would sell all its data to all other firms. In reality, that does not happen. Instead, we assume that when a firm sells its data, it loses a fraction ι of the amount of data that it sells to each other firm. Thus if a firm sells an amount of data $\delta_{i,t} < 0$ to other firms, then the firm has $n_{i,t} + \iota\delta_{i,t}$ data points left to add to its own stock of knowledge. Recall that for a data seller, $\iota\delta < 0$ so that the firm has less data than the $n_{i,t}$ points it produced. This loss of data could be a stand-in for the loss of market power that comes from sharing one's own data. It can also represent the extent of privacy regulations that prevent multiple organizations from using some types of personal data. Another interpretation of this assumption is that there is a transaction cost of trading data, proportional to the data value. If the firm buys $\delta_{i,t} > 0$ units of data, it adds $n_{i,t} + \delta_{i,t}$ units of data to its stock

[3]This formulation prohibits firms from both buying and selling data in the same period.

of knowledge.

Data Adjustment and the Stock of Knowledge The information set of firm i when it chooses its technique $a_{i,t}$ is[4] $\mathcal{I}_{i,t} = [\{A_{i,\tau}\}_{\tau=0}^{t-1}; \{\{s_{i,\tau,m}\}_{m=1}^{n_{i,\tau}}\}_{\tau=0}^{t-1}]$. To make the problem recursive and to define data adjustment costs, we construct a helpful summary statistic for this information, called the "stock of knowledge."

Each firm's flow of $n_{i,t}$ new data points allows it to build up a stock of knowledge $\Omega_{i,t}$ that it uses to forecast future economic outcomes. We define the stock of knowledge of firm i at time t to be $\Omega_{i,t}$. We use the term "stock of knowledge" to mean the precision of firm i's forecast of θ_t, which is formally:

$$\Omega_{i,t} := \mathbb{E}_i[(\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t)^2]^{-1}. \tag{5}$$

Note that the conditional expectation on the inside of the expression is a forecast. It is the firm's best estimate of θ_t. The difference between the forecast and the realized value, $\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t$, is therefore a forecast error. An expected squared forecast error is the variance of the forecast. It's also called the variance of θ, conditional on the information set $\mathcal{I}_{i,t}$, or the posterior variance. The inverse of a variance is a precision. Thus, this is the precision of firm i's forecast of θ_t.

Data adjustment costs capture the idea that a firm that does not store or analyze any data cannot freely transform itself to a big-data machine learning powerhouse. That transformation requires new computer systems, new workers with different skills, and learning by the management team. As a practical matter, data adjustment costs are important because they make dynamics gradual. If data is tradeable and there is no adjustment cost, a firm would immediately purchase the optimal amount of data, just as in models of capital investment without capital adjustment costs. Of course, the optimal amount of data might change as the price of data changes. But such adjustment would mute some of the dynamics we are interested in.

We assume that, if a firm's data stock was $\Omega_{i,t}$ and becomes $\Omega_{i,t+1}$, the firm's period-t output is diminished by $\Psi(\Delta\Omega_{i,t+1}) = \psi(\Delta\Omega_{i,t+1})^2$, where ψ is a constant parameter and Δ represents the percentage change: $\Delta\Omega_{i,t+1} = (\Omega_{i,t+1} - \Omega_{i,t})/\Omega_{i,t}$. The percentage change formulation is helpful because it makes doubling one's stock of knowledge equally costly, no matter what units data is

[4]We could include aggregate output and price in this information set as well. We explain in the model solution why observing aggregate variables makes no difference in the agents' beliefs. Therefore, for brevity, we do not include these extraneous variables in the information set.

measured in.

Firm's Problem A firm chooses a sequence of production, quality and data-use decisions $k_{i,t}, a_{i,t}, \delta_{i,t}$ to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \left(\frac{1}{1+r} \right)^t \left(P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t} \right) \tag{6}$$

Firms update beliefs about θ_t using Bayes' law. Each period, firms observe last period's revenues and data, and then choose capital level k and production technique a. The information set of firm i when it chooses its technique $a_{i,t}$ and its investment $k_{i,t}$ is $\mathcal{I}_{i,t}$.

As in Solow (1956), we take the rental rate of capital as given. This reveals the data-relevant mechanisms as clearly as possible. It could be that this is an industry or a small open economy, facing a world rate of interest r.

Equilibrium P_t denotes the equilibrium price per quality unit of goods. In other words, the price of a good with quality A is AP_t. The inverse demand function and the industry quality-adjusted supply are:

$$P_t = \bar{P} Y_t^{-\gamma}, \tag{7}$$

$$Y_t = \int_i A_{i,t} k_{i,t}^{\alpha} di. \tag{8}$$

Firms take the industry price P_t as given and their quality-adjusted outputs are perfect substitutes.

1.2 Solution

The state variables of the recursive problem are the prior mean and variance of beliefs about θ_{t-1}, last period's revenues, and the new data points. However, we can simplify this to one sufficient state variables to solve the model simply. The next steps explain how.

Optimal Technique and Expected Quality Taking a first order condition with respect to the technique choice, we find that the optimal technique is $a_{i,t}^* = \mathbb{E}_i[\theta_t | \mathcal{I}_{i,t}]$. Thus, expected quality of firm i's good at time t in (2) can be rewritten as $\mathbb{E}[A_{i,t}] = E\left[g\left((\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t - \epsilon_{a,i,t})^2 \right) \right]$. The

squared term is a squared forecast error. It's expected value is a conditional variance, of $\theta_t + \epsilon_{a,i,t}$. That conditional variance is denoted $\Omega_{i,t}^{-1} + \sigma_a^2$.

To compute expected quality, we first take a second-order Taylor approximation of the quality function, expanding around the expected value of its argument: $g(v) \approx g(E[v]) + g`(v) \cdot (v - E[v]) + (1/2)g''(v) \cdot (v - E[v])^2$. Next, we take an expectation of this approximate function: $E[g(v)] \approx g(E[v]) + g`(v) \cdot 0 + (1/2)g''(v) \cdot var(v)$. Recognizing that the argument v is a chi-square variable with mean $\Omega_{i,t}^{-1} + \sigma_a^2$ and variance $2(\Omega_{i,t}^{-1} + \sigma_a^2)$, the expected quality of firm i's good at time t in (2) can be approximated as

$$\mathbb{E}_i[A_{i,t}] \approx g\left(\Omega_{i,t}^{-1} + \sigma_a^2\right) + g''\left(\Omega_{i,t}^{-1} + \sigma_a^2\right) \cdot \left(\Omega_{i,t}^{-1} + \sigma_a^2\right). \tag{9}$$

Assume that the g function is not too convex, so that quality is a deceasing function of expected forecast errors. Or put simply, more data precision increases quality. We will return to the question of highly convex, unbounded g functions in the next section.

Notice that the way signals enter in expected utility, only the variance (or precision) matters, not the prior mean or signal realization. As in Morris and Shin (2002), precision, which in this case is the stock of knowledge, is a sufficient statistic for expected utility and therefore, for all future choices. The quadratic loss, which eliminates the need to keep track of signal realizations, simplifies the problem greatly.

The Stock of Knowledge Since the stock of knowledge $\Omega_{i,t}$ is the sufficient statistic to keep track of information and its expected utility, we need a way to update or keep track of how much of this stock there is. Lemma 1 is just an application of Bayes' law, or equivalently, a modified Kalman filter, that tell us how the stock of knowledge evolves from one period to the next.

Lemma 1 Evolution of the Stock of Knowledge *In each period t,*

$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(n_{i,t} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0})\right)\sigma_\epsilon^{-2} \tag{10}$$

The proof of this lemma and of all the lemmas and propositions that follow are in Appendix A. Lemma 1 says that the stock of knowledge is the depreciated stock from the previous period t,

plus new data inflows.

The inflows of data are new pieces of data that are generated by economic activity. The number of new data points $n_{i,t}$ is assumed to be data mining ability times end of period physical output: $z_i k_{i,t}^{\alpha}$. By Bayes' law for normal variables, the total precision of that information is the sum of the precisions of all the data points: $n_{i,t}\sigma_{\epsilon}^{-2}$. The term σ_a^{-2} in (10) is the additional information learned from seeing one's own realization of quality $A_{i,t}$, at the end of period t. That information also gets added to the stock of knowledge. At the firm level, we need to keep track of whether a firm buys or sells data. Thus the newly added stock of data $n_{i,t}$ has to be adjustd for data trade. That is the role of the indicator functions at the end of (10).

One might wonder why firms do not also learn from seeing aggregate price and the aggregate output. These obviously reflect something about what other firms know. But what they reflect is the squared difference between θ_t and other firms' technique a_{jt}. That squared difference reflects how much others know, but not the content of what others know. Because the mean and variance of normal variables are independent, knowing others' forecast precision reveals nothing about θ_t. Seeing one's own outcome $A_{i,t}$ is informative only because a firm also knows its own production technique choice $a_{i,t}$. Other firms' actions are not observable. Therefore, aggregate prices or quantities reveal what other firms predicted well, which conveys no useful information about whether θ_t is high or low.

How does data flow out or depreciate? Data depreciates because data generated at time t is about next period's optimal technique θ_{t+1}. That means that data generated s periods ago is about θ_{t-s+1}. Since θ is an AR(1) process, it is constantly evolving. Data from many periods ago, about a θ realized many periods ago, is not as relevant as more recent data. So, just like capital, data depreciates. Mathematically, the depreciated amount of data carried forward from period t is the first term of (10): $\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$. The $\Omega_{i,t}+\sigma_a^{-2}$ term represents the stock of knowledge at the start of time t plus the information about period t technique revealed to a firm by observing its own output. This stock of knowledge is multiplied by the persistence of the AR(1) process squared, ρ^2. If the process for optimal technique θ_t was perfectly persistent then $\rho = 1$ and this term would not discount old data. If the θ process is i.i.d. $\rho = 0$, then old data is irrelevant for the future. Next, the formula says to invert the precision, to get a variance and add the variance of the AR(1) process innovation σ_θ^2. This represents the idea that volatile θ innovations

make knowledge about past θ's less relevant. Finally, the whole expression is inverted again so that the variance is transformed back into a precision. This precision represents a (discounted) stock of knowledge. The depreciation of knowledge is the period-t stock of knowledge, minus the discounted stock.

At the aggregate level, an economy as a whole cannot buy or sell data. Therefore, for the aggregate economy,

$$\text{Inflows:} \qquad \Omega_t^+ = \sigma_\epsilon^{-2} \int_i z_i k_{i,t}^\alpha di + \sigma_a^{-2} \tag{11}$$

$$\text{Outflows:} \qquad \Omega_t^- = \Omega_t + \sigma_a^{-2} - \int_i \left[(\rho^2 (\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2 \right]^{-1} di. \tag{12}$$

A One-State-Variable Problem We can now express expected firm value recursively, with the stock of knowledge as the single state variable in the following lemma.

Lemma 2 *The optimal sequence of capital investment choices $\{k_{i,t}\}$ and data sales $\{\delta_{i,t} \geq -n_{i,t}\}$ solve the following recursive problem:*

$$\begin{aligned} V(\Omega_{i,t}) = \max_{k_{i,t}, \delta_{i,t}} \; & P_t E[A_{i,t}] k_{i,t}^\alpha - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t} \\ & + \left(\frac{1}{1+r} \right) V(\Omega_{i,t+1}) \end{aligned} \tag{13}$$

where $E[A_{i,t}]$ is an increasing function of $\Omega_{i,t}$, given by (9), $n_{i,t} = z_i k_{i,t}^\alpha$, and the law of motion for $\Omega_{i,t}$ is given by (10).

This result greatly simplifies the problem by collapsing it to a deterministic problem with choice variables k and δ and one state variable, $\Omega_{i,t}$. In expressing the problem this way, we have already substituted in the optimal choice of production technique. The quality $A_{i,t}$ that results from the optimal technique depends on the conditional variance of θ_t. Because the information structure is similar to that of a Kalman filter, that sequence of conditional variances is deterministic.

The non-rivalry of data acts like a kinked price of data, or a negative transactions cost in (10).[5]

[5]To see the kinked price interpretation more clearly, redefine the choice variable to be ω, the amount of data added to a firm's stock of knowledge Ω. Then, $\omega = n_{i,t} + \delta_{i,t}$ for data purchases ($\delta_{i,t} > 0$) and $\omega = n_{i,t} + \iota\delta_{i,t}$ for data sales when $\delta_{i,t} < 0$. Then, we could re-express this problem as a choice of ω and a corresponding price that depends on whether $\omega \geq n_{i,t}$ or $\omega < n_{i,t}$.

Valuing Data Since $\Omega_{i,t}$ can be interpreted as a discounted stock of data, $V(\Omega_{i,t})$ captures the value of this data stock. $V(\Omega_{i,t}) - V(0)$ is the present discounted value of the net revenue the firm receives because of its data. Therefore, the marginal value of one additional piece of data, of precision 1, is simply $\partial V_t / \partial \Omega_{i,t}$. When we consider markets for buying and selling data, $\partial V_t / \partial \Omega_{i,t}$ represents the firm's demand, its marginal willingness to pay for data.

2 Long-Run Featues of a Data Economy

In this section, we show the various ways in which the long-run in this data economy is surprisingly similar to a capital-based production economy. The following section emphasizes the contrasts. The first set of results show that within the model, there is no long run growth. We then move to results that describe general conditions under which data used for forecasting can sustain infinite growth. If one believes that the accumulation of data for process innovation can sustain growth forever without innovation, there are some logically equivalent statements that one must also accept. Next, we show how to break this result. To sustain long-run growth, data must be an input into idea creation. Thus, just like capital, data can sustain growth only if it is an input into research and development. Third, we explore the pattern of data production and data trade that arise when some firms are more data savvy than others. Finally, we consider welfare and find that, despite being non-rival and a by-product of economic activity, long-run data production is not a source of inefficiency.

2.1 Diminishing Returns and Zero Long Run Growth

Just like we typically teach the Solow (1956) model by examining the inflows and outflows of capital, we can gain insight into our data economy growth model by exploring the inflows and outflows of data. Figure 1 illustrates the inflows and outflows (eq.s 11 and 12), in a form that looks just like the traditional Solow model with capital accumulation. What we see on the left is the large distance between inflows and outflows of data, when data is scarce. This is a period of fast data accumulation and fast growth in the quality and value of goods. What we see on the right is the distance between inflows and outflows diminishing, which represents growth slowing. Eventually, inflows and outflows cross at the steady state. If the stock of knowledge ever reached its steady

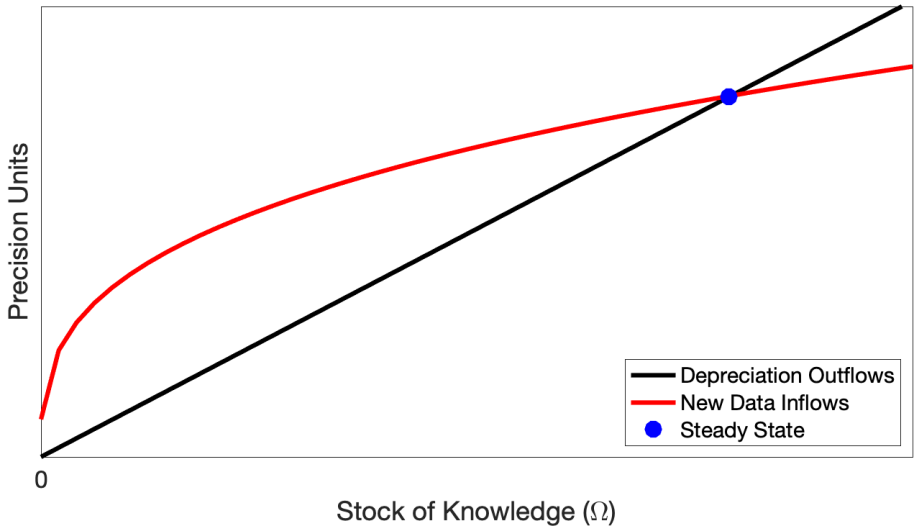


Figure 1: Economy converges to a data steady state: Aggregate inflows and outflows of data.
Line labeled inflows plots the quantity in (11) for the aggregate economy, for different levels of initial data stock. Line labeled outflows plots the quantity in (12). This is equivalent to the outflow and inflow for a representative firm i who operates in an economy populated with identical firms with no trade. The representative firm makes optimal capital decision $k^*_{i,t}$, with different levels of initial data stock. In this example and the ones that follow, we adopt a simple, linear quality function $g(z) = g(0) - z$.

state level, it would no longer change, as inflows and outflows just balance each other. Likewise, quality and GDP would stop growing.

One difference between data and capital accumulation is the nature and form of depreciation. In the Solow model of capital accumulation, depreciation is a fixed fraction of the capital stock, always linear. In the data accumulation model, depreciation is not linear, but is very close to linear. Lemma 5 in the Appendix shows that depreciation is approximately linear in the stock of knowledge, with an error bound that depends primarily on the variance of the innovation in θ.

What diminishing returns means for a data-accumulation economy is that, over time, the aggregate stock of knowledge and aggregate amount of output would have a time path that resembles the concave path in Figure 2. Without idea creation, data accumulation alone would generate slower and slower growth.

Conceptually, diminishing returns arise because we model data as information, not directly as an addition to productivity. Information has diminishing returns because its ability to reduce variance gets smaller and smaller as beliefs become more precise. Mathematically, diminishing returns comes from two distinct and independent sources: the finite-valued quality function and

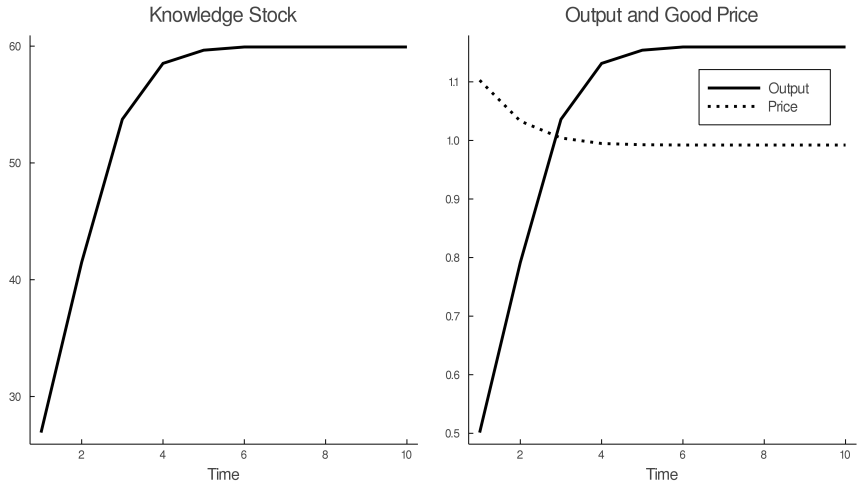


Figure 2: Aggregate growth dynamics: Data accumulation grows knowledge and output over time, with diminishing returns. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \gamma = 0.1, \overline{A} = 1, \overline{P} = 1, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 5, \iota = 1$. See appendix B for details of parameter selection and numerical solution of the model.

unlearnable risk. The next set of results explain why either feature bounds the growth from data.

Long Run Growth Impossibility Results Can data accumulation sustain growth in an economy without innovation? For this to be possible, two things must both be true: 1) Perfect one-period-ahead foresight implies infinite real output; and 2) the future is a deterministic function of today's observable data.[6] While empirical studies support the idea of decreasing returns to data (Bajari et al., 2018), it is not possible to prove the nature of a production function theoretically. What theory can do is tell us that if we believe data can sustain long-run growth, this logically implies other economic properties. In this case, long-run growth implies two conditions that are at odds with most theories. If a researcher does not believe either property to be true, they must then believe data-induced growth, without innovation, cannot be sustained.

In our economy, expected aggregate output is $\sum_i E[A_{i,t}]k_{i,t}^{\alpha}$. From the capital first order condition, we know that capital choice $k_{i,t}$ will be finite, as long as expected quality $E[A_{i,t}]$ is finite. Thus, the question of whether growth can be sustained becomes a question of whether $E[A_{i,t}]$ can become infinite in the limit, as all firms accumulate more and more data.

Definition 1 (Sustainable growth) *Let* $Y_t = \sum_i E[A_{i,t}]k_{i,t}^{\alpha}$ *such that* $ln(Y_{t+1}) - ln(Y_t)$ *is the*

[6]It is also true that inflow concavity comes from capital having diminishing returns. The exponent in the production function is $\alpha < 1$. But that is a separate force. Even if capital did not have diminishing marginal returns, inflows would still exhibit concavity.

aggregate growth rate of output. A data economy can sustain a minimum growth rate $\underline{g} > 0$ *if* $\exists\, T$ *such that in each period* $t > T$, $ln(Y_{t+1}) - ln(Y_t) > \underline{g}$.

Proposition 1 ***To Sustain Growth, Forecasts Must Make Infinite Output Possible*** *Sustainable growth in our data economy requires that there exists a* $\underline{v}$ *such that as* $v \to \underline{v}$ *the quality function approaches infinity* $g(v) \to \infty$.

From a mathematical perspective, this result is perhaps obvious. If the condition on g holds, then output cannot become infinite as forecast errors go to zero. If output cannot be infinite, then it cannot grow at any rate $\underline{g} > 0$ forever. But this simple idea is economically significant for two reasons. First, there are many models with perfect foresight. None generate infinite real economic value. Second, if society as a whole knows tomorrow's state, they can simply produce today what they would otherwise be able to produce tomorrow. Thus, imposing finite real output at zero forecast error is a sensible assumption. But this common-sense assumption then leads to the conclusion that data has diminishing returns.

The next result relates what is random or learnable to the potential for data to sustain growth. First, we formalize the notion of learnable. Recall that $\zeta_{i,t}$ is the set of all signals that nature draws for firm i. These are all potentially observable signals. Not all will be observed. Define Ξ_t to be the Borel σ-algebra generated by $\{\zeta_{i,t} \cup \mathcal{I}_{i,t}\}_{i=1}^{\infty}$. This is the set of all variables that can be perfectly predicted with some combination of prior information $\mathcal{I}_{i,t}$ and time-t observable data, somewhere in the economy.

Definition 2 (Fundamental randomness) *A variable* v *has time-*t *fundamental randomness if* $v \not\in \Xi_t$.

Fundamental randomness means future events that are not deterministic functions of observable events today. If they are not deterministic functions of something, like a signal, that can be observed today, they there is some risk that a signal cannot perfectly predict. In other words, fundamentally random variables are not perfectly learnable. In our model, fundamental randomness or unlearnable risk is present when $\sigma_a^2 > 0$.

Proposition 2 ***Data-Driven Growth Implies No Fundamental Randomness*** *Suppose the quality function* g *is finite almost everywhere, except* $g(0) \to \infty$. *Sustainable growth requires*

that future quality $g(a_{t+1} - \theta_{t+1}\epsilon_{a,t+1})$ has no time-t fundamental randomness.

The condition that g is finite-valued, except at zero, simply rules out the possibility that firms that have imperfect forecasts and still make mistakes can still achieve perfect, infinite quality. But this formulation allows what Proposition 1 does not. It says, even if you believe perfect one-period-ahead forecasts can produce infinite output, you still might get diminishing returns because of the existence of fundamental, unlearnable randomness.

An implication of Proposition 2 is that, for long-run data-driven growth, the economically-relevant state tomorrow must be a deterministic function of observable events today. If fundamental randomness means future random events that are not deterministic functions of observable events today, then there can be no fundamental randomness that affects the profitability of investment. If such randomness exists, it cannot be learned because it is not a deterministic function of signals, which are observable today. If it is not a deterministic function of a signal, it cannot be perfectly forecasted. If forecasts cannot be perfect, output cannot grow indefinitely. In other words, there cannot be sustainble growth.

Thus, if one believes some events tomorrow are fundamentally random, then even if perfect precision can potentially generate infinite output, data will still have diminishing returns. Conversely, even if one believes that nothing is truly random, but they believe that with one-period ahead knowledge, an economy can only produce the finite amount today that they would otherwise produce tomorrow, then data must also have diminishing returns. For process-optimizing data, without technological innovation, to produced sustained growth, one must embrace both the infinite output and no-fundamental-randomness assumptions. Keep in mind that this analysis holds technology fixed. This technology includes advances in prediction technology. So to sustain growth, without technological advance, what is required is that both perfect one-period-ahead forecasts enable infinite production, with current production technology, and that current prediction technology can, with sufficient data, achieve one-period-ahead forecasts with zero prediction error.

In section 2.2 we show how the model can naturally be extended to accommodate endogenous growth and why using data to accumulate ideas overcomes these limitations.

2.2 Endogenous Growth

Every result hinges on its assumptions. In this case, long-run stagnation comes from the assumption that transactions data is used for process optimization, as opposed to technological innovation. If we relax this assumption, we find that data can sustain growth, under the same circumstances as those in which capital can sustain growth. If data (or capital) is used for research and development, the constant addition to the stock of new ideas can sustain growth. This extension connects our main framework, which is a data economy version of Solow (1956), to a data economy version of endogenous growth with quality ladders, in the spirit of Grossman and Helpman (1991) and Aghion and Howitt (1992). Of course, for this extension to make sense, one needs to believe that information about who buys what can be used to discover growth-sustaining technologies.

Assume instead of equation (2), the evolution of quality follows

$$A_{i,t} = A_{i,t-1} + \max\{0, \Delta A_{i,t}\}, \tag{14}$$

where $\Delta A_{i,t}$ is a concave-valued function of some output relevant random variable. One can interpret $\Delta A_{i,t}$ as an uncertain technological-improvement opportunity. If adopted, this new technology will change quality of firm output at t. More data allows for more precisely targeted innovations, which increase the technology frontier by more:

$$\Delta A_{i,t} = \bar{A} - (a_{i,t} - \theta_t - \epsilon_{a,i,t})^2. \tag{15}$$

The solution follows exactly the same structure as before, $\mathbb{E}[\Delta A_{i,t}] = \bar{A} - E\left[(\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t - \epsilon_{a,i,t})^2\right]$. Therefore, the expected change in quality of firm i's good at time t can be rewritten again as $\mathbb{E}_i[\Delta A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$.

One version of this model might assume that $\bar{A} < (1+\rho^2)\sigma_a^2 + \sigma_\theta^2$. In this case, without any data, the expected value of the technological improvement is negative and the firm would not undertake it. If furthermore $\bar{A} > \sigma_a^2$, then with sufficient data, the expected value of the technological improvement turns positive. In such an economy, there would be no growth without data; data would make innovations viable.

But in a model where physical capital is an essential ingredient in research, the same would

be true of capital. The conclusion then is that, in the long run, the economic forces of the data economy are not very new at all.

2.3 Specialization in Data Production or Data Use

This framework can also give us insight into the organization of data markets. When some firms have better data mining ability (z_i), do they keep the data or sell most of it off? There are two possible ways a data-efficient firm might profit. First, it could retain the data, to make high-quality goods, to sell at a high price. Such firms are specialized in the production of high-quality goods. Alternatively, they could sell off most of their data and produce low-quality goods. Their goods would earn little or even no revenue. But their data sales would earn profits. We say that such a firm specializes in data production or data services.

When data is sufficiently non-rival, a version of comparative advantage emerges that resembles patterns of international trade: Firms that are better at data collection have a comparative (and absolute) advantage in data and specialize in data sales. Firms that are poor at data collection have the comparative advantage in high-quality goods production and specialize in that.

We consider a competitive market populated by a measure λ of low data-productivity firms ($z_i = z_L$, hereafter L-firms), and $1 - \lambda$ of high data-productivity firms ($z_i = z_H$, hereafter H-firms), in steady state. We are interested in the difference between the accumulated data of the H- and L-firms in the steady state. Firms who accumulate more data produce higher quality goods. In order to make this comparison, we define the concept of the *knowledge gap*.

Definition 3 (Knowledge Gap) *Knowledge gap denotes the equilibrium difference between knowledge level of a high and low data productivity firm,* $\Upsilon_t = \Omega_{Ht} - \Omega_{Lt}$.

When the knowledge gap is high, data-producing firms produce high-quality goods. When it is negative, data producers behave like data platforms, providing basic low-cost services and profiting mostly from their data. Regardless of the knowledge gap, high data-productivity firms would produce many units of goods and data. The question is whether they use data to produce high-quality goods or not.

Lemma 3 ***Data-Accumulation by Individual Efficient Data Producer*** *Suppose there is a single, measure-zero H-firm in the market with $z_i = z_H$ ($\lambda = 1$). In steady state, the knowledge gap is positive, $\Upsilon^{ss} > 0$, and increasing in H-firm data productivity, $\frac{d\Upsilon^{ss}}{dz_H} > 0$, $\forall \iota$ and z_H.*

When a single, high-productivity (H) firm enters a market populated by L-firms ($\lambda = 1$). The steady state outcome is what is intuitively expected. A positive knowledge gap means that the data-productive firm is larger, accumulates more data in the steady state, and specializes in high quality production. Furthermore, $\frac{d\Upsilon^{ss}}{dz_H} > 0$ means that the data productivity of the H-firm increases, it accumulates even more knowledge in steady state.

Next, consider a steady state in which there are many H-firms. Formally, the measure of L-firms, λ, is bounded away from one. In this case, when data is sufficiently non-rival, the reverse happens; the knowledge gap is negative. The next result shows that, when firms can retain most of the data they sell, high-productivity data producers sell more data; so much more that they are left with less knowledge.

Proposition 3 ***More Efficient Data Producers Accumulate Less Knowledge*** *Suppose that there is a strictly positive measure of high-data-productivity firms, $\lambda < 1$. If $\alpha < \frac{1}{2}$ and γ is sufficiently small then when data is sufficiently non-rival, $\iota < \bar{\iota}$, the steady state knowledge gap is negative: $\Upsilon^{ss} < 0$.*

The non-rivalry of data here is essential. The proof in the Appendix also shows that when data is sufficiently rival ($\iota > \bar{\iota}$), the knowledge gap becomes positive $\Upsilon^{ss} > 0$. Data non-rivalry acts like a negative bid-ask spread in the data market. It drives a wedge between the value of the data sold and the opportunity cost, the amount of data lost through the act of selling. While a bid-ask spread typically involves some loss from exchange of an asset, with non-rivalry, exchanging data results in more total data being owned. If the buyer pays a price π per unit of data gained, the seller earns more than π per unit of data forfeited, because they forfeit only a fraction of the data sold. This negative spread, or tax, on transactions incentivizes data producers to be prolific sellers of data. The incentive to sell data can be so great that these data producers are left with almost no data for themselves.

Since many economists and policy makers are concerned about concentration in data markets, we also explore what happens to data specialization when the the data market is more concentrated.

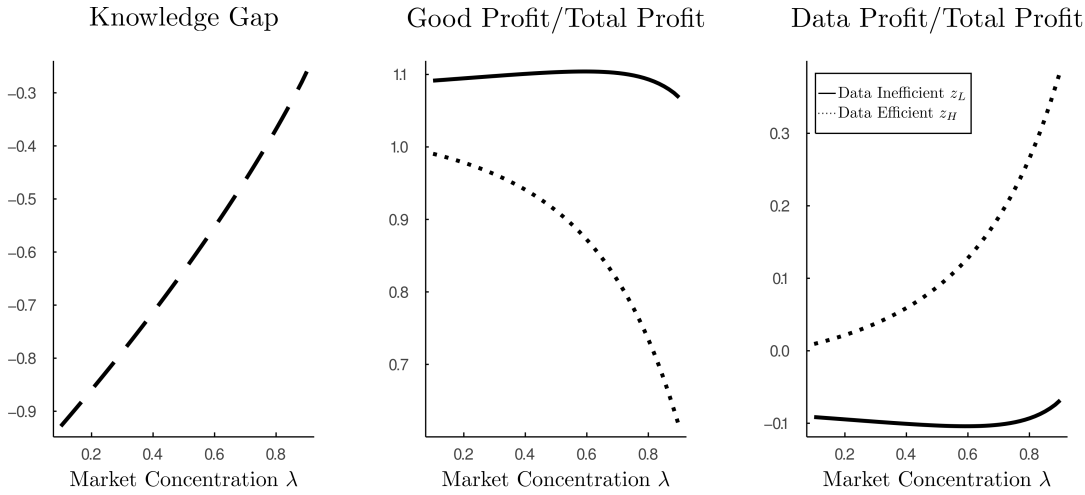


Figure 3: Data market concentration (λ) causes large (H) firms to derive most profits from data. Data market concentration is one minus the fraction of high-data-efficiency (H) firms. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \gamma = 0.09, \overline{A} = 1, \overline{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z_1 = 0.01, z_2 = 10$

The numerical example in Figure 3 illustrates the visible hallmarks of data specialization. Since data is multi-use (non-rival), the knowledge gap is negative. As a result, efficient data producers earn more of their profits from data sales. Low-efficiency producers earn negative data profits because they are data purchasers. We interpret λ close to 1, where there is a small measure of high-efficiency data producers, as being data market concentration. Figure 3 shows that data market concentration amplifies the specialization of data firms and high-quality goods producers.

Data specialization also grows as the gap between data-productivity of H- and L-type firms widens.

Corollary 1 ***Data Efficiency Divergence Amplifies Knowledge Gap*** *Suppose $\lambda < 1$, $\alpha < \frac{1}{2}$, γ is sufficiently small, and the economy is in steady state. For each λ, $\exists \bar{\iota}_2$, $\underline{z}_H$ such that*

$$\frac{d\Upsilon^{ss}}{dz_H} > 0 \qquad if \quad z_H > \underline{z}_H,\ \iota > \bar{\iota}_2$$
$$\frac{d\Upsilon^{ss}}{dz_H} < 0 \qquad if \quad z_H > \underline{z}_H,\ \iota < \bar{\iota}_2.$$

If ι is high, the knowledge gap was originally positive, and it becomes more positive when data processing efficiency diverges. If ι is low, meaning that data is not very rival, negative knowledge gap becomes more negative. But the cutoff $\bar{\iota}$ for positive knowledge gap is not the same as the cutoff $\bar{\iota}_2$ for growing knowledge gap. That means that for some intermediate levels of data rivalry,

the knowledge gap can shrink as the efficiency of the more efficient data producers improve.

Comparing Propositions 3 and 3 raises the question: How does a positive mass of high-data-productivity firms cause the result to change sign? The key is that the single, measure-zero H-firm cannot influence the amount of data held by the continuum or L-firms. The knowledge gap falls in Proposition 3, not because H-firms lose knowledge but because L-firms gain knowledge. That gain cannot happen when there is a single, measure-zero H-firm because that one firm is simply not large enough to sell data to all L-firms. By continuity, the knowledge gap also rises when data-productive firms in an industry are scarce ($\lambda \to 1$). Scarce data-efficient firms means that the data production market is very concentrated in a small number of firms.

Interpretation: Data Platforms and Data Services Large firms that sell most of their data are like data platforms (the second scenario of Corollary 1). That might appear contradictory because social networks and search engines do use lots of their own data. But they use that data primarily to sell data services to their business customers, which is a type of data sale. For example, Facebook revenue comes not from postings, but from advertising, which is a data service. A formal analysis of the equivalence between data services and data sales is in Admati and Pfleiderer (1990).

2.4 Welfare and Data Externalities

A key question for policy questions is what aspects of this economy are efficient or inefficient. Our original model gives us some tools to think about welfare, but lacks an important consideration: Data is not always used for a socially productive purpose. Firms can use data simply to steal customers away from other firms. Here, we explicitly model the household side of the economy and add a business stealing externality to examine the welfare properties of the data economy.

Consider an economy with two goods: a numeraire good, m_t, that will be produced using labor l_t, and a retail good c_t, that is produced using capital and data. Let P_t denote the price of the retail good in terms of the numeraire.

Households There is a continuum of homogeneous infinitely lived households, with quasi-linear preferences over consumption of the retail good c_t and the numeraire good m_t, and a convex dis-

utility of labor l_t. The representative household's optimization problem is

$$\max_{c_t, m_t, l_t} \sum_{t=0}^{+\infty} \beta^t (u(c_t) + m_t - v(l_t))$$
$$\text{s.t. } P_t c_t + m_t = \Phi_t + w_t l_t \qquad \forall t \tag{16}$$

The functional forms for retail good consumption and labor disutility take familiar forms: $u(c_t) = \bar{P}\frac{c_t^{1-\gamma}}{1-\gamma}$, $v(l_t) = \frac{l_t^{1+\nu}}{1+\nu}$. Equation (16) is the household budget constraint, which equates the expenditure on the two consumption goods to revenue. Household revenue is wages $w_t l_t$ and aggregate firm profits Φ_t.

Numeraire good production. The numeraire good is produced by a continuum of firms with a linear production technology that uses labor as the only input. Each unit of labor l_t produces $\bar{m}$ units of the numeraire. Since each unit of labor costs the numeraire firm w_t, the numeraire producer's objective is

$$\max_{l_t} \bar{m} l_t - w_t l_t.$$

Retail good production and using data for business stealing. There are two types of firms that produce the retail good in this economy. They are identical, except for the efficiency with which they convert transactions (produced units) into data. This data efficiency is captured by the parameter z_i for firm i. We consider two types of firms: low data efficiency firms with $z_i = z_L$ (measure λ), and high efficiency firms with $z_i = z_H$ (measure $1-\lambda$), where $z_L < z_H$.

The producers of the retail goods live forever and value their data as a long-lived asset. They use capital, rented at a constant exogenous cost r, trade data, and produce the retail good using their capital and data. Profit is revenue minus adjustment costs, minus data costs (if $\delta > 0$) or plus revenue from data sales (if $\delta < 0$), minus the cost of capital, $\Phi_{it} := P_t \mathbb{E}[A_{i,t}] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - r k_{i,t}$. The profit the households get is the aggregate firm profit $\Phi_t = \int \Phi_{it} di$. Firms maximize the present discounted value of their profit:

$$\max_{\{k_{i,t}, \delta_{i,t}\}_{t=0}^{\infty}} V(0) = \sum_{t=0}^{+\infty} \frac{1}{(1+r)^t} \left(P_t \mathbb{E}[A_{i,t}] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - r k_{i,t} \right).$$

Data governs the expected level of good quality $\mathbb{E}[A_{i,t}]$, described by equations (5) and (9), and

the law of motion of the stock of knowledge is expressed in equation (10).

Using data for business stealing can be represented as an externality that works through quality:

$$A_{i,t} = \bar{A} - \left(a_{i,t} - \theta_t - \epsilon_{a,i,t}\right)^2 + b \int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj \qquad \text{for } \; b \,\epsilon\, [0,1] \tag{17}$$

This captures the idea that when one firm uses data to reduce the distance between their chosen technique $a_{i,t}$ and the optimal technique $\theta + \epsilon$, that firm benefits, but all other firms lose a little bit. If $b = 0$, quality is the same as before. If $b = 1$, the losses in the second, externality term are such that, when added up to compute aggregate productivity, they cancel out the gains from data: $\int A_{i,t} di = \bar{A}$. The $b = 1$ assumption would represent an extreme view that data processing contributes absolutely nothing to social welfare. Moving b between 0 and 1 regulates the extent to which data adds to aggregate product quality. This formulation of the externality is inspired by Morris and Shin (2002).

Capital Market The retail sector represents an industry where consumption and data are industry-specific, but capital is rented from an inter-industry market, at rate r, paid in units of numeraire.[7]

Reformulating the problem this way makes very little difference for most of our conclusions. The business stealing externality does reduce productivity and welfare. But it does not change firms' choices because it does not enter in a firm's first order condition.[8] Therefore, it does not change data inflows, outflows, data sales or accumulation, at a given set of prices. The micro foundations are carefully chosen to deliver the same inverse demand as in (7).

Proposition 4 ***Welfare*** *If data is not used for business stealing ($b = 0$), then, for given investment choices k_{it}, the steady state equilibrium trade in data δ_{it} is socially optimal.*

If data is used for business stealing ($b > 0$), then the steady state equilibrium trade in data δ_{it}

[7] Equivalently, we can interpret this as a small, open economy where capital and numeraire goods are tradeable and retail goods are non-tradeable. The world rental rate of capital is r. This assumption makes the model tractable and allows us to focus on the economics of data. Nevertheless, it is not crucial for the result. An endogenously determined rental rate of capital leads to equilibrium effects associated with an increase in interest rates when firms are more productive, as well a wealth effect for capital owners. These equilibrium effects are well-studied in previous frameworks but are not directly related to economics of data.

[8] To see why this is the case, note that firm i's actions have a negligible effect on the average productivity term $\int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj$. So the derivative of that new externality term with respect to i's choice variables is zero. If the term is zero in the first order condition, it means it has no effect on choices of the firm.

is not socially optimal.

When data can be used for business stealing, there is a clear externality. Firms' use of data harms others. Thus, data is over-produced. That is not surprising. The contribution here is to show how to model that realistic and policy-relevant externality is a tractable way.

What is more surprising is the efficiency result. Without business stealing, the economy is not fully efficient because the capital choices are not optimal. With heterogeneous firms, one type of firm does not account for the effect of its capital choice k_{it} on the other type of firm. However, given the capital choice, the expressions for the amount of data firms choose to trade and the amount of data the social planner would want for them to trade, are identical. In other words, data is not a source of inefficiency in this market.

3 Short Run Features of a Data Economy

While the long run in the data economy looked familiar, the short run forces differ from the those in a standard capital accumulation economy. A key source of difference is short-run convexity of data accumulation at the firm level. The convexity is a form of increasing returns that arises from the data feedback loop: Firms with more data produce higher quality goods. The higher profit per unit from higher quality goods induces more production, which results in more transactions and more data. Thus more data begets more data. This convexity in the data flow explains why new firms often lose money and why goods might be sold at a zero price. These forces, in turn, affect the book-to-market or Tobin's Q of data firms. We also return to welfare and show how the capital inefficiencies that arise in competitive equilibrium may keep small firms small.

3.1 Increasing Returns in the Short Run

While the previous results focused on diminishing returns, the other force at work is increasing returns. Increasing returns arise from the data feedback loop: A firm with more data produces higher-quality goods, which induces them to invest more, produce more, and sell more. This, in turn, generates more data for them. That feedback causes aggregate knowledge accumulation to accelerate. The feedback loop competes against diminishing returns. Diminishing returns always dominate when data is abundant; the previous results about the long run were unambiguous. But

when firms are young, or data is scarce, increasing returns can be strong enough to create an increasing rate of growth. While that sounds positive, it also creates the possibility of a firm growth trap, with very slow growth, early on the in the lifecycle of a new firm.

While we have been talking about symmetric firms that do not trade data, we now relax the symmetry assumption to allow for data trade. From here on, we also adopt a linear formulation for the quality function, for simplicity. We assume that $g(x) = \bar{A} - x$. Since the second derivative of g is zero, this implies that expected quality is simply $E[A_{i,t}] = \bar{A} - \Omega_{i,t} - \sigma_a^2$.

We consider a setting were all firms are in steady state. Then, we drop in one, atomless, low-data (low $\Omega_{i,t}$) firm and observe its transition. From this exercise, we learn about barriers to new firm entrants.

Before stating the formal result, we need to define net data flow. Recall that aggregate data inflows Ω_t^+ are the total precision of all new data points at t (eq. 11). Aggregate data outflows Ω_t^- are the end-of-period-t stock of knowledge minus the discounted stock (eq. 12). We can define the data flows as the difference between data inflows and outflows. At the aggregate level, this is $d\Omega_t = d\Omega_t^+ - d\Omega_t^-$. At he individual level, data flows are defined using the individual version of equations (11)-(12), which incorporates data trade: $d\Omega_{i,t} = d\Omega_{i,t}^+ - d\Omega_{i,t}^-$. Proposition 5 states when a single firm entering faces increasing and then decreasing rates of net data flow.

Proposition 5 ***S-Shaped Accumulation of Knowledge*** *When all firms are in steady state, except for one firm i, then the firm's net data flow $d\Omega_{i,t}$*

1) *increases with the stock of knowledge $\Omega_{i,t}$ when that stock is low, $\Omega_{i,t} < \hat{\Omega}$, when goods production has sufficient diminishing marginal return, $\alpha < \frac{1}{2}$, adjustment cost Ψ is sufficiently low, $\bar{P}$ is sufficiently high, and the second derivative of the value function is bounded $V'' \in [\nu, 0)$; and*

2) *decreases with $\Omega_{i,t}$ when $\Omega_{i,t}$ is larger than $\hat{\hat{\Omega}}$.*

Entry dynamics and aggregate growth dynamics differ. The difference between one firm entering when all other firms are in steady state, and all firms growing together, is prices. When all firms are data-poor, all goods are low quality. Since quality units are scarce, prices are high. The high price of good induces these firms to produce goods, creating data. When the single firm enters, others

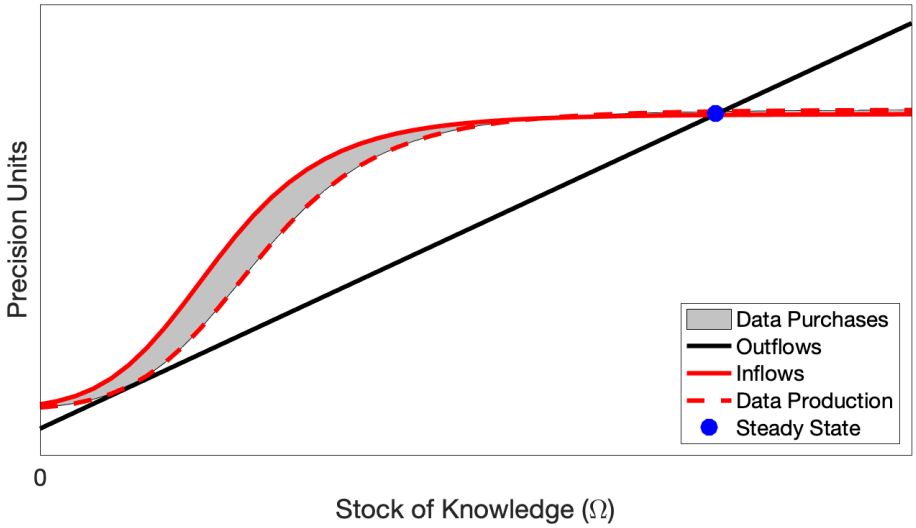


Figure 4: New firms grow slowly: inflows and outflows of data of a single firm.
Line labeled inflows plots the individual firm i version of the quantity in equation (11), that makes an optimal capital decision $k^*_{i,t}$ and data decision $\delta^*_{i,t}$, with different levels of initial data stock. This firm is in an economy where all other firms are in steady state. Line labeled outflows plots the individual firm i version of the quantity in (12). Data production is $z_i {k^*_{i,t}}^{\alpha} \sigma_\epsilon^{-2}$, which is inflows without the data purchases $\delta_{i,t}$.

are already data-rich. Quality goods are abundant, so prices are low. This makes it costlier for the single firm to grow. What works in the opposite direction is that data may also be abundant, keeping the price of data low.

For some parameter values, the diminishing returns to data is always stronger than the data feedback loop. Proposition 7 in the Appendix shows that, when learnable risk is abundant, knowledge accumulation is concave.[9] In such cases, each firm's trajectory looks like the concave path in Figure 2. But for other economies, the increasing returns of the data feedback loop is strong enough to make data inflows convex, at low levels of knowledge. The inflows, outflows and growth dynamics of such an economy are illustrated in Figure 4. This figure illustrates one possible economy. Data production may lie above or below the data outflow line.

The difference between data inflows (solid line) and data production (dashed line) is data purchases. These purchases push the inflows line above the outflows line and help speed up convergence.

What does an economy with this S-shaped knowledge accumulation look like? Figure 5 illustrates the growth path of a new entrant firm in this environment. On the left side of the time path,

[9] An additional proposition in the Appendix proves that the time-path of the stock of knowledge is s-shaped, implying a long low-return incubation period for new entrants. That result is not logically equivalent to proposition 5 because one also needs to show that the stock of knowledge Ω is always increasing.

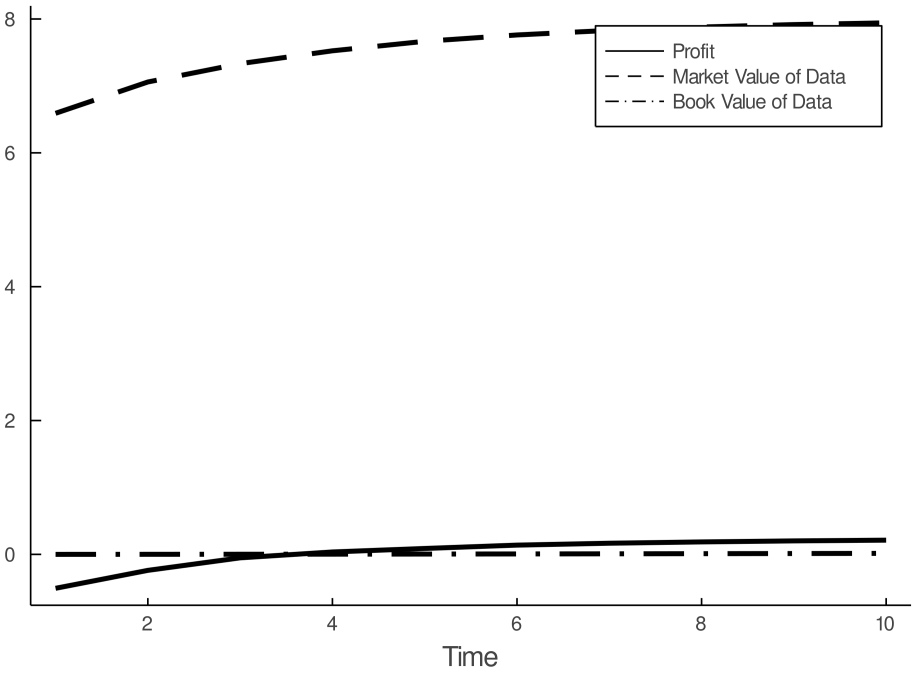


Figure 5: S-shaped growth can create initial profit losses and high market-to-book value of data. Knowledge stock defined in Lemma 1. Book value defined in (19). Parameters: $\beta = 0.7$, $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \overline{A} = 0.5, \overline{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 0.01, \pi = 0.002, P = 1, \iota = 1$

where the firm is young and the stock of data is low, increasing returns dominates. In this region, increasing returns in knowledge means low returns to production at low levels of knowledge.

We define the firm book value to be the discounted value of all purchased data. The reason is that the firm does not own any capital. It just rents capital each period. Of course, the firm also owns self-produced data. However, intangible assets created by the company, i.e. the firm's own data, are not counted. The indicator function $\mathbf{1}_{\delta_{i,t}>0}$ captures only data purchases, not sales of self-produced data. For simplicity, we equate that depreciation rate to the household's rate of time preference, $1 - \beta$.[10] Therefore, profits and book value for the new firm are

$$\text{Profits}_t = P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t} \tag{18}$$

$$\text{Book Value}_t = \sum_{\tau=0}^{t} \beta^{t-\tau} \pi_t \delta_t \mathbf{1}_{\delta_{i,\tau}>0} \tag{19}$$

The market value of the data is the Bellman equation value function $V(\Omega)$ in (13).

[10] data and software assets are typically amortized over three years. That amounts to accounting depreciation or 30% or $\beta = 0.7$ per year.

Since this firm's only asset is data, the market or book value of the data is the same as the value of the firm. However, for a firm that owns capital stock,

The difference between the market value of a firm and its book is used to measure intangible assets. This result connects our model to work measuring intangible capital as a gap between market and book values, as well as to work exploring financial barriers to firm entry.

3.2 Data Barter

Data barter arises when the goods are exchanged for customer data, at a zero price. While this is a knife-edge possibility in this model, it is an interesting outcome because it illustrates a phenomenon we see in reality. In many cases, digital products, like apps, are being developed at great cost to a company and then given away "for free." Free here means zero monetary price. But obtaining the app does involve giving one's data in return. That sort of exchange, with no monetary price attached, is a classic barter trade.

Proposition 6 ***Bartering Goods for Data*** *It is possible that a firm will optimally choose positive production $k^{\alpha}_{i,t} > 0$, even if its price per unit is zero: $P_t = 0$.*

The possibility of barter is not shocking, given the assumptions. But the result demonstrates the plausibility of the framework, by showing how it speaks to data-specific phenomena we see. The framework also allows us to value data, despite its zero monetary price.

At $P_t = 0$, the marginal benefit of investment is additional data that can be sold at price π_t. If the price of data is sufficiently high, and/or the firm is a sufficiently productive data producer (high z_i), then the firm should engage in costly production, even at a zero goods price, because it also produces data, which has a positive price.

Figure 5 illustrated an example where the firm makes negative profits for the first 1-2 periods because they sell goods at a loss. Producing goods at a loss eventually pays off for this firm. It generates data that allows the firm to become profitable. This situation looks like Amazon at its inception. During its first 17 quarters as a public company, Amazon lost $2.8 billion, before turning a profit. Today, it is one of the most valuable companies in the world.

4 Conclusions

The economics of transactions data bears some resemblance to technology and some to capital. It is not identical to either. Yet, when economies accumulate data alone, the aggregate growth economics are similar to an economy that accumulates capital alone. Diminishing returns set in the result is a higher level of income, but not sustained growth. Data's production process, with its feedback loop from data to production and back to data, also makes increasing returns a natural outcome. Thus, while the accumulation and analysis of data may be the hallmark of the "new economy," this new economy has many economic forces at work that are old and familiar.

In conclusion, we describe four possible directions of future research that our framework could be used for.

Optimal Data Policy. A benevolent government might adopt a data policy to promote the growth of small and mid-size firms. The policy solution to increasing returns growth traps is typically a form of big push investment. In the context of data investment, the government could collect data itself, from taxes or reporting requirements, and share it with firms. For example, China shares data with some firms, in a way that seems to facilitate their growth Beraja et al. (2020). Alternatively, the government might facilitate data sharing or act to prevent data from being exported to foreign firms.

Firm Size Dispersion. One of the biggest questions in macroeconomics and industrial organization is: What is the source of the changes in the distribution of firm size? One possible source is the accumulation of data. The S-shaped dynamic of firm growth implies that firm size first becomes more heterogeneous and then converges. During the convex, increasing returns portion of the growth trajectory, small initial differences in the initial data stock of firms get amplified.

Investing in data-savviness. The fixed data productivity parameter z_i represents the idea that certain industries will spin off more data than others. Credit card companies learn more than barber shops. We could allow a firm to do more to collect, structure and analyze the data that its transactions produce. It could choose its data-savviness z_i, at a cost. Endogenizing this choice might produce changes in the cross-section of firms' data, over time.

Data and product portfolio choice. We examined a model about how much data a firm produces and accumulates. Just as important a question is what type of data that is. Appendix B sketches

a model where different goods can be informative about different risks. In such a world, firms could invest in a portfolio of products to diversify and learn some about each or could specialize to become expert in producing one good with high quality. The choices, in turn, would shape the forces of market competition.

Measurement of economic activity. Since digital service providers often get compensated with data, rather than monetary payments, valuing digital goods and services according to their market price undervalues this economic activity. If we can estimate the value function $v(\cdot)$ to value data, we can then derive new measures of the value of digital goods and services that incorporate the valuable data involved in each transaction.

This simple framework enables research on many data-related phenomena. It can be a foundation for thinking about many more.

References

Acemoglu, Daron and Pascual Restrepo, "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, June 2018, *108* (6), 1488–1542.

Admati, Anat and Paul Pfleiderer, "Direct and Indirect Sale of Information," *Econometrica*, 1990, *58* (4), 901–928.

Aghion, Philippe and Peter Howitt, "A model of growth through creative destruction," 1992.

_ , Benjamin F. Jones, and Charles I. Jones, "Artificial Intelligence and Economic Growth," 2017. Stanford GSB Working Paper.

Agrawal, Ajay, John McHale, and Alexander Oettl, "Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth," in "The Economics of Artificial Intelligence: An Agenda," National Bureau of Economic Research, Inc, 2018.

Atkeson, Andrew and Patrick J Kehoe, "Modeling and Measuring Organization Capital," *Journal of political Economy*, 2005, *113* (5), 1026–1053.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki, "The Impact of Big Data on Firm Performance: An Empirical Investigation," Working Paper 24334, National Bureau of Economic Research February 2018.

Baumol, William J., "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *The American Economic Review*, 1967, *57* (3), 415–426.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp, "Big Data in Finance and the Growth of Large Firms," Working Paper 24550, National Bureau of Economic Research April 2018.

Beraja, Martin, David Y. Yang, and Noam Yuchtman, "Data-intensive Innovation and the State: Evidence from AI Firms in China," 2020. MIT Working Paper.

Caplin, Andrew and John Leahy, "Business as Usual, Market Crashes, and Wisdom After the Fact," *American Economic Review*, 1994, *84* (3), 548–565.

Cong, Lin William, Danxia Xie, and Longtian Zhang, "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Privacy, and Growth in a Data Economy (October 8, 2020)*, 2020.

Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel, "Uncertainty Traps," *The Quarterly Journal of Economics*, 2017, *132* (4), 1641–1692.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp, "Big Data and Firm Dynamics," *American Economic Association Papers and Proceedings*, May 2019.

Farmer, RE and J Benhabib, "Indeterminacy and Increasing Returns," *Journal of Economic Theory*, 1994, *63*, 19–41.

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

Grossman, Gene M and Elhanan Helpman, "Quality ladders in the theory of growth," *The review of economic studies*, 1991, *58* (1), 43–61.

Ilut, Cosmin and Martin Schneider, "Ambiguous Business Cycles," *American Economic Review*, August 2014, *104* (8), 2368–99.

Jones, Chad and Chris Tonetti, "Nonrivalry and the Economics of Data," 2018. Stanford GSB Working Paper.

Jovanovic, Boyan and Yaw Nyarko, "Learning by Doing and the Choice of Technology," *Econometrica*, 1996, *64* (6), 1299–1310.

Lorenzoni, Guido, "A Theory of Demand Shocks," *American Economic Review*, December 2009, *99* (5), 2050–84.

Lu, Chia-Hui, "The impact of artificial intelligence on economic growth and welfare," 2019. National Taipei University Working Paper.

Lucas, Robert, "On the mechanics of economic development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

Morris, Stephen and Hyun Song Shin, "Social value of public information," *The American Economic Review*, 2002, *92* (5), 1521–1534.

Oberfield, Ezra and Venky Venkateswaran, "Expertise and Firm Dynamics," 2018 Meeting Papers 1132, Society for Economic Dynamics 2018.

Ordonez, Guillermo, "The Asymmetric Effects of Financial Frictions," *Journal of Political Economy*, 2013, *121* (5), 844–895.

Solow, Robert M., "A Contribution to the Theory of Economic Growth," *The Quarterly Journal of Economics*, 02 1956, *70* (1), 65–94.

Veldkamp, Laura, "Slow Boom, Sudden Crash," *Journal of Economic Theory*, 2005, *124(2)*, 230–257.

A Appendix: Derivations and Proofs. Not For Publication.

A.1 Belief updating

The information problem of firm i about its optimal technique $\theta_{i,t}$ can be expressed as a Kalman filtering system, with a 2-by-1 observation equation, $(\hat{\mu}_{i,t}, \Sigma_{i,t})$.

We start by describing the Kalman system, and show that the sequence of conditional variances is deterministic. Note that all the variables are firm specific, but since the information problem is solved firm-by-firm, for brevity we suppress the dependence on firm index i.

At time t, each firm observes two types of signals. First, date $t-1$ output reveals -1 good quality $A_{i,t-1} = y_{i,t-1}/k^{\alpha}_{i,t-1}$. Good quality $A_{i,t-1}$ provides a noisy signal about θ_{t-1}. Let that signal be $s^{a}_{i,t-1} = (\bar{A} - A_{i,t-1})^{1/2} - a_{i,t-1}$. Note that, from equation 2, that the signal derived from observed output is equivalent to

$$s^{a}_{i,t-1} = \theta_{t-1} + \epsilon_{a,t-1}, \tag{20}$$

where $\epsilon_{a,t} \sim \mathcal{N}(0, \sigma^2_a)$.

The second type of signal the firm observes is data points. They are a by-product of economic activity. For firms that do not trade data, the number of new data points added to the firm's data set is $\omega_{i,t} = n_{i,t} = z k^{\alpha}_{i,t}$. For firms that do trade data, $\omega_{i,t} = n_{i,t} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0})$. The set of signals $\{s_{t,m}\}_{m\in[1:\omega_{i,t}]}$ are equivalent to an aggregate (cross-firm average) signal $\bar{s}_t$ such that:

$$\bar{s}_t = \theta_t + \epsilon_{s,t}, \tag{21}$$

where $\epsilon_{s,t} \sim \mathcal{N}(0, \sigma^2_{\epsilon}/\omega_{it})$. The state equation is

$$\theta_t - \bar{\theta} = \rho(\theta_{t-1} - \bar{\theta}) + \eta_t,$$

where $\eta_t \sim \mathcal{N}(0, \sigma^2_{\theta})$.

At time, t, the firm takes as given:

$$\widehat{\mu}_{t-1} = \mathbb{E}\big[\theta_t \mid s^{t-1}, y^{t-2}\big]$$
$$\Sigma_{t-1} = Var\big[\theta_t \mid s^{t-1}, y^{t-2}\big]$$

where $s^{t-1} = \{s_{t-1}, s_{t-2}, \dots\}$ and $y^{t-2} = \{y_{t-2}, y_{i,t-3}, \dots\}$ denote the histories of the observed variables, and $s_t = \{s_{t,m}\}_{m\in[1:\omega_{i,t}]}$.

We update the state variable sequentially, using the two signals. First, combine the priors with $s^a_{i,t-1}$:

$$\begin{aligned}
\mathbb{E}\big[\theta_{t-1} \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] &= \frac{\Sigma^{-1}_{t-1}\widehat{\mu}_{t-1} + \sigma_a^{-2} s^a_{i,t-1}}{\Sigma^{-1}_{t-1} + \sigma_a^{-2}} \\
V\big[\theta_{t-1} \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] &= \big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} \\
\mathbb{E}\big[\theta_{t} \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] &= \bar{\theta} + \rho \cdot \big(\mathbb{E}\big[\theta_{t-1} \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] - \bar{\theta}\big) \\
V\big[\theta_{t} \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] &= \rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma^2_\theta
\end{aligned}$$

Then, use these as priors and update them with $\bar{s}_t$:

$$\widehat{\mu}_t = \mathbb{E}\big[\theta_t \mid \mathcal{I}_t\big] = \frac{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma^2_\theta\Big]^{-1} \cdot \mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] + \omega_t \sigma_\epsilon^{-2} \bar{s}_t}{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma^2_\theta\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}} \tag{22}$$

$$\Sigma_t = Var\big[\theta_t \mid \mathcal{I}_t\big] = \Big\{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma^2_\theta\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}\Big\}^{-1} \tag{23}$$

Multiply and divide equation (22) by Σ_t as defined in equation (23) to get

$$\widehat{\mu}_{i,t} = (1 - \omega_t \sigma_\epsilon^{-2}\Sigma_t)\left[\bar{\theta}(1-\rho) + \rho\left((1-M_t)\mu_{t-1} + M_t s^a_{i,t-1}\right)\right] + \omega_t \sigma_\epsilon^{-2}\Sigma_t \bar{s}_t, \tag{24}$$

where $M_t = \sigma_a^{-2}(\Sigma_{t-1} + \sigma_a^{-2})^{-1}$.

Equations (23) and (24) constitute the Kalman filter describing the firm dynamic information problem. Importantly, note that Σ_t is deterministic.

A.2 Making the Problem Recursive: Proof of Lemma 1

Lemma. The sequence problem of the firm can be solved as a non-stochastic recursive problem with one state variable. Consider the firm sequential problem:

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\frac{1}{1+r}\right)^t (P_t A_t k_t^\alpha - r k_t)$$

We can take a first order condition with respect to a_t and get that at any date t and for any level of k_t, the optimal choice of technique is

$$a^*_t = \mathbb{E}[\theta_t | \mathcal{I}_t].$$

Given the choice of a_t's, using the law of iterated expectations, we have:

$$\mathbb{E}[(a_t - \theta_t - \epsilon_{a,t})^2 | \mathcal{I}_s] = \mathbb{E}[Var[\theta_t + \epsilon_{a,t} | \mathcal{I}_t] | \mathcal{I}_s] + \sigma_a^2,$$

for any date $s \leq t$. We will show that this object is not stochastic and therefore is the same for any information set that does not contain the realization of θ_t.

We can restate the sequence problem recursively. Let us define the value function as:

$$V(\{s_{t,m}\}_{m\in[1:\omega_t]}, y_{t-1}, \hat{\mu}_{t-1}, \Sigma_{t-1}) = \max_{k_t, a_t} \mathbb{E}\left[P_t A_t k_t^\alpha - rk_t + \left(\frac{1}{1+r}\right) V(\{s_{t+1,m}\}_{m\in[1:\omega_{t+1}]}, y_t, \hat{\mu}_t, \Sigma_t)|\mathcal{I}_{t-1}\right]$$

with $\omega_{i,t}$ being the net amount of data being added to the data stock. Taking a first order condition with respect to the technique choice conditional on $\mathcal{I}_t$ reveals that the optimal technique is $a_t^* = \mathbb{E}[\theta_t|I_t]$. We can substitute the optimal choice of a_t into A_t and rewrite the value function as

$$V(\{s_{t,m}\}_{m\in[1:\omega_t]}, y_{t-1}, \hat{\mu}_{t-1}, \Sigma_{t-1}) = \max_{k_t} \mathbb{E}\Big[P_t\big(\bar{A} - (\mathbb{E}[\theta_t|I_t] - \theta_t - \epsilon_{a,t})^2\big)k_t^\alpha - rk_t + \left(\frac{1}{1+r}\right) V(\{s_{t+1,m}\}_{m\in[1:\omega_{t+1}]}, y_t, \hat{\mu}_t, \Sigma_t)|\mathcal{I}_{t-1}\Big].$$

Note that $\epsilon_{a,t}$ is orthogonal to all other signals and shocks and has a zero mean. Thus,

$$V(\{s_{t,m}\}_{m\in[1:\omega_t]}, y_{t-1}, \hat{\mu}_{t-1}, \Sigma_{t-1}) = \max_{k_t} \mathbb{E}\Big[P_t\big(\bar{A} - ((\mathbb{E}[\theta_t|I_t] - \theta_t)^2 + \sigma_a^2)\big)k_t^\alpha - rk_t + \left(\frac{1}{1+r}\right) V(\{s_{t+1,m}\}_{m\in[1:\omega_{t+1}]}, y_t, \hat{\mu}_t, \Sigma_t)|\mathcal{I}_{t-1}\Big].$$

Notice that $\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2|I_{t-1}]$ is the time-t conditional (posterior) variance of θ_t, and the posterior variance of beliefs is $\mathbb{E}[(\mathbb{E}[\theta_t|I_t] - \theta_t)^2] := \Sigma_t$. Thus, expected productivity is $\mathbb{E}[A_t] = \bar{A} - \Sigma_t - \sigma_a^2$, which determines the within period expected payoff. Additionally, using the Kalman system equation (23), this posterior variance is

$$\Sigma_t = \left[P_t\left[\rho^2(\Sigma_{t-1}^{-1} + \sigma_a^2)^{-1} + \sigma_\theta^2\right]^{-1} + \omega_t\sigma_\epsilon^{-2}\right]^{-1}$$

which depends only on Σ_{t-1}, n_t, and other known parameters. It does not depend on the realization of the data. Thus, $\{s_{t,m}\}_{m\in[1:\omega_t]}, y_{t-1}, \hat{\mu}_t$ do not appear on the right side of the value function equation; they are only relevant for determining the optimal action a_t. Therefore, we can rewrite the value function as:

$$V(\Sigma_t) = \max_{k_t} \left[P_t(\bar{A} - \Sigma_t - \sigma_a^2)k_t^\alpha + \pi\delta i, t - \Psi(\Delta\Omega_{i,t+1}) - rk_t + \left(\frac{1}{1+r}\right) V(\Sigma_{t+1})\right]$$
$$\text{s.t.} \qquad \Sigma_{t+1} = \left[\left[\rho^2(\Sigma_t^{-1} + \sigma_a^2)^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2}\right]^{-1}$$

Data use is incorporated in the stock of knowledge through (10), which still represents one state variable.

A.3 Lemma 2: Equilibrium and Steady State Without Trade in Data

Capital Choice The first order condition for the optimal capital choice is

$$\alpha P_t A_{i,t} k_t^{\alpha-1} - \Psi'(\cdot)\frac{\partial\Delta\Omega_{t+1}}{\partial k_{i,t}} - r + \left(\frac{1}{1+r}\right) V'(\cdot)\frac{\partial\Omega_{t+1}}{\partial k_{i,t}} = 0$$

where $\frac{\partial \Omega_{t+1}}{\partial k_{i,t}} = \alpha z_i k_{i,t}^{\alpha-1} \sigma_\epsilon^{-2}$ and $\Psi'(\cdot) = 2\psi(\Omega_{i,t+1} - \Omega_{i,t})$. Substituting in the partial derivatives and for $\Omega_{i,t+1}$, we get

$$k_{i,t} = \left[\frac{\alpha}{r} \left(P_t A_{i,t} + z_i \sigma_\epsilon^{-2} (\left(\frac{1}{1+r} \right) V'(\cdot) - 2\psi(\cdot)) \right) \right]^{1/(1-\alpha)} \tag{25}$$

Differentiating the value function in Lemma 1 reveals that the marginal value of data is

$$V'(\Omega_{i,t}) = P_t A_{i,t} k_{i,t}^{\alpha} \frac{\partial A_{i,t}}{\partial \Omega_{i,t}} - \Psi'(\cdot) \left(\frac{\partial \Omega_{t+1}}{\partial \Omega_t} - 1 \right) + \left(\frac{1}{1+r} \right) V'(\cdot) \frac{\partial \Omega_{t+1}}{\partial \Omega_t}$$

where $\partial A_{i,t} / \partial \Omega_{i,t} = \Omega_{i,t}^{-2}$ and $\partial \Omega_{t+1} / \partial \Omega_t = \rho^2 \left[\rho^2 + \sigma_\theta^2 (\Omega_{i,t} + \sigma_a^{-2}) \right]^{-2}$.

To solve this, we start with a guess of V' and then solve the non-linear equation above for $k_{i,t}$. Then, update our guess of V.

Steady State The steady state is where capital and data are constant. For data to be constant, it means that $\Omega_{i,t+1} = \Omega_{i,t}$. Using the law of motion for Ω (eq 10), we can rewrite this as

$$\omega_{ss} \sigma_\epsilon^{-2} + \left[\rho^2 (\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} = \Omega_{ss} \tag{26}$$

This is equating the inflows of data $\omega_{i,t} \sigma_\epsilon^{-2}$ with the outflows of data $\left[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} - \Omega_{i,t}$. Given a number of new data points ω_{ss}, this pins down the steady state stock of data. The number of data points depends on the steady state level of capital. The steady state level of capital is given by (25) for A_{ss} depending on Ω_{ss} and a steady state level of $V'(\Omega_{ss})$. We use the term V'_{ss} to refer to the partial derivative $\partial V / \partial \Omega$, evaluated at the steady state value of Ω. We solve for that steady state marginal value of data next.

If data is constant, then the level and derivative of the value function are also constant. Equating $V'(\Omega_{i,t}) = V'(\Omega_{i,t+1})$ allows us to solve for the marginal value of data analytically, in terms of k_{ss}, which in turn depends on Ω_{ss}:

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r} \right) \frac{\partial \Omega_{t+1}}{\partial \Omega_t} |_{ss} \right]^{-1} P_t k_{ss}^{\alpha} \Omega_{ss}^{-2} \tag{27}$$

Note that the data adjustment term $\Psi'(\cdot)$ dropped out because in steady state $\Delta \Omega = 0$ and we assumed that $\Psi'(0) = 0$.

The equations (25), (26) and (27) form a system of 3 equations in 3 unknowns. The solution to this system delivers the steady state levels of data, its marginal value and the steady state level of capital.

A.4 Equilibrium With Trade in Data

To simplify our solutions, it is helpful to do a change of variables and optimize not over the amount of data purchased or sold $\delta_{i,t}$, but rather the closely related, net change in the data stock $\omega_{i,t}$. We also substitute in $n_{i,t} = z_i k_{i,t}^{\alpha}$ and

substitute in the optimal choice of technique $a_{i,t}$. The equivalent problem becomes

$$V(\Omega_{i,t}) = \max_{k_{i,t},\omega_{i,t}} P_t\left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2\right) k_{i,t}^{\alpha} - \pi\left(\frac{\omega_{i,t} - z_i k_{i,t}^{\alpha}}{\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}}}\right) - r k_{i,t} - \Psi\left(\Delta\Omega_{i,t+1}\right) + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1}) \tag{28}$$

$$\text{where} \quad \Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2} \tag{29}$$

Capital Choice The first order condition for the optimal capital choice is

$$FOC[k_{i,t}]: \quad \alpha P_t A_{i,t} k_{i,t}^{\alpha-1} + \frac{\pi\alpha z_i k_{i,t}^{\alpha-1}}{\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}}} - r = 0 \tag{30}$$

Solving for $k_{i,t}$ gives

$$k_{i,t} = \left(\frac{1}{r}\left(\alpha P_t A_{i,t} + \tilde{\pi}\alpha z_i\right)\right)^{\frac{1}{1-\alpha}} \tag{31}$$

where $\tilde{\pi} \equiv \pi/(\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}})$. That the adjusted price $\tilde{\pi}$ is higher when a firm sells data. We are dividing by $\iota < 1$, which raises the price. This idea is that a firm that sells δ units of data only gives up $\delta\iota$ units of data. So it's as if they are getting a higher price per unit of data they actually forfeit.

Note that a firm's capital decision is optimally static. It does not depend on the future marginal value of data (i.e., $V'(\Omega_{i,t+1})$) explicitly.

Data Use Choice The first order condition for the optimal $\omega_{i,t}$ is

$$FOC[\omega_{i,t}]: \quad -\Psi'(\cdot)\frac{\partial\Delta\Omega_{i,t+1}}{\partial\omega_{i,t}} - \tilde{\pi} + \left(\frac{1}{1+r}\right) V'(\Omega_{i,t+1})\frac{\partial\Omega_{i,t+1}}{\partial\omega_{i,t}} = 0 \tag{32}$$

where $\frac{\partial\Omega_{i,t,t+1}}{\partial\omega_{i,t}} = \sigma_\epsilon^{-2}$.

Steady State The steady state is where capital and data are constant. For data to be constant, it means that $\Omega_{i,t+1} = \Omega_{i,t}$. Using the law of motion for Ω (eq 10), we can rewrite this as

$$\omega_{ss}\sigma_\epsilon^{-2} + \left[\rho^2(\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} = \Omega_{ss} \tag{33}$$

This is equating the inflows of data $\omega_{i,t}\sigma_\epsilon^{-2}$ with the outflows of data $\left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} - \Omega_{i,t}$. Given a number of new data points ω_{ss}, this pins down the steady state stock of data. The number of data points depends on the steady state level of capital. The steady state level of capital is given by Equation 31 for A_{ss} depending on Ω_{ss} and a steady state level of V'_{ss}. We solve for that steady state marginal value of data next.

If data is constant, then the level and derivative of the value function are also constant. Equating $V'(\Omega_{i,t}) = V'(\Omega_{i,t+1})$ allows us to solve for the marginal value of data analytically, in terms of k_{ss}, which in turn depends on Ω_{ss}:

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r}\right)\frac{\partial\Omega_{t+1}}{\partial\Omega_t}|_{ss}\right]^{-1} P_{ss} k_{ss}^{\alpha} \Omega_{ss}^{-2} \tag{34}$$

Note that the data adjustment term $\Psi'(\cdot)$ dropped out because in steady state $\Delta\Omega = 0$ and we assumed that $\Psi'(0) = 0$.

From the first order condition for $\omega_{i,t}$ (eq 32), the steady state marginal value is given by

$$V'_{ss} = (1+r)\tilde{\pi}\sigma_\epsilon^2 \tag{35}$$

The equations (31), (32), (33) and (34) form a system of 4 equations in 4 unknowns. The solution to this system delivers the steady state levels of capital, knowledge, data, and marginal value data.

A.4.1 Characterization of Firm Optimization Problem in Steady State

At this point, from tractability, we switch notation slightly. Instead of optimizing over the net additions to data ω, we refer to the purchase/sale of data $\delta := \omega_{i,t} - n_{i,t}$.

Individual Optimization Problem

$$\begin{aligned}
V(\Omega_{i,t}) &= \max_{k_{i,t},\delta_{i,t}} P_t A_{i,t} k_{i,t}^\alpha - \psi\left(\frac{\Omega_{i,t+1}-\Omega_{i,t}}{\Omega_{i,t}}\right)^2 - \pi\delta_{i,t} - rk_{i,t} + \frac{1}{1+r}V(\Omega_{i,t+1}) \\
\Omega_{i,t+1} &= \left(\rho^2(\Omega_{i,t}+\sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_i k_{i,t}^\alpha + \left(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0}\right)\delta_{i,t}\right)\sigma_\epsilon^{-2} \\
A_{i,t} &= \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2
\end{aligned}$$

where i denotes the firm data productivity.

Thus the steady state is characterized by the following 8 equations:

$$\Omega_L = \left(\rho^2(\Omega_L+\sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_L k_L^\alpha + \delta_L\right)\sigma_\epsilon^{-2} \tag{36}$$

$$\Omega_H = \left(\rho^2(\Omega_H+\sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_H k_H^\alpha + \iota\delta_H\right)\sigma_\epsilon^{-2} \tag{37}$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha-1} + \pi\alpha z_L k_L^{\alpha-1} = r \tag{38}$$

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha-1} + \frac{\pi\alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{39}$$

$$P\sigma_\epsilon^{-2}k_L^\alpha = \pi\Omega_L^2\left(1 + r - \frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right) \tag{40}$$

$$\iota P\sigma_\epsilon^{-2}k_H^\alpha = \pi\Omega_H^2\left(1 + r - \frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)^2}\right) \tag{41}$$

$$P = \bar{P}\left(\lambda(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^\alpha + (1-\lambda)(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^\alpha\right)^{-\gamma} \tag{42}$$

$$\lambda\delta_L + (1-\lambda)\delta_H = 0 \tag{43}$$

A.5 Proof of Proposition 1: Perfect Foresight Must Imply Infinite Output

Suppose not. Then, for every firm $i \in I$, with $\int_{i\notin I} di = 0$, producing infinite data $n_{i,t} \to \infty$ implies finite firm output $y_{i,t} < \infty$. Thus $M_y \equiv \sup_i\{y_{i,t}\} + 1$ exists and is finite. By definition, $y_{i,t} < M_y$, $\forall i$. If the measure of all firms is

also finite, that is $\exists 0 < N < \infty$ such that $\int_i di < N$. As a result, the aggregate output is also finite in any period $t+s$, $\forall s > 0$:

$$Y_{t+s} = \int_i y_{i,t} di < M_y \int_i di < M_y N < \infty. \tag{44}$$

On the other hand, given that the aggregate growth rate of output $\ln(Y_{t+1}) - \ln(Y_t) > \underline{g} > 0$, we have that in period $t+s$, $\forall s > 0$,

$$\ln(Y_{t+s}) - \ln(Y_t) = [\ln(Y_{t+s}) - \ln(Y_{t+s-1})] + \cdots + [\ln(Y_{t+1}) - \ln(Y_t)] > \underline{g}s, \tag{45}$$

which implies

$$Y_{t+s} > Y_t e^{\underline{g}s}. \tag{46}$$

Thus for $\forall s > \underline{s} \equiv \lceil \frac{\ln(MN) - \ln(Y_t)}{\underline{g}} \rceil$,

$$Y_{t+s} > Y_t e^{\underline{g}s} > Y_t e^{\underline{g}\underline{s}} > Y_t e^{\underline{g} \frac{\ln(M_y N) - \ln(Y_t)}{\underline{g}}} = M_y N, \tag{47}$$

which contradicts (44).

A.6 Proof of Proposition 2: Perfect Foresight Implies a Deterministic Future

We break this result into two parts. Part (a) of the result is that in order to have infinite output in the limit, an economy will need an infinite forecast precisions. Forecasts with errors won't produce the maximum possible, infinite, output.

Part (b) of the result says that if signals are derived from the observations of past events, then infinite precision implies that the one-period-ahead future is deterministic. Allowing precesion to be infinite means there cannot be any fundamental randomness, any unlearnable risk, because that would cause forecasts to be imperfect. Infinite precision means zero forecast error with certainty. Such perfect forecasts can only exist if future events are perfectly forecastable with past data. Perfectly forecastable means that, conditional on past events, the future is not random. Thus, future events are conditionally deterministic.

Part a. Claim: Suppose aggregate output is a finite-valued function of each firm's forecast precision: $Y_t = f(\Gamma_t)$. A data economy can sustain an aggregate growth rate of output $ln(Y_{t+1}) - ln(Y_t)$ that is greater than any lower bound $\underline{g} > 0$, in each each period t, only if infinite data $n_{i,t} \to \infty$ for some firm i implies infinite precision $\Omega_{i,t} \to \infty$. *Proof part a*: From proposition 1, we know that sustaining aggregate growth above any lower bound $\underline{g} > 0$ arises only if a data economy achieves infinite output $Y_t \to \infty$ when some firm has infinite data $n_{i,t} \to \infty$. Since Y_t is a finite-valued function of Γ_t, it can only be that $Y_t \to \infty$ if some moment of Γ_t is also becoming infinite $\Gamma_t \to \pm\infty$. Moments of Γ_t cannot become negative infinite because Γ_t is a distribution over Ω_t which is a precision, defined to be non-negative. Thus for some moment, $\Gamma_t \to \infty$. If some amount of probability mass is being placed on Ω's that are approaching infinity, that means there is some measure of firms that are achieving perfect forecast precision:

$\Omega_{i,t} \to \infty$. $\square$

Suppose not. Then, for every firm $i \in I$, with $\int_{i \notin I} di = 0$, producing infinite data $n_{i,t} \to \infty$ implies finite precision $\Omega_{i,t} < \infty$, that is Γ_t is finite (except for zero-measure sets). Since $Y_t = f(\Gamma_t)$ is a finite-valued function, we must have $Y_t < \infty$, as $n_{i,t} \to \infty$. In other words, since Y_t is a finite-valued function of Γ_t, it can only be that $Y_t \to \infty$ if some moment of Γ_t is also becoming infinite $\Gamma_t \to \pm\infty$. Moments of Γ_t cannot become negative infinite because Γ_t is a distribution over Ω_t which is a precision, defined to be non-negative. Thus for some moment, $\Gamma_t \to \infty$. If some amount of probability mass is being placed on Ω's that are approaching infinity, that means there is some measure of firms that are achieving perfect forecast precision: $\Omega_{i,t} \to \infty$.

But finite limit output is inconsistent with sustained growth. From proposition 1, we know that sustaining aggregate growth above any lower bound $\underline{g} > 0$ arises only if a data economy achieves infinite output $Y_t \to \infty$ when some firm with positive measure has infinite data $n_{i,t} \to \infty$. This is a contradiction.

Part b. Claim: Suppose all data points $s_{i,t,m}$ are t-measurable signals about some future event θ_{t+1}. If infinite data $n_{i,t} \to \infty$ for some firm i implies infinite precision $\Omega_{i,t} \to \infty$, then future events θ_{t+1} are deterministic: θ_{t+1} is a deterministic function of the sigma algebra of past events.

We prove this statement by proving the contrapositive: If the future, θ_t is not deterministic at $t-1$, then the stock of knowledge must be finite.

Suppose θ_{t+1} is not a deterministic function of the sigma algebra of past events. Then θ_{t+1} is random with respect to the sigma algebra of the t-1 history of events. Let $\mathcal{F}_t$ be the sigma algebra derived from the history $\{\theta_\tau, s_{\tau,m}, s^a_{i,\tau}\}_{\tau=0}^{t-1}$.

If signals are measurable with respect to all past events, then they are a subset of the sigma algebra of past events. Formally, the information set of firm i when it chooses its technique $a_{i,t}$ is $\mathcal{I}_{i,t} = [\{s^a_{i,\tau}\}_{\tau=0}^{t-1}; \{\{s_{i,\tau,m}\}_{m=1}^{n_{i,\tau}}\}_{\tau=0}^{t-1}]$. By assumption, $s_{i,t-1,m}$ and $s^a_{i,t-1}$ are measurable with respect to $\mathcal{F}_t$, that is $\forall B \in \mathcal{B}, \{\omega : s_{i,t,m}(\omega) \in B\} \subset \mathcal{F}_t$. So $\sigma(\mathcal{I}_{i,t}) \subset \mathcal{F}_t$. This implies that $t-1$ measurable signals cannot contain information about the future event θ_t, other than what is already present in the history of events.

By construction, θ_t is not measurable with respect to $\mathcal{F}_{t-1}$, that is $\exists B' \in \mathcal{B}$ *s.t.* $\{\omega : \theta_t(\omega) \in B\} \not\subset \mathcal{F}_t$. Since $\sigma(\mathcal{I}_{i,t}) \subset \mathcal{F}_t$, we have that $\{\omega : \theta_t(\omega) \in B\} \not\subset \mathcal{I}_{i,t,m}$, and thus θ_t is not measurable with respect to $\mathcal{I}_{i,t,m}$. Therefore $\mathbb{V}ar(\theta_t \mid \mathcal{I}_{i,t,m}) > 0$. By the definition of Ω as the inverse of the conditional variance, this implies $\Omega_{i,t} < \infty$.

This showed that, if θ_t is random with respect to past events, it must be random with respect to all possible signals $s_{i,t,m}$. If θ_t is random with respect to the signals, there is strictly positive forecast variance. If forecast variance cannot be zero, then signal precision cannot be infinite.

Since we proved that the the stock of knowledge must be finite, therefore the contrapositive, that infinite precision implies a deterministic future, must also be true.

A.7 Proof of Proposition 5: S-shaped Accumulation of Knowledge

We proceed in two parts: convexity and then concavity.

Part a. Convexity at low levels of Ω_t. In this part, we first calculate the derivatives of data infow and outflow with respect to $\Omega_{i,t}$, combine them to form the derivative of data net flow, and then show that it is positive in given parameter regions for $\Omega_{i,t} < \hat{\Omega}$.

Since all other firms, besides firm i are in steady state, we take the prices π_t and P_t as given. Since data is sufficiently expensive, data purchases are small. We prove this for zero data trade. By continuity, the result holds for small amounts of traded data.

Recall that data inflow is $d\Omega^+_{i,t} = z_{i,t}k^{\alpha}_{i,t}\sigma_\epsilon^{-2}$ and its first derivative is $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} = \alpha z_{i,t}k^{\alpha-1}_{i,t}\sigma_\epsilon^{-2}\frac{\partial k_{i,t}}{\partial \Omega_{i,t}}$. We then need to find $\frac{\partial k_{i,t}}{\partial \Omega_{i,t}}$.

Since we assumed that Ψ is small, consider the case where $\psi = 0$. In this case, the data adjustment term in equation 25) drops out and it reduces to $k_{i,t} = \left[\frac{\alpha}{r}\left(P_tA_{i,t} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V'(\Omega_{i,t+1})\right)\right]^{1/(1-\alpha)}$, which implies

$$k^{1-\alpha}_{i,t} = \frac{\alpha}{r}\left(P_tA_{i,t} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V'(\Omega_{i,t+1})\right). \tag{48}$$

Differentiating with respect to $\Omega_{i,t}$ on both sides yields

$$\frac{\partial k^{1-\alpha}_{i,t}}{\partial \Omega_{i,t}} = \frac{\partial k^{1-\alpha}_{i,t}}{\partial k_{i,t}}\cdot\frac{\partial k_{i,t}}{\partial \Omega_{i,t}} = (1-\alpha)k^{-\alpha}_{i,t}\cdot\frac{\partial k_{i,t}}{\partial \Omega_{i,t}}$$

Differentiating (48) with respect to $\Omega_{i,t}$ and using the relationships $\frac{\partial A_{i,t}}{\partial \Omega_{i,t}} = {\Omega_{i,t}}^{-2}$ and $\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}} = \rho^2[\rho^2 + \sigma^2_\theta(\Omega_{i,t} + \sigma_a^{-2})]^{-2}$, yields

$$\frac{\partial k_{i,t}}{\partial \Omega_{i,t}} = k^{\alpha}_{i,t}\frac{\alpha}{(1-\alpha)r}\left(P_t{\Omega_{i,t}}^{-2} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_\theta(\Omega_{i,t} + \sigma_a^{-2})]^{-2}\right).$$

Therefore,

$$\begin{aligned}\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}\left(P_t{\Omega_{i,t}}^{-2} + z_i\sigma_\epsilon^{-2}\frac{1}{1+r}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_\theta(\Omega_{i,t} + \sigma_a^{-2})]^{-2}\right)\\ &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z^2_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_\theta(\Omega_{i,t} + \sigma_a^{-2})]^{-2}.\end{aligned} \tag{49}$$

Next, take the derivative of data outflow $d\Omega^-_{i,t} = \Omega_{i,t} + \sigma_a^{-2} - \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma^2_\theta\right]^{-1}$ with respect to $\Omega_{i,t}$:

$$\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t} + \sigma_a^{-2})^2(\sigma^2_\theta + \rho^{-2}(\Omega_{i,t} + \sigma_a^{-2})^{-1})^2}. \tag{50}$$

The derivatives of net data flow is then

$$\begin{aligned}\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} &= z_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} + z^2_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2 + \sigma^2_\theta(\Omega_{i,t} + \sigma_a^{-2})]^{-2}\\ &\quad + \frac{1}{\rho^2(\Omega_{i,t} + \sigma_a^{-2})^2(\sigma^2_\theta + \rho^{-2}(\Omega_{i,t} + \sigma_a^{-2})^{-1})^2} - 1.\end{aligned} \tag{51}$$

For notational convenience, denote the first term in (51) as $M_1 = z_{i,t}k^{2\alpha-1}_{i,t}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2} > 0$, the sec-

ond term as $M_2 = z_{i,t}^2 k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2} \leq 0$ and the third term as $M_3 = \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2} > 0$. Notice that $M_3 - 1 < 0$ always holds, and thus $M_2 + M_3 - 1 < 0$. $\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^-}{\partial\Omega_{i,t}} > 0$ only holds when P_t is sufficiently large so that M_1 dominates. P_t is sufficiently large when $\bar{P}$ is sufficiently large.

Assume that $V'' \in [\nu, 0)$. Let $h(\Omega_{i,t}) \equiv M_1(\bar{P}) + M_2(\nu)$. Then

$$\begin{aligned} h'(\Omega_{i,t}) &= (2\alpha-1)z_{i,t}k_{i,t}^{3\alpha-2}\alpha\left(\frac{\alpha}{r(1-\alpha)}\right)^2\sigma_\epsilon^{-2}\left[\bar{P}{\Omega_{i,t}}^{-2} + z_{i,t}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2}\right]^2 \\ &\quad + z_{i,t}k_{i,t}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{i,t}^{-3} - z_{i,t}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{(\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2}))^3}\right]. \end{aligned}$$

The first term is positive when $\alpha > \frac{1}{2}$, and negative when $\alpha < \frac{1}{2}$. And the second term is positive when $\bar{P} < f(\Omega_{i,t})$, and negative when $\bar{P} > f(\Omega_{i,t})$. To see this, note that

$$z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3}\right] > 0 \tag{52}$$

if and only if $\bar{P} < f(\Omega_{i,t})$, where

$$f(\Omega_{i,t}) := -z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\Omega_{it}^3\frac{\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3} \tag{53}$$

Notice by inspection that $f'(\Omega_{i,t}) < 0$.

Let $\hat{\Omega}$ be the first root of

$$h(\Omega_{i,t}) = 1 - M_3, \tag{54}$$

then if $\alpha < \frac{1}{2}$, when $\Omega_{i,t} < \hat{\Omega}$ and $\bar{P} > f(\hat{\Omega})$, we have that $h(\Omega_{i,t})$ is decreasing in $\Omega_{i,t}$ and $h(\Omega) \geq 1 - M_3$. Since $\nu \leq V''$, we then have $M_1 + M_2 \geq 1 - M_3$, that is $\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^-}{\partial\Omega_{i,t}} > 0$. By the same token, if $\alpha > \frac{1}{2}$ and $\bar{P} < f(\Omega_{i,t})$, then $\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^-}{\partial\Omega_{i,t}} < 0$.

Part b. Concavity at high levels of Ω_t. In this part, we first calculate limit of the derivatives of net data flow with respect to $\Omega_{i,t}$ is negative when $\Omega_{i,t}$ goes to infinity and then prove that when $\Omega_{i,t}$ is large enough, $\frac{\partial d\Omega_{i,t}}{\partial\Omega_{i,t}}$ is negative.

For $\rho \leq 1$ and $\sigma_\theta^2 \geq 0$, data outflows are bounded below by zero. But note that outflows are not bounded above. As the stock of knowledge $\Omega_{i,t} \to \infty$, outflows are of $O(\Omega_{i,t})$ and approach infinity. We have that $\frac{\partial d\Omega_{i,t}^-}{\partial\Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2}$. It is easy to see that $\lim_{\Omega_{i,t}\to\infty}\frac{\partial d\Omega_{i,t}^-}{\partial\Omega_{i,t}} = 1$.

For the derivative of data inflow (49), note that $\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} \leq z_{i,t}k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t{\Omega_{i,t}}^{-2}$ because $0 < \alpha < 1$ and $V'' < 0$. Thus $\lim_{\Omega_{i,t}\to\infty}\frac{\partial d\Omega_{i,t}^+}{\partial\Omega_{i,t}} \leq 0$.

Therefore, $\lim_{\Omega_{i,t}\to\infty}\frac{\partial d\Omega_{i,t}+}{\partial\Omega_{i,t}} - \frac{\partial d\Omega_{i,t}-}{\partial\Omega_{i,t}} \leq -1$. Since data outflows and inflows are continuously differentiable, $\exists\hat{\hat{\Omega}} > 0$ such that $\forall\Omega_{i,t} > \hat{\hat{\Omega}}$, we have $\frac{\partial d\Omega_{i,t}+}{\partial\Omega_{i,t}} - \frac{\partial d\Omega_{i,t}-}{\partial\Omega_{i,t}} < 0$, which is the decreasing returns to data when data is abundant.

A.8 Proof of Proposition 6: Firms Sell Goods at Zero Price (Data Barter)

Proof: Proving this possibility requires a proof by example. Suppose the price goods is $P_t = 0$. We want to show that an optimal production/ investment level K_t can be optimal in this environment. Consider a price of data π_t is such that firm i finds it optimal to sell a fraction $\chi > 0$ of its data produced in period t: $\delta_{i,t} = -\chi n_{i,t}$. In this case, differentiating the value function (13) with respect to k yields $(\pi_t/\iota)\chi z_i \alpha k^{\alpha-1} = r + \frac{\partial \Psi(\Delta\Omega_{i,t+1})}{\partial k_{i,t}}$. Can this optimality condition hold for positive investment level k? If $k^{1-\alpha} = \frac{\pi_t \chi z_i \alpha}{\left(r + \frac{\partial \Psi(\Delta\Omega_{i,t+1})}{\partial k_{i,t}}\right)\iota} > 0$, then the firm optimally chooses $k_{i,t} > 0$, at price $P_t = 0$. □

A.9 Proof of Lemma 3: Knowledge Gap When High Data Productivity is Scarce

When there is a single z_H firm, $\delta_L = 0$ in steady state and (k_L, Ω_L) and (P, π) are determined by the following 4 equations:

$$\Omega_L = \left(\rho^2(\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + z_L k_L^\alpha \sigma_\epsilon^{-2} \tag{55}$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha-1} + \pi\alpha z_L k_L^{\alpha-1} = r \tag{56}$$

$$P\sigma_\epsilon^{-2} k_L^\alpha = \pi\Omega_L^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) \tag{57}$$

$$P = \bar{P}\left((\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^\alpha\right)^{-\gamma} \tag{58}$$

While $(k_H, \Omega_H, \delta_H)$ are determined by the following 3 equations, taking the above (k_L, Ω_L, P, π) as given:

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha-1} + \frac{\pi\alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{59}$$

$$\iota P\sigma_\epsilon^{-2} k_H^\alpha = \pi\Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2})\right)^2}\right) \tag{60}$$

$$\Omega_H = \left(\rho^2(\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_H k_H^\alpha + \iota\delta_H\right)\sigma_\epsilon^{-2} \tag{61}$$

Manipulate to get

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \frac{\pi\alpha z_H}{\iota} = r k_H^{1-\alpha} \tag{62}$$

$$k_H^\alpha = \left(k_H^{1-\alpha}\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{\iota P\sigma_\epsilon^{-2}}\Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2})\right)^2}\right) \tag{63}$$

$$\iota^{\frac{1-2\alpha}{1-\alpha}} \frac{1}{r^{\frac{\alpha}{1-\alpha}}} \left(\iota\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \pi\alpha z_H\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}}\Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2})\right)^2}\right) \tag{64}$$

Next we show three steps:

a. For $\iota < \bar{\iota}$, more data productivity makes the "more data productive firm" (z_H firm) both larger, and retaining

more data.

$$\exists \bar{\iota} \text{ s.t. } \iota < \bar{\iota} \Rightarrow \frac{dk_H}{dz_H} > 0, \frac{d\Omega_H}{dz_H} > 0.$$

b. For $\iota < \bar{\iota}$ and $\forall z_H$, the "more data productive firm" (z_H firm) retains more data when ι increases.

$$\exists \bar{\iota} \text{ s.t. } \iota < \bar{\iota} \Rightarrow \frac{d\Omega_H}{d\iota} > 0.$$

c. $\bar{\iota} > 1$.

This completes the proof.

Part a. Take the total derivative of equation (64) wrt to z_H and simplify. It implies

$$\begin{aligned}\frac{d\Omega_H}{dz_H} &= \frac{\frac{\alpha^2\pi\iota^{2-\frac{1}{1-\alpha}}\left(\alpha\left(\iota P\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_H}\right)+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)}}{\frac{2\pi\sigma_\epsilon^2\Omega_H\left(1+r-\frac{\rho^4+\rho^2\sigma_a^2\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right)}{P} - \frac{\alpha^2 P\iota^{\frac{1}{\alpha-1}+3}\left(\alpha\left(\bar{A}\iota P-\frac{\iota P(\sigma_a^2\Omega_H+1)}{\Omega_H}+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_H^2}} \\ &= \frac{\pi\Omega_i^2 A(H)}{B(H)-\iota P A(H)}\end{aligned}$$

Note that $\iota_i\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_i}\right)+\pi z_i = \iota_i r k_i^{1-\alpha}$. Use that to simplify $\frac{d\Omega_H}{dz_H}$ by letting

$$A(i) = \frac{\alpha^2\iota_i^{2-\frac{1}{1-\alpha}}\left(\iota_i r k_i^{1-\alpha}\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2} \tag{65}$$

$$B(i) = \frac{2\pi\sigma_\epsilon^2\Omega_i\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^3}\right)}{P} = \pi\frac{dC(i)}{d\Omega_i} \tag{66}$$

$$C(i) = \frac{\sigma_\epsilon^2\Omega_i^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^2}\right)}{P} = \frac{\iota_i k_i^\alpha}{\pi} \tag{67}$$

where $i = L, H$, $\iota_L = 1$ and $\iota_H = \iota$.

In $\frac{d\Omega_H}{dz_H}$ the numerator is positive. Thus "more data productive firms retains more data", or $\frac{d\Omega_H}{dz_H} > 0$ iff the denominator is positive, which is the case if

$$\begin{aligned}&\frac{2\pi\sigma_\epsilon^2\Omega_H\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right)}{P} - \iota_i P\frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2} > 0 \\ &2\pi\sigma_\epsilon^2(1-\alpha)\Omega_H^3\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right) > \iota_i P^2\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}\end{aligned}$$

which leads to $\bar{\iota}$:

$$\bar{\iota} = \frac{2\pi\sigma_\epsilon^2(1-\alpha)\Omega_H^3\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right)}{\alpha^2 P^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}$$

Furthermore, consider equation (41). Keeping the prices constant, the left hand side is increasing in k_H. Alternatively, the derivative of the right hand side with respect to Ω_H is given by

$$2\Omega_H \left(1 + r - \frac{\rho^2(\rho^2 + \sigma_\theta^2 \sigma_a^{-2})}{\left(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2})\right)^3}\right).$$

$\frac{(\rho^2+\sigma_\theta^2\sigma_a^{-2})}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)} < 1$, thus equation (41) implies that the term in the parenthesis is positive, thus the derivative is positive. Thus Ω_H and k_H move in the same direction.

Since the high data productivity firm is atomistic, so Ω_L and k_L are unchanged. Thus the proposition also implies that surprisingly, both H-L size ratio and H-L knowledge gap of the two firms is increasing in data productivity of the more productive firm if $\iota < \bar{\iota}$:

$$\frac{d(k_H - k_L)}{dz_H} > 0, \frac{d(\Omega_H - \Omega_L)}{dz_H} > 0$$

Equation (63) implies that fixing ι, k_H moves in the same direction as Ω_H.

Part b. The proof is the same as the previous step The derivative $\frac{d\Omega_H}{d\iota}$ is more complicated but it simplifies to the exact same expression. Furthermore, let $\hat{\iota}$ denote the smallest ι for which an equilibrium exists. We have $\Omega_H(\hat{\iota}) > \Omega_L$. Since Ω_L is independent of ι, this implies that whenever an equilibrium exist, $\forall z_H$,

$$\Omega_H - \Omega_L > 0 \qquad \iota < \bar{\iota}$$

Part c. It is straight forward to show that $\bar{\iota} > 1$, i.e. the proposition holds for $\forall\ \iota \leq 1$. Note that $\iota > 1$ would mean that selling data would result in more data for the seller, which is not economonically meaningful economic. We have thus restricted $\iota \leq 1$ from the start. As such, the result holds for every ι.

A.10 Proof of Proposition 3: Negative Knowledge Gap with Non-rival Data When High Data Productivity is Abundant

The proof proceeds in a few steps. We will do the proof for $\gamma = 0$, which implies $P = \bar{P}$. Then, by continuity, the same result holds for γ sufficiently small.

Part a. $\mathbf{z_H}$ *firms are data sellers while* $\mathbf{z_L}$ *firms are data buyers (*$\delta_\mathbf{H} < \mathbf{0}$ *and* $\delta_\mathbf{L} > \mathbf{0}$*).* The marginal benefit of selling data is the same for both firms, data price π. The marginal cost of producing data is lower for the z_H firms at the same level of capital. Thus the z_H firm produces more data in equilibrium. Furthermore, recall that each firm can only buy or sell data.

Now assume that in equilibrium $\sigma_L < 0$. This means that the H firm prefers to buy the last unit of data rather than to produce it, while the L firm prefers to produce it and sell it. This would imply that the marginal benefit of selling data is larger than marginal cost of producing it for a small firm, but smaller than marginal cost of its production for a large firm, a contradiction.

Part b. $\frac{\mathbf{d}\pi}{\mathbf{z_H}} < \mathbf{0}$. Step 1 shows that the H firm is always the data seller. Thus higher z_H corresponds to an upward shift of supply curve, which in turn implies a lower dat aprice.

Part c. Negative knowledge gap: $\exists \overline{\iota} \mid \iota \leq \overline{\iota} \Rightarrow \mathbf{\Upsilon^{ss}} < \mathbf{0}$. Merge equations (38)-(41) and use $P = \bar{P}$ to write Ω_H and Ω_L:

$$\frac{\pi}{\bar{P}\sigma_\epsilon^{-2}}\Omega_L^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right) = \frac{1}{r^{\frac{\alpha}{1-\alpha}}}\left(\alpha\bar{P}(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi\alpha z_L\right)^{\frac{\alpha}{1-\alpha}} \tag{68}$$

$$\frac{\pi}{\bar{P}\sigma_\epsilon^{-2}}\Omega_H^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)^2}\right) = \frac{\iota^{\frac{1-2\alpha}{1-\alpha}}}{r^{\frac{\alpha}{1-\alpha}}}\left(\iota\alpha\bar{P}(\bar{A}-\Omega_H^{-1}-\sigma_a^2)+\pi\alpha z_H\right)^{\frac{\alpha}{1-\alpha}}. \tag{69}$$

Consider equation (69). Since $\alpha < \frac{1}{2}$, $\iota \to 0$ implies that the first term on the right hand since goes to zero. Every other term in the left and right hand side of the equation is finite and bounded away from zero, except Ω_H^2, so $\Omega_H \to 0$. By continuity, as ι gets small, keeping everything else constant Ω_H has to decline while there is no effect in equation (68) on Ω_L. Thus $\exists \overline{\iota}$ such that $\iota \leq \overline{\iota} \Rightarrow \Upsilon^{ss} < 0$.

A.11 Proof of Corollary 1: Change in Knowledge Gap $\frac{\mathbf{d\Upsilon^{ss}}}{\mathbf{dz_H}}$

Similar to proof of Proposition 3, we do the proof for $\gamma = 0$, which implies $P = \bar{P}$. Then, by continuity, the same result holds for γ sufficiently small.

Equations (38) and (40) can be solved to get (k_L, Ω_L) in terms of data price π:

$$k_L^{1-\alpha} = \frac{\alpha}{r}\left(\bar{P}(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi z_L\right)$$

$$\Omega_L^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right)\frac{1}{\left(\bar{P}(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi z_L\right)^{\frac{\alpha}{1-\alpha}}} = \frac{\bar{P}\sigma_\epsilon^{-2}}{\pi}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}}$$

The second equation implies

$$\Omega_L^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right) = \frac{\left(\bar{P}(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi z_L\right)^{\frac{\alpha}{1-\alpha}}}{\pi}\bar{P}\sigma_\epsilon^{-2}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}} \tag{70}$$

The same argument as in proposition 3 shows that using equation (40), the derivative of the left hand side with respect to Ω_L is positive. Next, using implicit function theorem on both sides of equation (70) implies that if $\alpha \leq \frac{1}{2}$, the equation is only consistent with $\frac{d\Omega_L}{d\pi} < 0$. Note that $\alpha \leq \frac{1}{2}$ is a sufficient (not necessary) condition. As such, $\pi \downarrow \Leftrightarrow \Omega_L \uparrow$. Using this in the first equation implies k_L increases as well, $k_L \uparrow$.

Next, merge equations ((38), (40)) and ((39), (41)) to get:

$$\frac{1}{r^{\frac{\alpha}{1-\alpha}}}\left(\alpha P(\bar{A}-\Omega_L^{-1}-\sigma_a^2)+\pi\alpha z_L\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}}\Omega_L^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_L+\sigma_a^{-2})\right)^2}\right) \tag{71}$$

$$\frac{\iota^{\frac{1-2\alpha}{1-\alpha}}}{r^{\frac{\alpha}{1-\alpha}}}\left(\iota\alpha P(\bar{A}-\Omega_H^{-1}-\sigma_a^2)+\pi\alpha z_H\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}}\Omega_H^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\Omega_H+\sigma_a^{-2})\right)^2}\right). \tag{72}$$

Again, $\gamma = 0$ implies $P = \bar{P}$, thus taking the derivatives we have

$$\frac{d\Omega_L}{dz_H} = \frac{-\frac{\sigma_\epsilon^2 \Omega_L^2 \left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_L)\right)^2}\right)}{\bar{P}} + \frac{z_L \alpha^2 \left(\alpha\left(\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_L}\right)+\pi z_L\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)}}{\frac{2\pi\sigma_\epsilon^2\Omega_L\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_L)\right)^3}\right)}{\bar{P}} - \frac{\bar{P}\alpha^2\left(\alpha\left(\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_L}\right)+\pi z_L\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_L^2}} \frac{d\pi}{dz_H}$$

$$\frac{d\Omega_H}{dz_H} = \frac{\frac{\pi\alpha^2\iota^{2-\frac{1}{1-\alpha}}\left(\alpha\left(\iota\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_H}\right)+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)}}{\frac{2\pi\sigma_\epsilon^2\Omega_H\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right)}{\bar{P}} - \frac{\bar{P}\alpha^2\iota^{3-\frac{1}{1-\alpha}}\left(\alpha\left(\iota\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_H}\right)+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_H^2}}$$

$$+ \frac{-\frac{\sigma_\epsilon^2\Omega_H^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^2}\right)}{\bar{P}} + \frac{z_H\alpha^2\iota^{2-\frac{1}{1-\alpha}}\left(\alpha\left(\iota\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_H}\right)+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)}}{\frac{2\pi\sigma_\epsilon^2\Omega_H\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_H)\right)^3}\right)}{\bar{P}} - \frac{\bar{P}\alpha^2\iota^{3-\frac{1}{1-\alpha}}\left(\alpha\left(\iota\bar{P}\left(\bar{A}-\sigma_a^2-\frac{1}{\Omega_H}\right)+\pi z_H\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_H^2}} \frac{d\pi}{dz_H}$$

Using definition (65)-(67) the above expressions simplify to:

$$\frac{d\Omega_L}{dz_H} = \frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)}\frac{d\pi}{dz_H}$$
$$\frac{d\Omega_H}{dz_H} = \frac{\pi\Omega_H^2 A(H)}{B(H) - \iota\bar{P}A(H)} + \frac{z_H\Omega_H^2 A(H) - C(H)}{B(H) - \iota\bar{P}A(H)}\frac{d\pi}{dz_H}$$

Thus the derivative of the knowledge gap is given by

$$\frac{d\Upsilon}{dz_H} = \frac{d(\Omega_H - \Omega_L)}{dz_H} = \frac{\pi\Omega_H^2 A(H)}{B(H) - \iota\bar{P}A(H)} + \left(\frac{z_H\Omega_H^2 A(H) - C(H)}{B(H) - \iota\bar{P}A(H)} - \frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)}\right)\frac{d\pi}{dz_H}$$

We have already shown that $\frac{d\pi}{dz_H} < 0$. Using that, we first show that fixing the parameters, $\exists\hat{\iota}$ such that if and only if $\iota > \hat{\iota}$, the knowledge gap is increasing in z_H.

$$\exists\hat{\iota} \text{ s.t. } \iota > \hat{\iota} \Leftrightarrow \frac{d\Upsilon}{dz_H} > 0.$$

Note that

$$\frac{d(\Omega_H - \Omega_L)}{dz_H} = \frac{\pi\Omega_H^2 A(H)}{B(H) - \iota\bar{P}A(H)} + \left(\frac{z_H\Omega_H^2 A(H) - C(H)}{B(H) - \iota\bar{P}A(H)} - \frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)}\right)\frac{d\pi}{dz_H} \quad \Longrightarrow$$
$$\frac{d(\Omega_H - \Omega_L)}{dz_H} > 0 \Leftrightarrow \frac{\pi\Omega_H^2 A(H) + \left(z_H\Omega_H^2 A(H) - C(H)\right)\frac{d\pi}{dz_H}}{B(H) - \iota\bar{P}A(H)} > \frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)}\frac{d\pi}{dz_H}$$

Multiply both sides by the denominator on the left hand side, which is positive as $\iota < 1$. Divide both sides by the right hand side expression which is also positive. Since both expressions are positive, the inequality sign does not

change

$$\frac{\pi\Omega_H^2 A(H) + \left(z_H\Omega_H^2 A(H) - C(H)\right)\frac{d\pi}{dz_H}}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L)-\bar{P}A(L)}\frac{d\pi}{dz_H}} > B(H) - \iota\bar{P}A(H)$$

$$\iota > \frac{1}{\bar{P}A(H)}\left(B(H) - \frac{\pi\Omega_H^2 A(H) + \left(z_H\Omega_H^2 A(H) - C(H)\right)\frac{d\pi}{dz_H}}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L)-\bar{P}A(L)}\frac{d\pi}{dz_H}}\right)$$

$$A(i) = \frac{\alpha^{\frac{1}{1-\alpha}}\iota_i^{2-\frac{1}{1-\alpha}}\left(\left(\iota_i\bar{P}\left(\bar{A} - \sigma_a^2 - \frac{1}{\Omega_i}\right) + \pi z_i\right)\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2}$$

$$\iota > \frac{1}{\bar{P}\frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2}}\left(B(H) + \frac{C(H)}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L)-\bar{P}A(L)}}\right)$$

$$\iota > \bar{\iota}_{1H} = \left(\frac{(1-\alpha)\Omega_i^2}{\bar{P}\alpha^{\frac{1}{1-\alpha}}(\pi z_i)^{\frac{1}{1-\alpha}-2}}\left(B(H) + \frac{C(H)}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L)-\bar{P}A(L)}}\right)\right)\frac{1-\alpha}{2-3\alpha}$$

Since $\alpha < \frac{1}{2}$, and Ω_i and k_i, $i = L, H$ are finite, sufficiently large z_H insures $\bar{\iota}_{1H} < 1$.

A.12 Proposition 7: Accumulation Can be Purely Concave

It turns out that data accumulation is not always S-shaped. The S-shaped results in the previous proposition hold only for some parameter values. For others, it can be that data accumulation is purely concave. In other words, even when $\Omega_{i,t}$ is small enough, there is no convex region. Instead, the net data flow (the slope) decreases with $\Omega_{i,t}$, right from the start.

Proposition 7 ***Concavity of Data Inflow*** *$\exists\epsilon > 0$ such that $\forall\Omega_{i,t} \in B_\epsilon(0)$, the net data flow decreases with $\Omega_{i,t}$ if $\sigma_\theta^2 > \sigma_a^2$.*

We proceed in two steps. In Step 1, we prove that data outflows are approximately linear when $\Omega_{i,t}$ is small. And then in Step 2, we first calculate the derivative of net data flow with respect to $\Omega_{i,t}$ and then characterize the parameter region where it is negative.

Step 1: Data outflows are approximately linear when $\Omega_{i,t}$ is small.

This is proven separately in Lemma 4.

Step 2: Characterize the parameter region where the derivative of net data flow with respect to $\Omega_{i,t}$ is negative. A negative least upper bound is sufficient for it be negative.

Recall that the derivative of data inflows with respect to the current stock of knowledge Ω_t is

$\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} = \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} > 0$ (see the Proof of Proposition 5 for details). Thus

$$\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} \approx \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} - 1 + \rho^2(1 + \rho^2\sigma_\theta^2\sigma_a^{-2})^{-2}. \tag{73}$$

Since this derivative increases in ρ^2 and decreases in $\Omega_{i,t} = 0$, so its least upper bound $\frac{2}{1+\sigma_\theta^2\sigma_a^{-2}} - 1$ is achieved when $\rho^2 = 1$ and $\Omega_{i,t} = 0$. A non-negative least upper bound requires $\sigma_a^2 \geq \sigma_\theta^2$. That means, if $\sigma_\theta^2 > \sigma_a^2$, the supreme of $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}} - \frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}}$ is negative, so it will always be negative $\forall \Omega_{i,t} \in B_\epsilon(0)$.

A.13 Proposition 8: S-shaped Stock of Knowledge Over Time

This proposition shows the S-shape of $\Omega_{i,t}$ in the time domain. The result differs from Proposition 5 because instead of establishing that data flows are increasing or decreasing in Ω, this result establishes that flows increase and then decrease in time.

Proposition 8 *Let $\zeta \equiv \frac{\sigma_\theta^2}{\sigma_a^2}$. If ζ is sufficiently small, σ_ϵ sufficiently small, z_i sufficiently large, or σ_a sufficiently large, then*

1) *$\frac{\partial^2 \Omega_{i,t}}{\partial t^2} > 0$ when $\Omega_{i,t}$ is small enough, there is sufficient diminishing return to scale $\alpha < \frac{1}{2}$, price is sufficiently large $P_t > f(\hat{\Omega})$ and the value function is not too concave $V'' \in [\nu, 0)$, where $\hat{\Omega}$ is the first root of (54), and f is defined by (53);*

2) *and $\frac{\partial^2 \Omega_{i,t}}{\partial t^2} < 0$ when $\Omega_{i,t}$ is large enough.*

Proof: We have established how net data flows change with $\Omega_{i,t}$. To map it to concavity and convexity of $\Omega_{i,t}$ with respect to t, we need to find the regions where net data flow, $d\Omega^+_{i,t} - d\Omega^-_{i,t} = z_{i,t}k_{i,t}^\alpha\sigma_\epsilon^{-2} - \Omega_{i,t} - \sigma_a^{-2} + \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}$ is positive. If $d\Omega^+_{i,t} - d\Omega^-_{i,t} > 0$, then $\frac{\partial d\Omega^+_{i,t} - d\Omega^-_{i,t}}{\partial \Omega_{i,t}} > 0$ maps to $\frac{\partial^2 \Omega_{i,t}}{\partial t^2} > 0$ and $\frac{\partial d\Omega^+_{i,t} - d\Omega^-_{i,t}}{\partial \Omega_{i,t}} < 0$ maps to $\frac{\partial^2 \Omega_{i,t}}{\partial t^2} < 0$. The rest of the proof proceeds in two steps.

Step 1: Prove that net data flows are positive when $\Omega_{i,t} \in (0, \bar{\Omega}_f)$ and negative when $\Omega_{i,t} \in (\bar{\Omega}_l, \infty)$.

We can sign the second derivative of outflows with respect to $\Omega_{i,t}$ easily: $\frac{\partial^2 d\Omega^-_{i,t}}{\partial \Omega_{i,t}^2} = \frac{2\rho^4\sigma_\theta^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2}+\rho^2\sigma_\theta^2\Omega_{i,t})^3} > 0$. The first derivative of outflows with respect to $\Omega_{i,t}$ is $\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t}+\sigma_a^{-2})^2(\sigma_\theta^2+\rho^{-2}(\Omega_{i,t}+\sigma_a^{-2})^{-1})^2}$. Since $\frac{\partial^2 d\Omega^-_{i,t}}{\partial \Omega_{i,t}^2} > 0$, we have that $\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}}$ is monotonically increasing and its minimum value is obtained when $\Omega_{i,t} = 0$: $\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}}|_{\Omega_{i,t}=0} = 1 - \frac{1}{\rho^2\sigma_a^{-4}(\sigma_\theta^2+\rho^{-2}\sigma_a^2)^{-1})^2}$, which is always positive given $\sigma_a, \sigma_\theta > 0$ and $\rho^2 \leq 1$. Therefore $\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} > 0$. On the other hand, we know from the proof of Proposition 8 that $\frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}}|_{\Omega_{i,t}=0} > 0$ and $lim_{\Omega_{i,t}\to\infty} \frac{\partial d\Omega^+_{i,t}}{\partial \Omega_{i,t}}|_{\Omega_{i,t}=0} \leq 0$.

When $\Omega_{i,t} = 0$, the data infow and outflow are $d\Omega^+_{i,t}|_{\Omega_{i,t}=0} = z_i(k_0^*)^\alpha\sigma_\epsilon^{-2}$, where k_0^* is the optimal invesment when $\Omega_{i,t} = 0$, and $d\Omega^-_{i,t}|_{\Omega_{i,t}=0} = \sigma_a^{-2}(1 - \frac{\rho^2}{1+\rho^2\zeta})$, respectively. If $d\Omega^+_{i,t}|_{\Omega_{i,t}=0} \geq d\Omega^-_{i,t}|_{\Omega_{i,t}=0}$, then the data outflow and inflow curves must have intersection(s) in the region $(0, \infty)$. Let's denote the first intersection by $\bar{\Omega}_f$ and the last by $\bar{\Omega}_l$. When there is a unique intersection, $\bar{\Omega}_f$ and $\bar{\Omega}_l$ coincide. Then net data flows are positive when $\Omega_{i,t} \in (0, \bar{\Omega}_f)$ and negative when $\Omega_{i,t} \in (\bar{\Omega}_l, \infty)$.

Step 2: Find out the parameter regions where Step 1 holds.

Since $d\Omega^-_{i,t}|_{\Omega_{i,t}=0}$ is monotonic in ρ^2 and $\rho^2 \in [0,1]$, we have $d\Omega^-_{i,t}|_{\Omega_{i,t}=0} \in [\frac{\zeta}{1+\zeta}\sigma_a^{-2}, \sigma_a^{-2}]$. $d\Omega^+_{i,t}|_{\Omega_{i,t}=0} \geq d\Omega^-_{i,t}|_{\Omega_{i,t}=0}$ is guaranteed when $d\Omega^+_{i,t}|_{\Omega_{i,t}=0} \geq \frac{\zeta}{1+\zeta}\sigma_a^{-2}$, that is $k_0^* \geq \frac{\sigma_\epsilon^2}{z_i\sigma_a^2}\frac{\zeta}{1+\zeta}$. The last inequality is only true when ζ is sufficiently small, σ_ϵ sufficiently small, z_i sufficiently large, or σ_a sufficiently large.

We can then apply Proposition 8 and map concavity and convexity of $\Omega_{i,t}$ to the time domain.

A.14 Lemma 4, 5, 6: Linearity of Data Depreciation

One property of the model that comes up in a few different places is that the depreciation of knowledge (outflows) is approximately a linear function of the stock of knowledge $\Omega_{i,t}$. There are a few different ways to establish this approximation formally. The three results that follow show that the approximation error from a linear function is small i) when the stock of knowledge is small; ii) when the state is not very volatile; and iii) when the stock of knowledge is large.

Lemma 4 ***Linear Data Outflow with Low Knowledge*** *$\exists \epsilon > 0$ such that $\forall \Omega_{i,t} \in B_\epsilon(0)$, data outflow is approximately linear and the approximation error is bounded from above by $\frac{\rho^4\sigma_\theta^2}{1+\rho^2\sigma_\theta^2\sigma_a^{-2}}\frac{\epsilon^2}{1+\rho^2\sigma_\theta^2(\epsilon+\sigma_a^{-2})}$. The approximation error is small when ρ or σ_θ is small, or when $\Omega_{i,t}$ is very close to 0.*

Proof:
Recall that data outflows are $d\Omega^-_{i,t} = \Omega_{i,t}+\sigma_a^{-2}-\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$. Let $g(\Omega_{i,t}) \equiv \left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$ be the nonlinear part of data outflows. Its first-order Taylor expansion around 0 is $g(\Omega_{i,t}) = g(0)+g'(0)\Omega_{i,t}+o(\Omega_{i,t})$, with $g'(0) = \frac{\rho^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})^2}$. Thus $\frac{\partial d\Omega^-_{i,t}}{\partial \Omega_{i,t}} = 1-g'(\Omega_{i,t}) \approx 1-g'(0)$ for $\Omega_{i,t}$ in a small open ball $B_\epsilon(0)$, $\epsilon > 0$, around 0. And the approximation error is $|o(\Omega_{i,t})| = \frac{\rho^4\sigma_\theta^2\Omega_{i,t}^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})[1+\rho^2\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]}$, which increases with $\Omega_{i,t}$ and thus is bounded from above by error term evaluated at ϵ, that is $\frac{\rho^4\sigma_\theta^2}{1+\rho^2\sigma_\theta^2\sigma_a^{-2}}\frac{\epsilon^2}{1+\rho^2\sigma_\theta^2(\epsilon+\sigma_a^{-2})}$.

Lemma 5 ***Linear Data Outflow with Small State Innovations*** *$\exists \epsilon_\sigma > 0$ such that $\forall \sigma_\theta \in B_{\epsilon_\sigma}(0)$, data outflows are approximately linear and the approximation error is bounded from above by $\frac{\rho^4\tilde{\epsilon}^2(\Omega_{i,t}+\sigma_a^{-2})^2}{1+\rho^2\epsilon_\sigma^2(\Omega_{i,t}+\sigma_a^{-2})}$. The approximation error is small when ρ is small, or when σ_θ is close to 0.*

Proof:
Recall that data outflows are $d\Omega^-_{i,t} = \Omega_{i,t}+\sigma_a^{-2}-\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$. The non-linear term $g(\Omega_{i,t}) = [(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2]^{-1}$ is linear when $\sigma_\theta = 0$. Therefore, $\exists \epsilon_\sigma > 0$ such that $\forall \sigma_\theta \in B_{\epsilon_\sigma}(0)$, $g(\Omega_{i,t})$ is approximately linear. The approximation error $|g(\Omega_{i,t})-\rho^2(\Omega_{i,t}+\sigma_a^{-2})|$ is increasing with ϵ_σ and reaches its maximum value at $\sigma_\theta = \epsilon_\sigma$, with value $\frac{\rho^4\epsilon_\sigma^2(\Omega_{i,t}+\sigma_a^{-2})^2}{1+\rho^2\tilde{\epsilon}^2(\Omega_{i,t}+\sigma_a^{-2})}$.

Lemma 6 ***Linear Data Outflow with Abundant Knowledge*** *When $\Omega_{i,t} \gg \sigma_\theta^{-2}$, discounted data stock is very small relative to $\Omega_{i,t}$, so that data outflows are approximately linear. The approximation error is small when ρ is small or when σ_θ is sufficiently large.*

Proof:

Recall that data outflows are $d\Omega_{i,t}^{-} = \Omega_{i,t} + \sigma_a^{-2} - \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}$. Let $g(\Omega_{i,t}) \equiv \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}$ be the nonlinear part of data outflows. Since $(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} \geq 0$, we have $g(\Omega_{i,t}) \leq \sigma_\theta^{-2}$. Since $\Omega_{i,t} \geq 0$, we have $g(\Omega_{i,t}) \geq (\rho^{-2}\sigma_a^2 + \sigma_\theta^2)^{-1}$. That is $g(\Omega_{i,t}) \in [(\rho^{-2}\sigma_a^2 + \sigma_\theta^2)^{-1}, \sigma_\theta^{-2}]$. For high levels of $\Omega_{i,t}$, $\Omega_{i,t} \gg \sigma_\theta^{-2}$ generally holds. And for low levels of $\Omega_{i,t}$, it holds when σ_θ is very large. The approximation error is $|\sigma_\theta^{-2} - \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}|$ and decreases with $\Omega_{i,t}$, reaching its minimum at $\Omega_{i,t} = 0$ with a value of $\frac{\rho^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})^2}$.

A.15 Proof of Proposition 4 Welfare

We begin by characterizing competitive equilibrium. Then we characterize the solution to the social planner problem. Finally, we compare the two solutions to determine the efficiency of the equilibrium outcome.

Competitive Equilibrium

Definition 4 *The competitive equilibrium consists of household optimal choice of demand for retail and numeraire good, and labor supply, $\{c_t, m_t, l_t\}$, numeraire good producer optimal choice of supply of numeraire good and labor demand, $\{m_t, l_t\}$, retail good producer i's optimal choice of capital and traded data, $\{k_{i,t}, \delta_{i,t}\}$, $\forall$ i, and price of retail good, P_t, data, π_t, and wages, w_t , in terms of the numeraire good, for each period t, such that the market for retail goods, numeraire good, data, and labor clears.*

First order conditions.

Let Γ_t denote the Lagrange multiplier on the household budget constraint, multiplied by $(1 + r)$. Then we can write the household's first order conditions with respect to c_t, m_t, and l_t are given by:

$$\begin{aligned} c_t: &\quad u'(c_t) = P_t\Gamma_t, \\ m_t: &\quad 1 = \Gamma_t, \\ l_t: &\quad v'(l_t) = \Gamma_t w_t, \end{aligned}$$

which together imply

$$P_t = u'(c_t) \tag{74}$$

$$w_t = v'(l_t). \tag{75}$$

The first order condition for the numeraire good producers pins down the wage to be a constant equal to the marginal product of labor,

$$w_t = \bar{m}.$$

The sequential optimization problem of retail producers takes a recursive form,

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t\mathbb{E}[A_{i,t}]k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - rk_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r}.$$

Note that in the steady state, $\Delta\Omega_{i,t+1} = 0$. Furthermore, for the representative firm, net data trade is also zero, $\delta_{i,t}$.

Market clearing conditions. The market clearing conditions for retail goods and data are given by

$$\begin{aligned} c_t &= \lambda A_{L,t}k_{L,t}^{\alpha} + (1-\lambda)A_{H,t}k_{H,t}^{\alpha}, \\ \lambda\delta_{L,t} &+ (1-\lambda)\delta_{H,t} = 0. \end{aligned}$$

We solve for the competitive equilibrium in the following steps.

Optimal retail good consumption. The first order condition for retail consumption tells us that price equals the marginal utility,

$$P_t = u'(Y_t) = \bar{P}Y_t^{-\gamma}.$$

Data market optimum. The optimal choice of data purchases or sales for a retail firm equates the marginal cost of data π with the future discounted value of data. Thus absent adjustment cost of data, which is the relevant case in the steady state,

$$\begin{aligned} \pi_t &= \frac{1}{1+r}V'(\Omega_{i,t+1})\sigma_{\varepsilon}^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \tau\mathbb{1}_{\delta_{i,t}<0}) \\ \implies V'(\Omega_{i,t+1}) &= \frac{(1+r)\pi_t}{\sigma_{\varepsilon}^{-2}(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})} \end{aligned} \tag{76}$$

Optimal capital choices. Take the derivative of the Bellman equation for retail firm producers with respect to $k_{i,t}$

$$P_t\mathbb{E}[A_{i,t}]\alpha k_{i,t}^{\alpha-1} - \Psi'(\Delta\Omega_{i,t+1})\sigma_{\varepsilon}^{-2}z_i\alpha k_{i,t}^{\alpha-1} - r + \frac{1}{1+r}V'(\Omega_{i,t+1})\sigma_{\varepsilon}^{-2}z_i\alpha k_{i,t}^{\alpha-1} = 0$$

Substitute equation (76) to get

$$\left[\alpha P_t\mathbb{E}[A_{i,t}] - \alpha z_i\Psi'(\Delta\Omega_{i,t+1})\sigma_{\varepsilon}^{-2} + \frac{\pi\alpha z_i}{(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})}\right]k_{i,t}^{\alpha-1} = r,$$

With the specific form of $g(.)$, in the steady state we have,

$$\alpha\left[P_t(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2) + \frac{\pi z_i}{(\mathbb{1}_{\delta_{i,t}>0} + \iota\mathbb{1}_{\delta_{i,t}<0})}\right]k_{i,t}^{\alpha-1} = r. \tag{77}$$

Note that the term $\Psi'(\Delta\Omega_{i,t+1}) = 0$ because the stock of knowledge does not change in the steady state.

Next, differentiate the value function with respect to $\Omega_{i,t}$ and use the envelope condition to hold the choice

variables constant,

$$V'(\Omega_{i,t}) = P_t k_{i,t}^{\alpha} \Omega_{i,t}^{-2} - \underbrace{\Psi'(\Delta\Omega_{i,t+1})\left(\frac{\partial\Omega_{i,t+1}}{\partial\Omega_{i,t}} - 1\right)}_{=0 \text{ in SS}} + \frac{1}{1+r} V'(\Omega_{i,t+1}) \frac{\partial\Omega_{i,t+1}}{\Omega_{i,t}}, \tag{78}$$

Equate equations (76) and (78) for $V'(\Omega_{i,t})$ to get:

$$\frac{\partial\Omega_{i,t+1}}{\Omega_{i,t}} = \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right)^2}$$

The rest of the derivation is the same as we have done for optimality conditions (40) and (41).

We can rewrite conditions (38) and (39) as:

$$r = \alpha k_{i,t}^{\alpha-1} \left[P_t \underbrace{(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2)}_{\mathbb{E}[A_{i,t}]} + \frac{z_i \sigma_\varepsilon^{-2}}{1+r} P_{t+1} k_{i,t+1}^{\alpha} \Omega_{i,t+1}^{-2} \right] \tag{79}$$

First, note that equations (77) and (79) are both in steady state, and are the same. The only difference is that the first one uses the recursive formulations while the second one uses the non-recursive formulation. Thus, they are not two independent equations.

We will use equation (79) to derive one of the independent equilibrium optimality conditions. Removing the prices from this optimality conditions facilitates the comparison with the planner problem. Thus use the household first order condition $P_t = u'(Y_t) = \bar{P} Y_t^{-\gamma}$ to replace P_t,

$$r = \alpha k_{i,t}^{\alpha-1} \left[u'(c_t)\mathbb{E}[A_{i,t}] + \frac{z_i \sigma_\varepsilon^{-2}}{1+r} u'(c_{t+1}) k_{i,t+1}^{\alpha} \Omega_{i,t+1}^{-2} \right] \tag{80}$$

Finally, to get the last optimality condition in equilibrium, use equation (76) for $V'(\Omega_{i,t}$ in (78). Then assume we are in steady state and solve for $K_{i,t}^{\alpha}\Omega_{i,t}^{-2}$ for the L and H firms. Then, taking the ratio of the two allows us to cancel out P_t or $u'(c_t)$, so in *steady state* we have

$$\frac{\frac{K_{L,t}^{\alpha}}{\Omega_{L,t}^2}}{\frac{\iota k_{H,t}^{\alpha}}{\Omega_{H,t}^2}} = \frac{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{L,t} + \sigma_a^{-2}))^2}}{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{H,t} + \sigma_a^{-2}))^2}} \tag{81}$$

Steady state with two types of firms There are 7 equations in 7 unknowns that constitute the solution to the decentralized equilibrium. The 7 unknowns are Ω_L, Ω_H, k_L, k_H, δ_L, δ_H, and l. Two of the 7 equations are the law of motion for the stock of knowledge for the high and low retail firms (10). The third and fourth equations are the optimal choice of capital in equilibrium (80) for high and low retail firms. The fifth equation is (81). The sixth equation is the data market clearing condition. The last equation determines the equilibrium amount of labor, $v'(l_t) = \bar{m}$.

Steady state with one type of firm When there is a single type of firm i, data market clearing condition implies that $\delta_{i,t} = 0$. Thus it will be 3 equations in 3 unknowns that constitute the solution to the decentralized

equilibrium. The 3 unknowns are $\Omega_{\imath,t}, k_{i,t}$, and l_t. One equation is the law of motion for the stock of knowledge for type i retail firms (10), with $\delta_{i,t} = 0$. The second equation is the optimal choice of capital in equilibrium (80) for i firms. The last equation determines the equilibrium amount of labor, $v'(l_t) = \bar{m}$.

Optimum. Social Planner Problem

As all firm profits are redistributed to consumers, the social planner maximizes aggregate household utility,

$$\max \sum_{t=0}^{\infty} \frac{1}{(1+r)^t} \left(u(c_t) + m_t - v(l_t) \right).$$

Use the numeraire good resource constraint, $\bar{m} l_t = m_t + r k_t$, to rewrite the problem of the optimum in recursive form,

$$V^P(\Omega_{L,t}, \Omega_{H,t}) = \max u(c_t) + \bar{m} l_t - r k_t - v(l_t) + \frac{1}{1+r} V^P(\Omega_{L,t+1}, \Omega_{H,t+1}),$$

subject to two resource constraints in each period, one for the retail good and one for traded data

$$c_t = \lambda A_{L,t} k_{i,t}^{\alpha} + (1-\lambda) A_{H,t} k_{i,t}^{\alpha}$$
$$\lambda \delta_{L,t} + (1-\lambda)\delta_{H,t} = 0$$

Let Xi_t and η_t denote the Lagrangian multipliers on the retail good the data recourse constraints, multiplied by $(1+r)$, respectively. The optimum is characterized by first order conditions with respect to $k_{L,t}, k_{H,t}, \delta_{L,t}, \delta_{H,t}$, l_t, as well as the laws of motion for $\Omega_{L,t}$ and $\Omega_{H,t}$. We describe each in turn.

Let λ_i denote the measure of firms of type i in the economy, i.e. $\lambda_L = \lambda$ and $\lambda_H = 1 - \lambda$. The first order condition with respect to $k_{i,t}$ is

$$r = \lambda_i \alpha k_{i,t}^{\alpha - 1} \left[u'(c_t) \mathbb{E}[A_{i,t}] + \frac{z_i \sigma_{\varepsilon}^{-2}}{1+r} u'(c_{t+1}) k_{i,t+1}^{\alpha} \Omega_{i,t+1}^{-2} \right] \tag{82}$$

Alternatively, the planner's first order conditions with respect to consumption choice and the data choices $\delta_{i,t}$ are given by

$$\Xi_t = u'(c_t)$$
$$\eta_t = \frac{1}{1+r} V^{P'}(\Omega_{L,t+1}, \Omega_{H,t+1}) \sigma_{\varepsilon}^{-2} (\mathbb{1}_{\delta_{i,t}>0} + \iota \mathbb{1}_{\delta_{i,t}<0}).$$

Let V_i^P denote the derivative of the social planner value function with respect to Ω_i, $i = L, H$. To solve for V_i^P, differentiate the value function and apply the envelope condition to get:

$$\begin{aligned} V_i^P(\Omega_{i,t}, \Omega_{-i,t}) =& u'(c_t) k_{i,t}^{\alpha} \Omega_{i,t}^{-2} \lambda_i - \overbrace{\lambda_i \Psi'(\Delta \Omega_{i,t+1}) \left(\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}}, 1 \right)}^{=0 \text{ in SS}} \\ &+ \frac{1}{1+r} V_i^{P'}(\Omega_{i,t+1}, \Omega_{-i,t+1}) \frac{\rho^2}{(\sigma_{\theta}^2 + \rho^2(\Omega_{i,t} + \sigma_a^{-2}))^2}. \end{aligned}$$

Equating $V_i^p(\Omega_{i,t}, \Omega_{-i,t})$ and $V_i^P(\Omega_{i,t+1}, \Omega_{-i,t+!})$ in the *steady state* gives

$$\overset{SS}{\Rightarrow} \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2}))^2}\right) = \frac{\lambda_i u'(c_t) k_{i,t}^\alpha \Omega_{i,t}^{-2}(1+r)}{\frac{\eta(1+r)}{(\mathbb{1}_{\delta_{i,t}>0} + \iota \mathbb{1}_{\delta_{i,t}<0})}} \sigma_\epsilon^{-2}$$

Rearrange to get

$$1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2}))^2} = \frac{\lambda_i u'(c_t) k_{i,t}^\alpha (\mathbb{1}_{\delta_{i,t}>0} + \iota \mathbb{1}_{\delta_{i,t}<0})}{\eta \Omega_{i,t}^2} \sigma_\varepsilon^{-2}.$$

This is one condition for each type of firm H and L. We know that H firms sell data, $\delta_{H,t} < 0$. Thus we have $(\mathbb{1}_{\delta_{H,t}>0} + \iota \mathbb{1}_{\delta_{H,t}<0}) = \iota$. Alternatively, L firmd buy data, $\delta_{L,t} > 0$. Thus, $(\mathbb{1}_{\delta_{L,t}>0} + \iota \mathbb{1}_{\delta_{L,t}<0}) = 1$.

As with the decentralized problem, take the ratio of the H and L conditions. Thus the marginal value $u'()$ and the Lagrange multiplier η both drop out of the resulting equation and in the *steady state*

$$\frac{\frac{k_{L,t}^\alpha}{\Omega_{L,t}^2}}{\frac{\iota K_{H,t}^\alpha}{\Omega_{H,t}^2}} = \frac{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{L,t} + \sigma_a^{-2}))^2}}{1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_{H,t} + \sigma_a^{-2}))^2}} \tag{83}$$

Finally, the first order condition with respect to labor is

$$v'(l_t) = \bar{m}. \tag{84}$$

Steady state with two types of firms There are 7 equations in 7 unknowns that constitute the solution to the optimum. The 7 unknowns are Ω_L, Ω_H, k_L, k_H, δ_L, δ_H, and l. Two of the 7 equations are the law of motion for the stock of knowledge for the high and low retail firms (10). The third and fourth equations are the optimal choice of capital in optimum (82) for high and low retail firms. The fifth equation is (83). The sixth equation is the data resource constraint condition. The last equation determines the optimal choice of labor, equation (84).

Steady state with one type of firm When there is a single type of firm i, data resource constraint condition implies that $\delta_{i,t} = 0$. Thus it will be 3 equations in 3 unknowns that constitute the solution to the optimum. The 3 unknowns are $\Omega_{\imath,t}, k_{i,t}$, and l_t. One equation is the law of motion for the stock of knowledge for type i retail firms (10), with $\delta_{i,t} = 0$. The second equation is the optimal choice of capital in optimum (82) for i firms with $\lambda_i = 1$. The last equation determines the optimal choice of labor, $v'(l_t) = \bar{m}$.

The main result here is that the optimality conditions for data, (81) in the decentralized equilibrium and (83) in the social planner problem, are the same. Furthermore, the equilibrium price of data, π_t, is the same as the Lagrangian multiplier for the recourse constraint for data in the social planner problem. As such, there is no externality in the market for data trade.

The inefficiency is in the optimal choice of capital, comparing equilibrium equations (80) with optimum equations (82), for $i = L, H$. The inefficiency comes from heterogeneity across firms, and the fact that the planner takes the heterogeneity into account while firms do not. This implies that with a single type of firm, the equilibrium is efficient.

A.15.1 Implementing the Optimum Steady State in Equilibrium

Consider a given set of parameters Θ. Let $P^* = \Xi^{ss}$ in and $\pi^* = \eta^{ss}$ denote the Lagrangian multipliers that correspond to the retail good and data resource constraints in the social planner's problem, respectively.

Furthermore, consider a tax-subsidy policy that implements the optimum steady state in as a decentralized equilibrium allocation. This policy consists of a per unit consumption tax/subsidy rate τ^c and a per unit data tax/subsidy τ^δ rate, along with lump-sum transfers T^h to households.[11] τ^c, τ^δ, and T^h can be positive or negative. The equilibrium problem with tax-subsidy interventions can be written as

Household Problem

$$\begin{aligned}
&\max_{c_t, m_t, l_t} \sum_{t=0}^{+\infty} \beta^t (u(c_t) + m_t - v(l_t)) \\
&\text{s.t. } P_t(1+\tau^c)c_t + m_t = \Phi_t + w_t l_t + T^h \qquad \forall t \\
&c_t : \quad u'(c_t) = P_t \Gamma_t, \\
&m_t : \quad 1 = \Gamma_t, \\
&l_t : \quad v'(l_t) = \Gamma_t w_t, \\
&\Rightarrow \quad P_t(1+\tau^c) = u'(c_t)
\end{aligned}$$

Firm Problem

$$\max_{\{k_{i,t}, \delta_{i,t}\}_{t=0}^{\infty}} V(0) = \sum_{t=0}^{+\infty} \frac{1}{(1+r)^t} \left(P_t(1+\tau^c)\mathbb{E}[A_{i,t}]k_{i,t}^\alpha - \Psi(\Delta\Omega_{i,t+1}) - \pi(1+\tau^\delta)\delta_{i,t} - rk_{i,t} \right).$$

Equivalently

$$\begin{aligned}
V(\Omega_{i,t}) = \max_{k_{i,t}, \delta_{i,t}} \; & P_t(1+\tau^c)\mathbb{E}[A_{i,t}]k_{i,t}^\alpha - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - rk_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r} \\
\text{s.t. } & \Omega_{i,t+1} = \left(\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_i k_{i,t}^\alpha + \left(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0}\right)\delta_{i,t}\right)\sigma_\epsilon^{-2} \\
& A_{i,t} = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2
\end{aligned}$$

Market Clearing

$$\begin{aligned}
c_t &= \lambda A_{L,t} k_{L,t}^\alpha + (1-\lambda) A_{H,t} k_{H,t}^\alpha, \\
\lambda\delta_{L,t} &+ (1-\lambda)\delta_{H,t} = 0.
\end{aligned}$$

[11] Lump-sum transfers to firms is irrelevant as they can make negative profits.

Thus the equilibrium steady state equations that implement prices (P^*, π^*) given policy tuple $(\tau^c, \tau^\delta, T^h)$ are given by

$$\Omega_i^* = \left[\rho^2(\Omega_i^* + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i k_i^{*\alpha} + \delta_i^*(\mathbf{1}_{\delta_i^*>0} + \iota\mathbf{1}_{\delta_i^*<0})\right)\sigma_\epsilon^{-2} \qquad i = L, H \tag{85}$$

$$r = \alpha P^* k_i^{*(\alpha-1)}\left[\mathbb{E}[A_i^*] + \frac{z_i\sigma_\varepsilon^{-2}}{1+r}k_i^{*\alpha}\Omega_i^{*-2}\right] \qquad i = L, H$$

$$P^*\sigma_\epsilon^{-2}(\mathbf{1}_{\delta_i^*>0} + \iota\mathbf{1}_{\delta_i^*<0})k_i^{*\alpha} = (1+\tau^\delta)\pi^*\Omega_i^{*2}\left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_i + \sigma_a^{-2})\right)^2}\right) \qquad i = L, H$$

$$P^*(1+\tau^c) = u'(c^*)$$

$$v'(l^*) = \bar{m}$$

$$\lambda\delta_L^* + (1-\lambda)\delta_H^* = 0.$$

This is 9 equations in 9 unknowns. Note that we take (P^*, π^*), from the solution to the optimum steady state as given, and the two policy tools (τ^c, τ^d) are unknowns instead. So the 9 unknowns are: $(\Omega_i^*, k_i^*, \delta_i^*)$ for $i = L, H$, l_t, (τ^c, τ^δ).

Note that with the particular form of $g(.)$ that we are using,

$$\mathbb{E}[A_i^*] = \bar{A} - \Omega_i^{-1} - \sigma_a^2.$$

A.15.2 Short Run Dynamics

Assume an individual firm enters an economy in a steady state. Assume $z = z_i$, $i = H, L$. The new entrant faces market prices (P^*, π^*), and policy tools (τ^c, τ^δ). In the long run, the firm has to converge to $(\Omega_i^*, k_i^*, \delta_i^*)$. The three dynamic equations that characterize the new entrant's optimal dynamic path are given by

$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i k_{i,t}^\alpha + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})\right)\sigma_\epsilon^{-2} \tag{86}$$

$$r = \alpha k_{i,t}^{\alpha-1}\left[P^*\mathbb{E}[A_{i,t}] - \left(\Psi'(\Delta\Omega_{i,t+1}) - \Psi'(\Delta\Omega_{i,t+2})\right)z_i\sigma_\epsilon^{-2} + \frac{z_i\sigma_\varepsilon^{-2}}{1+r}P^*k_{i,t+1}^\alpha\Omega_{i,t+1}^{-2}\right] \tag{87}$$

$$\frac{(1+\tau^\delta)\pi^*}{\sigma_\epsilon^{-2}}\left(\frac{(1+r)}{(\mathbf{1}_{\delta_{i,t-1}>0} + \iota\mathbf{1}_{\delta_{i,t-1}<0})} - \frac{1}{(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})}\frac{\rho^2}{\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})}\right)$$
$$= P^*k_{i,t}^\alpha\Omega_{i,t-1}^{-2} + \Psi'(\Delta\Omega_{i,t+1})\left(1 - \frac{\rho^2}{\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})}\right) \tag{88}$$

Equation (86) is the dynamic evolution of firm knowledge, the same as equation (10). Equation (87) is the first order condition with respect to capital $k_{i,t}$, i.e. the generalized version of (79), that explicitly allows for dynamics. It can be replaced by

$$r = \alpha k_{i,t}^{\alpha-1}\left[P^*\mathbb{E}[A_{i,t}] - \Psi'(\Delta\Omega_{i,t+1})z_i\sigma_\epsilon^{-2} + \frac{(1+\tau^\delta)\pi z_i}{(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})}\right]. \tag{89}$$

Equation (89) is the generalized version of (77) that explicitly allows for dynamics and incorporates the data tax rate, i.e. the first order condition with respect to capital $k_{i,t}$ that uses the first order condition with respect to delta

choice $\delta_{i,t}$.

Lastly, equation (88) is the envelope condition, the derivative with respect to the state variable $\Omega_{i,t}$. It would be simplified if one can assume that $\delta_{i,t} = \delta_{i,t-1}$, i.e. a data seller is always a data seller and a buyer is always a buyer. However, I am *not* going to impose that. The reason is that in the very short run, I think it is violated at least for $i = H$ firms.

When the new entrant reaches the steady state, equations (86), (87), and (88) reduce to

$$\Omega_i^* = \left[\rho^2(\Omega_i^* + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i k_i^{*\alpha} + \delta_i^*(\mathbf{1}_{\delta_i^*>0} + \iota\mathbf{1}_{\delta_i^*<0})\right)\sigma_\epsilon^{-2}$$
$$r = \alpha k_i^{*(\alpha-1)}\left[P^*\mathbb{E}[A_i^*] - + \frac{z_i\sigma_\varepsilon^{-2}}{1+r}P^* k_i^{*\alpha}\Omega_i^{*-2}\right]$$
$$\frac{(1+\tau^\delta)\pi^*}{\sigma_\epsilon^{-2}(\mathbf{1}_{\delta_i^*>0} + \iota\mathbf{1}_{\delta_i^*<0})}\left((1+r) - \frac{\rho^2}{\rho^2 + \sigma_\theta^2(\Omega_i + \sigma_a^{-2})}\right) = P^* k_i^{*\alpha}\Omega_i^{*-2}$$

which are the exact same equations as the first 3 equations of the system (85) above, which characterizes the equilibrium steady state.

B Numerical Examples and Model Extensions

The section contains computational details, additional comparative statics and steady state numerical analyses that illustrate how our data economy responds to changes in parameter values for one or more firms.

Parameter Selection The results below are not calibrated.[12] However, the share of aggregate income paid to capital is commonly thought to be about 0.4. Since this is governed by the exponent α, we set $\alpha = 0.4$. For the rental rate on capital, we use a riskless rate of 3% , which is an average 3-month treasury rate over the last 40 years. The inverse demand curve parameters determine the price elasticity of demand. We take γ and $\bar{P}$ from the literature. Finally, we model the adjustment cost for data ψ in the same was as others have the adjust cost of capital.This approach makes sense because adusting one's process to use more data typically involves the purchase of new capital, like new computing and recording equipment and involves disruptive changes in firm practice, similar to the disruption of working with new physical machinery.

Finally, we normalize the noise in each data point $\sigma_\epsilon = 1$. We can do this without loss of generality because it is effectively a re-normalization of all the data-savviness parameter for all firms $\{z_i\}$. This is because for normal variables, having twice as many signals, each with twice the variance, makes no difference to the mean or variance of the agent's forecast. As long as we ignore any integer problems with the number of signals, the amount of information conveyed per signal is irrelevant. What matters is the total amount of information conveyed.

[12]To calibrate the model, one could match the following moments of the data. The capital-output ratio tells us something about the average productivity, which would be governed by a parameter like $\bar{A}$, among others. The variance of GDP and the capital stock, each relative to its mean, $var(K_t)/mean(K_t)$ and $var(Y_t)/mean(Y_t)$, are each informative about variance of the shocks to the model, such as σ_θ^2 and σ_a^2.

B.1 Computational Procedure

Figure 2 solves for the dynamic transition path when firms do not trade data.

Value Function Iteration: To solve for the value function, make a grid a values for Ω (state variable) and k (choice variable). Guess functions $V_0(\Omega)$ and $P_0(\Omega)$ on this grid. Guess a vector of ones for each. In an outer loop, iterate until the pricing function approximation converges. In an inner loop, given a candidate pricing function, iterate until the value function approximation converges.

Forward Iteration: Solving for the value function as described above also gives a policy function for $k(\Omega)$ and price function $P(\Omega)$. Linearly interpolate the approximations to these functions. Specify some initial condition Ω_0. For each t until T: Determine the choice of k_t and price at this state Ω_t. Calculate Ω_{t+1} from Ω_t and k_t.

Trade Value Function Approximation: Figure 5 solves for dynamic transition path when firms are allowed to buy/sell data for fixed final goods and data prices. We take the same steps as written above, but now optimize over ω rather then k.

Heterogeneous Firm Steady State Calculation: Figure 3 solves for the steady state equilibrium with two types of firms, in which both P and π are endogenous.

B.2 Data Portfolio Choice

A useful extension of the model would be to add a choice about what type of data to purchase or process. Firms that make different data choices would then naturally occupy different locations in a product space or operate in different industries.

The relevant state θ_t becomes an $n \times 1$ vector of variables. The stock of knowledge would then be the inverse variance-covariance matrix, $\Omega_{i,t} := \mathbb{E}_i[(\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t)(\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t)']^{-1}$, which is $n \times n$. The choice variables $\{k_{i,t}, \delta_{i,t}\}$ are $n \times 1$ vectors of investments in different sectors, projects or attributes and the corresponding data sales. The multi-dimensional recursive problem becomes

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t' \left(\mathbb{1}'(\bar{A} - \sigma_a^2)\mathbb{1} - \Omega_{i,t}^{-1}\right) k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t'\delta_{i,t} - rk_{i,t}'\mathbb{1} + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1}) \tag{90}$$

where $k_{i,t}^{\alpha}$ means that each element is raised to the power α, $\mathbb{1}$ is an $n \times 1$ vector of ones, and the law of motion for $\Omega_{i,t}$ is given by (10).

In such a model, locating in a crowded market space presents a trade-off. Abundant production of goods in that market will make goods profits low. However, for a firm that is a data purchaser, the abundance of data in this market will allow them to acquire the data they need to operate efficiently, at a low price. If many data purchasers locate in this product space and demand data about a particular risk $\theta_t(j)$, then efficient data producers might also want to produce goods that load on risk j, in order to produce high-demand data.