A Growth Model of the Data Economy

Maryam Farboodi* and Laura Veldkamp†

December 15, 2020

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 growth 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 entrants are unprofitable and why some of the biggest firms in the economy profit primarily from selling data. While these transition dynamics differ from those of traditional growth models, the long run features diminishing returns. Just like accumulating capital, accumulating predictive data, by itself, cannot sustain long-run growth.

1 Introduction

Does the new information economy have new economics? In the information age, production increasingly revolves around information and, specifically, data. Many firms, particuarly 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 (AI) allow us to use more data more efficiently. How will this new data economy evolve? Because data is non-rival, increases productivity and is freely replicable (has returns to scale), current thinking equates data growth with idea or technological growth. This article uses a simple framework to

*MIT Sloan School of Business and NBER; 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.

argue that data accumulation has forces of increasing and decreasing returns, as well as returns to specialization. But in the long run, data accumulation is more like capital accumulation, which, by itself, cannot sustain growth.

Data is information that can be encoded as a binary sequence of zeroes and ones. That broad definition includes literature, visual art and technological breakthroughs. We are focusing more narrowly on big data, defined as that data which is being analyzed by machine learning, AI or new big data technologies. The reason we focus on big data is because of claims that new big data technologies will spawn a new information age or economy, and the fact that an explosion of data accumulation has coincided with the arrival of this technology. Big data technologies, like machine learning and artificial intelligence, are prediction algorithms. Such algorithms predict the probability of a high demand for a good on a day, a picture being a cat, or advertisement resulting in a sale. Much of the data firms use for these predictions is transactions data. It is personal information about online buyers, satellite images of traffic patterns near stores, textual analysis of user reviews, click through data, and other evidence of economic activity. Such data is used to forecast sales, earnings and the future value of firms and their product lines. Data is also used to advertise, which may create social value or might simply steal business from other firms. We will consider both possibilities. But the essential features of the data production economy modeled in Section 1 are user-generated data that is a long-lived asset, used to predict uncertain future outcomes.

Section 2 performs a thought experiment, which is a data parallel to the question Solow (1956) asks about capital: Can data sustain growth, in the absence of any technological progress? To answer this question, the model will shut down all sources of technological change. Of course, this is unrealistic. Of course, data can be an input into research, just like capital can be an input into research, and thereby boost growth. But understanding whether data alone can sustain growth is fundamental to building our understanding of data, just like Solow's finding shaped our understanding of the role of capital accumulation in economic development.

We prove and trace out the consequences of three properties of data as an asset: 1) decreasing returns, 2) increasing returns, and 3) returns to specialization. Diminishing returns comes 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. In addition, unforecastable randomness 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, in a broad class of models.

However, 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 more transactions, which generates more data, and further increases 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. One reason this increasing returns force is significant is that it can generate a data poverty trap. Firms, industries, or countries may have low levels of data, which confine them to low levels of production and transactions, which make profits low, or even negative.

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. Many digital services, like apps, which were costly to develop, are given away to customers at zero price. Of course, companies are not being generous. 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.

Finally, a data economy may feature specialization. In some circumstances, large firms that have a comparative advantage in data production, derive most of their profit from data sales. Meanwhile, small firms have a comparative advantage in high-quality goods production. Therefore, the large firms produce high volumes of low-price goods, in order to produce data and sell it to small firms, that produce higher-quality goods. The business model of these large firms is to do lots of transactions at a low price and earn more revenue from data sales. While we know that many firms execute a strategy like this, it is different from a capital accumulation economy and suprising that such a strategy arises from simple economic properties of data as information.

The primary contribution of the paper is not the particular predictions we explore. Some of those predictions are more obvious, some more suprising. 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. Section 3 describes applications ranging from the distribution of firm size, to economic development, to finance.

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 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.

Our result should not be interpreted to mean that data does not contribute to growth. It absolutely does, in the same way that capital investment does. If ideas continue to improve, then data will help us find the most efficient uses of these new ideas. The accumulation of data may even reduce the costs of technological innovation by reducing its uncertainty, or increase the incentives for innovation by increasing the payoffs. The point is that being non-rival, freely replicable and productive is not enough for data to sustain growth. If it is used for forecasting, as most big data is, the data functions like capital. It increases output and can be an input in technology creation, but cannot sustain infinite growth all by itself. We still need innovation for that.

Related literature. Work on information frictions in business cycles, (Veldkamp (2005), Ordonez (2013) 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, data produced was used to forecast the state of the business cycle. In the data economy, that is not primarily what firms are using data for. Therefore, we adapt the modeling structure so that production generates firm- or industry-specific information that is private property of the firm.

In the growth literature, our model builds on Jones and Tonetti (2018). They explore how different data ownership models affect the rate of growth of the economy. In both models, data is a by-product of economic activity and therefore grows endogenously over time. What is different here is that data is information, used to forecast a random variable. In Jones and Tonetti (2018), 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, not technology, Jones and Tonetti (2018)'s conclusions about the benefits of data privacy may still hold. But instead of long-run growth, there is long-run stagnation.

Our work holds technology fixed. Other work complements ours by exploring the interactions of data and innovation. Agrawal et al. (2018) develop a combinatorial-based knowledge production function and embed it in the classic Jones (1995) growth model to explore how breakthroughs in AI could enhance discovery rates and economic growth. Lu (2019) embeds self-accumulating AI in a Lucas (1988) growth model and examines growth transition paths and welfare. Aghion et al. (2017) explore the reallocative effects of AI, as Baumol (1967)'s cost disease leads to the declining share of traditional industries' GDP. Furthremore, they note that AI may discourage future innovation for fear of imitation, undermining incentives to innovate in the first place. Our contribution is to understand the consequence of big data and the new prediction algorithms alone, for economic growth. Once we understand this foundation, one can layer these insights about data and innovation on top.

In the finance literature, Begenau et al. (2018) explore growth in the data processing capacity of financial investors and how data affects firm size through financial markets. They do not model firms' use of their own data. Such studies complement this work by illustrating other ways in which abundant data is re-shaping the economy.

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 that we solve analytically, models data in a more realistic way and explores different questions. The new features, including a market for data, non-rival data, and adjustment costs are not mere whistles and bells. These new margins shape the answers to our main questions about aggregate dynamics and long-run outcomes.

2 A Data Economy Growth Model

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.

2.1 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

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

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.[1] The transitory shock $\epsilon_{a,i,t}$ 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 $\theta_t + \epsilon_{a,i,t}$:

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

[1]One might consider different possible correlations of $\eta_{i,t}$ across firms i. An independent θ processes ($corr(\eta_{i,t}, \eta_{j,t}) = 0, \ \forall i \neq j$) 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.

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 marginal values but they are constantly changing and 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. The role of the temporary shock ε_a is that 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.

The next assumption captures the idea that data is a by-product of economic activity. The number of data points n observed by firm i at the end of period t depends on their production $k_{i,t}^{\alpha}$:

$$n_{i,t} = z_i k_{i,t}^{\alpha}, \tag{3}$$

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.

Each data point $m \in [1 : n_{i,t}]$ reveals

$$s_{i,t,m} = \theta_{t+1} + \epsilon_{i,t,m}, \tag{4}$$

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 $\epsilon_{i,t,m} \sim N(0, \sigma_\epsilon^2)$, and the unlearnable quality shock is $\epsilon_{a,i,t} \sim N(0, \sigma_a^2)$.

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.[2] 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 of knowledge.

Data Adjustment and the Stock of Knowledge The information set of firm i when it chooses its technique $a_{i,t}$ is[3] $\mathcal{I}_{i,t} = [\{A_{i,\tau}\}_{\tau=0}^{t-1}; \{\{s_{i,\tau,m}\}_{m=1}^{n_{i,\tau}}\}_{\tau=0}^{t-1}]$. 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

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

[3] 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.

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 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.

2.2 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 $\mathbb{E}[A_{i,t}] = \bar{A} - E\left[(\mathbb{E}_i[\theta_t|\mathcal{I}_{i,t}] - \theta_t - \epsilon_{a,i,t})^2\right]$. The second term is an expected squared forecast error, or equivalently, a conditional variance, of $\theta_t + \epsilon_{i,t}$. That conditional variance is denoted $\Omega_{i,t}^{-1} + \sigma_a^2$. Therefore, the expected quality of firm i's good at time t in (2) can be rewritten again as $\mathbb{E}_i[A_{i,t}] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2$.

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. Dispensing with the need to keep track of signal realizations, only signal precisions, 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,*

$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(n_{i,t} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})\right)\sigma_\epsilon^{-2} \tag{9}$$

Details of the proof are in Appendix A. However, we can understand Lemma 1 as a sum of the initial stock, outflows and inflows. When updating to time $t+1$, the initial stock of knowledge is $\Omega_{i,t}$. Outflows of data are data lost due to depreciation.

The inflows of data are new pieces of data that are generated by economic activity. The number of new data points $n_{i,t}$ was 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: $n_{i,t}\sigma_\epsilon^{-2}$. σ_a^{-2} 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. That is the role of the indicator functions at the end of (9). But at the aggregate level, an economy as a whole cannot buy or sell data. Therefore, for the aggregate economy,

$$\text{Inflows:} \qquad \Omega_{i,t}^+ = z_i k_{i,t}^\alpha \sigma_\epsilon^{-2} + \sigma_a^{-2} \tag{10}$$

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 prodcution technique choice $a_{i,t}$. Other firms' actions are not observable. The fact that other firms predicted well or poorly conveys no 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}. But 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.

The first term of the law of motion is the amount of data carried forward from period t: $\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1}$. 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. The information precision 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 add noise to old data and make it less valuable in the future. Finally, the whole expression is inverted again so that the variance is transformed back into a precision and once again, represents a (discounted) stock of knowledge. The depreciation of knowledge is the period-t stock of knowledge, minus the discounted stock:

$$\text{Outflows:} \qquad \Omega_{i,t}^{-}=\Omega_{i,t}+\sigma_a^{-2}-\left[(\rho^2(\Omega_{i,t}+\sigma_a^{-2}))^{-1}+\sigma_\theta^2\right]^{-1} \tag{11}$$

A one-state-variable problem We can thus express expected firm value recursively, with the stock of knowledge as the single state variable. The form of that recursive problem is described 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:*

$$\begin{aligned} V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} \; & P_t\left(\bar{A}-\Omega_{i,t}^{-1}-\sigma_a^2\right)k_{i,t}^{\alpha}-\Psi(\Delta\Omega_{i,t+1})-\pi_t\delta_{i,t}-rk_{i,t} \\ & +\left(\frac{1}{1+r}\right)V(\Omega_{i,t+1}) \end{aligned} \tag{12}$$

where $n_{i,t}=z_i k_{i,t}^{\alpha}$ and the law of motion for $\Omega_{i,t}$ is given by (9).

See Appendix for the proof. This result greatly simplifies the problem by collapsing it to a deterministic problem with only one choice variable k and one state variable, $\Omega_{i,t}$. The reason we can do this is that quality $A_{i,t}$ depends on the conditional variance of θ_t, and because the information structure is similar to that of a Kalman filter, where the sequence of conditional variances is generally deterministic. In expressing the problem this way, we have already substituted in the optimal choice of production technique.

Notice here that the non-rivalry of data does not change our problem substantially. It is equivalent to a kinked price of data, or to a transactions cost. We could 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 In this formulation of the problem, $\Omega_{i,t}$ can be interpreted as the amount of data a firm has. Technically, it is the precision of the firm's posterior belief. But according to Bayes' rule for normal variables, posterior precision is the discounted precision of prior beliefs plus the precision of each signal observed. $V(\Omega_{i,t})$ captures the value of observed data. $V(\Omega_{i,t}) - V(0)$ is the present discounted value of the additional net revenue the firm receives because of its stock of data.

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.

3 Long-Run and Short-Run Growth of a Data Economy

In this section, we establish key properties of growth in this data economy. This results explain the long-run and short-run behavior of the economy.

The first set of results involve the long-run growth of the data economy. We start by showing that within the model, there is no long run growth. We then move to broader results that are not model-specific. They describe general conditions under which data used for forecasting can sustain infinite growth. They do not prove that data-driven growth is not possible. Rather, they show that, if one believes that the accumulation of data for forecasting can sustain growth forever, there are some logically equivalent statements that one must also accept.

The second set of results demonstrate that data can create firm-level increasing returns, in the short run.

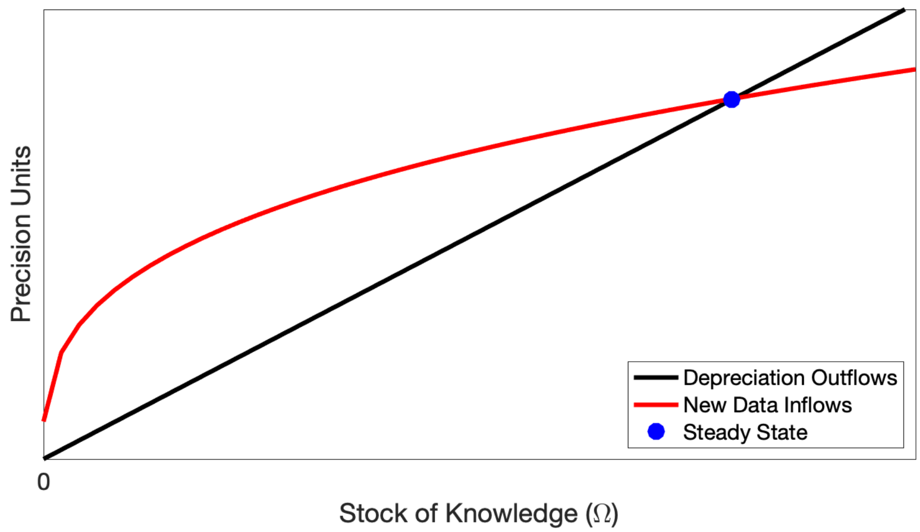

Figure 1: Economy converges to a data steady state: Aggregate inflows and outflows of data. Line labeled inflows plots $z_i k_{i,t}^{\alpha} \sigma_{\epsilon}^{-2}$ for a firm i, that makes an optimal capital decision $k_{i,t}^{*}$, with different levels of initial data stock. Line labeled outflows plots the quantity in (11).

3.1 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 10 and 11), 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 state level, it would no longer rise or diminish. Instead, data, quality and GDP would be constant as inflows and outflows just balance each other.

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 4 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.

Inflows have diminishing returns. As stated in the previous section, the reason is that the returns to data are bounded. With infinite data, all learnable uncertainty about θ can be resolved. With a perfect forecast of θ, the expected good quality is $\left(\bar{A} - \sigma_a^2\right)$, which is finite.[4] Thus, the optimal capital investment is finite. Since a function that is continuous and not concave will always cross any finite upper bound, productivity, investment and data inflows must all be concave in the stock of knowledge Ω.

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. Of course, mathematically, diminishing returns is hard-wired into the model by imposing $\bar{A}$ as the upper bound on productivity. That raises the question of how general this result is. The next set of results speak to the generality and justify the mathematical form we adopt.

Long Run Growth Impossibility Results How general is this idea that data accumulation cannot sustain positive growth? One might be concerned that diminishing returns is a remnant of the quadratic loss assumption, or some other functional form. Here, we consider an abstract economy, where the only assumptions we impose are that data is used to forecast future outcomes and that productivity is not growing from other sources. Then, we imagine that there is sustained growth arising from the economy. From this, we derive two assumptions: 1) Perfect foresight implies infinite real output; and 2) the future is a deterministic function of today's observable data. Neither assumption is standard is economics. Both appear implausible. But both would need to be satisfied for data accumulation to sustain growth.

The point is not that data is unproductive. The point is that absent any technological progress, better forecasts are not a tool that can logically sustain long-run growth.

Consider an economy, where data is used to forecast random outcomes. Let Γ_t represent the distribution of firms' stock of knowledge $\Omega_{i,t}$ at time t. For a continuum of firms, Γ_t encapsulates exactly what measure of firms have how much data, or more specifically, how much forecast precision, when they forecast θ_t. There is a finite-valued mapping $f(\Gamma_t)$ that maps the forecast precision of all firms into output (or, it could be a mapping into quality-adjusted output $A_t Y_t$). $f(\Gamma_t)$ defines

[4] i,t 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 second force. Even if capital did not have diminishing marginal returns, inflows would still exhibit concavity.

a "data economy." The fixed nature of the f function represents the assumption that productivity is fixed. As before, we are explore whether data alone sustains growth. Then the following results must hold for data accumulation to sustain a minimal positive rate of aggregate output growth.

Proposition 1 ***Perfect foresight must deliver infinite output.*** *A data economy with output $Y_t = f(\Gamma_t)$ 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 period t, only if perfect foresight $\Omega_{i,t} = \infty$ for some firm i implies infinite output $y_{i,t} = \infty$.*

Proof: Suppose not. Then, for every firm i, producing infinite data implies finite firm output $y_{i,t}$. If each firm's output is finite and the measure of all firms is finite, aggregate output, an integral of these individual firm outputs, is finite: Y_t has a finite upper bound in each period t. If Y_t has a finite upper bound in each period t and $ln(Y_{t+1}) - ln(Y_t) > \underline{g} > 0$ in every period t, then there exists a finite time at which Y_t will surpass its upper bound. Since only data is growing, aggregate output can only be infinite if some firm's output is infinite, $y_{i,t} = \infty$ for some i, t. That is a contradiction.
□

From a mathematical perspective, this result is perhaps obvious. But it 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.

Furthremore, the feasibility of perfect foresight is dubious, as the next result shows.

Proposition 2 ***Data-Driven Growth Implies a Deterministic Future.*** *Suppose aggregate output is a finite-valued function of each firm's forecast precision: $Y_t = f(\Gamma_t)$, and all data points $s_{i,t,m}$ are t-measurable signals about some future event θ_{t+1}.Sustained aggregate growth $ln(Y_{t+1}) - ln(Y_t) \geq \underline{g} > 0$ requires that future events θ_{t+1} are t-deterministic: θ_{t+1} is a deterministic function of the time-t sigma algebra of past events.*

Proof in the appendix. This result is significant because it illustrates why forecast precisions should not become infinite (no perfect foresight). Inifinite precision implies no un-forecastable noise. In other words, perfect foresight ($\Omega_{i,t} = \infty$) is not consistent with the existence of fundamental randomness. If there is not un-forecastable component of the future state, that means the future

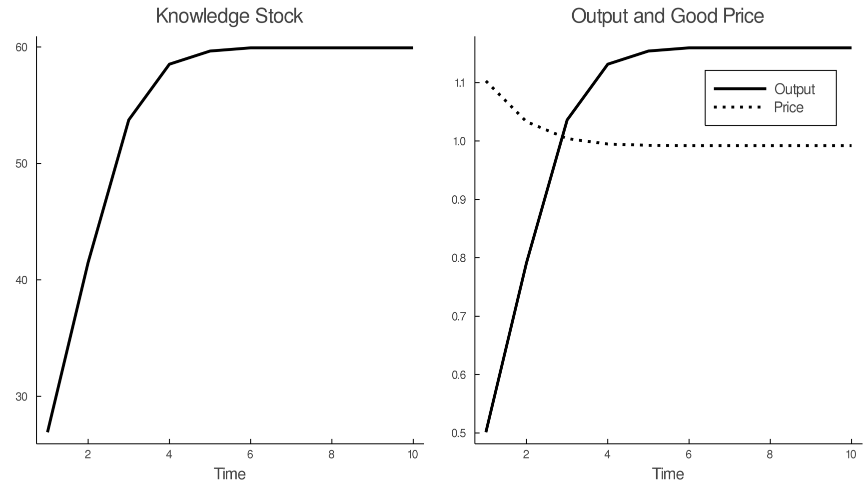

Figure 2: Aggregate growth dynamics: Data accumulation grows knowledge and output over time, with diminishing returns. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \gamma = 0.1, \overline{A} = 1, \overline{P} = 1, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2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

state is deterministic. Perfect forecasting requires that the future event is a deterministic function of observable information today. If the mapping is not deterministic, then the forecast has randomness, or noise and is not perfect.

Many think the idea of future determinism is a priori implausible. If one believes some events tomorrow are truly random today, then one must logically believe in a finite upper bound on precision. A finite upper bound implies a form of diminishing marginal returns, because any function without this property would violate the upper precision bound.

Both these observations are independent of the model. They simply start from the premise that data is used to forecast tomorrow's state. If data does not have diminishing returns, then that implies one of these awkward conclusions.

Taken together, these results suggests two distinct reasons why one might be skeptical of long-run, data-driven growth. The first reason is that perfect forecasts might not create infinite output. The second reason is that a perfect forecast implies that future events are deterministic functions of past events. Proving such determinism is well beyond the scope of this paper, and beyond the scope of economics.

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.

3.2 Increasing Returns in the Short Run

There are two main competing forces in this model. While the previous results focused on diminishing returns, the other force is one for increasing returns. Increasing returns arise from the data feedback loop: Firms with more data produce higher-quality goods, which induces them to invest more, produce more, and sell more, which, in turn, generates more data for them. That is a force that causes aggregate knowledge accumulation to accelerate. The feedback loop competes against diminishing returns. Diminishing returns always dominates when data is abundant. That's why 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.

The next result describes when increasing returns arise. While we have been talking about symmetric firms that do not trade data, we now relax the symmetry assumption. 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 how data accumulation can create barriers to new firm entrants.

Before we state the formal result, we need to define net data flow. Recall that data inflows $\Omega^+_{i,t}$ are the total precision of all new data points at t (eq. 10). Data outflows $\Omega^-_{i,t}$ are the end-of-period-t stock of knowledge minus the discounted stock (eq. 11). Net data flows are the difference between data inflows and outflows: $d\Omega_{i,t} = d\Omega^+_{i,t} - d\Omega^-_{i,t}$. Given these definitions, Proposition 3 states when a single firm entering faces increasing and then decreasing returns.

Proposition 3 ***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, goods and data prices are 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 and thus data. When the single firm enters, others 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 knowledge accumulation is concave, with no increasing returns region, when learnable risk is abundant.[5] In such cases, each firm's trajectory for output is also concave, as 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 3. This figure represents one possibility. Data production may lie above or below the data outflow line.

Notice that 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 4 illustrates the growth path of a new entrant firm in this environment. On the left side of the time path, 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. Early on, this firm has negative profits.

We define the firm book value to be the cumulated, discounted sum of profits, plus the cost of any purchased data.

$$\text{Book value} = \sum_{\tau=0}^{t} (1+r)^{t-\tau} \left(P_t A_{i,t} k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} \not\Vdash_{\delta_{i,t}<0} - r k_{i,t} \right) \tag{13}$$

[5] 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 3 because one also needs to show that the stock of knowledge Ω is always increasing.

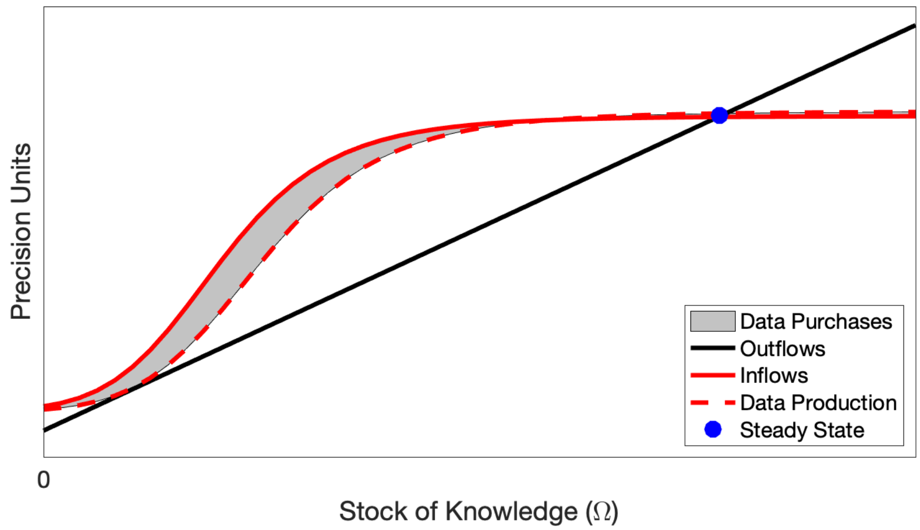

Figure 3: New firms grow slowly: inflows and outflows of data of a single firm.
Line labeled inflows plots $z_i k_{i,t}^{\alpha} \sigma_{\epsilon}^{-2} + \delta_{i,t}$ for a firm i, that makes an optimal capital decision $k_{i,t}^{*}$ and data decision $\delta_{i,t}$, with different levels of initial data stock. This firm is in an economy where all other firms are in steady state. Line labeled outflows plots the quantity in (11). Data production is $z_i k_{i,t}^{\alpha} \sigma_{\epsilon}^{-2}$, which is inflows, without the data purchases $\delta_{i,t}$.

The indicator function $\mathbb{1}_{\delta_{i,t}<0}$ does not subtract any cost of purchased data because, according to GAAP accounting rules, purchased intangible assets add to the book value of a firm. However, intangibles created by the company, i.e. the firm's own data, is not counted. The market value of the firm is the Bellman equation value function $V(\Omega)$ in (15). The difference between the market value of a firm and its book is used to measure intangible assets. In our numerical example, the new entrant has negative book value for 9 periods. Such a long period of negative value is significant because it suggests that new entrants with financing constraints might face a systematic entry barrier. However, working out the details of entry is scope for additional research.

4 Applications

Frameworks like this are only as important as the questions they can be used to answer. The benefit of a simple framework is that it can be extended in many directions to answer other questions. In this section, we investigate a diverse set of questions through the lens of the model.

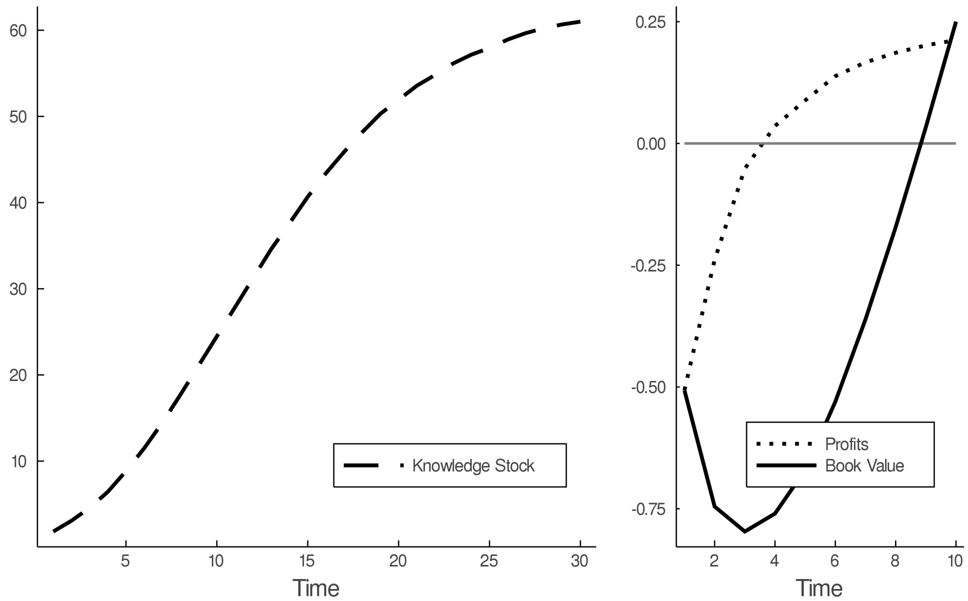

Figure 4: S-shaped growth creates initial profit losses. Knowledge stock defined in Lemma 1. Book value defined in (13). Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \overline{A} = 0.5, \overline{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z = 0.01, \pi = 0.002, P = 1, \iota = 1$

4.1 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.

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 the fact that data may be exchanged at zero monetary price.

Proposition 4 ***Firms barter goods for data.*** *It is possible that a firm will optimally choose positive production* $k_{i,t}^{\alpha} > 0$*, even if its price per unit is zero:* $P_t = 0$.

Investment $k_{i,t} > 0$ has a marginal cost r and a marginal benefit, which is that more 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 4 illustrates an example where the firm makes negative profits for the first 1-2 periods because they price their goods at less than marginal cost. 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. For its first 17 straight quarters as a public company, Amazon consistently made losses. It lost $2.8 billion before turning a profit. Today, it is one of the most valuable companies in the world.

4.2 Specialization in Data Sales

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 1 (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 still produce a lot of goods and data. The question is whether they use data to produce high-quality goods or not.

Proposition 5 ***If a single firm is more efficient at data production, that firm keeps more data for itself.*** *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 $\frac{d\Upsilon^{ss}}{dz_H} > 0$, $\forall \iota$ 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 knolwedge 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 both sell and retain data, high-productivity data miners sell more data; so much more that they are left with less knowledge.

Proposition 6 ***More efficient data producers may accumulate less knowledge.*** *Suppose that there is a strictly positive measure of high-data-productivity firms ($\lambda < 1$), then there exist a ball of price and production parameters $\gamma \epsilon B(0)$, $\alpha < \frac{1}{2}$ such that, 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 forfeitted, 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.

The next result describes how the intensity of data specialization grows as high data-productivity increases.

Corollary 1 ***If a measure of firms becomes more efficient at data production, they keep more (less) data for themselves when data is single (multi) use.*** *Suppose $\lambda < 1$, price and production parameters are $\gamma \epsilon B(0)$, $\alpha < \frac{1}{2}$ and the economy is in steady state. For each λ, $\exists \bar{\iota}_2$, $\underline{z}_H$ such that*

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

If ι is high, the knowledge gap was originally positive, and it becomes more positive when data processing efficient diverges. If ι is low, meaning that data is not very rival, negative knowledge gaps become more negative. But the cutoff ι_1 for positive knowledge gaps is not the same as the cutoff ι_2 for growing knowledge gaps. That means that for some intermediate levels of data rivalry, the knowledge gap can shrink.

Comparing Propositions 5 and 6 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 6, 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.

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. The numerical example in Figure 5 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

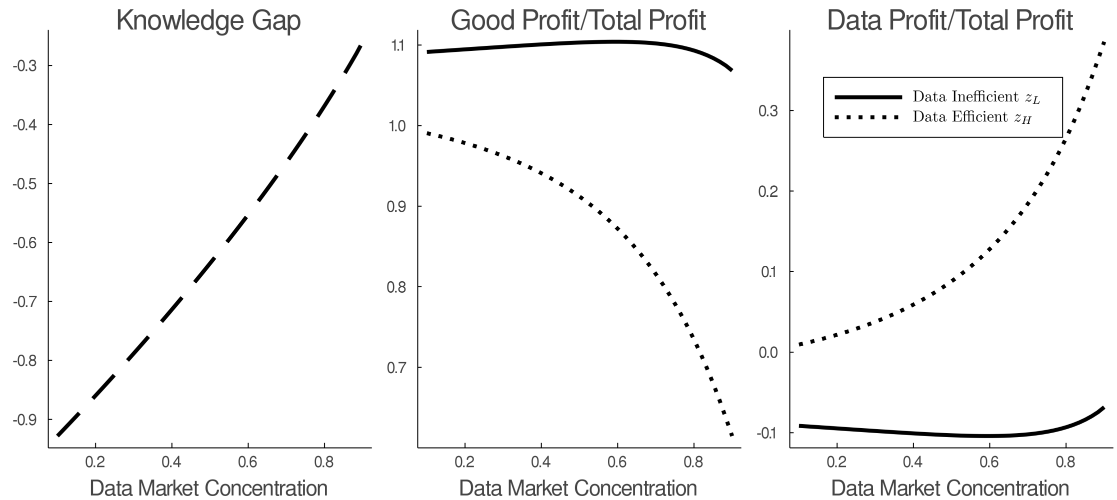

Figure 5: Big Firms Specialize in Selling Data When Data is Multi-Use. Small Firms Specialize in High-Quality Good Production.Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \gamma = 0.09, \overline{A} = 1, \overline{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z_1 = 0.01, z_2 = 10$

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 5 shows that data market concentration amplifies the specialization of data firms and high-quality goods producers.

Interpretation: Data Platforms and Data Services We interpret the circumstances when large firms sell most of their data as the emergence of 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 much of what they use that data for is to sell data services to their business customers. That may look different from what is going on in this model, but in fact, it is a type of data sale. An example is Facebook. Most of its valuation comes not from postings. Rather Facebook's revenue and valuation come mostly from the value of the data. Data is what allows it to sell advertising, which is a data service.

Data can be sold both directly and indirectly. A data vendor can sell you a data set directly, and transfer the binary code that constitutes the data. But they can also offer data services, which are an indirect sale of data. Such a service might entail using their data to place your ad on the screen of a particular user type; it could entail using their data to choose assets to invest your money in, or it could involve using proprietary data to provide a firm strategic business consulting advice. Such services monetize data, without transferring the underlying data used in the service. Admati and Pfleiderer (1990) examine when each type of data sale is optimal for the seller. Future work might explicitly model indirect sale. But advertising services is an interpretation of what our data sales represent.

4.3 Data for Business Stealing

Data is not always used for a socially productive purpose. One might argue that many firms use data simply to steal customers away from other firms. So far, we've modeled data as something that enhances a firm's productivity. But what if it increases profitability, in a way that detracts from the profitability of other firms? Using an idea from Morris and Shin (2002), we can model such business-stealing activity as an externality that works through productivity:

$$A_{i,t} = \bar{A} - \left(a_{i,t} - \theta_t - \epsilon_{a,i,t}\right)^2 + \int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj \tag{14}$$

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. These gains and losses are such that, when added up to compute aggregate productivity, they cancel out: $\int A_{i,t} = \bar{A}$. This represents an extreme view that data processing contributes absolutely nothing to social welfare. While that is unlikely, examining the two extreme cases is illuminating.

What we find is that reformulating the problem this way makes very little difference for most of our conclusions. The externality does reduce the productivity of firms and does reduce welfare, relative to the case without the externality. But it does not change firms' choices. Therefore, it does not change data inflows, outflows or accumulation. It does not change firm dynamics. The reason there is so little change is that the externality does not enter in a firm's first order condition. It does not change its optimal choice of anything. Firm i's actions have an infinitesimal, negligible effect on the average productivity term $\int_{j=0}^{1} \left(a_{j,t} - \theta_t - \epsilon_{a,j,t}\right)^2 dj$. Because firm i is massless in a competitive industry, its actions do not affect that aggregate term. So the derivative of that 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.

Whether data is productivity-enhancing or not matters for welfare and the price per good, but does not change our conclusions that a firm's growth from data alone is bounded, that firms can be stuck in data poverty traps, or that markets for data will arise to partially mitigate the unequal effects of data production.

4.4 Data-Driven Innovation

The result that data accumulation cannot sustain growth is a statement about what happens when we hold ideas fixed, and data is used as a prediction tool. However, if there is technological progress and ideas are accumulating, data could facilitate that accumulation of ideas, and thereby promote growth. For example, suppose that the maximum quality level $\bar{A}$ in equation 2 is endogenous and time-varying.

The simplest way to accommodate data-driven innovation in the model is to allow $\bar{A}_t$ to be a function of firm's stock of data at time t, $\bar{A}_t(\Omega_{it})$, or firm's newly collected data at time $t-1$, $\bar{A}_t\left(n_{i,t} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota\mathbf{1}_{\delta_{i,t}<0})\right)$. More generally, advances in the quality frontier $\bar{A}_t$ might depend on the size of the labor force engaged in research L^R, some research capital K^R, and research data D^R:

$$\bar{A}_t - \bar{A}_{t-1} = (L_t^R)^\alpha (K_t^R)^\gamma (D_t^R)^{1-\alpha-\gamma}$$

In this formulation, if $1-\alpha-\gamma > 0$, then it means that data is an essential ingredient in idea production. If that is true, then data not only contributes to growth, it is essential for growth.

Of course, this raises the question of whether we are talking about the same kinds of data. Perhaps transactions records are essential for marketing and product development, but true innovations make use of some other kind of data, perhaps developed with labor and capital, in a lab. If that is true, then D^R is a function of L^R and K^R, in which case, we could write the idea production function in terms of L^R and K^R, without mentioning data at all. That model's role for data would be similar to our baseline model. The alternative, where production or transactions generates data looks like learning-by-doing models (Jovanovic and Nyarko, 1996). Either could be integrated into this framework, perhaps to explore optimal data policy.

4.5 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. Then, the multi-dimensional recursive problem becomes

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t' \left(\mathbb{1}'(\bar{A} - \sigma_a^2)\mathbb{1} - \Omega_{i,t}^{-1} \right) k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t'\delta_{i,t} - rk_{i,t}'\mathbb{1}$$
$$+ \left(\frac{1}{1+r} \right) V(\Omega_{i,t+1}) \tag{15}$$

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 (9).

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.

4.6 Other Possible Applications and Extensions

Below we mention other extensions, which we do not explore in detail.

Optimal Government Data Policy Figure 3 shows how a lack of data can slow the growth of a firm. Given this problem, 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 invesment, 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 firms sharing data or act to prevent data from being exported to foreign firms. Firms who share data are like a larger firm. They can operate efficiently with the abundant data. This increases the returns to producing more goods and more data. The current policy debates around data privacy, for example in the European Union, could be partly about countries fighting for their ability to keep up in the data race, to prevent being stuck in relative data poverty.

Firm Size Dispersion: Bigger then Smaller. 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. Big firms do more transactions, which allows them to be more productive and grow bigger.

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. A firm with a larger stock of knowledge acumulates additional data at a faster pace. That drives divergence. Later on in the growth path, diminishing returns sets in. Firms with a large stock of knowledge accumulate more data, but less additional knowledge. In this region, small initial differences in the stock of knowledge fade in importance.

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 an enormous amount about their clients, per dollar value of service they provide. Barber shops or other cash-based service providers may accumulate little hard data. At the same time, every firm can do more to collect, structure and analyze the data that its transactions produce. We could allow a firm to increase its data-savviness z_i, at a cost. By duality, for any exogenously assumed set of z's, there exists a cost structure on the choice of z that would give rise to that set of z as an optimal outcome. However, endogenizing these choices, with a fixed cost structure, might produce changes in the cross-section of firms' data, over time.

5 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 and the gains are bounded. There can also be regions of increasing returns that create a possible entry barrier or firm size dispersion. Such traps arise with capital externalities as well. Data's production process, with its feedback loop from data to production and back to data, makes such 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.

This simple framework speaks to many data-related phenomena. It can be a foundation for thinking about many more.

References

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

Aghion, Philippe, 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.

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.

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.

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

Jones, Charles I., "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 1995, *103* (4), 759–784.

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

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.

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

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

$$s^a_{i,t-1} = \theta_{t-1} + \epsilon_{a,t-1}, \tag{16}$$

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

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:

$$\bar{s}_t = \theta_t + \epsilon_{s,t}, \tag{17}$$

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 $s^a_{i,t-1}$:

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

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

$$\widehat{\mu}_t = \mathbb{E}\big[\theta_t \mid \mathcal{I}_t\big] = \frac{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} \cdot \mathbb{E}\big[\theta_t \mid \mathcal{I}_{t-1}, s^a_{i,t-1}\big] + \omega_t \sigma_\epsilon^{-2} \bar{s}_t}{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}} \tag{18}$$
$$\Sigma_t = Var\big[\theta_t \mid \mathcal{I}_t\big] = \Big\{\Big[\rho^2\big[\Sigma^{-1}_{t-1} + \sigma_a^{-2}\big]^{-1} + \sigma_\theta^2\Big]^{-1} + \omega_t \sigma_\epsilon^{-2}\Big\}^{-1} \tag{19}$$

Multiply and divide equation (18) by Σ_t as defined in equation (19) to get

$$\widehat{\mu}_{i,t} = (1 - \omega_t \sigma_\epsilon^{-2} \Sigma_t)\left[\bar{\theta}(1-\rho) + \rho\left((1 - M_t)\mu_{t-1} + M_t s^a_{i,t-1}\right)\right] + \omega_t \sigma_\epsilon^{-2} \Sigma_t \bar{s}_t, \tag{20}$$

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

Equations (19) and (20) 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:

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

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

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

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 (19), this posterior variance is

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

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:

$$V(\Sigma_t) = \max_{k_t}\Big[P_t(\bar{A} - \Sigma_t - \sigma_a^2)k_t^\alpha + \pi\delta i, t - \Psi(\Delta\Omega_{i,t+1}) - rk_t + \left(\frac{1}{1+r}\right)V(\Sigma_{t+1})\Big]$$

$$\text{s.t.} \qquad \Sigma_{t+1} = \Big[\big[\rho^2(\Sigma_t^{-1} + \sigma_a^2)^{-1} + \sigma_\theta^2\big]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2}\Big]^{-1}$$

Data use is incorporated in the stock of knowledge through (9), 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

$$k_{i,t} = \left[\frac{\alpha}{r}\left(P_t A_{i,t} + z_i \sigma_\epsilon^{-2}(\left(\frac{1}{1+r}\right)V'(\cdot) - 2\psi(\cdot))\right)\right]^{1/(1-\alpha)} \tag{21}$$

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 9), we can rewrite this as

$$\omega_{ss}\sigma_\epsilon^{-2} + \left[\rho^2(\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} = \Omega_{ss} \tag{22}$$

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 (21) 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}:

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r}\right)\frac{\partial \Omega_{t+1}}{\partial \Omega_t}|_{ss}\right]^{-1} P_t k_{ss}^{\alpha}\Omega_{ss}^{-2} \tag{23}$$

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 (21), (22) and (23) 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

$$V(\Omega_{i,t}) = \max_{k_{i,t},\omega_{i,t}} P_t \left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2\right) k_{i,t}^{\alpha} - \pi \left(\frac{\omega_{i,t} - z_i k_{i,t}^{\alpha}}{\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}}}\right) - r k_{i,t}$$
$$- \Psi\left(\Delta\Omega_{i,t+1}\right) + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1}) \quad (24)$$

$$\text{where} \quad \Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{i,t}\sigma_\epsilon^{-2} \quad (25)$$

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

$$FOC[k_{i,t}]: \quad \alpha P_t A_{i,t} k_{i,t}^{\alpha-1} + \frac{\pi \alpha z_i k_{i,t}^{\alpha-1}}{\mathbf{1}_{\omega_{i,t}>n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t}<n_{i,t}}} - r = 0 \quad (26)$$

Solving for $k_{i,t}$ gives

$$k_{i,t} = \left(\frac{1}{r}\left(\alpha P_t A_{i,t} + \tilde{\pi}\alpha z_i\right)\right)^{\frac{1}{1-\alpha}} \quad (27)$$

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

$$FOC[\omega_{i,t}]: \quad -\Psi'(\cdot)\frac{\partial \Delta\Omega_{i,t+1}}{\partial \omega_{i,t}} - \tilde{\pi} + \left(\frac{1}{1+r}\right) V'(\Omega_{i,t+1})\frac{\partial \Omega_{i,t+1}}{\partial \omega_{i,t}} = 0 \tag{28}$$

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 9), we can rewrite this as

$$\omega_{ss}\sigma_\epsilon^{-2} + \left[\rho^2(\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} = \Omega_{ss} \tag{29}$$

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 27 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}:

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r}\right)\frac{\partial \Omega_{t+1}}{\partial \Omega_t}|_{ss}\right]^{-1} P_{ss}k_{ss}^\alpha \Omega_{ss}^{-2} \tag{30}$$

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 28), the steady state marginal value is given by

$$V'_{ss} = (1+r)\tilde{\pi}\sigma_\epsilon^2 \tag{31}$$

The equations (27), (28), (29) and (30) 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.

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

where i denotes the firm data productivity.

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

$$
\Omega_L = \left(\rho^2 (\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right)^{-1} + \left(z_L k_L^{\alpha} + \delta_L \right) \sigma_\epsilon^{-2} \tag{32}
$$

$$
\Omega_H = \left(\rho^2 (\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right)^{-1} + \left(z_H k_H^{\alpha} + \iota \delta_H \right) \sigma_\epsilon^{-2} \tag{33}
$$

$$
\alpha P (\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^{\alpha-1} + \pi \alpha z_L k_L^{\alpha-1} = r \tag{34}
$$

$$
\alpha P (\bar{A} - \Omega_H^{-1} - \sigma_a^2) k_H^{\alpha-1} + \frac{\pi \alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{35}
$$

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

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

$$
P = \bar{P} \left(\lambda (\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^{\alpha} + (1-\lambda)(\bar{A} - \Omega_H^{-1} - \sigma_a^2) k_H^{\alpha} \right)^{-\gamma} \tag{38}
$$

$$
\lambda \delta_L + (1-\lambda) \delta_H = 0 \tag{39}
$$

A.5 Proof of Proposition 1: Data without Diminishing Returns Implies 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 $t+s$, $\forall s>0$:

$$Y_{t+s}=\int_i y_{i,t}di<M_y\int_i di<M_yN<\infty. \tag{40}$$

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 $t+s$, $\forall s>0$,

$$\ln(Y_{t+s})-\ln(Y_t)=[\ln(Y_{t+s})-\ln(Y_{t+s-1})]+\cdots+[\ln(Y_{t+1})-\ln(Y_t)]>\underline{g}s, \tag{41}$$

which implies

$$Y_{t+s}>Y_te^{\underline{g}s}. \tag{42}$$

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

$$Y_{t+s}>Y_te^{\underline{g}s}>Y_te^{\underline{g}\underline{s}}>Y_te^{\underline{g}\frac{\ln(M_yN)-\ln(Y_t)}{\underline{g}}}=M_yN, \tag{43}$$

which contradicts (40).

A.6 Proof of Proposition 2: Data Growth Implies a Deterministic θ in 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: $\Omega_{i,t} \to \infty$. □

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 3: S-shaped Accumulation of Knowledge

We proceed in two parts: convexity and then concavity.

Part 1: 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 21) drops out and it reduces to $k_{i,t} = \left[\frac{\alpha}{r}\left(P_t A_{i,t} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V'(\Omega_{i,t+1})\right)\right]^{1/(1-\alpha)}$, which implies

$$k_{i,t}^{1-\alpha} = \frac{\alpha}{r}\left(P_t A_{i,t} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V'(\Omega_{i,t+1})\right).$$

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

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

Plugging in $\frac{\partial A_{i,t}}{\partial \Omega_{i,t}} = {\Omega_{i,t}}^{-2}$ and $\frac{\partial \Omega_{i,t+1}}{\partial \Omega_{i,t}} = \rho^2[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})]^{-2}$, we have

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

Therefore,

$$\begin{aligned} \frac{\partial d\Omega_{i,t}^{+}}{\partial \Omega_{i,t}} &= z_{i,t} k_{i,t}^{2\alpha-1} \sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r}\left(P_t {\Omega_{i,t}}^{-2} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V''(\Omega_{i,t+1}) \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})]^{-2}\right) \\ &= z_{i,t} k_{i,t}^{2\alpha-1} \sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r} P_t {\Omega_{i,t}}^{-2} + z_{i,t}^2 k_{i,t}^{2\alpha-1} \sigma_\epsilon^{-4} \frac{\alpha^2}{1-\alpha} \frac{1}{r(1+r)} V''(\Omega_{i,t+1}) \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})]^{-2}. \end{aligned} \tag{44}$$

Next, we calculate the derivative of data outflow $d\Omega_{i,t}^{-} = \Omega_{i,t} + \sigma_a^{-2} - \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}$ with respect to $\Omega_{i,t}$. We have

$$\frac{\partial d\Omega_{i,t}^{-}}{\partial \Omega_{i,t}} = 1 - \frac{1}{\rho^2(\Omega_{i,t} + \sigma_a^{-2})^2(\sigma_\theta^2 + \rho^{-2}(\Omega_{i,t} + \sigma_a^{-2})^{-1})^2}. \tag{45}$$

The derivatives of net data flow is then

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

For notational convenience, denote the first term in (2) as $M_1 = z_{i,t} k_{i,t}^{2\alpha-1} \sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r} P_t {\Omega_{i,t}}^{-2} > 0$, the second term as $M_2 = z_{i,t}^2 k_{i,t}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{i,t+1})\rho^2[\rho^2+\sigma_\theta^2(\Omega_{i,t}+\sigma_a^{-2})]^{-2} \le 0$ 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.

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

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

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

$$
z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3}\right] \quad > 0 \tag{47}
$$

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

$$
f(\Omega_{i,t}) \quad := -z_{it}\sigma_\epsilon^{-2}\tfrac{1}{1+r}\nu\rho^2\Omega_{it}^3\tfrac{\sigma_\theta^2}{\left(\rho^2+\sigma_\theta^2\left(\Omega_{it}+\sigma_a^{-2}\right)\right)^3} \tag{48}
$$

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

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

$$
h(\Omega_{i,t}) = 1 - M_3, \tag{49}
$$

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

Part 2: 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 $\rho \le 1$ and $\sigma_\theta^2 \ge 0$, 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 (44), 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 4: 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 (15) 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 Proposition 5: 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:

$$\Omega_L = \left(\rho^2(\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + z_L k_L^\alpha \sigma_\epsilon^{-2} \tag{50}$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^{\alpha-1} + \pi \alpha z_L k_L^{\alpha-1} = r \tag{51}$$

$$P \sigma_\epsilon^{-2} k_L^\alpha = \pi \Omega_L^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) \tag{52}$$

$$P = \bar{P}\left((\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^\alpha\right)^{-\gamma} \tag{53}$$

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:

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) k_H^{\alpha-1} + \frac{\pi \alpha z_H k_H^{\alpha-1}}{\iota} = r \tag{54}$$

$$\iota P \sigma_\epsilon^{-2} k_H^\alpha = \pi \Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2 (\Omega_H + \sigma_a^{-2})\right)^2} \right) \tag{55}$$

$$\Omega_H = \left(\rho^2 (\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right)^{-1} + \left(z_H k_H^\alpha + \iota \delta_H\right) \sigma_\epsilon^{-2} \tag{56}$$

Manipulate to get

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \frac{\pi \alpha z_H}{\iota} = r k_H^{1-\alpha} \tag{57}$$

$$k_H^\alpha = \left(k_H^{1-\alpha}\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{\iota P \sigma_\epsilon^{-2}} \Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2 (\Omega_H + \sigma_a^{-2})\right)^2} \right) \tag{58}$$

$$\iota^{\frac{1-2\alpha}{1-\alpha}} \frac{1}{r^{\frac{\alpha}{1-\alpha}}} \left(\iota \alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \pi \alpha z_H\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P \sigma_\epsilon^{-2}} \Omega_H^2 \left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2 (\Omega_H + \sigma_a^{-2})\right)^2} \right) \tag{59}$$

Next we show three steps:

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

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

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

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

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

This completes the proof.

Step 1. Take the total derivative of equation (59) wrt to z_H and simplify. It implies

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

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

$$A(i) = \frac{\alpha^2\iota_i^{2-\frac{1}{1-\alpha}}\left(\iota_i r k_i^{1-\alpha}\right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}}k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2} \tag{60}$$

$$B(i) = \frac{2\pi\sigma_\epsilon^2\Omega_i\left(1+r-\frac{\rho^2(\rho^2+\sigma_a^2\sigma_\theta^2)}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^3}\right)}{P} = \pi\frac{dC(i)}{d\Omega_i} \tag{61}$$

$$C(i) = \frac{\sigma_\epsilon^2\Omega_i^2\left(1+r-\frac{\rho^2}{\left(\rho^2+\sigma_\theta^2(\sigma_a^2+\Omega_i)\right)^2}\right)}{P} = \frac{\iota_i k_i^\alpha}{\pi} \tag{62}$$

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

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

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

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

Furthermore, consider equation (37). 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 (37) 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 (58) implies that fixing ι, k_H moves in the same direction as Ω_H.

Step 2. 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, $\forall z_H$,

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

Step 3. 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 6: Negative Knowledge Gap with Non-rival Data When High Data Productivity is Abundant

The proof proceeds in a few steps. First note that $\gamma = 0$ implies that $P = \bar{P}$.

Step 1. $\mathbf{z_H}$ 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.

Step 2. $\frac{\mathbf{d}\pi}{\mathbf{z_H}} < \mathbf{0}$. 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.

Step 3. Negative knowledge gap: $\exists \bar{\iota} \mid \iota \leq \bar{\iota} \Rightarrow \boldsymbol{\Upsilon}^{\mathbf{ss}} < \mathbf{0}$. Merge equations (34)-(37) and use $P = \bar{P}$ to write Ω_H and Ω_L:

$$\frac{\pi}{\bar{P}\sigma_\epsilon^{-2}}\Omega_L^2\left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) = \frac{1}{r^{\frac{\alpha}{1-\alpha}}}\left(\alpha\bar{P}(\bar{A} - \Omega_L^{-1} - \sigma_a^2) + \pi\alpha z_L\right)^{\frac{\alpha}{1-\alpha}} \tag{63}$$

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

Consider equation (64). 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 (63) on Ω_L. Thus $\exists\bar{\iota}$ such that $\iota \leq \bar{\iota} \Rightarrow \Upsilon^{ss} < 0$.

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

Equations (34) and (36) 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

$$\Omega_L^2\left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) = \frac{\left(\bar{P}(\bar{A} - \Omega_L^{-1} - \sigma_a^2) + \pi z_L\right)^{\frac{\alpha}{1-\alpha}}}{\pi}\bar{P}\sigma_\epsilon^{-2}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}} \tag{65}$$

The same argument as in proposition 5 shows that using equation (36), the derivative of the left hand side with respect to Ω_L is positive. Next, using implicit function theorem on both sides of equation (65) 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 ((34), (36)) and ((35), (37)) to get:

$$\frac{1}{r^{\frac{\alpha}{1-\alpha}}}\left(\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2) + \pi\alpha z_L\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}}\Omega_L^2\left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2})\right)^2}\right) \tag{66}$$
$$\frac{\iota^{\frac{1-2\alpha}{1-\alpha}}}{r^{\frac{\alpha}{1-\alpha}}}\left(\iota\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \pi\alpha z_H\right)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}}\Omega_H^2\left(1 + r - \frac{\rho^2}{\left(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2})\right)^2}\right). \tag{67}$$

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 (60)-(62) the above expressions simplify to:

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

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

Thus the derivative of the knowledge gap is given by

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

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

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

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

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

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

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

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

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

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

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 Data 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 3.

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 3 for details). Thus

$$\frac{\partial d\Omega_{i,t}^+}{\partial \Omega_{i,t}} - \frac{\partial d\Omega_{i,t}^-}{\partial \Omega_{i,t}} \approx \rho^2 \left[\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})\right]^{-2} - 1 + \rho^2(1 + \rho^2\sigma_\theta^2\sigma_a^{-2})^{-2}. \tag{68}$$

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 9: 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 3 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 (49), and* f *is defined by (48);*

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, $d\Omega^+_{i,t} - d\Omega^-_{i,t} = z_{i,t} k^\alpha_{i,t} \sigma_\epsilon^{-2} - \Omega_{i,t} - \sigma_a^{-2} + \left[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2\right]^{-1}$ 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: $\frac{\partial^2 d\Omega^-_{i,t}}{\partial \Omega^2_{i,t}} = \frac{2\rho^4 \sigma_\theta^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2}+\rho^2\sigma_\theta^2\Omega_{i,t})^3} > 0$. 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 $\frac{\partial^2 d\Omega^-_{i,t}}{\partial \Omega^2_{i,t}} > 0$, 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 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 3 (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 4 (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 $\frac{\rho^4\bar{\epsilon}^2(\Omega_{i,t}+\sigma_a^{-2})^2}{1+\rho^2\epsilon_\sigma^2(\Omega_{i,t}+\sigma_a^{-2})}$. 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 $\frac{\rho^4\epsilon_\sigma^2(\Omega_{i,t}+\sigma_a^{-2})^2}{1+\rho^2\bar{\epsilon}^2(\Omega_{i,t}+\sigma_a^{-2})}$.

Lemma 5 (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 $\frac{\rho^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})^2}$.

B Numerical Examples

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.[6] 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

[6] 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.

irrelevant. What matters is the total amount of information conveyed.

B.1 Computational Procedure

No-Trade Value Function Approximation

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 4 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 5 solves for the steady state equilibrium with two types of firms, in which both P and π are endogenous.