

A Growth Model of the Data Economy

Maryam Farboodi* and Laura Veldkamp†

October 5, 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 classic growth model where firms accumulate data. Data has three key features: 1) Data is a by-product of economic activity; 2) data is information used for resolving uncertainty, and 3) uncertainty reduction enhances firm productivity. The model can explain why data-intensive goods or services, like apps, are given away for free, why many new firms are unprofitable and why the biggest firms in the economy profit primarily from data. While these transition dynamics differ from those of traditional growth models, the long run features diminishing returns. Just like capital accumulation, data accumulation alone cannot sustain growth. Without other improvements in productivity, data-driven growth will grind to a halt.

Does the new information economy have new economics, in the long run? When the economy shifted from agrarian to industrial, economists focused on capital accumulation and removed land from production functions. As we shift from an industrial to a knowledge economy, the nature of inputs is changing again. In the information age, production increasingly revolves around information and, specifically, data. Many firms, particularly the most valuable U.S. firms, are valued primarily for the data they have accumulated. Collection and use of data is as old as book-keeping. But recent innovations in computing and artificial intelligence (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

*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.

or technological growth. This article uses a simple framework to argue that data accumulation has forces of increasing and decreasing returns at work. 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 because that is where the technological breakthroughs have taken place that have spawned talk of a new information age or economy. Machine learning or 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 big 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 explores the logical consequences of this simple set of assumptions. Specifically, 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 scale. We start with decreasing returns. Data cannot sustain long-run growth, because ultimately the more data a firm has, the less it benefits from additional data. The key is that data, like all information, is a means of reducing uncertainty. Uncertainty is bounded below by zero. Unless a perfect forecast gives a firm access to a pure, real, limitless arbitrage, the perfect forecast generates finite payoff. If the payoff to infinite data is finite, the returns, at some point, must diminish, so as not to exceed that upper bound. Furthermore, perfect forecasts imply that the future is not random. It must be a perfectly predictable function of past events; otherwise, a data set could not predict it perfectly. While both of these are mathematical possibilities, they go well beyond the fictions typically used in economic modelling.

At the same time, the way in which data is produced means that, at low levels, data has increasing returns. 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. But because data is a long-lived asset, firms may choose to produce with negative profits, as a form of costly investment in data.

Finally, like all information, data has returns to scale. Large firms' use of data can be quite different from small firms'. Our results show how large firms have a comparative advantage in data production. Small firms have a comparative advantage in high-quality goods production. Therefore, in some circumstances, the large firms produce high volumes of low-quality 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 low price and earn more revenue from data sales. While we know that many firms execute a strategy like this, it is surprising that such a strategy arises from these simple economic properties of data as information.

The primary contribution of the paper is a tool to think clearly about the economics of aggregate data accumulation. It answers the question: When we think about the data economy, how should we adjust our thinking from existing aggregate frameworks like Solow (1956), which lies at the foundation of modern DSGE models? 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 non-data-technology (referred to hereafter as "productivity" or "technology") continues to improve, data helps us find the most efficient uses of these new technologies. 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, 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 is a by-product of economic activity; firms use this information to reduce uncertainty and guide their decision-making. These authors did not call the information data. But it has all the hallmarks of modern transactions data. The data produced was used by firms to forecast the state of the business cycle. Better forecasting enabled the firms to invest more wisely and be more profitable. These models restricted attention to data about aggregate productivity. In the data economy, that is not primarily what firms are using data for. But such modeling structures can be adapted so that production can also generate firm- or industry-specific information. As such, they provide useful equilibrium frameworks on which to build a data economy.

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.

Other authors consider the interaction of artificial intelligence (AI) and innovation. Agrawal et al. (2018) develops a combinatorial-based knowledge production function and embeds 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 from an economy without AI to an economy with AI and how employment and welfare evolves. Aghion et al. (2017) explore the role of AI for the growth

process and its reallocative effects. The authors argue that Baumol (1967)'s cost disease leads to the declining share of traditional industries' GDP, as they become automated. This decline is offset by the growing fraction of automated industry. In such an environment, AI may discourage future innovation for fear of imitation, undermining incentives to innovate in the first place.

While some big data is used to facilitate innovation, most of the “new economy” data is web searches, shopping behavior and other evidence of economic transactions. While the existence of such data has inspired innovations such as the sharing economy and recommendations engines, those new ideas are distinct from the data itself. The contents of transactions data does not, by itself, reveal a breakthrough technology. Whether data and innovation are complements is a separate question that these studies shed light on. The accumulation of data used solely for prediction is driving a large and growing sector of the economy. Our contribution is to understand the consequence of big data and the new prediction algorithms alone, for economic growth.

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.

1 A Data Economy Growth Model

The model looks much like a simple Solow (1956) model. There are inflows of data from new economic activity and outflows, as data depreciates. The depreciation comes from the fact that the state is constantly evolving. 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. When data is scarce, little is lost due to depreciation. As data stocks grow large, depreciation losses are substantial. The point at which data depreciation equals the inflow of new data is a data steady state. Firms with less data than their steady state grow in data, and therefore in productivity and investment. If a firm ever had more data than its steady state level, it should shrink. But without any other source of growth in the model, data-driven growth, like capital-driven growth eventually grinds to a halt.

The difference between data accumulation and capital accumulation appears in the source of diminishing returns, the increasing returns from the data feedback loop, and the returns to scale that arise when one piece of data can guide a large firm, just as well as a small one.

1.1 Model

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

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

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

¹One might consider different possible correlations of $\eta_{i,t}$ across firms i . An independent θ processes ($\text{corr}(\eta_{i,t}, \eta_{j,t}) = 0, \forall i \neq j$) would effectively shut down any trade in data. Since buying and selling data happens and is worth exploring, we consider an aggregate θ process ($\text{corr}(\eta_{i,t}, \eta_{j,t}) = 1, \forall i, j$). It is also possible to achieve an imperfect, but non-zero correlation.

technique θ_t :

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

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 accordingly. Transactions help to reveal these preferences 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 many establishments with a range of practices and many customers provides useful information for optimizing business practices.

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 $k_{i,t}^\alpha$:

$$n_{i,t} = z_i k_{i,t}^\alpha, \quad (3)$$

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

$$s_{i,t,m} = \theta_{t+1} + \epsilon_{i,t,m}, \quad (4)$$

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 In order to model data as a tradeable asset, we need to allow for the possibility that the amount of data a firm produces is not the same as the amount of data they use. The difference between the two is the amount of data purchased from or effectively lost due to its sale to other firms.² Let δ_{it} be the amount of data traded by firm i a time t . If $\delta_{it} < 0$, the firm is selling data. If $\delta_{it} > 0$, the firm purchased data. We restrict $\delta_{it} \geq -n_{it}$ 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 then 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 $\delta_{it} < 0$ to other firms, then the firm has $n_{it} + \iota\delta_{it}$ data points left to add to its own stock of knowledge. (Note that $\iota\delta < 0$ so that the firm has less data than the n_{it} 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. If the firm buys $\delta_{it} > 0$ units of data, it adds $n_{it} + \delta_{it}$ units of data to its stock of knowledge. Another interpretation of this assumption is that there is a transaction cost of trading data, proportional to the data value.

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

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

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

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

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

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

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_{it} is $\mathcal{I}_{i,t}$.

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

Equilibrium

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

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

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

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

1.2 Solution

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

Optimal technique and expected quality. Taking a first order condition with respect to the technique choice, we find that the optimal technique is $a_{i,t}^* = \mathbb{E}_i[\theta_t | \mathcal{I}_{i,t}]$. Thus, expected quality of firm i 's good at time t in (2) can be rewritten as $\mathbb{E}[A_{i,t}] = \bar{A} - E [(\mathbb{E}_i[\theta_t | \mathcal{I}_{i,t}] - \theta_t - \epsilon_{a,i,t})^2]$. 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 $\Omega_{it}^{-1} + \sigma_a^2$. 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 ,*

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

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,

$$\text{Inflows} = z_i k_{i,t}^\alpha \sigma_\epsilon^{-2} + \sigma_a^{-2} \quad (10)$$

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 : $[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$. 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:

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

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:*

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

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, $\omega = n_{it} + \delta_{it}$ for data purchases ($\delta_{it} > 0$) and $\omega = n_{it} + \iota \delta_{it}$ for data sales when $\delta_{it} < 0$. Then, we could re-express this problem as a choice of ω and a corresponding price that depends on whether $\omega \geq n_{it}$ or $\omega < n_{it}$.

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. In other words, the precision of beliefs, Ω_{it} is a linear transformation of the number of all past used data points, $\{\omega_{is}\}_{s=1}^t$. Ω_{it} captures the value of past observed data through the term for the discounted prior precision, $\Omega_{i,t-1}$.

The marginal value of one additional piece of data, of precision 1, is simply $\partial V_t / \partial \Omega_{it}$. When we consider markets for buying and selling data, $\partial V_t / \partial \Omega_{it}$ represents the firm's demand, its marginal willingness to pay for data.

2 Properties of a Data Economy

We begin by establishing a few key properties of this data economy. The first three results are not model-specific. They describe conditions under which data 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, here are some logically equivalent statements that one must therefore also accept.

The remainder of the results describe comparative statics and dynamic properties: Who buys or sells data, at what price; how does the behavior of one firm affect the incentives of others, and finally, what are the dynamics of data accumulation?

2.1 Diminishing Returns and Zero Long Run Growth

Just like we typically teach the Solow (1956) model by examining the inflows and outflows of capital, we can gain insight into our data economy growth model by exploring the inflows and outflows of data. Figure 1 illustrates the inflows and outflows (eq.s 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 that distance diminishing, which represents growth slowing, and eventually a crossing of inflows and outflows that we label steady state. If the stock of knowledge ever reached its steady state level,

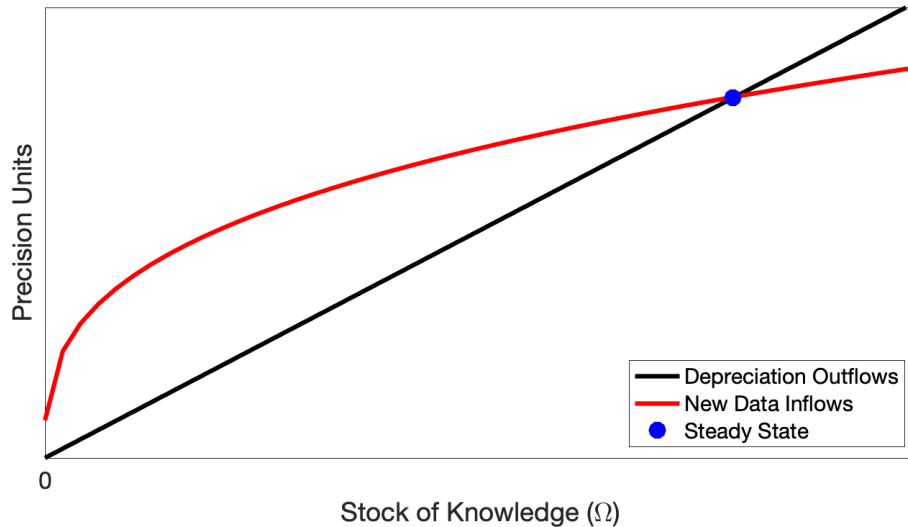


Figure 1: Economy converges to a data steady state: Aggregate inflows and outflows of data. Line labeled inflows plots $z_i k_{i,t}^\alpha \sigma_\epsilon^{-2}$ for a firm i , that makes an optimal capital decision $k_{i,t}^*$, with different levels of initial data stock. Line labeled outflows plots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

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. The proof of Proposition 4 (step 2) 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 $(\bar{A} - \sigma_a^2)$, which is finite.³ 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

³It 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.

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? 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, so that we can see what growth can possibly come from data alone.

Consider an economy, where productivity \bar{A} is fixed. Only data n_{it} grows. Data is used to forecast random outcomes. Call that a “data economy.” Then the following results must hold for data accumulation to sustain a minimal positive rate of aggregate output growth.

Proposition 1 *Data and Infinite Output.* *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 period t , only if infinite data $n_{it} \rightarrow \infty$ for some firm i implies infinite output $y_{it} \rightarrow \infty$.*

Proof: Suppose not. Then, for every firm i , producing infinite data implies finite firm output y_{it} . 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 and not the measure of firms, aggregate output can only be infinite if some firm’s output is infinite, $y_{it} = \infty$ for some i, t . That is a contradiction. \square

This result is significant for two reasons. First, a perfect forecast is does not generate infinite real economic value. An individual who knows what will yield profit tomorrow can leverage that information and become very rich. But if society as a whole knows tomorrow’s state, they can simply produce today what they would otherwise be able to produce tomorrow. Second, the feasibility of perfect foresight is dubious, as the next two results show. 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.

Of course, data can fuel growth temporarily. It can also be an input in idea production, which does fuel growth. But both of these statements are true for capital accumulation as well. The point is not that data is unproductive. The point is that absent any technological progress, data alone will not sustain growth.

Let Γ_t represent the distribution of firms' stock of knowledge Ω_{it} at time t . For a continuum of firms, Γ_t encapsulates exactly what measure of firms have how much data, or more specifically, what forecast precision, when they forecast θ_t .

Proposition 2 *Data and Infinite Precision.* *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 $n_{it} \rightarrow \infty$ for some firm i implies infinite precision $\Omega_{it} \rightarrow \infty$.*

Proof: 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 $Y_t \rightarrow \infty$ when some firm has infinite data $n_{it} \rightarrow \infty$. Since Y_t is a finite-valued function of Γ_t , it can only be that $Y_t \rightarrow \infty$ if some moment of Γ_t is also becoming infinite $\Gamma_t \rightarrow \pm\infty$. 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, $\Gamma_t \rightarrow \infty$. 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: $\Omega_{it} \rightarrow \infty$. \square

The next 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 precision 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.

Proposition 3 *Infinite Precision Implies a Deterministic Future.* *Suppose all data points $s_{i,t,m}$ are t -measurable signals about some future event θ_{t+1} . If infinite data $n_{it} \rightarrow \infty$ for some firm i*

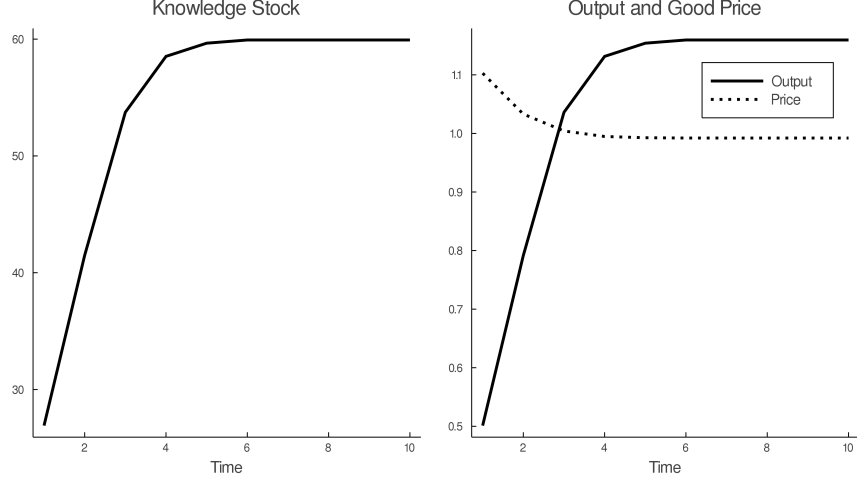


Figure 2: Aggregate growth dynamics: Data accumulation grows knowledge and output over time, with diminishing returns. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \gamma = 0.1, \bar{A} = 1, \bar{P} = 1, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\epsilon^2 = 0.1, z =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

implies infinite precision $\Omega_{it} \rightarrow \infty$, then future events θ_{t+1} are deterministic: θ_{t+1} is a deterministic function of the sigma algebra of past events.

Proof in the appendix. This result is significant because it illustrates why forecast precisions should have a finite upper bound. If they did not, that would mean precision could become infinite. Infinite precision implies no un-forecastable noise. If there is not un-forecastable component of the future state, then the future state is deterministic. Many think the idea of future determinism is a priori implausible. If you believe some events tomorrow are truly random today, then you 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.

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.

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

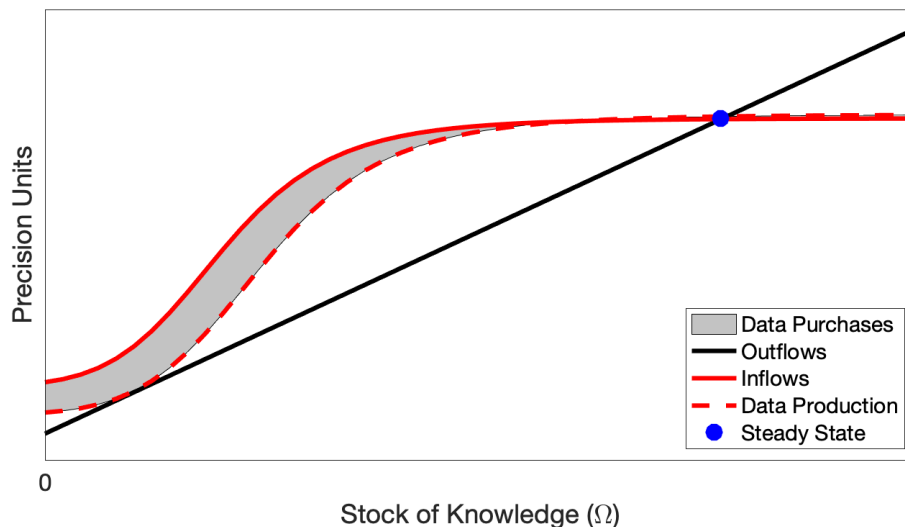


Figure 3: New firms grow slowly: inflows and outflows of data of a single firm.

Line labeled inflows plots $z_i k_{i,t}^\alpha \sigma_\epsilon^{-2} + \delta_{it}$ for a firm i , that makes an optimal capital decision $k_{i,t}^*$ and data decision δ_{it} , 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 δ_{it} .

scope of economics.

2.2 Increasing Returns, Data Barter and Entry Barriers

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 already abundant. That's why the previous results about the long run were about concavity and 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 in the lifecycle of a new firm.

For some parameter values, the diminishing returns to data is always stronger than the data feedback loop. Each firm's trajectory for output and knowledge is concave, as in Figure 2. But for other types of 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.

The next set of results describe what properties of an economy determine which regime prevails. We learn what makes increasing returns stronger or diminishing returns weaker. While we have been talking about symmetric firms that do not trade data, we now relax the symmetry assumption. We consider a setting where all firms are in steady state. Then, we drop in one atomless firm and observe its transition. From this exercise, we learn about how data accumulation can create barriers to new firm entrants.

Just to be clear, these results only hold for a single firm entrant, not for the transition path of the entire economy of firms growing together. Why does this S-shape arise for one firm, but not for the aggregate economy? 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 is also abundant, keeping the price of data low. 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.

While the diminishing returns to data was quite general, this S-shape is not robust. It arises only when lots of risk is unlearnable. Before we state the formal result, we need to define terms. First, let $\zeta \equiv \frac{\sigma_\theta^2}{\sigma_a^2}$. This is the ratio of variance of the learnable component of the state to the variance of the unlearnable component. If this ratio is high, most uncertainty can be resolved by accumulating enough data. Next, we define data inflows $d\Omega_{it}^+$, data outflows $d\Omega_{it}^-$, and net data flows $d\Omega_{it}$. Data inflows are the total precision of all new data points at t : $d\Omega_{it}^+ = z_{i,t} k_{i,t}^\alpha \sigma_\epsilon^{-2}$. Data outflows are the end-of-period- t stock of knowledge minus the discounted stock: $d\Omega_{it}^- = \Omega_{i,t} + \sigma_a^{-2} - [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$. Net data flows are the difference between data inflows and outflows: $d\Omega_{it} = d\Omega_{it}^+ - d\Omega_{it}^-$. Give these definitions, Proposition 4 states precisely when a single firm entering faces increasing and then decreasing returns.

Proposition 4 *S-shaped accumulation of aggregate stock of knowledge.* With symmetric firms and sufficiently low adjustment cost Ψ , the net data flows $d\Omega_{it}$

- 1) increase with the stock of knowledge Ω_{it} when that stock is low, $\Omega_{it} < \hat{\Omega}$, when goods production has sufficient diminishing marginal return, $\alpha < \frac{1}{2}$, price is above a threshold $P_t > f(\hat{\Omega})$ and the second derivative of the value function is bounded $V'' \in [\nu, 0)$; and
- 2) decrease with Ω_{it} when Ω_{it} is larger than $\hat{\Omega}$.

As we saw in the last section, the concave region of knowledge accumulation (part 2) is always present. However, the convex region (part 1) is not. Proposition 8 in the Appendix proves the converse of this result: when learnable risk is abundant, relative to unlearnable risk, the S-shape disappears; knowledge accumulation is concave.

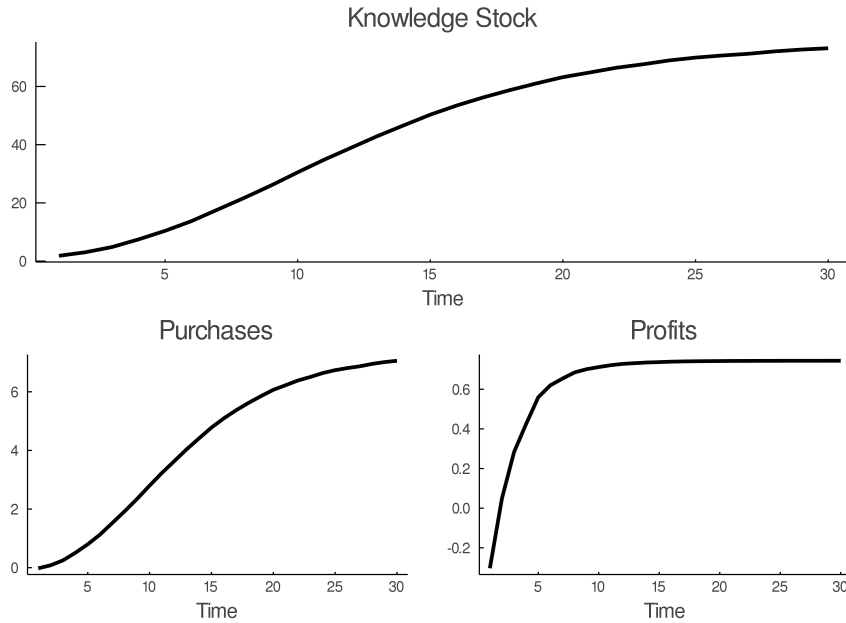


Figure 4: S-shaped growth creates initial profit losses, a possible barrier to entry. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.4, \bar{A} = 1, \bar{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 0.5, \sigma_\varepsilon^2 = 0.1, z = 0.05, \pi = 0.002, P = 1, \iota = 1$

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. Since the returns to producing and accumulating knowledge are low, data inflows are low. This makes the firm grow slowly and be less profitable. When this force is strong, such firms will be losing

money throughout this early phase of their growth.

This S-shape is important for firm dynamics. It is the same shape that the output takes in a growth model with poverty traps (Easterly, 2002). The same dynamic forces that give rise to poverty traps can give rise to data traps. Increasing returns implies high returns, for high levels of asset accumulation, but also implies low return to investment at low levels of asset accumulation. The asset here is data. Firms invest in data by producing goods. Thus, this low returns region is one with low profitability. Small firms can get stuck for a while in the convex region as they slowly produce more and accumulate more data to escape the trap.⁴

Data Barter In the model, barter arises when the goods are exchanged for customer data, at a zero price. This is a possibility result and a knife-edge case. But it is an interesting case 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 perhaps not surprising, given the value of data as an asset. But the result demonstrates the value of the framework, by showing how it speaks to data-specific phenomena we see.

Proposition 5 *Firms barter goods for data.* *If $\frac{\pi_t z_i \alpha}{\left(r + \frac{\partial \Psi(\Delta \Omega_{i,t+1})}{\partial k_{i,t}}\right)^\iota} > 0$, then it is possible that a firm will optimally choose positive production $k_{it}^\alpha > 0$, even if its price per unit is zero: $P_t = 0$.*

Proof: Proving this possibility requires a proof by example. Suppose the price of goods is zero P_t and the price of data π is such that firm i finds it optimal to sell all its data produced in period t : $\delta_{it} = -n_{it}$. In this case, differentiating the value function (12) with respect to k yields $(\pi_t/\iota)z_i \alpha k^{\alpha-1} = r + \frac{\partial \Psi(\Delta \Omega_{i,t+1})}{\partial k_{i,t}}$. Can this optimality condition hold for positive k ? If $k^{1-\alpha} = \frac{\pi_t z_i \alpha}{\left(r + \frac{\partial \Psi(\Delta \Omega_{i,t+1})}{\partial k_{i,t}}\right)^\iota} > 0$, then the firm optimally chooses $k_{it} > 0$, at price $P_t = 0$. \square

The idea of this argument is that costly investment $k_{it} > 0$ has a marginal benefit: more data that can be sold at price π_t and a marginal cost r . If the price of data is sufficiently high, and/or

⁴With additional parameter restrictions, 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. The additional parameter restrictions ensure that the stock of knowledge Ω is always increasing.

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.

In some economies, the optimal price of a firms' goods are initially well below marginal cost, and even close to zero. Figure 4 illustrates an example of this where the firm makes negative profits for the first 1-2 periods because they price their goods at less than marginal cost. The reason for producing at the near-zero price is that the firm wants to sell many units, in order to accumulate data. Data will boost the productivity of future production and enable future profitable goods sales. It could also be that the firm wants to accumulate data, in order to sell it. Lowering the price is a costly investment in a productive asset that yields future value. This is what makes a zero optimal price, or data barter, possible.

In this example, that strategy works for the firm because it faces no financing constraint. In reality, many firms that make losses for years on end lose their financing and exit.

Data purchases and data poverty traps. On the left side of Figure 3, there is a point where data outflows equal data production. If data were not traded, or if the available data was not the relevant data for this firm, the stock of data would not grow past that point. The optimal capital investment is not high enough to generate enough transactions to replace the number of data points lost due to data depreciation. This no-growth phase, in which the firm produces low-quality goods, can be interpreted as a barrier to entry in global data markets. Without data markets, if a firm could manage to accumulate enough data, it could escape the trap and then sustain a high stock of knowledge. However, the ability to buy data from other firms avoids this trap.

These results suggest that financially constrained young firms may never enter or make the investment in data systems, unless there is relevant data available for purchase. By not collecting and using data, they never escape the trap of poor data and low value-added.

2.3 Returns to Scale: Firm Size and Data Strategy

Large firms generate lots of data. There are two possible ways a firm might profit from this situation. First, they could retain the data, use it to make even higher-quality goods, and make profits from good production. In this case, we say data-productive firms are specialized in production of goods and services. Alternatively, they could profit by selling the data and produce a large volume of

low-quality goods. Little of their revenue would come from goods sales. But their data sales would earn their profits. We say that they are *data platforms* if specialize in data (services) production.

Regardless of their strategy, high data-productivity firms would produce a lot of data. The difference between the two strategies is what the firm specializes in and how it earns profits: the former strategy implies that the large firm makes profits from high quality goods and services it produces, while the latter strategy involves profits from data services.

One might think that allowing firms to trade data would eliminate data dispersion. Presumably, firms that do not produce much data could simply pay to acquire it. This would eliminate data dispersion. But that is not what happens. This section explores the conditions under which a large firm would optimally choose each strategy. We consider a 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). H-firms' superior efficiency at extracting or processing data allows them to be more efficient and produce higher-quality goods, which in turn, induces them to rent more capital and produce at a larger scale. Despite the two types, there is still an infinite number of price-taking firms. Since firms have different data productivity, there is room for data trade among firms. Data price adjusts endogenously so that the data market clears.

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. We will show that counter-intuitively, higher data-productivity firms can own less data in equilibrium. Furthermore, as their data-productivity grows, they might choose to keep even less data compared to L-firms. In order to make this comparison clear, it is useful to 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, $\Upsilon_t = \Omega_{Ht} - \Omega_{Lt}$.*

We use the knowledge gap to guide our thinking about the systematic emergence of data users and data platforms. Consider a single H-firm entering a market populated by L-firms ($\lambda = 1$), and focus on the steady state. In this case, the outcome is what is intuitively expected. The data-productive firm is always larger, accumulates more data in the steady state, and specializes in high quality production. Thus the knowledge gap is positive. Furthermore, as the data productivity of the H-firm increases, it accumulates even more knowledge in steady state, i.e. the knowledge

gap is increasing in z_H . Note that the single H-firm does not influence the good or data price. Furthermore, since it is infinitesimal it does not affect the allocation of capital or data for the measure of L-firms. By continuity, the same pattern arises when the measure of data-productive firms is tiny, i.e. $\lambda \rightarrow 1$. The following result describes the parameter conditions under which this scenario arises.

Proposition 6 *If a single firm is more efficient at data production, that firm keeps more data for itself. Suppose $\lambda = 1$, but there is a single, measure-zero H-firm in the market with $z_i = z_H$. If the economy is in steady state, then $\forall \iota$ and z_H , $\Upsilon^{ss} > 0$ and $\frac{d\Upsilon^{ss}}{dz_H} > 0$.*

Next, consider a steady state in which the measure of L-firms, λ , is bounded away from one. H-firms, which are more efficient in data production, may retain more data and use it to improve their products, as does the single firm in the previous result. In this case, the knowledge gap is positive, $\Upsilon_t > 0$. One might think of this as the Uber model, named after the ride-sharing service that uses users' demand for rides to direct drivers to high-demand areas and price the rides to equate supply and demand. Uber's main business is not the sale of the data they collect.

In other circumstances, the reverse happens. Firms that are better data producers find that their comparative advantage is in *data distribution*, not in high-quality goods production. Therefore, they retain less data, which reduces the quality of the goods they produce. Such firms derive most of their revenue from selling data, not from using the data. In this case, knowledge gap is negative, $\Upsilon_t < 0$. The next result describes the circumstances in which data-productive firms specialize in using data for high-quality production and what circumstances they specialize in data sales.

Proposition 7 *If a measure of firms becomes more efficient at data production, they keep more (less) data for themselves when data is more (less) private. Suppose $\lambda < 1$, $\gamma = 0$, $\alpha \leq \frac{1}{2}$, and the economy is in steady state. For each λ , there exists $\bar{\iota}_1$ such that $\forall z_H$*

$$\begin{cases} \Upsilon^{ss} > 0 & \iota > \bar{\iota}_i \\ \Upsilon^{ss} < 0 & \iota < \bar{\iota}_1. \end{cases}$$

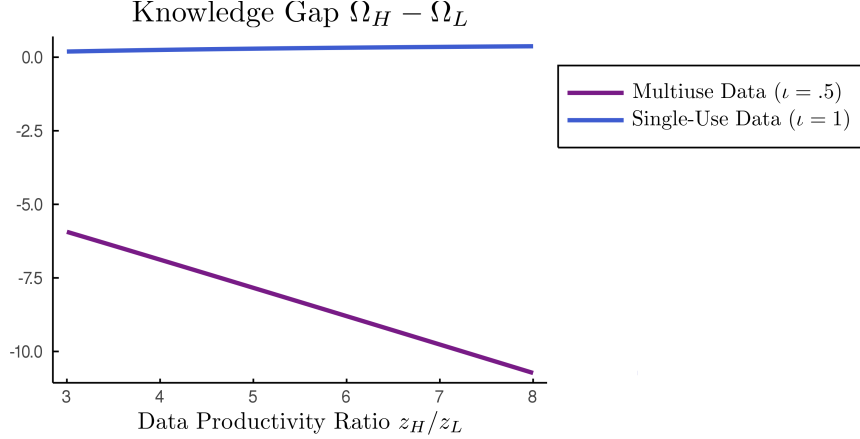


Figure 5: Big Firms Are Data Distributors When Privacy is Weak. Parameters: $\rho = 1, r = 0.2, \beta = 0.97, \alpha = 0.3, \psi = 0.3, \gamma = 0.1, \bar{A} = 1.4, \bar{P} = 0.5, \sigma_a^2 = 0.05, \sigma_\theta^2 = 1, \sigma_\epsilon^2 = 0.1, z_1 = 1, \lambda = 0.75$

Furthermore, $\exists \bar{\iota}_2, \bar{z}_H$ such that

$$\begin{cases} \frac{d\Upsilon^{ss}}{dz_H} > 0 & z_H > \bar{z}_H, \iota > \bar{\iota}_2 \\ \frac{d\Upsilon^{ss}}{dz_H} < 0 & z_H > \bar{z}_H, \iota < \bar{\iota}_2. \end{cases}$$

In the second scenario, the data-productive firm that keeps less data for themselves looks like a data platform. An example is Google. Most of its valuation comes not from the profit margins on internet searches. Rather, Google'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.

When data-productivity is a very scarce ability, $\lambda \rightarrow 1$, the more data productive firms always specialize in high quality production. As such, the rise of data platforms is a pure general equilibrium phenomena. However, even when high data productivity is not scarce, data should not be single-use in order for the data platforms to emerge. In this case, large supply of data by data-productive firms encourages the low data-productivity ones to demand a lot of it. When data is multi-use, highly data productive firms find it less costly to sell a lot of their data and capture the corresponding profits, i.e. become data platforms.

In either scenario, an increase in data productivity among high-type firms, z_H , prompts more specialization. In one case, the H firm specializes in high-quality goods production. In the other case, they specialize more in data production. This specialization is most pronounced when the demand curve for final goods is less elastic, $\gamma \approx 0$.

To visually interpret these results, Figure 5 shows how a larger, more data-productive firm could end up producing higher-quality goods and service, i.e. knowledge gap above zero, or choose to produce lower-quality goods and instead specialize in data production and distribution, depending on the privacy or rivalry of data. This figure demonstrates a second interesting phenomena. As one would expect, when data-productivity of H-firms increases, they might choose to accumulate more data, and increase the knowledge gap. However, they might choose to specialize even more in data distribution, in which case the knowledge gap becomes even more negative. In the latter case, higher data-productivity does lead to more data for both group of firms, i.e. both Ω_H and Ω_L increase. However, the data productive firms choose to specialize even more than before in data distribution. As such, an increase in z_H leads to more data accumulation by L firms than H firms, and knowledge gap gets more negative.

Data Platforms and Data Services One interpretation of the large firms that sell most of their data is as data platforms. 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 cutomers. That may look different from what is going on in this model, but in fact, it is a type of data sale. 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.

3 Extending the Data Growth Framework

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. Below are seven extensions that could allow the model to address a variety of interesting data-

related phenomena.

Data as an Input in Idea Production The previous result that data accumulation cannot sustain growth is a statement about what happens when we hold ideas fixed. 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. 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.

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:

$$A_{i,t} = \hat{A} - (a_{i,t} - \theta_t - \epsilon_{a,i,t})^2 + \int_{j=0}^1 (a_{j,t} - \theta_t - \epsilon_{a,j,t})^2 dj \quad (13)$$

This captures the idea that when one firm uses data to reduce the distance between their chosen technique a_{it} 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: $\int A_{it} = \hat{A}$. 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 $\int_{j=0}^1 (a_{j,t} - \theta_t - \epsilon_{a,j,t})^2 dj$. 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.

Investing in Data-Savviness So far, the data savviness parameter z_i has been fixed. To some extent, this 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 service providers may be learning soft information from their clients, but 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. Our previous results still hold for a given set of z choices. Put differently, 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.

Data allocation choice. A useful extension of the model would be to add a choice about what type of data to purchase or process. To do that, one needs to make the relevant state θ_t a vector of variables. Then, use rational inattention.

Why is rational inattention a natural complement to this model? Following Sims (2003), rational inattention problems consider what types of information or data is most valuable to process, subject to a constraint on the mutual information of the processed information and the underlying uncertain economic variables. The idea of using mutual information as a constraint, or the basis of a cost function, comes from the computer science literature on information theory. The mutual information of a signal and a state is an approximation to the length of the bit string or binary code necessary to transmit that information (Cover and Thomas (1991)). While the interpretation of rational inattention in economics has been mostly as a cognitive limitation on processing information, the tool was originally designed to model computers' processing of data. Therefore, it would be well suited to explore the data processing choices of firms.

Measuring data. The model suggests two possible ways of measuring data. One is to measure output or transactions. If we think data is a by-product of economic activity, then a measure of that activity should be a good indicator of aggregate data production. At the firm level, a firm's data use could differ from their data production, if data is traded. But one can adjust data production for data sales and purchases, to get a firm-level flow measure of data. Then, a stock of data is a discounted sum of data flows. The discount rate depends on the persistence of the market. If the data is about demand for fashion, then rapidly changing tastes imply that data has a short longevity and a high discount rate. If the data is mailing addresses, that is quite persistent, with small innovations. An AR(1) coefficient and innovation variance of the variable being forecasted are sufficient to determine the discount rate.

The second means of measuring data is to look at what actions it allows firms to choose. A firm with more data can respond more quickly to market conditions than a firm with little data to guide them. To use this measurement approach, one needs to take a stand on what actions firms are using data to inform, what variable firms are using the data to forecast, and to measure both the variable and the action. One example is portfolio choice in financial markets Farboodi et al. (2018). Another example is firms' real investment David et al. (2016). Both measure the

covariance between investment choices and future returns. That covariance between choices and unknown states reveals how much data investors have about the future unknown state. A similar approach could be to use the correlation between consumer demand and firm production, across a portfolio of goods, to infer firms' data about demand. To check the power of the resulting measure, one might look at the covariance with Solow residuals, under the assumption that more data makes firms look more productive. One could also compare that measure to imputed values of intangible capital.

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

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. This force is present in our model and does generate dispersion in firm size. However, that dispersion is a transitory phenomenon.

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. Increasing returns means that a firm with a larger stock of knowledge acumulates additional data at a faster pace. That drives divergence. Later on in the growth path, the trajectory becomes concave. This is where diminishing returns sets in. In this region, firms with a larger stock of knowledge accumulate more data, but less additional knowledge. As diminishing returns to data set in, firms that have varying degrees of abundance in their stock of knowlege become similar in their productivity and their output. This is the region of convergence. Small initial differences in the stock of knowledge fade in importance.

Data traps as a barrier to national growth The data trap at a firm level, described in Section 2, might also manifest itself as a data trap at the country level. Instead of one small, transitioning firm facing a static market, the interpretation would be that this is a small open economy, facing stable world prices. If that economy cannot purchase data, or the data for sale is not relevant to their forecasting problem, a strong data feedback loop could create a barrier to growth.

The policy solution would be a form of big push for data investment. 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.

Finally, there could be a trap that arises from the lack of complementary skills. In a country where data science skills are scarce, the labor cost of hiring a data analyst may make this data adjustment cost very high. Thus scarce data-skilled labor might condemn an economy of firms to this data poverty trap.

4 Conclusions

The economics of transactions data bears some resemblance to technology and some to capital. It is not identical to either. Yet, when economies accumulate data alone, the aggregate growth economics are similar to an economy that accumulates capital alone. Diminishing returns set in and the gains are bounded. Yet, the transition paths differ. There can be regions of increasing returns that create possible poverty traps. 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. When markets for data exist, some of the effects are mitigated, but the diminishing returns persist. Even if data does not increase output at all, but is only a form of business stealing, the dynamics are unchanged. 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.

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.
- Cover, Thomas and Joy Thomas**, *Elements of information theory*, first ed., John Wiley and Sons, New York, New York, 1991.
- David, Joel M, Hugo A Hopenhayn, and Venky Venkateswaran**, “Information, misallocation, and aggregate productivity,” *The Quarterly Journal of Economics*, 2016, 131 (2), 943–1005.
- Easterly, William R.**, *The Elusive Quest for Growth: Economists’ Adventures and Misadventures in the Tropics*, The MIT Press, 2002.
- Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel**, “Uncertainty Traps,” *The Quarterly Journal of Economics*, 2017, 132 (4), 1641–1692.
- Farboodi, Maryam, Adrien Matray, and Laura Veldkamp**, “Where has all the big data gone?,” 2018.
- , **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.
- Sims, Christopher**, “Implications of rational inattention,” *Journal of Monetary Economics*, 2003, *50*(3), 665–90.
- 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 $A_{i,t-1} = y_{i,t-1}/k_{i,t-1}^\alpha$. Good quality $A_{i,t-1}$ provides a noisy signal about θ_{t-1} . Let that signal be $s_{i,t-1}^a = (\bar{A} - A_{i,t-1})^{1/2} - a_{i,t-1}$. Note that, from equation 2, that the signal derived from observed output is equivalent to

$$s_{i,t-1}^a = \theta_{t-1} + \epsilon_{a,t-1}, \quad (14)$$

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 $\omega_{it} = n_{it} = z k_{it}^\alpha$. For firms that do trade data, $\omega_{it} = n_{it} + \delta_{it}(\mathbf{1}_{\delta_{it} > 0} + \iota \mathbf{1}_{\delta_{it} < 0})$. 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:

$$\bar{s}_t = \theta_t + \epsilon_{s,t}, \quad (15)$$

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

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

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

At time, t , the firm takes as given:

$$\begin{aligned}\hat{\mu}_{t-1} &= \mathbb{E}[\theta_t \mid s^{t-1}, y^{t-2}] \\ \Sigma_{t-1} &= \text{Var}[\theta_t \mid s^{t-1}, y^{t-2}]\end{aligned}$$

where $s^{t-1} = \{s_{t-1}, s_{t-2}, \dots\}$ and $y^{t-2} = \{y_{t-2}, y_{t-3}, \dots\}$ 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 $s_{i,t-1}^a$:

$$\begin{aligned}\mathbb{E}[\theta_{t-1} \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] &= \frac{\Sigma_{t-1}^{-1} \hat{\mu}_{t-1} + \sigma_a^{-2} s_{i,t-1}^a}{\Sigma_{t-1}^{-1} + \sigma_a^{-2}} \\ V[\theta_{t-1} \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] &= [\Sigma_{t-1}^{-1} + \sigma_a^{-2}]^{-1} \\ \mathbb{E}[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] &= \bar{\theta} + \rho \cdot (\mathbb{E}[\theta_{t-1} \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] - \bar{\theta}) \\ V[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] &= \rho^2 [\Sigma_{t-1}^{-1} + \sigma_a^{-2}]^{-1} + \sigma_\theta^2\end{aligned}$$

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

$$\hat{\mu}_t = \mathbb{E}[\theta_t \mid \mathcal{I}_t] = \frac{\left[\rho^2 [\Sigma_{t-1}^{-1} + \sigma_a^{-2}]^{-1} + \sigma_\theta^2 \right]^{-1} \cdot \mathbb{E}[\theta_t \mid \mathcal{I}_{t-1}, s_{i,t-1}^a] + \omega_t \sigma_\epsilon^{-2} \bar{s}_t}{\left[\rho^2 [\Sigma_{t-1}^{-1} + \sigma_a^{-2}]^{-1} + \sigma_\theta^2 \right]^{-1} + \omega_t \sigma_\epsilon^{-2}} \quad (16)$$

$$\Sigma_t = \text{Var}[\theta_t \mid \mathcal{I}_t] = \left\{ \left[\rho^2 [\Sigma_{t-1}^{-1} + \sigma_a^{-2}]^{-1} + \sigma_\theta^2 \right]^{-1} + \omega_t \sigma_\epsilon^{-2} \right\}^{-1} \quad (17)$$

Multiply and divide equation (16) by Σ_t as defined in equation (17) to get

$$\hat{\mu}_{it} = (1 - \omega_t \sigma_\epsilon^{-2} \Sigma_t) \left[\bar{\theta} (1 - \rho) + \rho \left((1 - M_t) \mu_{t-1} + M_t s_{i,t-1}^a \right) \right] + \omega_t \sigma_\epsilon^{-2} \Sigma_t \bar{s}_t, \quad (18)$$

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

Equations (17) and (18) 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:

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

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 ω_{it} 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 $a_t^* = \mathbb{E}[\theta_t | \mathcal{I}_t]$. We can substitute the optimal choice of a_t into A_t and rewrite the value function as

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

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

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

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 (17), this posterior variance is

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

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

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

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

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

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

$$k_{it} = \left[\frac{\alpha}{r} \left(P_t A_{it} + z_i \sigma_\epsilon^{-2} \left(\frac{1}{1+r} \right) V'(\cdot) - 2\psi(\cdot) \right) \right]^{1/(1-\alpha)} \quad (19)$$

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

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

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

To solve this, we start with a guess of V' and then solve the non-linear equation above for k_{it} . 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 $\Omega_{i,t+1} = \Omega_{it}$. Using the law of motion for Ω (eq 9), we can rewrite this as

$$\omega_{ss} \sigma_\epsilon^{-2} + [\rho^2 (\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2]^{-1} = \Omega_{ss} \quad (20)$$

This is equating the inflows of data $\omega_{it} \sigma_\epsilon^{-2}$ with the outflows of data $[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2]^{-1} - \Omega_{it}$. 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 (19) 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 $V'(\Omega_{it}) = V'(\Omega_{i,t+1})$ allows us to solve for the marginal value of data analytically, in terms of k_{ss} , which in turn depends on Ω_{ss} :

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r} \right) \frac{\partial \Omega_{t+1}}{\partial \Omega_t} \Big|_{ss} \right]^{-1} P_t k_{ss}^\alpha \Omega_{ss}^{-2} \quad (21)$$

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 (19), (20) and (21) 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 δ_{it} , but rather the closely related, net change in the data stock ω_{it} . We also substitute in $n_{it} = z_i k_{it}^\alpha$ and substitute in the optimal choice of technique a_{it} . The equivalent problem becomes

$$V(\Omega_{i,t}) = \max_{k_{i,t}, \omega_{i,t}} P_t \left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2 \right) k_{i,t}^\alpha - \pi \left(\frac{\omega_{it} - z_i k_{it}^\alpha}{\mathbf{1}_{\omega_{it} > n_{it}} + \iota \mathbf{1}_{\omega_{it} < n_{it}}} \right) - r k_{i,t} - \Psi(\Delta \Omega_{i,t+1}) + \left(\frac{1}{1+r} \right) V(\Omega_{i,t+1}) \quad (22)$$

$$\text{where } \Omega_{i,t+1} = [\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2]^{-1} + \omega_{it} \sigma_\epsilon^{-2} \quad (23)$$

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

$$FOC[k_{it}] : \quad \alpha P_t A_{it} k_{it}^{\alpha-1} + \frac{\pi \alpha z_i k_{it}^{\alpha-1}}{\mathbf{1}_{\omega_{it} > n_{it}} + \iota \mathbf{1}_{\omega_{it} < n_{it}}} - r = 0 \quad (24)$$

Solving for k_{it} gives

$$k_{it} = \left(\frac{1}{r} (\alpha P_t A_{it} + \tilde{\pi} \alpha z_i) \right)^{\frac{1}{1-\alpha}} \quad (25)$$

where $\tilde{\pi} \equiv \pi / (\mathbf{1}_{\omega_{it} > n_{it}} + \iota \mathbf{1}_{\omega_{it} < n_{it}})$. 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 ω_{it} is

$$FOC[\omega_{it}] : \quad -\Psi'(\cdot) \frac{\partial \Delta \Omega_{i,t+1}}{\partial \omega_{it}} - \tilde{\pi} + \left(\frac{1}{1+r} \right) V'(\Omega_{i,t+1}) \frac{\partial \Omega_{i,t+1}}{\partial \omega_{it}} = 0 \quad (26)$$

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

Steady state The steady state is where capital and data are constant. For data to be constant, it means that $\Omega_{i,t+1} = \Omega_{it}$. Using the law of motion for Ω (eq 9), we can rewrite this as

$$\omega_{ss} \sigma_\epsilon^{-2} + [\rho^2 (\Omega_{ss} + \sigma_a^{-2})^{-1} + \sigma_\theta^2]^{-1} = \Omega_{ss} \quad (27)$$

This is equating the inflows of data $\omega_{it} \sigma_\epsilon^{-2}$ with the outflows of data $[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2]^{-1} - \Omega_{it}$. 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 25 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 $V'(\Omega_{it}) = V'(\Omega_{i,t+1})$ allows us to solve for the marginal value of data analytically, in terms of k_{ss} , which in turn depends on Ω_{ss} :

$$V'_{ss} = \left[1 - \left(\frac{1}{1+r} \right) \frac{\partial \Omega_{t+1}}{\partial \Omega_t} \Big|_{ss} \right]^{-1} P_{ss} k_{ss}^\alpha \Omega_{ss}^{-2} \quad (28)$$

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 ω_{it} (eq 26), the steady state marginal value is given by

$$V'_{ss} = (1+r) \tilde{\pi} \sigma_\epsilon^2 \quad (29)$$

The equations (25), (26), (27) and (28) 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 $\delta := \omega_{it} - n_{it}$.

Individual Optimization Problem.

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

where i denotes the firm data productivity.

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

$$\Omega_L = (\rho^2 (\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2)^{-1} + (z_L k_L^\alpha + \delta_L) \sigma_\epsilon^{-2} \quad (30)$$

$$\Omega_H = (\rho^2 (\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2)^{-1} + (z_H k_H^\alpha + \iota \delta_H) \sigma_\epsilon^{-2} \quad (31)$$

$$\alpha P (\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^{\alpha-1} + \pi \alpha z_L k_L^{\alpha-1} = r \quad (32)$$

$$\alpha P (\bar{A} - \Omega_H^{-1} - \sigma_a^2) k_H^{\alpha-1} + \frac{\pi \alpha z_H k_H^{\alpha-1}}{\iota} = r \quad (33)$$

$$P \sigma_\epsilon^{-2} k_L^\alpha = \pi \Omega_L^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\Omega_L + \sigma_a^{-2}))^2} \right) \quad (34)$$

$$\iota P \sigma_\epsilon^{-2} k_H^\alpha = \pi \Omega_H^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\Omega_H + \sigma_a^{-2}))^2} \right) \quad (35)$$

$$P = \bar{P} (\lambda (\bar{A} - \Omega_L^{-1} - \sigma_a^2) k_L^\alpha + (1 - \lambda) (\bar{A} - \Omega_H^{-1} - \sigma_a^2) k_H^\alpha)^{-\gamma} \quad (36)$$

$$\lambda \delta_L + (1 - \lambda) \delta_H = 0 \quad (37)$$

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 $n_{it} \rightarrow \infty$ implies finite firm output $y_{it} < \infty$. Thus $M_y \equiv \sup_i \{y_{i,t}\} + 1$ exists and is finite. By definition, $y_{i,t} < M_y$, $\forall i$. If the measure of all firms is also finite, that is $\exists 0 < N < \infty$ such that $\int_i di < N$. As a result,

the aggregate output is also finite in any period $t + s$, $\forall s > 0$:

$$Y_{t+s} = \int_i y_{i,t} di < M_y \int_i di < M_y N < \infty. \quad (38)$$

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

$$\ln(Y_{t+s}) - \ln(Y_t) = [\ln(Y_{t+s}) - \ln(Y_{t+s-1})] + \dots + [\ln(Y_{t+1}) - \ln(Y_t)] > \underline{g}s, \quad (39)$$

which implies

$$Y_{t+s} > Y_t e^{\underline{g}s}. \quad (40)$$

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

$$Y_{t+s} > Y_t e^{\underline{g}s} > Y_t e^{\underline{g}s} > Y_t e^{\frac{\underline{g}(\ln(MN) - \ln(Y_t))}{\underline{g}}} = M_y N, \quad (41)$$

which contradicts (38).

A.6 Proof of Proposition 2: Data without Diminishing Returns Implies Infinite Precision.

Suppose not. Then, for every firm $i \in I$, with $\int_{i \notin I} di = 0$, producing infinite data $n_{it} \rightarrow \infty$ implies finite precision $\Omega_{it} < \infty$, 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 $n_{it} \rightarrow \infty$. In other words, since Y_t is a finite-valued function of Γ_t , it can only be that $Y_t \rightarrow \infty$ if some moment of Γ_t is also becoming infinite $\Gamma_t \rightarrow \pm\infty$. 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, $\Gamma_t \rightarrow \infty$. 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: $\Omega_{it} \rightarrow \infty$.

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 $Y_t \rightarrow \infty$ when some firm with positive measure has infinite data $n_{it} \rightarrow \infty$. This is a contradiction.

A.7 Proof of Proposition 3: Infinite Precision Implies Deterministic Future.

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 $\{\theta_\tau, s_{\tau,m}, s_{i,\tau}^a\}_{\tau=0}^{t-1}$.

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 $\mathcal{I}_{i,t} = [\{s_{i,\tau}^a\}_{\tau=0}^{t-1}; \{\{s_{i,\tau,m}\}_{m=1}^{n_{i,\tau}}\}_{\tau=0}^{t-1}]$. By assumption, $s_{i,t-1,m}$ and $s_{i,t-1}^a$ are measurable with respect to \mathcal{F}_t , that is $\forall B \in \mathcal{B}, \{\omega : s_{i,t-1,m}(\omega) \in B\} \subset \mathcal{F}_t$. 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 $\text{Var}(\theta_t | \mathcal{I}_{i,t,m}) > 0$. By the definition of Ω as the inverse of the conditional variance, this implies $\Omega_{it} < \infty$.

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.8 Proof of Proposition 4: S-shaped accumulation of knowledge.

Before we proceed to the proof, we define data inflows $d\Omega_{it}^+$, data outflows $d\Omega_{it}^-$, and net data flows $d\Omega_{it}$. Data inflows are the total precision of all new data points at t : $d\Omega_{it}^+ = z_{i,t} k_{i,t}^\alpha \sigma_\epsilon^{-2}$. Data outflows are the end-of-period- t stock of knowledge minus the discounte stock: $d\Omega_{it}^- = \Omega_{i,t} + \sigma_a^{-2} -$

$[(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$. Net data flows are the difference between data inflows and outflows:
 $d\Omega_{it} = d\Omega_{it}^+ - d\Omega_{it}^-$.

Proof: We proceed in two parts: convexity and then concavity.

Part 1: Convexity at low levels of Ω_{it} . In this part, we first calculate the derivatives of data inflow and outflow with respect to Ω_{it} , combine them to form the derivative of data net flow, and then show that it is positive in given parameter regions for $\Omega_{it} < \hat{\Omega}$.

Recall that data inflow is $d\Omega_{it}^+ = z_{i,t}k_{i,t}^\alpha\sigma_\epsilon^{-2}$ and its first derivative is $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{i,t}} = \alpha z_{i,t}k_{i,t}^{\alpha-1}\sigma_\epsilon^{-2}\frac{\partial k_{i,t}}{\partial \Omega_{i,t}}$. 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 19) drops out and it reduces to $k_{it} = \left[\frac{\alpha}{r} \left(P_t A_{it} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V'(\Omega_{it+1}) \right) \right]^{1/(1-\alpha)}$, which implies

$$k_{it}^{1-\alpha} = \frac{\alpha}{r} \left(P_t A_{it} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V'(\Omega_{it+1}) \right).$$

Differentiating with respect to Ω_{it} 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 $\frac{\partial A_{it}}{\partial \Omega_{it}} = \Omega_{it}^{-2}$ and $\frac{\partial \Omega_{it+1}}{\partial \Omega_{it}} = \rho^2[\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2}$, we have

$$\implies \frac{\partial k_{it}}{\partial \Omega_{it}} = k_{it}^\alpha \frac{\alpha}{(1-\alpha)r} \left(P_t \Omega_{it}^{-2} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V''(\Omega_{it+1}) \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2} \right).$$

Therefore,

$$\begin{aligned} \frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} &= z_{it}k_{it}^{2\alpha-1}\sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r} \left(P_t \Omega_{it}^{-2} + z_i \sigma_\epsilon^{-2} \frac{1}{1+r} V''(\Omega_{it+1}) \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2} \right) \\ &= z_{it}k_{it}^{2\alpha-1}\sigma_\epsilon^{-2} \frac{\alpha^2}{(1-\alpha)r} P_t \Omega_{it}^{-2} + z_{it}^2 k_{it}^{2\alpha-1} \sigma_\epsilon^{-4} \frac{\alpha^2}{1-\alpha} \frac{1}{r(1+r)} V''(\Omega_{it+1}) \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2}. \end{aligned} \tag{42}$$

Next, we calculate the derivative of data outflow $d\Omega_{it}^- = \Omega_{i,t} + \sigma_a^{-2} - [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$

with respect to Ω_{it} . We have

$$\frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} = 1 - \frac{1}{\rho^2(\Omega_{it} + \sigma_a^{-2})^2(\sigma_\theta^2 + \rho^{-2}(\Omega_{it} + \sigma_a^{-2})^{-1})^2}. \quad (43)$$

The derivatives of net data flow is then

$$\begin{aligned} \frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} &= z_{it}k_{it}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t\Omega_{it}^{-2} + z_{it}^2k_{it}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{it+1})\rho^2[\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2} \\ &+ \frac{1}{\rho^2(\Omega_{it} + \sigma_a^{-2})^2(\sigma_\theta^2 + \rho^{-2}(\Omega_{it} + \sigma_a^{-2})^{-1})^2} - 1. \end{aligned} \quad (44)$$

For notational convenience, denote the first term in (2) as $M_1 = z_{it}k_{it}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t\Omega_{it}^{-2} > 0$, the second term as $M_2 = z_{it}^2k_{it}^{2\alpha-1}\sigma_\epsilon^{-4}\frac{\alpha^2}{1-\alpha}\frac{1}{r(1+r)}V''(\Omega_{it+1})\rho^2[\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2} \leq 0$ and the third term as $M_3 = \frac{1}{\rho^2(\Omega_{it} + \sigma_a^{-2})^2(\sigma_\theta^2 + \rho^{-2}(\Omega_{it} + \sigma_a^{-2})^{-1})^2} > 0$. Notice that $M_3 - 1 < 0$ always holds, and thus $M_2 + M_3 - 1 < 0$. $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} > 0$ only holds when P_t is sufficiently large so that M_1 dominates.

Assume that $V'' \in [\nu, 0)$. Let $h(\Omega_{it}) \equiv M_1(\bar{P}) + M_2(\nu)$. Then

$$\begin{aligned} h'(\Omega_{it}) &= (2\alpha - 1)z_{it}k_{it}^{3\alpha-2}\alpha\left(\frac{\alpha}{r(1-\alpha)}\right)^2\sigma_\epsilon^{-2}\left[\bar{P}\Omega_{it}^{-2} + z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2[\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2})]^{-2}\right]^2 \\ &+ z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{(\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2}))^3}\right]. \end{aligned}$$

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

$$z_{it}k_{it}^{2\alpha-1}\frac{\alpha^2}{(1-\alpha)r}\sigma_\epsilon^{-2}\left[-2\bar{P}\Omega_{it}^{-3} - z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\frac{2\sigma_\theta^2}{(\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2}))^3}\right] > 0 \quad (45)$$

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

$$f(\Omega_{i,t}) := -z_{it}\sigma_\epsilon^{-2}\frac{1}{1+r}\nu\rho^2\Omega_{it}^3\frac{\sigma_\theta^2}{(\rho^2 + \sigma_\theta^2(\Omega_{it} + \sigma_a^{-2}))^3} \quad (46)$$

Notice by inspection that $f'(\Omega_{it}) < 0$.

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

$$h(\Omega_{it}) = 1 - M_3, \quad (47)$$

then if $\alpha < \frac{1}{2}$, when $\Omega_{it} < \hat{\Omega}$ and $P_t > f(\hat{\Omega})$, we have that $h(\Omega_{it})$ is decreasing in Ω_{it} and $h(\Omega) \geq 1 - M_3$. Since $\nu \leq V''$, we then have $M_1 + M_2 \geq 1 - M_3$, that is $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} > 0$. By the same token, if $\alpha > \frac{1}{2}$ and $P_t < f(\Omega_{it})$, then $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} < 0$.

Part 2: Concavity at high levels of Ω_{it} . In this part, we first calculate limit of the derivatives of net data flow with respect to Ω_{it} is negative when Ω_{it} goes to infinity and then prove that when Ω_{it} is large enough, $\frac{\partial d\Omega_{it}}{\partial \Omega_{it}}$ is negative.

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

For the derivative of data inflow (42), note that $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} \leq z_{it}k_{it}^{2\alpha-1}\sigma_\epsilon^{-2}\frac{\alpha^2}{(1-\alpha)r}P_t\Omega_{it}^{-2}$ because $0 < \alpha < 1$ and $V'' < 0$. Thus $\lim_{\Omega_{it} \rightarrow \infty} \frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} \leq 0$.

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

A.9 Proof of Proposition 6: 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:

$$\Omega_L = (\rho^2(\Omega_L + \sigma_a^{-2})^{-1} + \sigma_\theta^2)^{-1} + z_L k_L^\alpha \sigma_\epsilon^{-2} \quad (48)$$

$$\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^{\alpha-1} + \pi \alpha z_L k_L^{\alpha-1} = r \quad (49)$$

$$P\sigma_\epsilon^{-2}k_L^\alpha = \pi\Omega_L^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2}))^2} \right) \quad (50)$$

$$P = \bar{P}((\bar{A} - \Omega_L^{-1} - \sigma_a^2)k_L^\alpha)^{-\gamma} \quad (51)$$

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:

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2)k_H^{\alpha-1} + \frac{\pi\alpha z_H k_H^{\alpha-1}}{\iota} = r \quad (52)$$

$$\iota P \sigma_\epsilon^{-2} k_H^\alpha = \pi \Omega_H^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2}))^2} \right) \quad (53)$$

$$\Omega_H = (\rho^2(\Omega_H + \sigma_a^{-2})^{-1} + \sigma_\theta^2)^{-1} + (z_H k_H^\alpha + \iota \delta_H) \sigma_\epsilon^{-2} \quad (54)$$

Manipulate to get

$$\alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \frac{\pi\alpha z_H}{\iota} = r k_H^{1-\alpha} \quad (55)$$

$$k_H^\alpha = (k_H^{1-\alpha})^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{\iota P \sigma_\epsilon^{-2}} \Omega_H^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2}))^2} \right) \quad (56)$$

$$\iota^{\frac{1-2\alpha}{1-\alpha}} \frac{1}{r^{\frac{\alpha}{1-\alpha}}} (\iota \alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \pi \alpha z_H)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P \sigma_\epsilon^{-2}} \Omega_H^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2}))^2} \right) \quad (57)$$

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 (57) wrt to z_H and simplify. It implies

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

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

$$A(i) = \frac{\alpha^2 \iota_i^{2-\frac{1}{1-\alpha}} \left(\iota_i r k_i^{1-\alpha} \right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha) \Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}} k_i^{2\alpha-1}}{(1-\alpha) \Omega_i^2} \quad (58)$$

$$B(i) = \frac{2\pi \sigma_\epsilon^2 \Omega_i \left(1 + r - \frac{\rho^2 (\rho^2 + \sigma_a^2 \sigma_\theta^2)}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_i))^3} \right)}{P} = \pi \frac{dC(i)}{d\Omega_i} \quad (59)$$

$$C(i) = \frac{\sigma_\epsilon^2 \Omega_i^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_i))^2} \right)}{P} = \frac{\iota_i k_i^\alpha}{\pi} \quad (60)$$

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

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

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

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

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

Furthermore, consider equation (35). 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

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

$\frac{(\rho^2 + \sigma_\theta^2\sigma_a^{-2})}{(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2}))} < 1$, thus equation (35) 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 (56) 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, $\forall z_H$,

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

Step 3. It is straight forward to show that $\bar{\iota} > 1$. $\iota > 1$ would mean that selling data would result in more data for the seller. That makes no economic sense. We have thus restricted $\iota \leq 1$, from the start. The interpretation of this result is therefore that the conditions that guarantee Proposition 6 holds are satisfied for any ι .

A.10 Proof of Proposition 7: Knowledge Gap When High Data Productivity is Abundant.

When $\gamma = 0$, $P = \bar{P}$. The proof proceeds in a few steps.

\mathbf{z}_H firms are data sellers while \mathbf{z}_L firms are data buyers ($\delta_H < 0$ and $\delta_L > 0$).

$$\frac{d\pi}{dz_H} < 0.$$

Change in knowledge gap: $\frac{d\Upsilon^{ss}}{dz_H}$. Equations (32) and (34) can be solved to get (k_L, Ω_L) in terms of data price π :

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

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

The second equation implies

$$\Omega_L^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2}))^2} \right) = \frac{(\bar{P}(\bar{A} - \Omega_L^{-1} - \sigma_a^2) + \pi z_L)^{\frac{\alpha}{1-\alpha}}}{\pi} \bar{P}\sigma_\epsilon^{-2} \left(\frac{\alpha}{r} \right)^{\frac{\alpha}{1-\alpha}} \quad (61)$$

The same argument as in proposition 6 shows that using equation (34), the derivative of the left hand side with respect to Ω_L is positive. Next, using implicit function theorem on both sides of equation (61) 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 ((32), (34)) and ((33), (35)) to get:

$$\frac{1}{r^{\frac{\alpha}{1-\alpha}}} (\alpha P(\bar{A} - \Omega_L^{-1} - \sigma_a^2) + \pi \alpha z_L)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}} \Omega_L^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_L + \sigma_a^{-2}))^2} \right) \quad (62)$$

$$\frac{\iota^{\frac{1-2\alpha}{1-\alpha}}}{r^{\frac{\alpha}{1-\alpha}}} (\iota \alpha P(\bar{A} - \Omega_H^{-1} - \sigma_a^2) + \pi \alpha z_H)^{\frac{\alpha}{1-\alpha}} = \frac{\pi}{P\sigma_\epsilon^{-2}} \Omega_H^2 \left(1 + r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2(\Omega_H + \sigma_a^{-2}))^2} \right). \quad (63)$$

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

$$\begin{aligned} \frac{d\Omega_L}{dz_H} &= \frac{-\frac{\sigma_\epsilon^2 \Omega_L^2 \left(1+r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_L))^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)}}{2\pi \sigma_\epsilon^2 \Omega_L \left(1+r - \frac{\rho^2 (\rho^2 + \sigma_a^2 \sigma_\theta^2)}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_L))^3}\right) - \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)}}{2\pi \sigma_\epsilon^2 \Omega_H \left(1+r - \frac{\rho^2 (\rho^2 + \sigma_a^2 \sigma_\theta^2)}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_H))^3}\right) - \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}} \\ &\quad + \frac{-\frac{\sigma_\epsilon^2 \Omega_H^2 \left(1+r - \frac{\rho^2}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_H))^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)}}{2\pi \sigma_\epsilon^2 \Omega_H \left(1+r - \frac{\rho^2 (\rho^2 + \sigma_a^2 \sigma_\theta^2)}{(\rho^2 + \sigma_\theta^2 (\sigma_a^2 + \Omega_H))^3}\right) - \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} \end{aligned}$$

Using definition (58)-(60) the above expressions simplify to:

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

Thus the derivative of the knowledge gap is given by

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

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

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

Note that

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

Multiply both sides by the denominator on the left hand side, which is positive as $\iota < 1$.

and divide both sides by the right hand side expression which is also positive. Since both expressions are positive, the inequality sign does not change

$$\begin{aligned} &\frac{\pi\Omega_H^2 A(H) + (z_H\Omega_H^2 A(H) - C(H)) \frac{d\pi}{dz_H}}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)} \frac{d\pi}{dz_H}} > B(H) - \iota\bar{P}A(H) \\ \iota &> \frac{1}{\bar{P}A(H)} \left(B(H) - \frac{\pi\Omega_H^2 A(H) + (z_H\Omega_H^2 A(H) - C(H)) \frac{d\pi}{dz_H}}{\frac{z_L\Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)} \frac{d\pi}{dz_H}} \right) \\ A(i) &= \frac{\alpha^{\frac{1}{1-\alpha}} \iota_i^{2-\frac{1}{1-\alpha}} \left(\left(\iota_i \bar{P} \left(\bar{A} - \sigma_a^2 - \frac{1}{\Omega_i} \right) + \pi z_i \right) \right)^{\frac{1}{1-\alpha}-2}}{(1-\alpha)\Omega_i^2} = \frac{\alpha^2 r^{\frac{2\alpha-1}{1-\alpha}} k_i^{2\alpha-1}}{(1-\alpha)\Omega_i^2} \end{aligned}$$

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

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

Next we show that fixing the parameters, $\exists \bar{\iota}, \underline{\iota}$ such that if $\iota < \bar{\iota}$, the knowledge gap is negative.

$$\exists \bar{\iota} \text{ s.t. } \underline{\iota} < \iota < \bar{\iota} \rightarrow \Omega_H - \Omega_L < 0.$$

This is done in the following steps:

1. compute $\frac{d\Omega_H}{d\iota}$ and $\frac{d\Omega_L}{d\iota}$.

2. show that $\exists \underline{\iota}_1$ such that $\frac{d(\Omega_H - \Omega_L)}{d\iota} > 0$ when $\iota > \underline{\iota}_1$. (we know no equilibrium exist when $\iota = 0$, so $\underline{\iota}$ keeps us away from the non-existence limit for sure).
3. show that $\Omega_H - \Omega_L > 0$ for $\iota = 1$.
4. show that $\exists \underline{\iota} \geq \underline{\iota}_1$ such that $\Omega_H - \Omega_L < 0$ for $\iota = \underline{\iota}$.

Step 1. Taking the derivative, we have

$$\begin{aligned} \frac{d\Omega_L}{d\iota} &= \frac{z_L \Omega_L^2 A(L) - C(L)}{B(L) - \bar{P}A(L)} \frac{d\pi}{d\iota} \\ \frac{d\Omega_H}{d\iota} &= \frac{\bar{P}(\bar{A} - \Omega_H^{-1} - \sigma_a^2) A(H) - D(H)}{B(H) - \iota \bar{P}A(H)} + \frac{z_H \Omega_H^2 A(H) - C(H)}{B(H) - \iota \bar{P}A(H)} \frac{d\pi}{d\iota} \end{aligned}$$

where

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

A.11 Proposition 8: 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 Ω_{it} is small enough, there is no convex region. Instead, the net data flow (the slope) decreases with Ω_{it} , right from the start.

Proposition 8 $\exists \epsilon > 0$ such that $\forall \Omega_{it} \in B_\epsilon(0)$, the net data flow decreases with Ω_{it} if $\sigma_\theta^2 > \sigma_a^2$.

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

Step 1: Data outflows are approximately linear when Ω_{it} 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 Ω_{it} 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 $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{i,t}} = \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})]^{-2} > 0$ (see the Proof of Proposition 4 for details). Thus

$$\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} \approx \rho^2 [\rho^2 + \sigma_\theta^2(\Omega_{i,t} + \sigma_a^{-2})]^{-2} - 1 + \rho^2(1 + \rho^2\sigma_\theta^2\sigma_a^{-2})^{-2}. \quad (64)$$

Since this derivative increases in ρ^2 and decreases in $\Omega_{it} = 0$, 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 $\frac{\partial d\Omega_{it}^+}{\partial \Omega_{it}} - \frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}}$ is negative, so it will always be negative $\forall \Omega_{it} \in B_\epsilon(0)$.

A.12 Proposition 9: S-shaped stock of knowledge over time.

This proposition shows the S-shape of Ω_{it} in the time domain. The result differs from Proposition 4 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 9 *Let $\gamma \equiv \frac{\sigma_\theta^2}{\sigma_a^2}$. If γ is sufficiently small, σ_ϵ sufficiently small, z_i sufficiently large, or σ_a sufficiently large, then*

- 1) $\frac{\partial^2 \Omega_{it}}{\partial t^2} > 0$ when Ω_{it} 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 (47), and f is defined by (46);
- 2) and $\frac{\partial^2 \Omega_{it}}{\partial t^2} < 0$ when Ω_{it} is large enough.

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

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

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

When $\Omega_{it} = 0$, the data infow and outflow are $d\Omega_{it}^+|_{\Omega_{it}=0} = z_i (k_0^*)^\alpha \sigma_\epsilon^{-2}$, where k_0^* is the optimal
investment when $\Omega_{it} = 0$, and $d\Omega_{it}^-|_{\Omega_{it}=0} = \sigma_a^{-2} (1 - \frac{\rho^2}{1+\rho^2\gamma})$, respectively. If $d\Omega_{it}^+|_{\Omega_{it}=0} \geq d\Omega_{it}^-|_{\Omega_{it}=0}$,
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 $\Omega_{it} \in (0, \bar{\Omega}_f)$ and negative when $\Omega_{it} \in (\bar{\Omega}_l, \infty)$.

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

Since $d\Omega_{it}^-|_{\Omega_{it}=0}$ is monotonic in ρ^2 and $\rho^2 \in [0, 1]$, we have $d\Omega_{it}^-|_{\Omega_{it}=0} \in [\frac{\gamma}{1+\gamma} \sigma_a^{-2}, \sigma_a^{-2}]$.
 $d\Omega_{it}^+|_{\Omega_{it}=0} \geq d\Omega_{it}^-|_{\Omega_{it}=0}$ is guaranteed when $d\Omega_{it}^+|_{\Omega_{it}=0} \geq \frac{\gamma}{1+\gamma} \sigma_a^{-2}$, that is $k_0^* \geq \frac{\sigma_\epsilon^2}{z_i \sigma_a^2} \frac{\gamma}{1+\gamma}$. The
last inequality is only true when γ 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 Ω_{it} to the time domain.

A.13 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 (Low levels of Ω_{it}) $\exists \epsilon > 0$ such that $\forall \Omega_{it} \in B_\epsilon(0)$, 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 approx-
imation error is small when ρ or σ_θ is small, or when Ω_{it} is very close to 0.

Proof:

Recall that data outflows are $d\Omega_{it}^- = \Omega_{i,t} + \sigma_a^{-2} - [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$. Let $g(\Omega_{i,t}) \equiv [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$ be the nonlinear part of data outflows. Its first-order Taylor expansion around 0 is $g(\Omega_{i,t}) = g(0) + g'(0)\Omega_{it} + o(\Omega_{it})$, with $g'(0) = \frac{\rho^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})}$. Thus $\frac{\partial d\Omega_{it}^-}{\partial \Omega_{it}} = 1 - g'(\Omega_{i,t}) \approx 1 - g'(0)$ for Ω_{it} in a small open ball $B_\epsilon(0)$, $\epsilon > 0$, around 0. And the approximation error is $|o(\Omega_{it})| = \frac{\rho^4\sigma_\theta^2\Omega_{it}^2}{(1+\rho^2\sigma_\theta^2\sigma_a^{-2})[1+\rho^2\sigma_\theta^2(\Omega_{it}+\sigma_a^{-2})]}$, which increases with Ω_{it} 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 $\exists \epsilon_\sigma > 0$ such that $\forall \sigma_\theta \in B_{\epsilon_\sigma}(0)$, data outflows are approximately linear and the approximation error is bounded from above by $\frac{\rho^4\epsilon^2(\Omega_{it}+\sigma_a^{-2})^2}{1+\rho^2\epsilon_\sigma^2(\Omega_{it}+\sigma_a^{-2})}$. The approximation error is small when ρ is small, or when σ_θ is close to 0.

Proof:

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

Lemma 5 When $\Omega_{it} \gg \sigma_\theta^{-2}$, discounted data stock is very small relative to Ω_{it} , 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 $d\Omega_{it}^- = \Omega_{i,t} + \sigma_a^{-2} - [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$. Let $g(\Omega_{i,t}) \equiv [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}$ be the nonlinear part of data outflows. Since $(\rho^2(\Omega_{it} + \sigma_a^{-2}))^{-1} \geq 0$, we have $g(\Omega_{it}) \leq \sigma_\theta^{-2}$. Since $\Omega_{it} \geq 0$, we have $g(\Omega_{it}) \geq (\rho^{-2}\sigma_a^2 + \sigma_\theta^2)^{-1}$. That is $g(\Omega_{it}) \in [(\rho^{-2}\sigma_a^2 + \sigma_\theta^2)^{-1}, \sigma_\theta^{-2}]$. For high levels of Ω_{it} , $\Omega_{it} \gg \sigma_\theta^{-2}$ generally holds. And for low levels of Ω_{it} , it holds when σ_θ is very large. The approximation error is $|\sigma_\theta^{-2} - [(\rho^2(\Omega_{i,t} + \sigma_a^{-2}))^{-1} + \sigma_\theta^2]^{-1}|$ and decreases with Ω_{it} , reaching its minimum at $\Omega_{it} = 0$ 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.⁵ 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 way as others have the adjust cost of capital. This approach makes sense because adjusting one's process to use more data typically involves the purchase of new capital, like new computing and recording equipment and involves disruptive changes in firm practice, similar to the disruption of working with new physical machinery.

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

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 of values for Ω (state variable) and k (choice variable). Guess functions $V_0(\Omega)$ and $P_0(\Omega)$ on this grid. Guess a vector

⁵To 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 .

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 than 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.